

*IBM SPSS Statistics Base 29*



**Poznámka**

Před použitím těchto informací a produktu, který podporují, si přečtěte informace v tématu [“Upozornění” na stránce 263](#).

**Informace o produktu**

Toto vydání se vztahuje k verzi 29, vydání 0, modifikaci 1 produktu IBM® SPSS Statistics a ke všem následujícím vydáním a modifikacím, dokud nebude v nových vydáních uvedeno jinak.

© Copyright International Business Machines Corporation .

---

# Obsah

<b>Kapitola 1. Základní funkce.....</b>	<b>1</b>
Analýza napájení.....	1
Prostředky.....	2
rozměry.....	15
Korelace.....	25
Regrese.....	32
Analýza napájení: Přesnost.....	35
Analýza napájení: Hodnoty mřížky.....	36
Metaanalýza.....	36
Souvislá meta-analýza.....	36
Velikost souvislé analýzy meta-analýzy.....	45
Binární meta-analýza.....	52
Velikost binárního efektu metaanalýzy.....	61
Regrese meta-analýzy.....	69
Meta-analýza-volby.....	73
Kódová kniha.....	79
Karta Výstup z kódové knihy.....	80
Karta Statistika kódové knihy.....	81
Frekvence.....	82
Statistika frekvencí.....	83
Grafy četnosti.....	84
Formát frekvence.....	84
Popisovače.....	84
Deskriptivní volby.....	85
Další funkce příkazu DESCRIPTIVA.....	86
percentily.....	86
Percentily Chybějící hodnoty.....	87
Prozkoumat.....	87
Prozkoumat statistiku.....	88
Prozkoumat paty.....	88
Volby prozkoumání.....	89
Dodatečné funkce příkazu EXAMINE.....	89
Kontingenční tabulky.....	89
Vrstvy kontingenční tabulky.....	90
Sloupcové pruhové grafy kontingenční tabulky.....	90
Kontingenční tabulky zobrazující proměnné vrstvy v vrstvách tabulky.....	91
Statistika kontingenčních tabulek.....	91
Zobrazení kontingenční tabulky.....	92
Formát tabulky kontingenčních tabulek.....	94
Shrnout.....	94
Volby sumarizace.....	94
Shrnout statistiku.....	95
Prostředky.....	96
Možnosti.....	97
Krychle OLAP.....	98
Statistika datových krychlí OLAP.....	99
Rozdíly krychlí OLAP.....	100
Nadpis datových krychlí OLAP.....	101
rozměry.....	101
Úvod do porcí.....	101
Jednovzorkový proporce.....	101

Paired-rozměry-porce.....	104
Nezávisle-části vzorků.....	106
Testy T.....	109
Testy T.....	109
Nezávislý-Test č. T.....	109
Paired-test s ukázkami T.....	111
Jednovzorkový test T.....	112
Dodatečné funkce příkazu T TEST.....	113
Jednosměrný test ANOVA.....	113
Kontrasty ANOVA One-Way.....	114
Jednorázové testy ANOVA jednofaktorové.....	115
Možnosti jednofaktorové ANOVA.....	117
Další funkce příkazu ONEWAY.....	117
Analýza proměnné GLM Univariate.....	118
Model GLM.....	119
Kontrasty GLM.....	120
Plosy profilu GLM.....	121
Porovnání post hoc GLM.....	123
Uložení GLM.....	126
Odhadované mezní hodnoty GLM.....	127
Volby GLM.....	127
Další funkce příkazu UNIANOVA.....	128
Bivariantní koráže.....	129
Volby relace korelace bivariate.....	130
Interval spolehlivosti Bivariate Correlations.....	130
Další funkce CORVZTAS a NONPAR CORR.....	131
Částečné korelace.....	131
Volby dílčích korelací.....	132
Další funkce příkazu PARTIAL CORR.....	132
Vzdálenosti.....	132
Měřítka přidružení k dispoďobnosti.....	133
Měřítka podobnosti odpojení.....	133
Další funkce příkazu PROXIMITIES.....	134
Lineární modely.....	134
Získání lineárního modelu.....	134
Cíle .....	134
Základní .....	135
Výběr modelu .....	136
Komplety .....	136
Rozšířené .....	137
Volby modelu .....	137
Souhrn modelu .....	137
Automatická příprava dat .....	137
Důležitost produktu Predictor .....	138
Predicted By Observed .....	138
Residuals .....	138
Odhadované prostředky .....	138
Efekty .....	138
Koeficienty .....	139
Odhadované prostředky .....	139
Souhrn budovy modelu .....	139
Lineární regrese.....	140
Metody výběru lineární regrese regresní proměnné.....	141
Pravidlo lineární regresní sady.....	141
Lineární regresní paty.....	141
Lineární regrese: Uložení nových proměnných.....	142
Lineární regresní statistika.....	143
Volby lineární regrese.....	144

Další funkce příkazu REGRESSION.....	145
Ordinální regrese.....	145
Ordinální regresní volby.....	146
Ordinální regresní výstup.....	146
Model ordinální regrese.....	147
Model Ordinální regresní stupnice.....	147
Další funkce příkazu PLUM.....	148
Lineární regrese elastické sítě.....	148
Lineární regrese elastické sítě: Volby.....	149
Lineární greso Regression.....	151
Lineární Lasso Regrese: Volby.....	152
Regrese lineárního Ridge.....	154
Lineární regrese Ridge: Volby.....	155
Odhad křivky.....	157
Modely odhadu křivky.....	158
Uložit odhad křivky.....	159
Regrese částečných nejmenších čtverců.....	159
Model .....	161
Volby .....	161
Analýza nejbližšího souseda.....	161
Sousedé .....	163
Funkce .....	164
Oblasti .....	164
Uložit .....	165
Výstup .....	165
Možnosti .....	166
Zobrazení modelu .....	166
Analýza diskriminátoru.....	169
Definovat rozsah analýzy diskriminátoru.....	170
Výběr případů analýzy diskriminátoru.....	170
Statistika analýzy diskriminátoru.....	170
Metoda selektivní analýzy diskriminátoru.....	171
Klasifikace analýzy diskriminátoru.....	171
Uložení analýzy diskriminátoru.....	172
Dodatečné funkce příkazu DISCRIMINANT.....	172
Analýza faktoru.....	172
Výběr případů analýzy faktoru.....	173
Deskriptory analýzy faktoru.....	173
Extrakce analýzy faktoru.....	174
Rotace analýzy faktoru.....	175
Skóre analýzy faktoru.....	175
Volby analýzy faktoru.....	175
Další funkce příkazu FACTOR.....	176
Výběr procedury pro klastrování.....	176
Analýza klastru TwoStep.....	176
Volby analýzy klastru TwoStep.....	178
Výstup analýzy klastru TwoStep.....	179
Prohlížeč klastrů.....	179
Analýza hierarchického klastru.....	184
Hierarchická metoda analýzy klastrů.....	184
Hierarchická statistika analýzy klastru.....	185
Plot pro analýzu hierarchického klastru.....	185
Hierarchická analýza klastru-Uložit nové proměnné.....	185
Další funkce syntaxe příkazu CLUSTER.....	185
Analýza K-Means Cluster.....	185
Efektivita analýzy klastru KK.....	186
Iterace analýzy K-Means Iterate.....	187
K-Means Uložení analýzy klastru.....	187

Volby analýzy klastrů K-Means.....	187
Další funkce příkazu QUICK CLUSTER.....	187
Neparametrické testy.....	188
Jedna-ukázka neparametrických testů.....	188
Nezávislí-Ukázkové neparametrické testy.....	191
Související-Neparametrické testy ukázek.....	194
Další funkce příkazu NPTESTS.....	196
Starší dialogová okna.....	196
Analýza vícenásobné odezvy.....	207
Analýza vícenásobné odezvy.....	207
Definovat více sad odpovědí.....	207
Frekvence více odpovědí.....	208
Kontingenční tabulky s více odpovědí.....	209
Výsledky vykazování.....	210
Výsledky vykazování.....	210
Souhrny sestav v řádcích.....	210
Souhrny sestav ve sloupcích.....	213
Další funkce příkazu REPORT.....	214
Analýza spolehlivosti.....	215
Analýza spolehlivosti: Statistika.....	216
Dodatečné funkce příkazu RELIABILITY.....	218
Vážené Kappy.....	218
Vážené Kappa: Kritéria.....	219
Vážené Kappa: Tisk.....	220
Multidimenzionální škálování.....	220
Vícerozměrný tvar měřítka dat.....	221
Ukazatel vytvoření vícerozměrného škálování.....	221
Vícerozměrný model škálování.....	222
Vícerozměrné volby škálování.....	222
Další funkce příkazu ALSCAL.....	222
Poměr statistik.....	223
Poměrové statistiky: Statistika.....	223
Poměr statistik.....	225
Poměrové statistiky: Statistika.....	225
Výkres P-P.....	226
Zakreslení.....	229
Analýza ROC.....	231
Analýza ROC: Volby.....	233
Analýza ROC: Zobrazit.....	233
Analýza ROC: Definovat skupiny (řetězec).....	234
Analýza ROC: Definovat skupiny (číselné).....	234
ROC Křivky.....	234
Volby křivky ROC.....	235
Simulace.....	235
Chcete-li navrhnout simulaci na základě souboru modelu.....	236
Chcete-li navrhnout simulaci na základě vlastních rovnic.....	237
Chcete-li navrhnout simulaci bez prediktivního modelu.....	237
Chcete-li spustit simulaci ze simulačního plánu.....	238
Tvůrce simulace.....	238
Spustit dialogové okno simulace.....	248
Práce s výstupem grafu ze simulace.....	250
Geoprostorové modelování.....	251
Výběr map.....	252
Zdroje dat.....	253
Geoprostorová pravidla přidružení.....	255
Prostorová časová předpověď.....	258
Dokončit.....	261

<b>Upozornění.....</b>	<b>263</b>
Ochranné známky.....	264
<b>Rejstřík.....</b>	<b>267</b>





---

# Kapitola 1. Základní funkce

Do produktu IBM SPSS Statistics Base Edition jsou zahrnuty následující základní funkce.

---

## Analýza napájení

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Produkt IBM SPSS Statistics poskytuje následující postupy pro analýzu napájení:

### Jeden vzorek T-test

V jedné analýze vzorku se zjištěná data shromažďují jako jediný náhodný vzorek. Předpokládá se, že ukázková data nezávisle a identicky odpovídají normální distribuci s pevnou střední hodnotou a rozptylem a odvozuje statistické odvozování o střední hodnotu parametru.

### Zkušební vzorek T-test

V párování-ukázková analýza, zjištěná data obsahují dva spárované a souvztažné vzorky a každý případ má dvě měření. Předpokládá se, že data v každém vzorku nezávisle a stejně sledují normální rozdělení s pevně stanovenou střední hodnotou a rozptylem a odvozuje statistické inference o rozdílu mezi oběma prostředky.

### Nezávislý vzorek T-test

V nezávislé analýze vzorku obsahují zjištěná data dva nezávislé vzorky. Předpokládá se, že data v každém vzorku nezávisle a stejně sledují normální rozdělení s pevně stanovenou střední hodnotou a rozptylem a odvozuje statistické inference o rozdílu mezi oběma prostředky.

### Jednosměrný test ANOVA

Analýza rozptylu (ANOVA) je statistická metoda odhadu prostředků pro několik populací, u nichž se často předpokládá, že jsou obvykle rozdělovány. Jednofaktorová ANOVA, obecný typ ANOVA, je rozšířením dvouvzorového  $t$ -testu.

**Příklad.** Síla testu hypotézy ke zjištění správné alternativní hypotézy je pravděpodobnost, že test zamítne testovací hypotézu. Vzhledem k tomu, že pravděpodobnost chyby typu II je pravděpodobnost přijetí nulové hypotézy, když je alternativní hypotéza pravdivá, může být vyjádřen jako  $(1 - \text{pravděpodobnost chyby typu II})$ , což je pravděpodobnost zamítnutí nulové hypotézy, když je alternativní hypotéza pravdivá.

**Statistika a grafy.** Jednostranná zkouška, dvoustranný test, úroveň významnosti, typ I-četnost chyb, předpoklady zkoušek, směrodatná odchylna počtu obyvatel, průměrná velikost populace podle velikosti vzorku, dvourozměrná síla podle velikosti vzorku, dvoudimenzionální výkon podle velikosti vzorku, trojrozměrný výkon podle velikosti vzorku, normalizační koeficient rotace, skupiny párů, Pearsonův korelační koeficient, průměrný rozdíl, zákresový rozsah účinku velikosti, směrodatná odchylna základního souboru, kontrast a párové rozdíly, kontrastní koeficienty, kontrastní test, BONERRONNI, SIDAK, LSD, výkon podle celkové velikosti vzorku, dvoudimenzionální výkon podle souhrnné směrodatné odchylny, troj-dimenzionální výkon podle celkového vzorku, troj-dimenzionální výkon podle celkové velikosti vzorku, souhrnné směrodatné odchylny, Studentova  $t$ -distribuce, necentrální distribuce  $t$ -rozdělení,

## Aspekty dat analýzy napájení

### Data

V jedné analýze vzorku se zjištěná data shromažďují jako jediný náhodný vzorek.

V párování-ukázková analýza, zjištěná data obsahují dva spárované a souvztažné vzorky a každý případ má dvě měření.

V nezávislé analýze vzorku obsahují zjištěná data dva nezávislé vzorky.

Analýza rozptylu (ANOVA) je statistická metoda odhadu prostředků pro několik populací, u nichž se často předpokládá, že jsou obvykle rozdělovány.

### Předpoklady

V jedné ukázkové analýze se předpokládá, že ukázková data nezávisle a identicky odpovídají normální distribuci s pevnou střední hodnotou a rozptylem a odvozuje statistické odvozování o střední hodnotu parametru.

V párované analýze vzorku se předpokládá, že data v každém vzorku nezávisle a stejně sledují normální rozdělení s pevně stanovenou střední hodnotou a rozptylem a odvozuje statistické inference o rozdílu mezi oběma prostředky.

V nezávislé analýze vzorku se předpokládá, že data v každém vzorku nezávisle a stejně sledují normální rozdělení s pevně stanovenou střední hodnotou a rozptylem a odvozuje statistické inference o rozdílu mezi oběma prostředky.

V jednocestné ANOVA se často předpokládá, že statistická metoda odhadu prostředků pro několik populací je obvykle rozdělena.

### Získání analýzy napájení

1. Z nabídky vyberte:

**Analyzovat > Analýza napájení > Porovnat prostředky > jednovzorkový T test** nebo **Paired-Vzorový T-Test**, nebo **Independent-Sample T-Test**, nebo **One-way ANOVA**

2. Definujte požadované předpoklady testu.

3. Klepněte na tlačítko **OK**.

### Prostředky

V produktu IBM SPSS Statistics Base Edition jsou zahrnuty následující statistické funkce.

#### Power Analysis of One-Ukázka T Test

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

V jedné analýze vzorku se zjištěná data shromažďují jako jediný náhodný vzorek. Předpokládá se, že ukázková data nezávisle a identicky odpovídají normální distribuci s pevnou střední hodnotou a rozptylem a odvozuje statistické odvozování o střední hodnotu parametru.

1. Z nabídky vyberte:

**Analyzovat > Analýza napájení > Prostředky > jednovzorkový T test**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku** nebo **Výkon)**.

3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Napájení** jako metoda odhadu testu **Odhad**, zadejte příslušnou hodnotu **Velikost vzorku** pro hodnotu odhadu výkonu. Hodnota musí být celé číslo větší než 1.

5. Volitelně můžete vybrat volbu z pole **Určit**.

### Hypotézmaticizované hodnoty

Výchozí nastavení poskytuje nastavení **Střední hodnota naplnění** a **Hodnota null**.

### Populační průměr

Zadejte hodnotu, která určuje střední hodnotu populace pod testováním. Hodnota musí být jedna číselná hodnota.

### Hodnota null

Volitelně zadejte hodnotu, která uvádí hodnotu hypotézy s hodnotou null, která má být testována. Hodnota musí být jedna číselná hodnota.

### Velikost efektu

Cohen je  $f$  se používá k odhadu velikosti účinku jako vstup na odhad o výkonu nebo vzorku velikosti. Definovaná hodnota efektu **Hodnota** se předá do intermediačního kroku procedury a vypočítá požadovanou velikost nebo velikost vzorku.

6. Zadejte hodnotu **směrodatné odchytky naplnění**. Hodnota musí být jedna číselná hodnota větší než 0.
7. Vyberte, zda je test jedna nebo dvě strany.

### Nesměrová (dvoustranná) analýza

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

### Analýza směrová (jednostranná) analýza

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

8. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.
9. Volitelně můžete klepnout na tlačítko **Zakreslit** a zadat nastavení "Power Analysis of One-Ukázka T Test: Zakreslit" na stránce 3 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

10. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části "Analýza napájení: Přesnost" na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Velikost vzorku** jako metoda odhadu testu **Odhad**, **Hypotézmaticizované hodnoty** je vybrána ze seznamu **Zadat** a analýza **Nesměrová (oboustranná)** je vybrána jako **Směr testu**.

### *Power Analysis of One-Ukázka T Test: Zakreslit*

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a tří grafů s rozsahem výkonu podle vzorku a efektu velikosti grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

### Dvourozměrný zákres

#### Odhad výkonu ve srovnání s velikostí vzorku

Je-li tato volba povolena, toto volitelné nastavení poskytuje volby pro řízení dvourozměrného výkonu podle vzorkového grafu velikosti. Nastavení je ve výchozím nastavení vypnuto.

#### Rozsah velikosti vzorku

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu velikosti vzorku.

#### Dolní mez

Řídí dolní mez pro dvourozměrný výkon podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než 1 a nesmí být větší než hodnota **Horní mez**.

### **Horní mez**

Řídí horní mez dvourozměrného výkonu podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Odhad výkonu v porovnání s velikostí efektu**

Při výchozím nastavení je toto volitelné nastavení zakázáno. Je-li povolena tato volba, graf se zobrazí ve výstupu. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou určeny žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu použité velikosti efektu.

### **Rozsah platnosti rozsahu účinku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze.

### **Dolní mez**

Řídí dolní mez pro dvourozměrný výkon podle velikosti efektu grafu. Hodnota musí být větší než nebo rovna hodnotě -5.0 a nesmí být větší než hodnota **Horní mez**.

### **Horní mez**

Řídí horní mez pro dvourozměrný výkon podle velikosti efektu grafu. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5.0.

## **Trojrozměrný zákres**

### **Odhad výkonu versus**

Poskytuje volby pro řízení trojrozměrného výkonu podle velikosti vzorku (osa X) a velikosti efektu (osa Y), nastavení svislého/vodorovného otočení a rozsah zkusné plochy velikosti vzorku nebo efektu. Toto nastavení je při výchozím nastavení zakázáno.

### **Velikost efektu na ose x a velikost vzorku na ose Y**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa X) a velikosti efektu (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Velikost efektu na ose Y a velikost vzorku na ose X**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa Y) a velikosti efektu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovně stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

### **Rozsah velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu velikosti vzorku.

### **Dolní mez**

Řídí dolní mez pro trojrozměrný výkon podle vzorkového grafu velikosti. Hodnota musí být větší než 1 a nesmí být větší než hodnota **Horní mez**.

### **Horní mez**

Řídí horní mez pro trojrozměrný výkon podle vzorkového grafu velikosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Rozsah platnosti rozsahu účinku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

### **Dolní mez**

Řídí dolní mez pro trojrozměrný rozměr příkonu podle velikosti efektu. Hodnota musí být větší než nebo rovna hodnotě -5.0 a nesmí být větší než hodnota **Horní mez** .

### **Horní mez**

Řídí horní mez pro trojrozměrný rozměr příkonu podle velikosti efektu. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5.0.

## **Power Analysis of Paired-Samples T Test**

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

V analýze párovaných ukázek se sledovaná data obsahují ve dvou párovaných a korelovaných vzorcích a každý případ má dvě měření. Předpokládá se, že data v každém vzorku nezávisle a stejně sledují normální rozdělení s pevně stanovenou střední hodnotou a rozptylem a odvozuje statistické inference o rozdílu mezi oběma prostředky.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Prostředky > Paired-test T Test**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku** nebo **Výkon)**.

3. Je-li vybrána volba **Velikost vzorku** , zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Napájení** jako metoda odhadu testu **Odhad** , zadejte příslušnou hodnotu **Velikost vzorku** . Hodnota musí být kladné celé číslo větší než, nebo rovno, 2.

5. Volitelně můžete vybrat volbu z pole **Určit** .

### **Hypotézmaticované hodnoty**

Výchozí nastavení poskytuje nastavení **Střední hodnota** a **Směrodatná odchylka** .

#### **Rozdíl počtu obyvatel**

Je-li požadována jediná populace, zadejte hodnotu **Rozdíl počtu obyvatel** . Je-li zadána jednotlivá hodnota, označuje střední hodnotu počtu obyvatel  $\mu_d$ .

**Poznámka:** Hodnota nemůže být 0, je-li vybrána **Velikost vzorku** .

#### **Populace střední hodnoty pro skupinu 1 a skupinu 2**

Pokud je pro zadané dvojice skupin požadováno více populačních prostředků, zadejte hodnoty pro **Populační průměr pro skupinu 1** a **Populační střední hodnota pro skupinu 2**. Je-li uvedeno více hodnot, označují populační průměrný rozdíl  $\mu_1$  a  $\mu_2$ .

**Poznámka:** Tyto dvě hodnoty nemohou být stejné, je-li vybrána volba **Velikost vzorku** .

#### **Odchylka populační směrodatné odchylky pro průměrný rozdíl**

Je-li určena jediná populace, zadejte směrodatnou odchylku populace pro střední hodnotu rozdílu. Je-li zadána jedna hodnota, označuje směrodatnou odchylku počtu obyvatel ve skupinovém rozdílu  $\sigma_d$ . Hodnota musí být jedna číselná hodnota větší než 0.

#### **Směrodatná odchylka populace pro skupinu 1 a skupinu 2**

Je-li uvedeno více populačních prostředků, zadejte směrodatnou odchylku populace pro hodnoty skupiny 1 a skupiny 2. Je-li uvedeno více hodnot, označují směrodatnou odchylku skupiny rozdílů ve skupině  $\sigma_1$  a  $\sigma_2$ . Hodnoty musí být jednoduché číselné hodnoty větší než 0.

### **Pearsonův korelační koeficient**

Volitelně zadejte hodnotu, která uvádí korelační koeficient Pearson-moment  $\rho$ . Hodnota musí být jedna číselná hodnota v rozsahu -1 až 1. Hodnota nemůže být 0.

**Poznámka:** Je-li zadána hodnota **Standardní odchylka střední hodnoty populace pro střední hodnotu**, bude toto nastavení ignorováno. Jinak se hodnoty **Populační směrodatná odchylka pro skupinu 1 a skupinu 2** použijí k výpočtu  $\sigma_d$ .

### **Velikost efektu**

Odhaduje velikost efektu jako vstup pro odhad velikosti energie nebo vzorku. Definovaná hodnota efektu **Hodnota** se předá do intermediačního kroku procedury a vypočítá požadovanou velikost nebo velikost vzorku.

6. Vyberte, zda je test jedna nebo dvě strany.

### **Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

7. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.

8. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení “Power Analysis of Paired-Samples T Test: Plot” na stránce 6 (výstup grafu, nastavení dvourozměrného výkresu, nastavení trojrozměrného výkresu a nástrojové tipy).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

9. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části “Analýza napájení: Přesnost” na stránce 35.

**Poznámka:** Volba **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Odhad velikosti** jako metoda odhadu testu **Odhad** a **Hypotická hodnota** je vybrána ze seznamu **Určit**.

## **Power Analysis of Paired-Samples T Test: Plot**

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a tří grafů s rozsahem výkonu podle vzorku a efektu velikosti grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

### **Dvourozměrný zakres**

#### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba povolena, toto volitelné nastavení poskytuje volby pro řízení dvourozměrného výkonu podle vzorkového grafu velikosti. Nastavení je ve výchozím nastavení vypnuto.

#### **Rozsah velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu velikosti vzorku.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný výkon podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než 1 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez dvourozměrného výkonu podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

#### **Odhad výkonu v porovnání s velikostí efektu**

Při výchozím nastavení je toto volitelné nastavení zakázáno. Je-li povolena tato volba, graf se zobrazí ve výstupu. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou určeny žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu použité velikosti efektu.

### **Rozsah platnosti rozsahu účinku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný výkon podle velikosti efektu grafu. Hodnota musí být větší než nebo rovna hodnotě -5.0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro dvourozměrný výkon podle velikosti efektu grafu. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5.0.

### **Trojrozměrný zakres**

#### **Odhad výkonu versus**

Poskytuje volby pro řízení trojrozměrného výkonu podle velikosti vzorku (osa X) a velikosti efektu (osa Y), nastavení svislého/vodorovného otočení a rozsah zkusné plochy velikosti vzorku nebo efektu. Toto nastavení je při výchozím nastavení zakázáno.

#### **Velikost efektu na ose x a velikost vzorku na ose Y**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa X) a velikosti efektu (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **Velikost efektu na ose Y a velikost vzorku na ose X**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa Y) a velikosti efektu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zakres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

#### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zakres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

#### **Rozsah velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu velikosti vzorku.

#### **Dolní mez**

Řídí dolní mez pro trojrozměrný výkon podle vzorového grafu velikosti. Hodnota musí být větší než 1 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro trojrozměrný výkon podle vzorového grafu velikosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

#### **Rozsah platnosti rozsahu účinku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro trojrozměrný rozměr příkonu podle velikosti efektu. Hodnota musí být větší než nebo rovna hodnotě -5.0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro trojrozměrný rozměr příkonu podle velikosti efektu. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5.0.

## Analýza napájení nezávislého-ukázkovém testu T

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

V nezávislé analýze vzorků obsahují zjištěná data dva nezávislé vzorky. Předpokládá se, že data v každém vzorku nezávisle a stejně sledují normální rozdělení s pevně stanovenou střední hodnotou a rozptylem a odvozuje statistické inference o rozdílu mezi oběma prostředky.

1. Z nabídky vyberte:

**Analýzovat > Analýza napájení > Prostředky > Nezávisle-Test T**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku nebo Výkon)**.

3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části “Analýza napájení: Hodnoty mřížky” na stránce 36.

Zadejte hodnotu **Poměr velikosti skupiny** pro uvedení poměru velikosti vzorku (hodnota musí být jedna hodnota mezi 0.01 a 100).

4. Když je vybrána volba **Napájení** jako metoda odhadu testu **Odhad**, zadejte hodnoty pro určení velikosti vzorku pro **Velikost vzorku pro skupinu 1 a skupinu 2** pro porovnání. Hodnoty musí být celá čísla větší než 1.

5. Volitelně můžete vybrat volbu z pole **Určit**.

### Hypotézmatizované hodnoty

Výchozí nastavení poskytuje nastavení **Průměrný počet obyvatel a směrodatné odchytky naplnění**.

#### Rozdíl počtu obyvatel

Je-li určena jediná populace, zadejte hodnotu rozdílu populačního rozdílu. Je-li zadána jedna hodnota, označuje skupinu průměrného rozdílu  $\sigma_d$  skupiny. Hodnota musí být jedna číselná hodnota větší než 0.

#### Populace střední hodnoty pro skupinu 1 a skupinu 2

Je-li uvedeno více populačních prostředků, zadejte počet obyvatel pro skupinu 1 a hodnoty skupiny 2. Je-li uvedeno více hodnot, označují populační průměr z rozdílu skupiny  $\sigma_1$  a  $\sigma_2$ . Hodnoty musí být jednoduché číselné hodnoty větší než 0.

#### Počet směrodatných odchylek obyvatelstva

Uveďte, zda jsou směrodatné odchytky populace **Rovno pro dvě skupiny**, nebo **Nerovnají se dvěma skupinám**.

- Pokud jsou směrodatné odchytky populace rovny pro dvě skupiny, zadejte hodnotu pro **Směrodatná odchytky ve fondu**, která označuje  $\sigma_a$  předpokládá se, že jsou dvě rozptylové skupiny stejné, nebo  $\sigma_1 = \sigma_2 = \sigma$ .
- Pokud se směrodatné odchytky počtu obyvatel nerovnají pro dvě skupiny, zadejte hodnoty do pole **Směrodatná odchytky pro skupinu 1 a skupinu 2**, která označuje  $\sigma_1$  a  $\sigma_2$ .

**Poznámka:** Jsou-li hodnoty **Směrodatná odchytky pro skupinu 1 a skupinu 2** identické, zachází se s nimi jako s jedinou hodnotou.



### **Velikost efektu**

Odhaduje velikost efektu jako vstup pro odhad velikosti energie nebo vzorku. Definovaná hodnota efektu **Hodnota** se předá do intermediačního kroku procedury a vypočítá požadovanou velikost nebo velikost vzorku.

6. Vyberte, zda je test jedna nebo dvě strany.

### **Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

7. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.

8. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení “Analýza napájení nezávislého-ukázkovém T Test: Zakreslit” na stránce 9 (výstup grafu, nastavení dvourozměrného výkresu, nastavení trojrozměrného výkresu a nástrojové tipy).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

9. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části “Analýza napájení: Přesnost” na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Velikost vzorku** jako metoda odhadu testu **Odhad**, **Hypotézmaticizované hodnoty** je vybrána ze seznamu **Zadat** a analýza **Nesměrové (oboustranné)** je vybrána jako **Směr testu**.

### **Analýza napájení nezávislého-ukázkovém T Test: Zakreslit**

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvourozměrného a třírozměrného výkonu podle poměru vzorku, velikosti efektu nebo středních odlišených grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

### **Dvourozměrný zakres**

#### **Odhad výkonu oproti poměru velikosti vzorku/odhadu výkonu oproti velikosti vzorku**

Je-li tato volba povolena, toto volitelné nastavení poskytuje volby pro řízení dvourozměrného výkonu vzorkového grafu poměru velikosti vzorku. Další nastavení nejsou k dispozici v případě, že bylo zadáno více hodnot napájení (**Odhad výkonu oproti velikosti vzorku**). Nastavení je ve výchozím nastavení vypnuto.

#### **Rozsah poměru velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu velikosti vzorku.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný výkon podle poměru velikosti vzorku. Hodnota musí být mezi 0.01 a 100 a nemůže být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní hranici pro dvourozměrný výkon podle poměru velikosti vzorku. Hodnota musí být mezi 0.01 a 100 a musí být větší než hodnota **Dolní mez**.

#### **Odhad výkonu v porovnání s velikostí efektu (nebo průměrný rozdíl)**

Při výchozím nastavení je toto volitelné nastavení zakázáno. Je-li povolena tato volba, graf se zobrazí ve výstupu. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu pro velikost efektu (nebo střední rozdíl).

#### **Rozsah platnosti účinku (nebo průměrný rozdíl)**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze.

### **Dolní mez**

Řídí dolní mez pro dvourozměrný výkon podle velikosti efektu grafu. Hodnota musí být větší než nebo rovna hodnotě -5.0 a nesmí být větší než hodnota **Horní mez** .

### **Horní mez**

Řídí horní mez pro dvourozměrný výkon podle velikosti efektu grafu. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5.0.

## **Trojrozměrný zákres**

### **Odhad výkonu versus**

Poskytuje volby pro řízení trojrozměrného výkonu podle poměru velikosti vzorku (osa X) a velikosti efektu (osa Y), nastavení svislého/vodorovného otočení a rozsah zkusné plochy velikosti vzorku nebo efektu. Toto nastavení je při výchozím nastavení zakázáno.

### **Velikost efektu (nebo průměrný rozdíl) na ose x a velikost vzorku na ose Y**

Volitelné nastavení řídí poměr trojrozměrného výkonu podle poměru velikosti vzorku (osa X) a velikosti efektu (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Velikost efektu (nebo střední hodnota rozdílu) na ose y a velikosti vzorku na ose x**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa Y) a velikosti efektu (osa x). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

### **Rozsah poměru velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu velikosti vzorku.

### **Dolní mez**

Určuje dolní mez pro trojrozměrný výkon podle poměru velikosti vzorku. Hodnota musí být mezi 0.01 a 100 a nemůže být větší než hodnota **Horní mez** .

### **Horní mez**

Řídí horní mez pro třírozměrný výkon podle poměru velikosti vzorku velikosti vzorku. Hodnota musí být mezi 0.01 a 100 a musí být větší než hodnota **Dolní mez** .

### **Rozsah platnosti účinku (nebo průměrný rozdíl)**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

### **Dolní mez**

Řídí dolní mez pro trojrozměrný rozměr příkonu podle velikosti efektu. Hodnota musí být větší než nebo rovna hodnotě -5.0 a nesmí být větší než hodnota **Horní mez** .

### **Horní mez**

Řídí horní mez pro trojrozměrný rozměr příkonu podle velikosti efektu. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

## **Analýza napájení jednofaktorové ANOVA**

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Analýza rozptylu (ANOVA) je statistická metoda odhadu prostředků pro několik populací, u nichž se často předpokládá, že jsou obvykle rozdělovány. Jednofaktorová ANOVA, obecný typ ANOVA, je rozšířením dvouvzorového *t*-testu. Procedura poskytuje přístupy k odhadu výkonu dvou typů hypotéz, které porovnávají více skupin prostředků, celkový test a test se zadanými kontrasty. Test over se zaměřuje na nulovou hypotézu, že všechny skupinové prostředky jsou stejné. Test se zadanými kontrasty rozkládá celkové hypotézy ANOVA na menší, ale více popsateľné a užitečné kousky prostředků.

1. Z nabídky vyberte:

#### **Analyzovat > Analýza napájení > Prostředky > Jednofaktorová ANOVA**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku nebo Výkon)**.

3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Volitelně můžete vybrat volbu z pole **Určit**.

#### **Hypotézmatizované hodnoty**

Výchozí nastavení poskytuje nastavení **směrodatné odchytky populační populace**. Zadejte jednu číselnou hodnotu větší než 0. Výchozí hodnota je 1.

#### **Efekt velikosti: Cohen je f**

Cohen je *f* se používá k odhadu velikosti účinku jako vstup na odhad o výkonu nebo vzorku velikosti. Definovaná hodnota efektu **Hodnota** se předá do intermediačního kroku procedury a vypočítá požadovanou velikost nebo velikost vzorku.

#### **Efekt velikosti: Eta na druhou**

Eta-na druhou ( $\eta^2$ ) se používá k odhadu velikosti účinku jako vstupu pro odhad velikosti energie nebo vzorku. Definovaná hodnota efektu **Hodnota** se předá do intermediačního kroku procedury a vypočítá požadovanou velikost nebo velikost vzorku.

5. Je-li vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**, zadejte hodnoty **Velikosti skupiny a Skupina**. Musí být uvedena alespoň dvě hodnoty velikosti skupiny a každá z hodnot velikosti skupiny musí být větší nebo rovna 2 (ale nesmí překročit 2,147, 483, 647). Musí být také uvedena alespoň dvě střední hodnoty skupiny (počet uvedených hodnot musí být roven hodnotám velikosti skupiny).

Když je vybrána volba **Velikost vzorku** jako nastavení odhadu testu **Odhad**, zadejte hodnoty **Váhy skupin** (volitelné) a **Seskupit hodnoty**. Váhy skupin přiřazují váhy velikosti skupiny při výběru volby **Napájení**.

**Poznámka:** Nastavení **Váhy skupin** se ignorují, jsou-li uvedeny hodnoty **Velikosti skupiny**.

Klepněte na tlačítko **Přidat**, chcete-li zahrnout více hodnot **Velikosti skupiny, Váhy skupina Skupina**. Klepnutím na tlačítko **Odstranit** odeberete existující hodnoty.

6. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.

7. Volitelně můžete klepnutím na volbu **Kontrast** zadat nastavení [“Analýza napájení úložiště One-way ANOVA: Srovnejte”](#) na stránce 12 (kontrastní test, rozdíly po dvojicích, velikost efektů a odhad velikosti vzorku na základě intervalů spolehlivosti).

8. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení [“Analýza napájení úložiště One-way ANOVA: Zakreslit”](#) na stránce 13 (výstup grafu, nastavení dvourozměrného výkresu, nastavení trojrozměrného výkresu a nástrojové typy).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že jsou zadány hodnoty **Velikosti skupiny** a je vybrána hodnota **Napájení** jako nastavení **Odhad** .

9. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části [“Analýza napájení: Přesnost”](#) na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je jako metoda odhadu testu **Odhad** vybrána volba **Velikost vzorku** .

## **Analýza napájení úložiště One-way ANOVA: Srovnejte**

Dialogové okno **Kontrast** poskytuje volby pro určení kontrastu, koeficientu, velikosti efektu, párových rozdílů a nastavení intervalů spolehlivosti polovičního šířky pro analýzu Power analýzy jednofaktorové ANOVA.

### **Kontrastní test**

#### **Test podle lineárních kontrastů**

Je-li povolena tato volba, jsou k dispozici nastavení kontrastu a koeficientu.

#### **Směr testu**

##### **Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

##### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

#### **Koeficienty**

V tabulce určete kontrastní koeficienty a vyžádejte kontrastní test. Hodnoty tabulky jsou nepovinné. Počet uvedených hodnot se musí rovnat hodnotám uvedeným pro **Velikosti skupiny** a **Skupina znamená**. Součet pro všechny zadané hodnoty musí být roven 0, jinak se poslední hodnota automaticky upraví.

#### **Velikost efektu**

Určete velikost efektu kontrastní testu, který se měří hodnotou  $\delta$ . Je třeba určit jedinou číselnou hodnotu. Je-li **Směr testu** nastaven na **Směrová (jednostranná) analýza**, zadaná hodnota musí být  $\geq 0$ . Je-li jako předpoklad testu **Odhad** zadána hodnota **Power** , zadaná hodnota nemůže být 0.

### **Rozdíly po dvojicích**

#### **Odhad výkonu testování pro porovnávání po dvojicích**

Řídí, zda se má či nemá odhadnout výkon testování pro párové rozdíly. Standardně je volitelné nastavení vypnuto, což potlačí výstup pro párování rozdílů.

#### **Korekce Bonferroniho**

Využívá Bonferroniho korekce na odhad síly párových rozdílů. Toto je výchozí nastavení.

#### **Korekce dav**

Používá opravu Sidka při odhadování síly párových rozdílů.

#### **Nejméně významný rozdíl (LSD)**

Použije se oprava LSD při odhadování síly párových rozdílů.

### **Zadejte polovinu intervalu spolehlivosti**

Odhaduje velikost vzorku na základě hodnoty intervalu spolehlivosti half-width. Zadejte hodnotu v rozsahu 0 až 1. U jednovzorového testu Binomického testu musí být hodnota v rozsahu 0-0.5.

**Poznámka:** Duplicitní hodnoty jsou ignorovány.

- Klepnutím na tlačítko **Přidat** přidáte zadanou hodnotu polovičaté šířky do seznamu.
- Chcete-li aktualizovat hodnotu, zvýrazněte existující hodnotu polovičního šířky a klepněte na tlačítko **Změnit** .
- Chcete-li odebrat hodnotu ze seznamu, zvýrazněte existující hodnotu polovičního šířky a klepněte na tlačítko **Odebrat** .

## **Analyza napájení úložiště One-way ANOVA: Zakreslit**

Dialogové okno **Zakreslit** ovládá grafy, které jsou výstupem, aby ilustrovaly dvě a tři rozměrové hodnoty podle grafů velikosti vzorku a efektu. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

### **Dvourozměrný zakres**

#### **Odhad výkonu v porovnání s celkovou velikostí vzorku**

Je-li tato volba povolena, toto volitelné nastavení poskytuje volby pro řízení dvourozměrného výkonu podle celkové velikosti vzorku grafu. Nastavení je ve výchozím nastavení vypnuto.

#### **Rozsah celkové velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu celkové velikosti vzorku.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než nebo rovna:

- 2 x počet celých čísel zadaných pro **Velikosti skupiny**
- 2 x součet celých čísel zadaných pro **Velikosti skupiny** /nejmenší celočíselnou hodnotou pro **Velikosti skupin**

Hodnota nesmí být větší než hodnota **Horní mez** .

#### **Horní mez**

Řídí horní mez dvourozměrného výkonu podle celkové velikosti vzorku grafu. Hodnota musí být menší než nebo rovna:

- 5000/podle největší celočíselné hodnoty uvedené pro **Velikosti skupiny** x součet celých čísel zadaných pro **Velikosti skupin**

Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 2147483647.

#### **Odhad energie oproti souhrnné směrodatné odchylce**

Při výchozím nastavení je toto volitelné nastavení zakázáno. Nastavení řídí dvourozměrnou mocnost seskupená směrodatná odchylka grafu. Je-li povolena tato volba, graf se zobrazí ve výstupu. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového rozsahu směrodatné odchylky ve střední hodnotě.

#### **Poznámka:**

Zákrasový graf je deaktivován, když jsou uvedené hodnoty **skupiny znamená** všechny stejné.

#### **Rozsah směrodatné odchylky ve fondu**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrnou mocnost seskupená směrodatná odchylka grafu. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez** .

#### **Horní mez**

Řídí horní mez pro dvourozměrnou mocnost seskupená směrodatná odchylka grafu. Hodnota musí být větší než hodnota **Dolní mez** .

### **Trojrozměrný zakres**

#### **Odhad výkonu versus**

Poskytuje volby pro řízení trojrozměrného výkonu podle celkové velikosti vzorku (osa X) a velikosti efektu (osa Y), nastavení svislého a vodorovného otočení a uživatelem zadané oblasti kreslení rozsahu vzorku a efektu. Toto nastavení je při výchozím nastavení zakázáno.

#### **Poznámka:**

Zákrasový graf je deaktivován, když jsou uvedené hodnoty **skupiny znamená** všechny stejné.

### **Směrodatná odchylka ve střední velikosti na ose x a celková velikost vzorku na ose Y**

Volitelné nastavení řídí trojrozměrnou mocnost podle celkové velikosti vzorku (osa Y) a souhrnné směrodatné odchylky (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Směrodatná odchylka ve fondu na ose y a celková velikost vzorku na ose x**

Volitelné nastavení řídí trojrozměrnou mocnost podle celkové velikosti vzorku (osa X) a souhrnné směrodatné odchylky (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Rozsah celkové velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu velikosti vzorku.

#### **Dolní mez**

Řídí dolní mez pro trojrozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro trojrozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než hodnota **Dolní mez**.

### **Rozsah směrodatné odchylky ve fondu**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro trojrozměrnou mocnost podle smíšeného směrodatného grafu odchylky. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro trojrozměrnou mocnost podle smíšeného směrodatného grafu odchylky. Hodnota musí být větší než hodnota **Dolní mez**.

### **Odhad výkonu ve srovnání s velikostí vzorku**

Poskytuje volby pro řízení trojrozměrného rozměrového výkonu podle velikosti efektu a uživatelem specifikovaného rozsahu vzorku a velikosti efektu. Toto nastavení je při výchozím nastavení zakázáno.

#### **Poznámka:**

Zákresový graf je deaktivován, když jsou uvedené hodnoty **skupiny znamená** všechny stejné.

### **Směrodatná odchylka ve střední velikosti na ose x a celková velikost vzorku na ose Y**

Volitelné nastavení řídí trojrozměrný výkon podle celkové velikosti vzorku velikosti vzorku (osa X) a velikosti efektu (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Směrodatná odchylka ve fondu na ose y a celková velikost vzorku na ose x**

Volitelné nastavení řídí trojrozměrný výkon podle celkové velikosti vzorku velikosti vzorku (osa Y) a velikosti efektu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Rozsah celkové velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Nejsou-li pro pole **Dolní mez** nebo **Horní mez** zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu velikosti vzorku.

#### **Dolní mez**

Řídí dolní mez pro trojrozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro trojrozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než hodnota **Dolní mez**.

### **Rozsah platnosti rozsahu účinku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro trojrozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro trojrozměrný výkon podle celkové velikosti vzorku grafu. Hodnota musí být větší než hodnota **Dolní mez**.

### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

## **rozměry**

V produktu IBM SPSS Statistics Base Edition jsou zahrnuty následující statistické funkce.

### **Power Analysis of One-Ukázka Binomial Test**

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Binomická distribuce je založena na pořadí Bernoulliho pokusy. Lze ji použít k modelování experimentů, včetně pevně stanoveného počtu celkových pokusů, které jsou považovány za nezávislé na sobě navzájem. Každý proces vede k dichotomnímu výsledku, se stejnou pravděpodobností pro úspěšný výsledek.

Jeden vzorek binomický test provádí statistické inferenze vůči parametru proporcionálně porovnáním s hypotetizovanou hodnotou. Metody pro odhad výkonu takové zkoušky jsou buď normální aproximace, nebo binomický výčet.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Proporce > jednovzorkový binární test**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku** nebo **Výkon)**.
3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**, zadejte příslušnou hodnotu **Celkový počet pokusů**. Hodnota musí být celé číslo větší než, nebo rovno, 1.

5. Zadejte hodnotu, která uvádí alternativní hodnotu hypotézy podílu parametru v poli **Podíl obyvatelstva** . Hodnota musí být jedna číselná hodnota.

**Poznámka:** Je-li zadána hodnota **Power** , hodnota **Podíl obyvatelstva** nemůže být rovna hodnotě **Nulová hodnota**.

6. Volitelně zadejte hodnotu, která určuje hodnotu nulové hypotézy parametru proporce, která má být testována, do pole **Hodnota null** . Hodnota musí být jedna číselná hodnota v rozsahu 0-1. Výchozí hodnota je 0,50.
7. Vyberte metodu odhadu napájení.

#### **Normální přibližování**

Umožňuje normální přibližování. Toto je výchozí nastavení.

#### **Provést opravu kontinuity**

Řídí, zda se pro normální aproximaci použije náprava kontinuity.

#### **Binomické výčet**

Povolí metodu binomického výčtového typu. Volitelně můžete do pole **Časový limit** zadat maximální počet minut, který může odhad velikosti vzorku odhadnout. Když je dosažen časový limit, analýza se ukončí a zobrazí se varovná zpráva. Je-li uvedeno, hodnota musí být jediné kladné celé číslo, které označuje počet minut. Výchozí nastavení je 5 minut.

**Poznámka:** Vybraný předpoklad odhadu napájení nebude mít žádný vliv, pokud hodnota **Celkový počet pokusů** překročí 500.

8. Vyberte, zda je test jedna nebo dvě strany.

#### **Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

#### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

9. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti** . Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.
10. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení “Power Analysis of One-Sample Binomické: Plot” na stránce 16 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Power** jako předpoklad testu **Odhad** a volba **Binomický výčet** není vybrána.

11. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části “Analýza napájení: Přesnost” na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Ukázková velikost** jako metoda odhadu testu **Odhad** a jako **Směr testuje** vybrána analýza **Nesměrové (dvoustranná)** .

### **Power Analysis of One-Sample Binomické: Plot**

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a tří dimenzí výkonu podle grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

#### **Dvourozměrný zakres**

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Toto nastavení je při výchozím nastavení zakázáno.

#### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon hodnotového grafu s hodnotou null. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.



### **Odhad výkonu versus alternativní hodnota Hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon v rámci grafu alternativní hodnoty. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad výkonu oproti rozdílu mezi hypotézovanými hodnotami**

Je-li tato volba povolena, toto volitelné nastavení řídí dvoudimenzionální výkon podle rozdílu mezi hypotetickým měřicím grafem. Nastavení je ve výchozím nastavení vypnuto.

### **Odhad energie oproti celkovému počtu pokusů**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrnou mocnost podle celkového počtu zkušebních grafů. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Rozsah zákresu celkového počtu pokusů**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní hranici pro dvourozměrný odhad výkonu oproti celkovému počtu zkušebních grafů. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro dvourozměrný odhad energie oproti celkovému počtu zkušebních grafů. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Trojrozměrný zákres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy. Toto nastavení je při výchozím nastavení zakázáno.

### **Odhad energie oproti celkovému počtu pokusů**

Je-li tato volba vybrána, je možné povolit následující volby.

#### **na ose x a rozdíl mezi hypotetizovanými hodnotami na ose Y**

Volitelné nastavení řídí tři rozměrová mocnost podle celkového počtu pokusů (osa X) a rozdílu mezi hodnotami hypotetické hodnoty (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **o ose Y a rozdílu mezi hypotetizovanými hodnotami na ose X**

Volitelné nastavení řídí tři rozměrová mocnost podle celkového počtu pokusů (osa Y) a rozdílu mezi hodnotami hypotetické hodnoty (osa x). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **Rozsah zákresu celkového počtu pokusů**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro odhad trojrozměrného výkonu oproti celkovému počtu zkušebních grafů. Hodnota musí být větší než 0 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro odhad trojrozměrného výkonu oproti celkovému počtu zkušebních grafů. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba vybrána, je možné povolit následující volby.

#### **na ose X a alternativní hodnota hypotézy na ose Y**

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa X) a v grafu alternativní hodnota (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **o ose Y a alternativní hypotéze na ose x**

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa Y) a alternativní hodnotu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### Svislé otočení

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný záznam. Hodnota musí být jedna celočíselná hodnota menší nebo rovná hodnotě 359. Výchozí hodnota je 10.

### Vodorovné otočení

Volitelné nastavení nastaví vodorovně stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný záznam. Hodnota musí být jedna celočíselná hodnota menší nebo rovná hodnotě 359. Výchozí hodnota je 325.

## Analýza Power souvisejících-Ukázka Binomial Test

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Binomická distribuce je založena na pořadí Bernoulliho pokusy. Lze ji použít k modelování experimentů, včetně pevně stanoveného počtu celkových pokusů, které jsou považovány za nezávislé na sobě navzájem. Každý proces vede k dichotomnímu výsledku, se stejnou pravděpodobností pro úspěšný výsledek.

Související-vzorek binomial odhaduje sílu McNemarova testu porovnat dva proporční parametry založené na porovnaných pár předmětů, které byly odebrány ze dvou souvisejících binomických populací.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Proporce > Související-Ukázka binomického testu**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku** nebo **Výkon**).

3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Napájení** jako metoda odhadu testu **Odhad**, zadejte příslušnou hodnotu **Celkový počet dvojic**.

5. Tuto volbu vyberte, chcete-li určit testovací hodnoty pro **proportions** nebo **counts**.

- Je-li vybrána volba **proporce**, zadejte hodnoty do polí **Proportion 1** a **Proportion 2**. Hodnoty musí být v rozsahu 0 až 1.

- Je-li vybrána volba **počet**, zadejte hodnoty do polí **Počet 1** a **Počet 2**. Hodnoty musí být mezi 0 a hodnotou uvedenou pro **Celkový počet dvojic**.

### Poznámky k porcím:

- Volba **Proporce** je jediná dostupná volba, je-li zadána hodnota **Power**.
- Není-li volba **Testovací hodnoty marginální** vybrána:  $0 < \text{Podíl 1} + \text{Podíl 2} \leq 1$
- Je-li vybrána volba **Testovací hodnoty jsou okrajové**:
  - **Proportion 1 \* Proportion 2** > 0
  - **Podíl 1** < 1
  - **Proporce 2** < 1
  - Hodnoty pro **Proportion 1** a **Proportion 2** nemohou být stejné.

### Počítá poznámky:

- Nastavení **počtů** jsou k dispozici pouze v případě, že je vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**.
  - Není-li vybrána volba **Testovací hodnoty nejsou okrajové** :  $0 < \text{Počet 1} + \text{Počet 2} \leq \text{Celkový počet dvojic}$
  - Je-li vybrána volba **Testovací hodnoty jsou okrajové** :
    - **Počet 1 \* Počet 2** > 0
    - **Počet 1** < **Celkový počet dvojic**
    - **Počet 2** < **Celkový počet dvojic**
6. Volitelně můžete vybrat volbu **Testovací hodnoty jsou mezní**, chcete-li určit, zda mají být zadané proporce nebo hodnoty počtu okrajové. Je-li povolena volba **Testovací hodnoty jsou okrajové**, musíte zadat hodnotu **Korelace mezi dvojicemi dvojic**. Hodnota musí být jednotlivá hodnota mezi -1 a 1.
7. Vyberte metodu odhadu napájení.

### Normální přibližování

Umožňuje normální přibližování. Toto je výchozí nastavení.

### Binomické výčet

Povolí metodu binomického výčtového typu. Volitelně můžete do pole **Časový limit** zadat maximální počet minut, který může odhad velikosti vzorku odhadnout. Když je dosažen časový limit, analýza se ukončí a zobrazí se varovná zpráva. Je-li uvedeno, hodnota musí být jediné kladné celé číslo, které označuje počet minut. Výchozí nastavení je 5 minut.

8. Vyberte, zda je test jedna nebo dvě strany.

### Nesměrová (dvoustranná) analýza

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

### Analýza směrová (jednostranná) analýza

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

9. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.
10. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení “Analýza Power souvisejících-Ukázka Binomial Test” na stránce 18 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Power** jako předpoklad testu **Odhad** a volba **Binomický výčet** není vybrána.

11. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části “Analýza napájení: Přesnost” na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Ukázková velikost** jako metoda odhadu testu **Odhad** a jako **Směr testu** je vybrána analýza **Nesměrové (dvoustranná)**.

### **Power Analysis of Related-Sample Binomial: Plot**

Dialogové okno **Zakreslit** ovládá grafy, které jsou výstupem, aby ilustrovaly grafy s dvěma a třemi dimenzemi s dimenzionálním výkonem. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

### Dvourozměrný zakres

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Toto nastavení je při výchozím nastavení zakázáno.

### Odhad výkonu versus celkový počet párů

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrnou mocnost podle celkového počtu dvojic. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Rozsah zákresu celkového počtu párů**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní hranici pro dvourozměrný odhad spotřeby energie oproti celkovému počtu dvojic. Hodnota musí být větší než 1 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro dvourozměrný odhad výkonu a celkový počet dvojic graf. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 2500.

### **Odhad výkonu versus rozdíl rizik**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrnou mocnost podle grafu rozdílu rizik. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad spotřeby energie a poměr rizik**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný graf poměru rizika a rizik. Nastavení je ve výchozím nastavení vypnuto.

#### **Zákresový rozsah poměru rizika**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní hranici pro dvourozměrný graf míry rizika a poměru rizika. Hodnota nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní hranici pro dvourozměrný graf míry rizika a poměru rizika. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 10.

### **Odhad výkonu versus poměr šancí**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný poměr výkonu podle poměru pravděpodobnosti. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Poměr nákresu poměru pravděpodobnosti**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 10.

### **Odhad výkonu oproti korelaci mezi odpovídajícími dvojicemi**

Řídí odhadovanou sílu napětí a korelace mezi porovnávanými párovými grafy. Graf je vytvořen pouze v případě, že jsou určeny mezní hodnoty nebo počty (spíše než proporce rozložení dat).

### **Trojrozměrný zákres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy. Toto nastavení je při výchozím nastavení zakázáno.

### **Odhad výkonu versus proporce s diskordantem**

Je-li tato volba povolena, toto volitelné nastavení řídí trojrozměrnou moc podle proporcí dimenzí diskordant. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad výkonu oproti okrajových proporcích**

Je-li tato volba povolena, toto volitelné nastavení řídí trojrozměrný výkon podle okrajových proporcí grafu. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

**Poznámka:** Toto nastavení je k dispozici pouze v případě, že je vybrána volba **Testovací hodnoty jsou okrajové**.

### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

## **Analýza Power of Independent-Ukázka Binomial Test**

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Binomická distribuce je založena na pořadí Bernoulliho testy. Lze ji použít k modelování těchto experimentů včetně pevně stanoveného počtu celkových pokusů, které jsou považovány za nezávislé na sobě navzájem. Každý proces vede k dichotomnímu výsledku, se stejnou pravděpodobností pro "úspěšný" výsledek. Nezávislý ukázkový test binomial porovnává dva nezávislé parametry proporce.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Proporce > Binomický test Independent-Samples Binomial**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku nebo Výkon)**.

3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části "[Analýza napájení: Hodnoty mřížky](#)" na stránce 36.

Volitelně můžete zadat hodnotu **Poměr velikosti skupiny**. Výchozí hodnota je 1.

4. Je-li vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**, zadejte hodnoty pro určení **Celkový počet pokusů pro skupinu 1 a skupinu 2**. Hodnoty musí být celá čísla větší než 1.

5. Uveďte parametry poměru pro dvě skupiny. Obě hodnoty musí být v rozsahu 0-1.

**Poznámka:** Tyto dvě hodnoty nemohou být stejné, je-li uvedena hodnota **Power**.

6. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.

7. Vyberte požadovanou testovací metodu:

#### **Test chí-kvadrát**

Odhaduje sílu na základě Pearsonova chí-kvadrát test. Toto je výchozí nastavení.

### **Směrodatná odchylka se sdruž**

Toto volitelné nastavení kontroluje, zda je odhad směrodatné odchylky sloučený nebo nefond. Nastavení je ve výchozím nastavení povoleno.

### **Provést opravu kontinuity**

Toto volitelné nastavení určuje, zda je použita oprava kontinuity. Nastavení je ve výchozím nastavení vypnuto.

### **T-test**

Odhaduje sílu na základě Studentova t-testu.

### **Směrodatná odchylka se sdruž**

Toto volitelné nastavení kontroluje, zda je odhad směrodatné odchylky sloučený nebo nefond. Nastavení je ve výchozím nastavení povoleno.

### **Zkouška poměru pravděpodobnosti**

Odhaduje výkon na základě testu poměru pravděpodobnosti.

### **Fisherův přesný test**

Odhaduje sílu na základě Fisher je přesný test.

### **Poznámky:**

- V některých případech, Fisher přesný test může trvat delší dobu k dokončení.
- Všechny zkusné plochy jsou zablokovány, když je vybrán Fisher přesný test.

8. Vyberte metodu odhadu napájení.

### **Normální přibližování**

Umožňuje normální přibližování. Toto je výchozí nastavení.

### **Binomické výčet**

Povolí metodu binomického výčtového typu. Volitelně můžete do pole **Časový limit** zadat maximální počet minut, který může odhad velikosti vzorku odhadnout. Když je dosažen časový limit, analýza se ukončí a zobrazí se varovná zpráva. Je-li uvedeno, hodnota musí být jediné kladné celé číslo, které označuje počet minut. Výchozí nastavení je 5 minut.

9. Vyberte, zda je test jedna nebo dvě strany.

### **Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

10. Volitelně můžete klepnout na tlačítko **Zakreslit** a zadat nastavení “Power Analysis of Independent-Samples Binomial Test: Plot” na stránce 22 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Power** jako předpoklad testu **Odhad** a volba **Binomický výčet** není vybrána.

11. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části “Analýza napájení: Přesnost” na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Ukázková velikost** jako metoda odhadu testu **Odhad** a jako **Směr testuje** vybrána analýza **Nesměrové (dvoustranná)**.

### **Power Analysis of Independent-Samples Binomial Test: Plot**

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a třírozměrných grafů odhadu výkonu. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

### **Dvourozměrný zakres**

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Toto nastavení je při výchozím nastavení zakázáno.

### **Odhad výkonu versus poměr velikosti skupiny/Odhad výkonu oproti velikosti skupiny**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon podle poměru velikosti skupiny. Je-li tato volba vybrána, zobrazí se graf. Další nastavení nejsou k dispozici, je-li uvedeno více hodnot napájení (**Odhad výkonu oproti velikosti skupiny**). Nastavení je ve výchozím nastavení vypnuto.

#### **Rozsah zákresové oblasti poměru velikosti skupiny**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

##### **Dolní mez**

Řídí dolní hranici pro dvourozměrný odhad spotřeby energie oproti celkovému počtu dvojic. Hodnota musí být větší než 0,01 a nesmí být větší než hodnota **Horní mez**.

##### **Horní mez**

Řídí horní mez pro dvourozměrný odhad výkonu a celkový počet dvojic graf. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 100.

### **Odhad výkonu versus rozdíl rizik**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrnou mocnost podle grafu rozdílu rizik. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad spotřeby energie a poměr rizik**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný graf poměru rizika a rizik. Nastavení je ve výchozím nastavení vypnuto.

#### **Zákresový rozsah poměru rizika**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

##### **Dolní mez**

Řídí dolní hranici pro dvourozměrný graf míry rizika a poměru rizika. Hodnota nesmí být větší než hodnota **Horní mez**.

##### **Horní mez**

Řídí horní hranici pro dvourozměrný graf míry rizika a poměru rizika. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 10.

### **Odhad výkonu versus poměr šancí**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný poměr výkonu podle poměru pravděpodobnosti. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Poměr nákresu poměru pravděpodobnosti**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

##### **Dolní mez**

Řídí dolní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota nesmí být větší než hodnota **Horní mez**.

##### **Horní mez**

Řídí horní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 10.

### **Trojrozměrný zákres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy. Toto nastavení je při výchozím nastavení zakázáno.

### **Odhad výkonu oproti proporcím**

Je-li toto volitelné nastavení vybráno, poskytuje následující možnosti poměru výkonu:

### **Část skupiny 1 na ose x a podíl skupiny 2 na ose Y**

Řídí trojrozměrnou moc podle podílu skupiny 1 (osa X) a podílu grafu skupiny 2 (osa Y). Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Část skupiny 1 na ose Y a část skupiny 2 na ose X**

Řídí trojrozměrnou moc podle podílu skupiny 2 (osa X) a podílu grafu skupiny 1 (osa Y). Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad výkonu oproti velikostem skupin**

Je-li toto volitelné nastavení vybráno, poskytuje následující volby velikosti skupiny:

#### **velikost skupiny 1 na ose x a velikost skupiny 2 na ose Y**

Řídí trojrozměrnou mocnost podle počtu pokusů ve skupině 1 (osa X) a počtu pokusů v grafu Group 2 (osa Y). Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **velikost skupiny 1 na ose Y a velikost skupiny 2 na ose x**

Řídí trojrozměrnou mocnost podle počtu pokusů ve skupině 2 (osa X) a počtu pokusů ve skupině 1 (osa Y). Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Rozsah zákresu velikosti skupiny 1**

Je-li vybrána tato volba, jsou k dispozici volby dolní a horní meze pro rozsah zákresového grafu skupiny 1. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota musí být větší než nebo rovna 2 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 2500.

### **Rozsah zákresu velikosti skupiny 2**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze pro zákresový rozsah skupiny 2. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota musí být větší než nebo rovna 2 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní hranici pro dvourozměrný graf odhadu spotřeby energie a poměru pravděpodobnosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 2500.

### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.



## Korelace

V produktu IBM SPSS Statistics Base Edition jsou zahrnuty následující statistické funkce.

### Power Analysis of One-Sample Pearson Correlation Test

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Pearson 's product-moment korelační koeficient měří sílu lineárního spojení mezi dvěma náhodných proměnných stupnice, které se předpokládá, že následovat bivariate normální distribuci. Konventu se jedná o bezrozměrné množství a získané standardizací kovariance mezi dvěma spjitými proměnnými, čímž se pohybuje mezi -1 a 1.

Test používá Fisher je asymptotická metoda k odhadu moci pro jednovzorku Pearson korelace.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Korelace > Pearson Product-Moment**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku nebo Výkon)**.
3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**, zadejte příslušnou hodnotu **Velikost vzorku v párech**. Hodnota musí být jediné celé číslo větší než 3.
5. Zadejte hodnotu, která určuje alternativní hodnotu hypotetické hodnoty korelačního parametru v poli **Pearsonův korelační parametr**. Hodnota musí být jednoduché číslo v rozsahu -1 až 1.

**Poznámka:** Je-li zadána hodnota **Napájení**, hodnota **Pearsonova korelačního parametru** nemůže být -1 nebo 1 a nemůže být rovna hodnotě **Nulová hodnota**.

6. Volitelně zadejte hodnotu, která uvádí hodnotu nulové hypotézy korelačního parametru, která má být testována, do pole **Hodnota null**. Hodnota musí být jednoduché číslo v rozsahu -1 až 1. Výchozí hodnota je 0.

**Poznámka:** Je-li zadána hodnota **Power**, hodnota **Hodnota Null** nesmí být -1 nebo 1.

7. Volitelně můžete vybrat volbu **Použit pro odhad napájení formalitu pro opravu zkreslení**, abyste určili, zda je úprava zkreslení zahrnuta nebo ignorována. Nastavení je ve výchozím nastavení povoleno, což zahrnuje termín úpravy zkreslení v odhadu výkonu. Není-li toto nastavení vybráno, je výraz zarovnání předpojatosti ignorován.

8. Vyberte, zda je test jedna nebo dvě strany.

**Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

**Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

9. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.
10. Volitelně můžete klepnout na tlačítko **Zakreslit** a zadat nastavení [“Power Analysis of One-Sample Pearson Correlation: Plot”](#) na stránce 26 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

11. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části [“Analýza napájení: Přesnost”](#) na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Ukázková velikost** jako metoda odhadu testu **Odhad** a jako **Směr testuje** vybrána analýza **Nesměřové (dvoustranná)**.

### **Power Analysis of One-Sample Pearson Correlation: Plot**

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a tří dimenzí výkonu podle grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

#### **Dvourozměrný zakres**

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Nastavení je standardně vypnuto.

##### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon hodnotového grafu s hodnotou null. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

##### **Odhad výkonu versus alternativní hodnota Hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon v rámci grafu alternativní hodnoty. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

##### **Odhad výkonu oproti rozdílu mezi hypotézovanými hodnotami**

Je-li tato volba povolena, toto volitelné nastavení řídí dvoudimenzionální výkon podle rozdílu mezi hypotetickým měřicím grafem. Nastavení je ve výchozím nastavení vypnuto.

##### **Odhad výkonu oproti velikosti vzorku (v párech)**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon podle vzorkového grafu velikosti vzorku. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

##### **Rozsah vzorku velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu.

##### **Dolní mez**

Řídí dolní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

##### **Horní mez**

Řídí horní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

#### **Trojrozměrný zakres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy. Toto nastavení je při výchozím nastavení zakázáno.

##### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba vybrána, je možné povolit následující volby.

##### **na ose x a rozdíl mezi hypotetizovanými hodnotami na ose Y**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa X) a rozdílu mezi hodnotami hypotetizovaných hodnot (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **o ose Y a rozdílu mezi hypotetizovanými hodnotami na ose X**

Volitelné nastavení určuje trojrozměrnou energii podle velikosti vzorku (osa Y) a rozdílu mezi hodnotami hypotetické hodnoty (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf

### **Rozsah vzorku velikosti vzorku (v párech)**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro odhad trojrozměrného výkonu podle vzorového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro odhad vzorkovací síly podle vzorového grafu velikosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba vybrána, je možné povolit následující volby.

#### **na ose X a alternativní hodnota hypotézy na ose Y**

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa X) a v grafu alternativní hodnota (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **o ose Y a alternativní hypotéze na ose x**

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa Y) a alternativní hodnotu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

## **Power Analysis of One-Sample Spearman Correlation Test**

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Spearman-order correlation coefficient is a rank-based nonparametric statistic to measure the monotonic relationship between two variables that are usually censored and not normally distributed. The Spearman rank-order correlation is equal to the Pearson correlation between the rank values of the two variables, thereand also ranging between -1 and 1. Detekování síly korelačního testu Spearman je důležitým tématem v analýze dat řady hydrologických časových řad.

Test používá Fisher je asymptotická metoda k odhadu moci pro jednu ukázkou Spearman pořadí-order correlation.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Korelace > Spearman Rank-Pořadí**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku nebo Výkon)**.

3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**, zadejte příslušnou hodnotu **Velikost vzorku v párech**. Hodnota musí být jediné celé číslo větší než 3.
5. Zadejte hodnotu, která uvádí alternativní hodnotu hypotetické hodnoty korelačního parametru v poli **Parametr korelace Spearman**. Hodnota musí být jednoduché číslo v rozsahu -1 až 1.

**Poznámka:** Je-li zadána hodnota **Napájení**, hodnota **Parametr korelace Spearman** nemůže být -1 nebo 1 a nemůže být rovna hodnotě **Hodnota Null**.

6. Volitelně zadejte hodnotu, která uvádí hodnotu nulové hypotézy korelačního parametru, která má být testována, do pole **Hodnota null**. Hodnota musí být jednoduché číslo v rozsahu -1 až 1. Výchozí hodnota je 0.

**Poznámka:** Je-li zadána hodnota **Power**, hodnota **Hodnota Null** nesmí být -1 nebo 1.

7. Volitelně můžete vybrat volbu určující, jak se odhaduje, jak bude asymptotická odchylka odhadnuta na analýzu výkonu.

#### **Bonett a Wright**

Odhaduje odchylku navrženou společností Bonett a Wright. Toto je výchozí nastavení.

#### **Fieller, Hartley a Pearson**

Odhaduje se rozptyl navržený Fieller, Hartley a Pearson.

#### **Caruso a Cliff**

Odhaduje odchylku navrženou Caruso a Cliff.

8. Vyberte, zda je test jedna nebo dvě strany.

#### **Nesměrová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

#### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

9. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.
10. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení [“Power Analysis of One-Sample Spearman Correlation: Plot”](#) na stránce 28 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

11. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části [“Analýza napájení: Přesnost”](#) na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Ukázková velikost** jako metoda odhadu testu **Odhad** a jako **Směr testu** je vybrána analýza **Nesměrové (dvoustranná)**.

### **Power Analysis of One-Sample Spearman Correlation: Plot**

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a tří dimenzí výkonu podle grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

#### **Dvourozměrný zakres**

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Nastavení je standardně vypnuto.

### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon hodnotového grafu s hodnotou null. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad výkonu versus alternativní hodnota Hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon v rámci grafu alternativní hodnoty. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### **Odhad výkonu oproti rozdílu mezi hypotézovanými hodnotami**

Je-li tato volba povolena, toto volitelné nastavení řídí dvoudimenzionální výkon podle rozdílu mezi hypotetickým měřicím grafem. Nastavení je ve výchozím nastavení vypnuto.

### **Odhad výkonu oproti velikosti vzorku (v párech)**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon podle vzorkového grafu velikosti vzorku. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Rozsah vzorku velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Trojrozměrný zákres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy. Toto nastavení je při výchozím nastavení zakázáno.

### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba vybrána, je možné povolit následující volby.

#### **na ose x a rozdíl mezi hypotetizovanými hodnotami na ose Y**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa X) a rozdílu mezi hodnotami hypotetizovaných hodnot (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **o ose Y a rozdílu mezi hypotetizovanými hodnotami na ose X**

Volitelné nastavení určuje trojrozměrnou energii podle velikosti vzorku (osa Y) a rozdílu mezi hodnotami hypotetické hodnoty (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **Rozsah vzorku velikosti vzorku (v párech)**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro odhad trojrozměrného výkonu podle vzorového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro odhad vzorkovací síly podle vzorového grafu velikosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba vybrána, je možné povolit následující volby.

### na ose X a alternativní hodnota hypotézy na ose Y

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa X) a v grafu alternativní hodnota (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### o ose Y a alternativní hypotéze na ose x

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa Y) a alternativní hodnotu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

### Svislé otočení

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zakres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### Vodorovné otočení

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zakres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

## Analýza napájení částečného Pearsonova korelačního testu

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.

Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Částečné korelace lze vysvětlit přidružením mezi dvěma náhodnými proměnnými po vyloučení efektu jiného nebo několika dalších proměnných. Je to užitečné měření v přítomnosti zmatení. Podobně jako u Pearsonova korelačního koeficientu je dílčí korelační koeficient také bezrozměrným množstvím v rozsahu -1 až 1.

Test používá Fisher je asymptotická metoda k odhadu moci pro jednovzorku Pearson korelace.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Korelace > Dílčí**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku** nebo **Výkon)**.
3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části "Analýza napájení: Hodnoty mřížky" na stránce 36.

4. Je-li vybrána volba **Napájení** jako nastavení odhadu testu **Odhad**, zadejte příslušnou hodnotu **Velikost vzorku**. Hodnota musí být jediné celé číslo větší než 1.
5. Zadejte hodnotu, která uvádí **Number of the variables assumed to be partialled out**. Hodnota musí být celé číslo větší než nebo rovné 0.
6. Zadejte hodnotu, která uvádí alternativní hodnotu hypotézy **Dílčí korelační parametr**. Hodnota musí být jednoduché číslo v rozsahu -1 až 1.

**Poznámka:** Je-li zadána hodnota **Napájení**, hodnota parametru **Dílčí korelační parametr** nemůže být -1 nebo 1 a nemůže být rovna hodnotě **Hodnota Null**.

7. Volitelně zadejte hodnotu, která uvádí hodnotu nulové hypotézy dílčího korelačního parametru, která má být testována, do pole **Hodnota null**. Hodnota musí být jednoduché číslo v rozsahu -1 až 1. Výchozí hodnota je 0.

**Poznámka:** Je-li zadána hodnota **Power**, hodnota **Hodnota Null** nesmí být -1 nebo 1.

8. Vyberte, zda je test jedna nebo dvě strany.

#### **Nesměřová (dvoustranná) analýza**

Je-li tato volba vybrána, použije se dvoustranný test. Toto je výchozí nastavení.

#### **Analýza směrová (jednostranná) analýza**

Je-li tato volba vybrána, napájení se vypočítá pro jednostranný test.

9. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.

10. Volitelně můžete klepnutím na volbu **Zakreslit** zadat nastavení "Analýza napájení částečné Pearsonovy korelace: Zakreslit" na stránce 31 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

11. Volitelně můžete klepnutím na volbu **Přesnost** odhadnout velikost vzorku na základě intervalů spolehlivosti tak, že určíte hodnoty pološířek intervalu spolehlivosti. Více informací naleznete v části "Analýza napájení: Přesnost" na stránce 35.

**Poznámka:** **Přesnost** je k dispozici pouze v případě, že je vybrána volba **Ukázková velikost** jako metoda odhadu testu **Odhad** a jako **Směr testu** je vybrána analýza **Nesměřové (dvoustranná)**.

### ***Analýza napájení částečné Pearsonovy korelace: Zakreslit***

Dialogové okno **Zakreslit** poskytuje volby pro řízení grafů, které jsou výstupem pro ilustraci dvou a tří dimenzí výkonu podle grafů. Dialogové okno také řídí vertikální a horizontální rotaci stupňů pro trojrozměrné grafy.

#### **Dvourozměrný zákres**

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Nastavení je standardně vypnuto.

#### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon hodnotového grafu s hodnotou null. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Odhad výkonu versus alternativní hodnota Hypotéza**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon v rámci grafu alternativní hodnoty. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Odhad výkonu ve srovnání s počtem proměnných, které byly odvoláno**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrnou hodnotu podle počtu oblastí s předvolanými proměnnými. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Odhad výkonu oproti rozdílu mezi hypotézovanými hodnotami**

Je-li tato volba povolena, toto volitelné nastavení řídí dvoudimenzionální výkon podle rozdílu mezi hypotetickým měřicím grafem. Nastavení je ve výchozím nastavení vypnuto.

#### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon podle vzorkového grafu velikosti vzorku. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Rozsah vzorku velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

### **Horní mez**

Řídí horní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Trojrozměrný zakres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy. Toto nastavení je při výchozím nastavení zakázáno.

#### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba vybrána, je možné povolit následující volby.

##### **na ose x a rozdíl mezi hypotetizovanými hodnotami na ose Y**

Volitelné nastavení řídí trojrozměrný výkon podle velikosti vzorku (osa X) a rozdílu mezi hodnotami hypotetizovaných hodnot (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

##### **o ose Y a rozdílu mezi hypotetizovanými hodnotami na ose X**

Volitelné nastavení určuje trojrozměrnou energii podle velikosti vzorku (osa Y) a rozdílu mezi hodnotami hypotetické hodnoty (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf

#### **Rozsah vzorku velikosti vzorku (v párech)**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu.

#### **Dolní mez**

Řídí dolní mez pro odhad trojrozměrného výkonu podle vzorového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro odhad vzorkovací síly podle vzorového grafu velikosti. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### **Hodnota odhadu spotřeby energie versus nulová hypotéza**

Je-li tato volba vybrána, je možné povolit následující volby.

#### **na ose X a alternativní hodnota hypotézy na ose Y**

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa X) a v grafu alternativní hodnota (osa Y). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **o ose Y a alternativní hypotéze na ose x**

Volitelné nastavení řídí trojrozměrný výkon hodnotou null (osa Y) a alternativní hodnotu (osa X). Ve výchozím nastavení je graf potlačen. Je-li zadán, zobrazí se graf.

#### **Svislé otočení**

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zakres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

#### **Vodorovné otočení**

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zakres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

## **Regrese**

V produktu IBM SPSS Statistics Base Edition jsou zahrnuty následující statistické funkce.

### **Power Analysis of Univariate Linear Regression Test**

Tato funkce vyžaduje produkt IBM SPSS Statistics Base Edition.



Analýza napájení hraje hlavní roli ve studijním plánu, návrhu a převodním vedení. Výpočet výkonu je obvykle před shromážděnými údaji o vzorku, s výjimkou případů, kdy je to možné, z malé pilotní studie. Přesný odhad moci může vyšetřovatelům sdělit, jak je pravděpodobné, že statisticky významný rozdíl bude zjištěn na základě konečné velikosti vzorku v rámci skutečné alternativní hypotézy. Je-li síla příliš nízká, existuje jen malá naděje na odhalení podstatného rozdílu a nevýznamné výsledky jsou pravděpodobné i tehdy, pokud skutečné rozdíly skutečně existují.

Univariate lineární regrese je základní a standardní statistický přístup, ve kterém výzkumní pracovníci používají hodnoty několika proměnných k vysvětlení nebo předpovídání hodnot výsledků stupnice.

Power Analysis of Univariate Linear Regression test estimates the power of the type III *F*-test in univariate multiple linear regression models. S velikostí efektu představovanou vícenásobnými (částečnými) korelacemi jsou k dispozici metody jak pro fixní prediktory, tak pro náhodné prediktory. U pevných prediktorů je odhad výkonu založen na necentrální distribuci *F*. Pro náhodné prediktory se předpokládá, že cílová proměnná a prediktory společně sledují vícerozměrné normální rozdělení. V takovém případě je odhad spotřeby energie založen na rozdělení vzorku násobného koeficientu korelace.

1. Z nabídky vyberte:

**Analýza > Analýza napájení > Regrese > Univariate Linear**

2. Vyberte nastavení odhadu testu **Odhad (Velikost vzorku nebo Výkon)**.
3. Je-li vybrána volba **Velikost vzorku**, zadejte buď hodnotu **Hodnota jednotlivého napájení** pro hodnotu odhadu velikosti vzorku (hodnota musí být jednotlivá hodnota mezi 0 a 1), nebo vyberte volbu **Hodnoty výkonu mřížky** a poté klepnutím na tlačítko **Mřížka** zobrazte projektované velikosti vzorku pro rozsah specifických hodnot napájení.

Více informací naleznete v části [“Analýza napájení: Hodnoty mřížky”](#) na stránce 36.

4. Je-li vybrána volba **Odhadnout výkon**, zadejte příslušnou hodnotu **Velikost vzorku** pro hodnotu odhadu napájení. Hodnota musí být celé číslo větší než nebo rovné celkovému počtu predikátů modelu + 2 (je-li povolena volba **Zahrnout výraz zachycení v modelu**). Jinak musí být hodnotou jediné celé číslo větší než nebo rovné celkovému počtu predikátů modelu +1.
5. Do pole **Populace více částečných korelací** zadejte hodnotu vícenásobného dílčího korelačního koeficientu. Hodnota musí být jednotlivá hodnota mezi -1 a 1.

**Poznámka:** Je-li zadána hodnota **Power**, hodnota **Populace vícenásobná dílčí korelace** nemůže být 0.

Následující nastavení jsou povolena, je-li vybrána volba **Populace více částečných korelace**:

#### **Celkový počet prediktorů v modelu**

Zadejte počet buďto celkového počtu prediktorů, nebo prediktory v úplném modelu (případně včetně zachyovače, pokud je to možné). Hodnota musí být jednotlivá celočíselná hodnota větší než nebo rovna 1.

#### **Počet prediktorů testu**

Zadejte počet prediktorů testu nebo prediktorů ve vnořeném modelu (nezahrnuje zachycení, je-li to možné). Hodnota musí být větší nebo rovna 1, ale nesmí být větší než hodnota **Celkový počet prediktorů v modelu**.

6. Určete **R-čtvercové hodnoty pro násobné korelační koeficienty pro Úplný model i Vnořený model**. Hodnoty musí být jednotlivé hodnoty v rozsahu 0 až 1.

**Poznámka:** Je-li zadána hodnota **Power**, musí být hodnota **Úplný model** větší než hodnota **Vnořený model**.

Následující nastavení jsou povolena, když je vybrána volba **R-čtvercové hodnoty pro**:

#### **Celkový počet prediktorů-Úplný model**

Zadejte počet celkových prediktorů pro celý model (kromě zachycení, je-li to možné). Hodnota musí být jednotlivá celočíselná hodnota větší než nebo rovna 1.

### **Celkový počet prediktorů-Vnořený model**

Určete počet všech prediktorů pro vnořený model (pokud jej nelze použít, včetně zachycení). Hodnota musí být větší než nebo rovna 1, ale menší než hodnota **Celkový počet prediktorů-Úplný model**.

7. Volitelně určete úroveň významnosti chyby typu I pro test v poli **Úroveň významnosti**. Hodnota musí být jednoduchá hodnota typu double mezi 0 a 1. Výchozí hodnota je 0,05.
8. Volitelně můžete vybrat nastavení **Zahrnout výraz zachycení v modelu**. Nastavení je ve výchozím nastavení povoleno. Není-li tato volba vybrána, je výraz zachycení vyloučen z analýzy výkonu.
9. Volitelně můžete vybrat, zda jsou prediktory modelu **Pevné** nebo **Náhodné**. **Pevné** je výchozí nastavení.
10. Volitelně můžete klepnout na tlačítko **Zakreslit** a zadat nastavení "Power Analysis of Univariate Linear Regression: Plot" na stránce 34 (výstup grafu, nastavení dvourozměrného výkresu a nastavení trojrozměrného výkresu).

**Poznámka:** Volba **Zakreslit** je k dispozici pouze v případě, že je vybrána volba **Napájení** jako předpoklad testu.

### **Power Analysis of Univariate Linear Regression: Plot**

Grafy, které jsou výstupem pro ilustraci dvou a třírozměrných grafů, můžete ovládat grafy. Můžete také řídit zobrazení rad k nástrojům a svislých/vodorovných stupňů otočení pro trojrozměrné grafy.

#### **Dvourozměrný zakres**

Poskytuje volby pro řízení dvourozměrného odhadu výkonu oproti diagramem. Nastavení je standardně vypnuto.

#### **Odhad výkonu v porovnání s více částečnými korelací**

Je-li povoleno, toto volitelné nastavení řídí dvourozměrnou mocnost s vícenásobným částečným součinitelem korelace. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba povolena, toto volitelné nastavení řídí dvourozměrný výkon podle vzorkového grafu velikosti vzorku. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **Rozsah vzorku velikosti vzorku**

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zakresového grafu.

#### **Dolní mez**

Řídí dolní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

#### **Horní mez**

Řídí horní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

#### **Trojrozměrný zakres**

Poskytuje volby pro řízení trojrozměrného odhadu výkonu v porovnání s grafy, nastavení svislého/vodorovného otočení a uživatelem zadaný rozsah vzorku velikosti vzorku. Toto nastavení je při výchozím nastavení zakázáno.

#### **Odhad výkonu ve srovnání s velikostí vzorku**

Je-li tato volba povolena, toto volitelné nastavení řídí tři rozměrové grafy podle velikosti grafů velikosti vzorku. Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

#### **na ose x a vícenásobné částečné korelace na ose y**

Toto volitelné nastavení řídí trojrozměrnou energii podle velikosti vzorku (osa X) a grafu s vícenásobným částečným korelačním koeficientem (osa Y). Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### na ose Y a vícenásobné dílčí korelace na ose x

Toto volitelné nastavení řídí trojrozměrnou energii podle velikosti vzorku (osa Y) a grafu s vícenásobným částečným korelačním koeficientem (osa x). Nastavení je ve výchozím nastavení vypnuto. Je-li tato volba vybrána, zobrazí se graf.

### Rozsah vzorku velikosti vzorku

Je-li tato volba vybrána, jsou k dispozici volby dolní a horní meze. Pokud pro pole **Dolní mez** nebo **Horní mez** nejsou zadány žádné celočíselné hodnoty, použije se výchozí rozsah zákresového grafu.

#### Dolní mez

Řídí dolní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než nebo rovna 4 a nesmí být větší než hodnota **Horní mez**.

#### Horní mez

Řídí horní mez pro dvourozměrný odhad energie podle vzorkového grafu velikosti vzorku. Hodnota musí být větší než hodnota **Dolní mez** a nesmí být větší než 5000.

### Svislé otočení

Volitelné nastavení nastaví svislý úhel otočení (ve směru hodinových ručiček zleva) pro trojrozměrný graf. Otočením grafu lze svisle otočit pomocí myši. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 10.

### Vodorovné otočení

Volitelné nastavení nastaví vodorovné stupně otočení (ve směru hodinových ručiček zepředu) pro trojrozměrný graf. Graf lze otáčet vodorovně pomocí myši a otáčení grafu. Nastavení se projeví, když se požaduje trojrozměrný zákres. Hodnota musí být jedna celočíselná hodnota menší nebo rovna hodnotě 359. Výchozí hodnota je 325.

## Analýza napájení: Přesnost

Dialogové okno **Přesnost** je k dispozici pro každou proceduru produktu Power Analysis (s výjimkou lineární regrese Univariate) a poskytuje volby pro odhad velikosti vzorku na základě zadané šířky intervalu spolehlivosti. Dialogové okno je k dispozici, je-li pro odhad **Velikost vzorku** zadána hodnota **Hodnota jednoho napájení** nebo **Hodnoty výkonu mřížky**, a je-li nastavení **Směr testu** nastaveno na hodnotu **Nesměrový (dvoustranná analýza)**.

**Poznámka:** Nastavení **Směr testu** se nevztahuje na jednofaktorovou ANOVA.

jsou-li vybrány volby **Odhadnout velikost vzorku** a **Hodnoty výkonu mřížky** (klepnutím na ovládací prvek **Mřížka** se zobrazí dialogové okno).

### Typ (y) intervalu spolehlivosti

Vyberte příslušné typy intervalů spolehlivosti a zda se použije oprava kontinuity.

**Poznámka: Typ (y) intervalu (y) důvěryhodnosti** jsou k dispozici pouze pro proporční procedury Power analýzy (One-Sample Proportion, Related-Samples Binomial Test, and Independent-Samples Binomial Test).

### Zadejte polovinu intervalu spolehlivosti

Odhaduje velikost vzorku na základě hodnoty intervalu spolehlivosti half-width. Zadejte hodnotu v rozsahu 0 až 1. U jednovzorového testu Binomického testu musí být hodnota v rozsahu 0-0.5.

**Poznámka:** Duplicitní hodnoty jsou ignorovány.

- Klepnutím na tlačítko **Přidat** přidáte zadanou hodnotu polovičaté šířky do seznamu.
- Chcete-li aktualizovat hodnotu, zvýrazněte existující hodnotu polovičité šířky a klepněte na tlačítko **Změnit**.
- Chcete-li odebrat hodnotu ze seznamu, zvýrazněte existující hodnotu polovičité šířky a klepněte na tlačítko **Odebrat**.

## Analýza napájení: Hodnoty mřížky

Dialogové okno **Hodnoty mřížky** poskytuje volby pro uvedení rozsahu hodnot **POWER** pro účely zobrazení projektovaných velikostí vzorku ve formátu mřížky pro každou uvedenou hodnotu rozsahu **POWER**.

Dialogové okno **Hodnoty mřížky** je k dispozici pro každou proceduru Power Analysis, jsou-li vybrány volby **Odhadnout velikost vzorku** a **Hodnoty napájení mřížky** (klepnutím na ovládací prvek **Mřížka** se zobrazí dialogové okno).

### Určete jednu mocnost

Je-li vybrána tato volba, je vyžadována alespoň jedna hodnota pro spuštění analýzy. Je povoleno více hodnot a každá hodnota musí patřit do  $[0, 1]$ . Můžete uvést více hodnot s každou hodnotou oddělenou mezerou (nebo prázdnými mezerami). Pomocí ovládacích prvků **Přidat**, **Změnit** a **Odebrat** můžete pracovat s hodnotami v seznamu hodnot napájení.

Všechny hodnoty musí být jedinečné (duplicitní hodnoty nejsou povoleny).

### Uveďte rozsah výkonu

Je-li tato volba vybrána, je možné zadat rozsah hodnot výkonu od hodnoty **Začátek** (value1) do hodnoty **Koncová** (value2) s přírůstkem **Podle** (value3). Je-li zadán, je povolen pouze jeden platný soubor  $[value1 \text{ TO } value2 \text{ BY } value3]$ . Musí se ujistit, že  $0 \leq value1 \leq value2 \leq 1$ . V případech, kdy  $value1 = value2$ , je ekvivalentní zadání jednoho value1, bez ohledu na value3.

**Poznámka:** Volby **Uvést jednotlivé napájení** a **Uvést rozsah výkonu** jsou nezávislé; můžete vybrat jednu volbu nebo obě volby.

## Metaanalýza

---

Metaanalýza je analýza údajů získaných z kolekce studií, které odpovídají na podobné výzkumné otázky. Tyto studie jsou známy jako primární studie. Metaanalýza používá statistické metody k vytvoření celkového odhadu účinku, zkoumání mezilaboratorní heterogenity a zkoumání dopadu zveřejňování předpojatosti nebo obecněji malých účinků na výsledky konečných výsledků.

Produkt IBM SPSS Statistics podporuje standardní velikosti efektu a generické (předvypočtené) velikosti pro obě binární data (jako např. poměr odchylky protokolu) a pro souvislá data (jako např. hrany Hwges). Při kroku deklarace meta analýzy jsou určeny informace z metaanalýzy (jako např. specifické velikosti výsledků studie a jejich odpovídající standardní chyby a model a metoda metaanalýzy). Informace se automaticky použijí všemi následnou metaanalýzou.

Jsou podporovány náhodné efekty, společný efekt a modely meta analýzy pevných efektů. V závislosti na zvoleném modelu metaanalýzy jsou k dispozici různé metody odhadu (například inverzní variance a Mantel-Haenszel) pro běžné efekty a modely s pevnými účinky. Pro parametr kurzového rozdílu pro model náhodných účinků je k dispozici několik různých odhadců.

Produkt IBM SPSS Statistics podporuje následující procedury meta analýzy:

- [“Souvislá meta-analýza” na stránce 36](#)
- [“Velikost souvislé analýzy meta-analýzy” na stránce 45](#)
- [“Binární meta-analýza” na stránce 52](#)
- [“Velikost binárního efektu metaanalýzy” na stránce 61](#)
- [“Regrese meta-analýzy” na stránce 69](#)

## Souvislá meta-analýza

Souvislý postup Metaanalýza provádí metaanalýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Příklad

Několik výzkumných studií bylo provedeno v historii, aby prošetřila faddish, ale sporný lék na pomoc při léčbě diabetu typu II. Perorální přípravek byl po jídle tvrdil, že je schopen snížit hladinu glukózy v krvi. Údaje byly shromažďovány z různých výzkumných míst od roku 1979 do roku 1986.

Hlavní vyšetřovatel by rád nakreslil statistické inference o účinku perorálního léku. Vzhledem k tomu, že údaje byly získány z různých studií, navrhla myšlenku syntetizovat výsledky v rámci studií za účelem dosažení celkového pochopení účinku a zjištění těchto základních zdrojů variace ve výsledcích.

## Statistika

Úroveň spolehlivosti, iterační tolerance, vzorová odchylka, směrodatná odchylka, vzorek odchylky, směrodatná odchylka, odhadovaná velikost účinku, Cohen's  $d$ , Hedges'  $g$ , Glass's Delta, mean difference, cumulative analysis, estimation method, trim-and-fill, regression-based test, random-effects model, fixed-effects estimator, empirical Bayes estimator, DerSimonian-Laird estimator, Siddik-Jonkman estimator, Knapp-Hartung standard-error adjustment, truncated, korektor Knapp-Hartung standard-chyba korektura Knapp-Hartung standard-error adjustment, coefficients, EGGER in regression-based test, intercept, multiplikativní model, multiplikativní disperzní parameter, quadratic estimator, homogeneity test, heterogeneity measures, prediction interval, estimated standard error, estimated  $k$ -value, celkový overall effect size, estimated study weight, estimated study weight.

## Získání průběžné analýzy meta-analýzy

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V sekci **Skupina opatření** vyberte proměnnou **Velikost studie**, která bude představovat velikost vzorku pro danou skupinu léčby. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
3. Vyberte proměnnou **Střední hodnota**, která bude představovat vzorek pro danou skupinu ošetření. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
4. Vyberte volbu **Směrodatná odchylka** pro určení směrodatné odchylky vzorku nebo **Odchylka** pro určení rozptylu vzorku a poté vyberte proměnnou, která bude představovat standardní odchylku/varpro danou skupinu ošetření.
5. V sekci **Řídicí skupina** vyberte proměnnou **Velikost studie**, která bude představovat velikost vzorku pro řídicí skupinu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
6. Vyberte proměnnou **Střední hodnota**, která bude znázorňovat ukázkové prostředky pro řídicí skupinu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
7. Vyberte volbu **Směrodatná odchylka** pro určení směrodatné odchylky vzorku nebo **Odchylka** pro určení rozptylu vzorku a poté vyberte proměnnou, která bude představovat standardní odchylku/varpro řídicí skupinu.
8. Volitelně můžete vybrat proměnné **ID studie** a/nebo **Štítek studie**. Zvolená proměnná **Study ID** nemůže být stejná jako vybraná proměnná **Štítek studie**.
9. Volitelně můžete vybrat nastavení **Velikost efektu**.

### Kráva

Výchozí nastavení odhaduje  $d$  Coheni. Je-li vybrána volba **Upravená standardní chyba**, nastavení odhaduje  $d$  a jeho rozptyl pomocí alternativního vzorce rozděleného  $2(N_{\text{treatment}} + N_{\text{control}} - 2)$ .

### Hrany $g$

Odhaduje  $H$  kraje  $g$ . Je-li vybrána volba **Upravená standardní chyba**, nastavení odhadu  $H$  hraty  $g$  a jeho rozptylu se použije s použitím alternativního vzorce rozděleného na  $2(N_{\text{treatment}} + N_{\text{control}} - 3.94)$ .

### Skleněná delta

Odhaduje Delta Delta na základě kontrolní skupiny. Je-li vybrána volba **Standardizováno na základě léčebné skupiny**, je rozdílová data skla standardizována na základě směrodatné odchylky exponované skupiny.

### Nestandardizovaný střední rozdíl

Odhaduje střední rozdíl tím, že za předpokladu, že dvě populace směrodatné odchylky jsou stejné. Při výběru volby **Unequal group variance** se průměrný rozdíl odhaduje za předpokladu, že dvě směrodatné odchylky počtu obyvatel se nerovnájí.

10. Volitelně můžete vybrat nastavení **Model** . Je-li povolena nastavení **Oříznout a-Vyplnit** , nastavení také řídí model, který je používán fondem v analýze obruby a výplně. Je-li nastavení **Bias** povoleno, nastavení také řídí model, který používá test založený na regresi.

#### **Náhodné účinky**

Výchozí nastavení vytváří model s náhodnými efekty.

#### **Fixní-účinky**

Sestaví model s neměnným účinkem.

11. Volitelně můžete:

- Klepněte na **Kritéria ...** specifikujte obecná kritéria.
- Klepnutím na volbu **Analýza** určete podskupinu a kumulativní analýzu.
- Chcete-li určit metody odhadu, klepněte na volbu **Inference** .
- Klepnutím na tlačítko **Kontrast** otestujte kontrast testu.
- Klepnutím na tlačítko **Bias** získáte přístup k předpojatosti publikování tím, že provedete test na regresní testu EGGER.
- Klepnutím na volbu **Oříznout-a-Vyplnit** implementujte analýzu obruby publikování a výplně.
- Chcete-li řídit výstupy tabulky, klepněte na tlačítko **Tisk** .
- Klepnutím na tlačítko **Uložit** uložíte odhadovanou statistiku do aktivní datové sady.
- Chcete-li určit vykreslení dat ve výstupu, klepněte na volbu **Zakreslit** .

12. Klepněte na tlačítko **OK**.

## **Souvislá analýza meta-analýzy: kritéria**

Dialog **Kritéria** poskytuje nastavení pro uvedení kritérií pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

#### **Interval spolehlivosti**

Volitelné nastavení určuje úroveň důvěry. Hodnota musí být číselná hodnota mezi 0 a 100. Předvolené nastavení je 95.

#### **Chybějící rozsah dat**

Volitelná nastavení řídí, jak procedura zpracovává chybějící data.

#### **Vyloučit analýzu případů podle analýzy**

Výchozí nastavení zahrnuje všechny případy s dostatkem dat na proměnných použitých v každé konkrétní analýze.

#### **Vyloučit případy jako litwise**

Zahrnuje všechny případy s dostatečnými daty pro všechny proměnné použité ve všech analýzách uvedených v rámci procedury.

#### **Uživatelské chybějící hodnoty**

Volitelná nastavení řídí způsob zpracování uživatelských chybějících hodnot.

#### **Vyloučit**

Výchozí nastavení považuje uživatele za platné chybějící hodnoty.

#### **Zahrnout**

Ignoruje uživatele, kterému chybí hodnota, a zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami.

#### **Iterace**

#### **Maximum iterací**

Volitelné nastavení určuje maximální počet iterací v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 100. Hodnota 0 znamená, že se neprovedou žádné iterace.

### Maximální krok-snížení

Volitelné nastavení určuje maximální krok-halving v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 5. Hodnota 0 znamená, že krok-hallování není použit.

### Konvergence

Volitelné nastavení určuje toleranci konvergence. Hodnota musí být jedna kladná hodnota. Nastavení je k dispozici při použití iterativních metod. Standardní hodnota je  $1E-6$ .

## Definování spojitých kritérií Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé meta-analýza** klepněte na **Kritéria**.

3. Vyberte a definujte příslušné nastavení kritérií.

4. Klepněte na tlačítko **Pokračovat**.

## Souvislá analýza meta-analýzy: Analýza

Dialogové okno **Analýza** poskytuje nastavení pro uvedení podskupiny a kumulativní analýzy pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Analýza podskupin

Vyberte proměnnou, která vyvolává analýzu podskupin. Proměnná nemůže být stejná jako proměnná definovaná pro **Kumulativní analýza**.

### Kumulativní analýza

Vyberte proměnnou, která vyvolává kumulativní analýzu, a na níž je provedena kumulativní analýza meta-analýzy. Proměnná nemůže být stejná jako proměnná definovaná pro **Analýza podskupiny**. Je-li vybrána volba **Vzestupně**, souhrnná analýza je založena na zadané proměnné ve vzestupném pořadí. Je-li vybrána volba **Sestupně**, souhrnná analýza je založena na zadané proměnné v sestupném pořadí.

### Kumulativní statistika

Poskytuje volby pro uložení odhadované kumulativní celkové velikosti efektu. Nastavení jsou k dispozici pouze v případě, že je vybrána proměnná **Kumulativní analýza**.

#### Velikost kumulativního účinku

Uloží odhadovanou kumulativní celkovou velikost celkového účinku.

#### Standardní chyba

Uloží odhadovanou standardní chybu kumulativní celkové velikosti efektu.

#### Dolní hranice intervalu spolehlivosti

Uloží odhadovaný interval spolehlivosti dolní hranice kumulativní celkové velikosti efektu.

#### Horní hranice intervalu spolehlivosti

Uloží odhadovanou horní mez intervalu spolehlivosti kumulativní celkové velikosti efektu.

#### Hodnota P

Uloží odhadovanou hodnotu  $p$ -hodnota kumulativní celkové velikosti efektu.

### Cíl

Poskytuje volby pro uvedení souboru dat typu save file nebo datového souboru. Je-li vybrána volba **Datová sada**, můžete zadat název nové datové sady (můžete zachovat výchozí název datové sady). Je-li vybrána volba **Datový soubor**, klepněte na tlačítko **Procházet ...**, abyste vybrali název souboru typu save file a umístění.

## Definování nastavení průběžné analýzy meta-analýzy

1. Z nabídky vyberte:

## **Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé meta-analýza** klepněte na **Analýza**.
3. Vyberte a definujte odpovídající nastavení analýzy.
4. Klepněte na tlačítko **Pokračovat**.

### **Souvislá-analýza metadat: inference**

Dialog **Inference** poskytuje nastavení pro uvedení metod odhadu pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

**Poznámka:** Dialogové okno **Inference** je k dispozici pouze v případě, že je vybrán model **Náhodné efekty**.

#### **Odhadce**

Poskytuje nastavení pro určení odhadu.

##### **Poznámky:**

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také odhad, který se používá při slučování v analýze obruby a výplně.
- Je-li zadána nastavení Bias, řídí se také odhad, který používá test založený na regresi.

##### **Omezená maximální pravděpodobnost (REML)**

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

##### **Maximální pravděpodobnost (ML)**

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

##### **Empirical Bayes**

Použije iterativní metodu a vypočítá empirický odhad Bayes.

##### **Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

##### **Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

##### **DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

##### **Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

#### **Úprava standardní chyby**

Poskytuje nastavení, které řídí, zda má být použita standardní úprava nastavení chyb Knapp-Hartung.

##### **Poznámky:**

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také standardní úprava chyb, která je použita při slučování v analýze obruby a výplně.
- Je-li zadána hodnota Bias, bude také použita standardní úprava chyb, kterou používá test založený na regresi.

##### **Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

##### **Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

##### **Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.



## Definování nastavení kontinuálního odvozování Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé analýzy meta-analýzy** klepněte na volbu **Inference**.

3. Vyberte a definujte odpovídající nastavení odvozování.

4. Klepněte na tlačítko **Pokračovat**.

## Souvislá analýza metadat: kontrast

Dialogové okno **Kontrast** poskytuje nastavení pro řízení kontrastního testu pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Proměnné

Seznam zobrazí všechny dostupné proměnné datové sady. Vyberte proměnné ze seznamu a přesuňte je do seznamu **Kontrast (s)**.

### Kontrast (y)

Seznam identifikuje koeficienty, které jsou uloženy jako proměnné v aktivní datové sadě. Je povoleno více proměnných. Řetězcové proměnné nejsou podporovány.

### User-Input Coefficient Values

Poskytuje nastavení pro určení koeficientů kontrastu uživatele. Povoleny jsou pouze číselné hodnoty. Chcete-li formulovat platnou kontrastní test, musí se počet uvedených hodnot shodovat s počtem platných studií.

## Definování nastavení kontinuálního kontrastu Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé analýzy meta-analýzy** klepněte na volbu **Kontrast**.

3. Vyberte a definujte vhodná nastavení kontrastu.

4. Klepněte na tlačítko **Pokračovat**.

## Souvislá analýza meta-analýzy: bias

Dialogové okno **Bias** poskytuje nastavení pro povolení předpojatosti tím, že provede test na bázi regrese Eggers pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Zkouška na regresní přímce Egger

Výběrem této volby povolíte předpojatost publikování provedením testu na regresní testu Eggers.

### Proměnné

Seznam poskytuje všechny dostupné proměnné datové sady.

### Kovariáta (proměnné)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako proměnné. Je povoleno více proměnných covariates.

### Faktor (y)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako faktory. Je povoleno více faktorů.

### Zahrnout zachycení v regresi

Řídí výraz zachycení v regresním testu.

### Zahrnout parametr disperze v modelu s pevnou efektům

Řídí nastavení multiplikačního modelu a zavádí multiplikační disperzní parametr pro analýzu.

Nastavení je k dispozici pouze v případě, že je vybrán model s pevným efektem.

### **Odhadnout statistiku na základě t-distribuce**

Řídí distribuci použitou v regresních testech. Nastavení je standardně povoleno, což odhaduje statistiku na základě *t*-rozdělení. Není-li toto nastavení vybráno, statistiky se odhadují na základě normální distribuce.

## **Definování nastavení průběžné analýzy funkce Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislá analýza meta-analýzy** klepněte na **Bias**.

3. Vyberte a definujte odpovídající nastavení předpojatosti.

4. Klepněte na tlačítko **Pokračovat**.

## **Souvislá analýza meta-analýzy: Trim-and-Fill**

Dialogové okno **Trim-and-Fill** poskytuje nastavení pro implementaci analýzy obruby publikování a výplně pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### **Odhadnout počet chybějících studií**

Řídí analýzu obruby-a-výplň publikování. Výběrem tohoto nastavení povolíte jiná nastavení dialogových oken.

### **Ze strany Impute Studies**

Poskytuje volby pro uvedení strany trychtýřového grafu, na které jsou přisouzena chybějící studie.

#### **Stanoveno svahy Eggerově testu**

Výchozí nastavení určuje stranu založenou na odhadovaném sklonu testu Egger.

#### **Vlevo**

Impuuje levou stranu trychtýřku grafu.

#### **Vpravo**

Impuuje pravou stranu trychtýřového grafu.

### **Metoda**

Uvádí metodu pro odhad počtu chybějících studií.

#### **Lineární**

Výchozí nastavení vypočítá lineární odhad.

#### **Spustit**

Vypočte odhad spuštění.

#### **Kvadratické**

Vypočte kvadratický odhad.

### **Proces iterace**

Poskytuje nastavení pro určení odhadu iterace a standardní úpravy chyb.

#### **Model s pevnou efecí**

Je-li vybrána tato volba, je použit model s pevným efektem a nejsou k dispozici standardní volby úpravy iterace a nejsou k dispozici standardní volby úpravy chyby.

#### **Model náhodných efektů**

Je-li vybrána tato volba, je použit model s náhodnými efekty a jsou k dispozici následující nastavení.

#### **Odhadce**

Poskytuje nastavení pro určení odhadu iterace.

#### **Omezená maximální pravděpodobnost (REML)**

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

### **Maximální pravděpodobnost (ML)**

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

### **Empirical Bayes**

Použije iterativní metodu a vypočítá empirický odhad Bayes.

### **Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

### **Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

### **DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

### **Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

### **Úprava standardní chyby**

Poskytuje nastavení pro řízení, zda má být použita standardní úprava chyb Knapp-Hartung na iterace algoritmu obruby a výplně.

#### **Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

#### **Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

#### **Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## **Definování kontinuálních nastavení trim-and-fill**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé analýzy meta-analýzy** klepněte na **Oříznout a-Vyplnit**.

3. Vyberte a definujte odpovídající nastavení obruby a výplně.

4. Klepněte na tlačítko **Pokračovat**.

## **Souvislá analýza metadat: Tisk**

Dialogové okno **Tisk** poskytuje nastavení pro řízení výstupů tabulky pro meta-analýzu s průběžným výsledkem na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### **Homogenita/heterogenita**

Poskytuje nastavení pro řízení homogenity a testů heterogenity.

#### **Zkouška homogenity**

Je-li tato volba vybrána, je ve výstupu uveden odpovídající test homogenity.

#### **Opatření heterogenity**

Je-li tato volba vybrána, jsou ve výstupu poskytnuta opatření heterogenity.

### **velikost účinku**

Poskytuje následující nastavení velikosti efektu.

#### **Jednotlivé studie**

Řídí zobrazení jednotlivých studií. Je-li tato volba vybrána, bude poskytnut odpovídající výstup.

#### **Velikosti kumulativního účinku**

Řídí zobrazení kumulativní analýzy. Je-li tato volba vybrána, bude poskytnut odpovídající výstup. Nastavení je k dispozici pouze v případě, že je v dialogovém okně **Analýza** vybrána proměnná

**Kumulativní analýza** .

### Interval předpovědí v modelu s náhodnými účinky

Řídí zobrazení intervalu předpovědi. Je-li tato volba vybrána, bude poskytnut odpovídající výstup. Nastavení je k dispozici pouze v případě, že je zadán model s náhodnými účinky.

## Definování nastavení průběžného tisku meta-analýzy

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé metaanalýzy** klepněte na volbu **Tisk**.

3. Vyberte a definujte odpovídající nastavení tisku.

4. Klepněte na tlačítko **Pokračovat**.

## Souvislá analýza meta-analýzy: Uložit

Dialogové okno **Uložit** poskytuje nastavení pro uložení odhadované statistiky na aktivní datovou sadu pro meta-analýzu s průběžným výsledkem na nezpracovaná data, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Jednotlivé studie

Poskytuje volby pro uložení odhadované velikosti efektu.

#### Velikost individuálního účinku

Uloží odhadovanou velikost efektu.

#### Standardní chyba

Uloží odhadovanou směrodatnou chybu velikosti efektu.

#### Dolní hranice intervalu spolehlivosti

Uloží odhadovanou mez intervalu spolehlivosti dolní meze velikosti efektu.

#### Horní hranice intervalu spolehlivosti

Uloží odhadovanou horní mez intervalu spolehlivosti o velikosti efektu.

#### Hodnota P

Uloží odhadovanou hodnotu  $p$ -value efektu velikosti.

#### (%)

Uloží odhadovanou váhu studie.

#### Procentní část tělesné hmotnosti

Uloží normalizovanou váhu studie jako procentní část.

## Definování nastavení průběžného ukládání meta-analýzy

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Nezpracovaná data ...**

2. V dialogovém okně **Souvislé analýzy meta-analýzy** klepněte na tlačítko **Uložit**.

3. Vyberte a definujte odpovídající odhadované nastavení pro uložení statistiky.

4. Klepněte na tlačítko **Pokračovat**.

## Souvislá analýza metadat: Zakreslit

Dialogové okno **Zakreslit** poskytuje nastavení pro následující typy výkresu:

- “[Výkres lesa](#)” na stránce 73
- “[Kumulativní vykreslení lesa](#)” na stránce 74
- “[Zákresový](#)” na stránce 75
- “[Trychtýř](#)” na stránce 76
- “[Galbraitní nákres](#)” na stránce 77

## Velikost souvislé analýzy meta-analýzy

Procedura Velikost spojitého efektu Meta-Analysis provádí meta-analýzu s kontinuálním výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného účinku.

### Příklad

Několik výzkumných studií bylo provedeno v historii, aby prošetřila faddish, ale sporný lék na pomoc při léčbě diabetu typu II. Perorální přípravek byl po jídle tvrdil, že je schopen snížit hladinu glukózy v krvi. Údaje byly shromažďovány z různých výzkumných míst od roku 1979 do roku 1986.

Hlavní vyšetřovatel by rád nakreslil statistické inference o účinku perorálního léku. Vzhledem k tomu, že údaje byly získány z různých studií, navrhla myšlenku syntetizovat výsledky v rámci studií za účelem dosažení celkového pochopení účinku a zjištění těchto základních zdrojů variace ve výsledcích.

### Statistika

Úroveň spolehlivosti, iterační tolerance, vzorová odchylka, směrodatná odchylka, vzorek odchylky, směrodatná odchylka, odhadovaná velikost účinku, Cohen's  $d$ , Hedges'  $g$ , Glass's Delta, mean difference, cumulative analysis, estimation method, trim-and-fill, regression-based test, random-effects model, fixed-effective estimator, empirical Bayes estimator, DerSimonian-Laird estimator, Siddik-Jonkman estimator, Knapp-Hartung standard-error adjustment, truncated, korektor Knapp-Hartung standard-chyba korektura Knapp-Hartung standard-error adjustment, coefficients, EGGER's regression-based test, intercept, multiplikative model, multiplikative dispersion parameter, quadratic estimator, homogeneity test, heterogeneity measures, prediction interval, estimated standard error, estimated  $k$ -value, celculy overall effect size, estimated study weight, estimated study weight

## Analýza analýzy velikosti kontinuálního efektu Meta-analýza

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. Vyberte proměnnou **Effect Size**, která označuje velikost efektu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
3. Vyberte buď volbu **Směrodatná odchylka**, chcete-li určit směrodatnou odchylku velikosti efektu, nebo volbu **Odchylka**, chcete-li určit odchylku velikosti efektu a poté vybrat proměnnou, která bude představovat standardní odchylku/varpro velikost efektu.
4. Volitelně můžete vybrat proměnné **ID studie** a/nebo **Štítek studie**. Zvolená proměnná **Study ID** nemůže být stejná jako vybraná proměnná **Štítek studie**.
5. Volitelně můžete vybrat nastavení **Model**. Je-li povolena nastavení **Oříznout a-Vyplnit**, nastavení také řídí model, který je používán fondem v analýze obruby a výplně. Je-li nastavení **Bias** povoleno, nastavení také řídí model, který používá test založený na regresii.

### Náhodné účinky

Výchozí nastavení vytváří model s náhodnými efekty.

### Fixní-účinky

Sestaví model s neměnným účinkem.

6. Volitelně můžete:

- Klepněte na **Kritéria ...** specifikujte obecná kritéria.
- Klepnutím na volbu **Analýza** určete podskupinu a kumulativní analýzu.
- Chcete-li určit metody odhadu, klepněte na volbu **Inference**.
- Klepnutím na tlačítko **Kontrast** otestujte kontrast testu.
- Klepnutím na tlačítko **Bias** získáte přístup k předpojatosti publikování tím, že provedete test na regresní testu EGGER.
- Klepnutím na volbu **Oříznout-a-Vyplnit** implementujte analýzu obruby publikování a výplně.
- Chcete-li řídit výstupy tabulky, klepněte na tlačítko **Tisk**.
- Klepnutím na tlačítko **Uložit** uložíte odhadovanou statistiku do aktivní datové sady.

- Chcete-li určit vykreslení dat ve výstupu, klepněte na volbu **Zakreslit**.

7. Klepněte na tlačítko **OK**.

## Velikost kontinuálního efektu meta-analýzy: Kritéria

Dialog **Kritéria** poskytuje nastavení pro uvedení kritérií pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Interval spolehlivosti

Volitelné nastavení určuje úroveň důvěry. Hodnota musí být číselná hodnota mezi 0 a 100. Předvolené nastavení je 95.

### Chybějící rozsah dat

Volitelná nastavení řídí, jak procedura zpracovává chybějící data.

### Vyloučit analýzu případů podle analýzy

Výchozí nastavení zahrnuje všechny případy s dostatkem dat na proměnných použitých v každé konkrétní analýze.

### Vyloučit případy jako litwise

Zahrnuje všechny případy s dostatečnými daty pro všechny proměnné použité ve všech analýzách uvedených v rámci procedury.

### Uživatelské chybějící hodnoty

Volitelná nastavení řídí způsob zpracování uživatelských chybějících hodnot.

#### Vyloučit

Výchozí nastavení považuje uživatele za platné chybějící hodnoty.

#### Zahrnout

Ignoruje uživatele, kterému chybí hodnota, a zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami.

### Iterace

#### Maximum iterací

Volitelné nastavení určuje maximální počet iterací v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 100. Hodnota 0 znamená, že se neprovedou žádné iterace.

#### Maximální krok-snížení

Volitelné nastavení určuje maximální krok-halving v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 5. Hodnota 0 znamená, že krok-hallování není použito.

#### Konvergence

Volitelné nastavení určuje toleranci konvergence. Hodnota musí být jedna kladná hodnota. Nastavení je k dispozici při použití iterativních metod. Standardní hodnota je  $1E-6$ .

## Definování kritérií průběžné analýzy Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Meta-Analysis Continuous Effect Size** klepněte na volbu **Criteria**.

3. Vyberte a definujte příslušné nastavení kritérií.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost kontinuálního efektu Meta-analýza: Analýza

Dialog **Analýza** poskytuje nastavení pro uvedení podskupiny a kumulativní analýzy pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předem vypočteného efektu.

### **Analýza podskupin**

Vyberte proměnnou, která vyvolává analýzu podskupin. Proměnná nemůže být stejná jako proměnná definovaná pro **Kumulativní analýza**.

### **Kumulativní analýza**

Vyberte proměnnou, která vyvolává kumulativní analýzu, a na niž je provedena kumulativní analýza meta-analýzy. Proměnná nemůže být stejná jako proměnná definovaná pro **Analýza podskupiny**. Je-li vybrána volba **Vzestupně**, souhrnná analýza je založena na zadané proměnné ve vzestupném pořadí. Je-li vybrána volba **Sestupně**, souhrnná analýza je založena na zadané proměnné v sestupném pořadí.

### **Kumulativní statistika**

Poskytuje volby pro uložení odhadované kumulativní celkové velikosti efektu. Nastavení jsou k dispozici pouze v případě, že je vybrána proměnná **Kumulativní analýza**.

#### **Velikost kumulativního účinku**

Uloží odhadovanou kumulativní celkovou velikost celkového účinku.

#### **Standardní chyba**

Uloží odhadovanou standardní chybu kumulativní celkové velikosti efektu.

#### **Dolní hranice intervalu spolehlivosti**

Uloží odhadovaný interval spolehlivosti dolní hranice kumulativní celkové velikosti efektu.

#### **Horní hranice intervalu spolehlivosti**

Uloží odhadovanou horní mez intervalu spolehlivosti kumulativní celkové velikosti efektu.

#### **Hodnota P**

Uloží odhadovanou hodnotu  $p$ -hodnota kumulativní celkové velikosti efektu.

### **Cíl**

Poskytuje volby pro uvedení souboru dat typu save file nebo datového souboru. Je-li vybrána volba **Datová sada**, můžete zadat název nové datové sady (můžete zachovat výchozí název datové sady). Je-li vybrána volba **Datový soubor**, klepněte na tlačítko **Procházet ...**, abyste vybrali název souboru typu save file a umístění.

## **Definování nastavení analýzy Velikost kontinuálního efektu analýzy Meta-Analýza**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost spojitého efektu meta-analýzy** klepněte na **Analýza**.

3. Vyberte a definujte odpovídající nastavení analýzy.

4. Klepněte na tlačítko **Pokračovat**.

## **Velikost souvislé analýzy meta-analýzy: Inference**

Dialog **Inference** poskytuje nastavení pro uvedení metod odhadu pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

**Poznámka:** Dialogové okno **Inference** je k dispozici pouze v případě, že je vybrán model **Náhodné efekty**.

### **Odhadce**

Poskytuje nastavení pro určení odhadu.

#### **Poznámky:**

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také odhad, který se používá při slučování v analýze obruby a výplně.
- Je-li zadána nastavení Bias, řídí se také odhad, který používá test založený na regresi.

#### **Omezená maximální pravděpodobnost (REML)**

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

**Maximální pravděpodobnost (ML)**

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

**Empirical Bayes**

Použije iterativní metodu a vypočítá empirický odhad Bayes.

**Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

**Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

**DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

**Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

**Úprava standardní chyby**

Poskytuje nastavení, které řídí, zda má být použita standardní úprava nastavení chyb Knapp-Hartung.

**Poznámky:**

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také standardní úprava chyb, která je použita při slučování v analýze obruby a výplně.
- Je-li zadána hodnota Bias, bude také použita standardní úprava chyb, kterou používá test založený na regresí.

**Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

**Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

**Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

**Definování nastavení inference velikosti spojitého efektu Meta-Analýza**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost souvislého efektu Meta-Analysis** klepněte na volbu **Inference**.

3. Vyberte a definujte odpovídající nastavení odvozování.

4. Klepněte na tlačítko **Pokračovat**.

**Velikost kontinuálního efektu metaanalýzy: kontrast**

Dialogové okno **Kontrast** poskytuje nastavení pro řízení kontrastní testu pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

**Proměnné**

Seznam zobrazí všechny dostupné proměnné datové sady. Vyberte proměnné ze seznamu a přesuňte je do seznamu **Contrast (s)**.

**Kontrast (y)**

Seznam identifikuje koeficienty, které jsou uloženy jako proměnné v aktivní datové sadě. Je povoleno více proměnných. Řetězcové proměnné nejsou podporovány.

**User-Input Coefficient Values**

Poskytuje nastavení pro určení koeficientů kontrastu uživatele. Povoleny jsou pouze číselné hodnoty. Chcete-li formulovat platnou kontrastní test, musí se počet uvedených hodnot shodovat s počtem platných studií.



## Definování kontrastu pro průběžné analýzy obsahu Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost spojitého efektu meta-analýzy** klepněte na **Kontrast**.

3. Vyberte a definujte vhodná nastavení kontrastu.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost kontinuálního efektu Meta-analýza: Bias

Dialogové okno **Bias** poskytuje nastavení pro povolení předpojatosti tím, že provede test založený na regresi Eggers pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Zkouška na regresní přímlce Egger

Výběrem této volby povolíte předpojatost publikování provedením testu na regresní testu Eggers.

### Proměnné

Seznam poskytuje všechny dostupné proměnné datové sady.

### Kovariáta (proměnné)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako proměnné. Je povoleno více proměnných covariates.

### Faktor (y)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako faktory. Je povoleno více faktorů.

### Zahrnout zachycení v regresi

Řídí výraz zachycení v regresním testu.

### Zahrnout parametr disperze v modelu s pevnou efektům

Řídí nastavení multiplikačního modelu a zavádí multiplikační disperzní parametr pro analýzu. Nastavení je k dispozici pouze v případě, že je vybrán model s pevným efektem.

### Odhadnout statistiku na základě t-distribuce

Řídí distribuci použitou v regresních testech. Nastavení je standardně povoleno, což odhaduje statistiku na základě *t*-rozdělení. Není-li toto nastavení vybráno, statistiky se odhadují na základě normální distribuce.

## Definování nastavení předpojatosti průběžné velikosti analýzy Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost spojitého efektu meta-analýzy** klepněte na **Bias**.

3. Vyberte a definujte odpovídající nastavení předpojatosti.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost kontinuálního efektu meta-analýzy: Trim-and-Fill

Dialogové okno **Trim-and-Fill** poskytuje nastavení pro implementaci analýzy obruby publikování a výplně pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Odhadnout počet chybějících studií

Řídí analýzu obruby-a-výplň publikování. Výběrem tohoto nastavení povolíte jiná nastavení dialogových oken.

### Ze strany Impute Studies

Poskytuje volby pro uvedení strany trychtýřového grafu, na které jsou přisouzena chybějící studie.

### Stanoveno svahy Eggerově testu

Výchozí nastavení určuje stranu založenou na odhadovaném sklonu testu Egger.

**Vlevo**

Impuuje levou stranu trychtýřku grafu.

**Vpravo**

Impuuje pravou stranu trychtýřového grafu.

**Metoda**

Uvádí metodu pro odhad počtu chybějících studií.

**Lineární**

Výchozí nastavení vypočítá lineární odhad.

**Spustit**

Vypočte odhad spuštění.

**Kvadratické**

Vypočte kvadratický odhad.

**Proces iterace**

Poskytuje nastavení pro určení odhadu iterace a standardní úpravy chyb.

**Model s pevnou efecí**

Je-li vybrána tato volba, je použit model s pevným efektem a nejsou k dispozici standardní volby úpravy iterace a nejsou k dispozici standardní volby úpravy chyby.

**Model náhodných efektů**

Je-li vybrána tato volba, je použit model s náhodnými efekty a jsou k dispozici následující nastavení.

**Odhadce**

Poskytuje nastavení pro určení odhadu iterace.

**Omezená maximální pravděpodobnost (REML)**

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

**Maximální pravděpodobnost (ML)**

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

**Empirical Bayes**

Použije iterativní metodu a vypočítá empirický odhad Bayes.

**Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

**Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

**DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

**Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

**Úprava standardní chyby**

Poskytuje nastavení pro řízení, zda má být použita standardní úprava chyb Knapp-Hartung na iterace algoritmu obruby a výplně.

**Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

**Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

**Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## Definování nastavení obruby a výplně velikosti spojitého efektu Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost spojitého efektu meta-analýzy** klepněte na **Oříznout-a-Vyplnit**.

3. Vyberte a definujte odpovídající nastavení obruby a výplně.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost kontinuálního efektu metaanalýzy: Tisk

Dialogové okno **Tisk** poskytuje nastavení pro řízení výstupů tabulky pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Homogenita/heterogenita

Poskytuje nastavení pro řízení homogenity a testů heterogenity.

#### Zkouška homogenity

Je-li tato volba vybrána, je ve výstupu uveden odpovídající test homogenity.

#### Opatření heterogenity

Je-li tato volba vybrána, jsou ve výstupu poskytnuta opatření heterogenity.

### velikost účinku

Poskytuje následující nastavení velikosti efektu.

#### Jednotlivé studie

Řídí zobrazení jednotlivých studií. Je-li tato volba vybrána, bude poskytnut odpovídající výstup.

#### Velikosti kumulativního účinku

Řídí zobrazení kumulativní analýzy. Je-li tato volba vybrána, bude poskytnut odpovídající výstup.

Nastavení je k dispozici pouze v případě, že je v dialogovém okně **Analýza** vybrána proměnná

**Kumulativní analýza**.

#### Interval předpovědi v modelu s náhodnými účinky

Řídí zobrazení intervalu předpovědi. Je-li tato volba vybrána, bude poskytnut odpovídající výstup.

Nastavení je k dispozici pouze v případě, že je zadán model s náhodnými účinky.

## Definování nastavení tisku průběžné velikosti analýzy Meta-Analýza

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost spojitého efektu meta-analýzy** klepněte na **Tisk**.

3. Vyberte a definujte odpovídající nastavení tisku.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost kontinuálního efektu meta-analýzy: Uložit

Dialogové okno **Uložit** poskytuje nastavení pro uložení odhadované statistiky na aktivní datovou sadu pro meta-analýzu s průběžným výsledkem, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Jednotlivé studie

Poskytuje volby pro uložení odhadované velikosti efektu.

#### Standardní chyba

Uloží odhadovanou směrodatnou chybu velikosti efektu. Nastavení je k dispozici pouze v případě, že proměnná **Effect Size** není určena.

#### Dolní hranice intervalu spolehlivosti

Uloží odhadovanou mez intervalu spolehlivosti dolní meze velikosti efektu.

### **Horní hranice intervalu spolehlivosti**

Uloží odhadovanou horní mez intervalu spolehlivosti o velikosti efektu.

### **Hodnota P**

Uloží odhadovanou hodnotu  $p$ -value efektu velikosti.

### **(%)**

Uloží odhadovanou váhu studie.

### **Procentní část tělesné hmotnosti**

Uloží normalizovanou váhu studie jako procentní část.

## **Definování nastavení uložení velikosti kontinuálního efektu Meta-Analýza**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Nepřetržité výstupy > Velikost předběžného vypočteného efektu**

2. V dialogovém okně **Velikost spojitého efektu meta-analýzy** klepněte na tlačítko **Uložit**.

3. Vyberte a definujte odpovídající odhadované nastavení pro uložení statistiky.

4. Klepněte na tlačítko **Pokračovat**.

## **Velikost kontinuálního efektu metaanalýzy: Zakreslit**

Dialogové okno **Zakreslit** poskytuje nastavení pro následující typy výkresu:

- [“Výkres lesa”](#) na stránce 73
- [“Kumulativní vykreslení lesa”](#) na stránce 74
- [“Zákresový”](#) na stránce 75
- [“Trychtýř”](#) na stránce 76
- [“Galbraithův nákres”](#) na stránce 77

## **Binární meta-analýza**

Binární procedura Metaanalýza provádí meta-analýzu s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### **Příklad**

Několik výzkumných studií bylo provedeno v historii, aby prošetřila faddish, ale sporný lék na pomoc při léčbě diabetu typu II. Perorální přípravek byl po jídle tvrdil, že je schopen snížit hladinu glukózy v krvi. Údaje byly shromažďovány z různých výzkumných míst od roku 1979 do roku 1986.

Hlavní vyšetřovatel by rád nakreslil statistické inference o účinku perorálního léku. Vzhledem k tomu, že údaje byly získány z různých studií, navrhla myšlenku syntetizovat výsledky v rámci studií za účelem dosažení celkového pochopení účinku a zjištění těchto základních zdrojů variace ve výsledcích.

### **Statistika**

Konfidenční interval, poměr log Odds, Poměr rizika protokolu, Poměr rizika protokolu, rozdíl rizik, náhodné účinky, fixní efekty, inverzní rozptyl, Mantel-Haenzel, iterace, krok-zastavení, konvergence, kumulativní statistika, kumulativní velikost účinku, maximální pravděpodobnost, maximální pravděpodobnost, ML, Empirical Bayes, Hedges, Hunter-Schmidt, DerSironian-Laird, Sidik-Jonkman, Knapp-Hartung, Egger's Test, Harbord's Test, Peters ' Test, intercept in regression, dispersion parameter, homogeniety, heterogeniety, exponentiated statistics, standard error,  $p$ -value, gravait weight.

## **Získání binární analýzy Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. Pod sekci **Skupina opatření** vyberte proměnnou **Úspěch** , která bude znázorňovat počet "úspěšnosti" pro danou skupinu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
3. Vyberte proměnnou **Selhání** , která bude představovat počet "selhání" pro skupinu ošetření. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
4. V sekci **Řídící skupina** vyberte proměnnou **Úspěch** , která bude představovat počet "úspěšných" pro řídicí skupinu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
5. Vyberte proměnnou **Selhání** představující počet "selhání" pro řídicí skupinu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
6. Volitelně můžete vybrat proměnné **ID studie** a/nebo **Štítek studie** . Zvolená proměnná **Study ID** nemůže být stejná jako vybraná proměnná **Štítek studie** .
7. Volitelně můžete vybrat nastavení **Velikost efektu** . Dostupné volby jsou **Poměr pravděpodobnosti protokolu**, **Poměr protokolu odds protokolu**, **Poměr rizika protokolua** **Rozdíl rizik**.
8. Volitelně můžete vybrat nastavení **Model** . Je-li povolena nastavení **Oříznout a-Vyplnit** , nastavení také řídí model, který je používán fondem v analýze obruby a výplně. Je-li nastavení **Bias** povoleno, nastavení také řídí model, který používá test založený na regresi.

#### Náhodné účinky

Výchozí nastavení vytváří model s náhodnými efekty.

#### Fixní-účinky

Sestaví model s neměnným účinkem. **Inverse-rozptyl** odhaduje váhu inverzního rozptylu.

**Mantel-Haenszel** odhaduje váhu Mantel-Haenszel.

9. Volitelně můžete:

- Klepněte na **Kritéria ...** specifikujte obecná kritéria.
- Klepnutím na volbu **Analýza** určete podskupinu a kumulativní analýzu.
- Chcete-li určit metody odhadu, klepněte na volbu **Inference** .
- Klepnutím na tlačítko **Kontrast** otestujte kontrast testu.
- Klepnutím na tlačítko **Bias** získáte přístup k předpojatosti publikování tím, že provedete test na regresní testu EGGER.
- Klepnutím na volbu **Oříznout-a-Vyplnit** implementujte analýzu obruby publikování a výplně.
- Chcete-li řídit výstupy tabulky, klepněte na tlačítko **Tisk** .
- Klepnutím na tlačítko **Uložit** uložíte odhadovanou statistiku do aktivní datové sady.
- Chcete-li určit vykreslení dat ve výstupu, klepněte na volbu **Zakreslit** .

10. Klepněte na tlačítko **OK**.

## Binární meta-analýza: kritéria

Dialog **Kritéria** poskytuje nastavení pro uvedení kritérií pro meta-analýzu s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

#### Interval spolehlivosti

Volitelné nastavení určuje úroveň důvěry. Hodnota musí být číselná hodnota mezi 0 a 100. Předvolené nastavení je 95.

#### Chybějící rozsah dat

Volitelné nastavení řídí, jak procedura zpracovává chybějící data.

#### Vyloučit analýzu případů podle analýzy

Výchozí nastavení zahrnuje všechny případy s dostatkem dat na proměnných použitých v každé konkrétní analýze.

#### Vyloučit případy jako litwise

Zahrnuje všechny případy s dostatečnými daty pro všechny proměnné použité ve všech analýzách uvedených v rámci procedury.

## Uživatelské chybějící hodnoty

Volitelná nastavení řídí způsob zpracování uživatelských chybějících hodnot.

### Vyloučit

Výchozí nastavení považuje uživatele za platné chybějící hodnoty.

### Zahrnout

Ignoruje uživatele, kterému chybí hodnota, a zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami.

## Iterace

### Maximum iterací

Volitelné nastavení určuje maximální počet iterací v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 100. Hodnota 0 znamená, že se neprovedou žádné iterace.

### Maximální krok-snížení

Volitelné nastavení určuje maximální krok-halving v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 5. Hodnota 0 znamená, že krok-hallování není použit.

### Konvergence

Volitelné nastavení určuje toleranci konvergence. Hodnota musí být jedna kladná hodnota. Nastavení je k dispozici při použití iterativních metod. Standardní hodnota je  $1E-6$ .

## Upravit Nulové počty

Volitelná nastavení řídí, jak upravit data nulového počtu, je-li to vhodné, v odhadu velikosti efektu.

### Upravte každou hodnotu pouze těch studií, které obsahují alespoň jednu nulu.

Výchozí nastavení upravuje pouze data s nulovým počtem.

### Upravit každou hodnotu všech studií pouze v případě, že existuje alespoň jedna studie obsahující nulovou hodnotu

Upravuje všechny frekvence, pokud ve studii existuje alespoň jeden nulový počet.

### Upravit každou hodnotu všech studií

Upraví všechna data bez ohledu na existenci nulového počtu.

### Neprovádět žádné úpravy a ponechat nula (s)

Neprovede žádnou úpravu dat.

### Přidaná hodnota

Volitelné nastavení uvádí hodnotu přidanou k datům s nulovým počtem. Výchozí hodnota je 0.5. Uvedená hodnota musí být jedna číselná hodnota větší než 0 a menší než nebo rovna 1.

## Definování binárních kritérií Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na volbu **Kritéria**.

3. Vyberte a definujte příslušné nastavení kritérií.

4. Klepněte na tlačítko **Pokračovat**.

## Binární meta-analýza: analýza

Dialogové okno **Analýza** poskytuje nastavení pro uvedení podskupiny a kumulativní analýzy pro meta-analýzu s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Analýza podskupin

Vyberte proměnnou, která vyvolává analýzu podskupin. Proměnná nemůže být stejná jako proměnná definovaná pro **Kumulativní analýza**.

## Kumulativní analýza

Vyberte proměnnou, která vyvolává kumulativní analýzu, a na níž je provedena kumulativní analýza meta-analýzy. Proměnná nemůže být stejná jako proměnná definovaná pro **Analýza podskupiny**. Je-li vybrána volba **Vzestupně**, souhrnná analýza je založena na zadané proměnné ve vzestupném pořadí. Je-li vybrána volba **Sestupně**, souhrnná analýza je založena na zadané proměnné v sestupném pořadí.

## Kumulativní statistika

Poskytuje volby pro uložení odhadované kumulativní celkové velikosti efektu. Nastavení jsou k dispozici pouze v případě, že je vybrána proměnná **Kumulativní analýza**.

### Velikost kumulativního účinku

Uloží odhadovanou kumulativní celkovou velikost celkového účinku.

### Velikost kumulativního účinku (exponovaná forma exponentu)

Uloží odhadovanou kumulativní celkovou velikost celkového účinku v exponenciální podobě.

### Standardní chyba

Uloží odhadovanou standardní chybu kumulativní celkové velikosti efektu.

### Dolní hranice intervalu spolehlivosti

Uloží odhadovaný interval spolehlivosti dolní hranice kumulativní celkové velikosti efektu.

### Horní hranice intervalu spolehlivosti

Uloží odhadovanou horní mez intervalu spolehlivosti kumulativní celkové velikosti efektu.

### Dolní hranice intervalu spolehlivosti (exponovaná forma exponent)

Uloží odhadovaný interval spolehlivosti dolní limit kumulativní celkové velikosti celkového účinku ve exponenciální podobě.

### Horní hranice intervalu spolehlivosti (exponovaná forma exponent)

Uloží odhadovaný interval spolehlivosti horní limit kumulativní celkové velikosti celkového účinku v exponenciálním tvaru.

### Hodnota P

Uloží odhadovanou hodnotu  $p$ -hodnota kumulativní celkové velikosti efektu.

## Cíl

Poskytuje volby pro uvedení souboru dat typu save file nebo datového souboru. Je-li vybrána volba **Datová sada**, můžete zadat název nové datové sady (můžete zachovat výchozí název datové sady). Je-li vybrána volba **Datový soubor**, klepněte na tlačítko **Procházet ...**, abyste vybrali název souboru typu save file a umístění.

## Definování nastavení binární analýzy Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na volbu **Analýza**.

3. Vyberte a definujte odpovídající nastavení analýzy.

4. Klepněte na tlačítko **Pokračovat**.

## Binární meta-analýza: inference

Dialogové okno **Inference** poskytuje nastavení pro uvedení metod odhadu pro meta-analýzu s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

**Poznámka:** Dialogové okno **Inference** je k dispozici pouze v případě, že je vybrán model **Náhodné efekty**.

### Odhadce

Poskytuje nastavení pro určení odhadu.

### Poznámky:

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také odhad, který se používá při slučování v analýze obruby a výplně.
- Je-li zadána nastavení Bias, řídí se také odhad, který používá test založený na regresi.

#### **Omezená maximální pravděpodobnost (REML)**

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

#### **Maximální pravděpodobnost (ML)**

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

#### **Empirical Bayes**

Použije iterativní metodu a vypočítá empirický odhad Bayes.

#### **Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

#### **Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

#### **DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

#### **Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

### **Úprava standardní chyby**

Poskytuje nastavení, které řídí, zda má být použita standardní úprava nastavení chyb Knapp-Hartung.

#### **Poznámky:**

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také standardní úprava chyb, která je použita při slučování v analýze obruby a výplně.
- Je-li zadána hodnota Bias, bude také použita standardní úprava chyb, kterou používá test založený na regresi.

#### **Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

#### **Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

#### **Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## **Definování nastavení binárních analýz Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na volbu **Inference**.

3. Vyberte a definujte odpovídající nastavení odvozování.

4. Klepněte na tlačítko **Pokračovat**.

## **Metaanalýza-binární: kontrast**

Dialogové okno **Kontrast** poskytuje nastavení pro řízení kontrastní testu pro meta-analýzu s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

#### **Proměnné**

Seznam zobrazí všechny dostupné proměnné datové sady. Vyberte proměnné ze seznamu a přesuňte je do seznamu **Contrast (s)**.



## Kontrast (y)

Seznam identifikuje koeficienty, které jsou uloženy jako proměnné v aktivní datové sadě. Je povoleno více proměnných. Řetězcové proměnné nejsou podporovány.

## User-Input Coefficient Values

Poskytuje nastavení pro určení koeficientů kontrastu uživatele. Povoleny jsou pouze číselné hodnoty. Chcete-li formulovat platnou kontrastní test, musí se počet uvedených hodnot shodovat s počtem platných studií.

## Zobrazit exponovanou statistiku

Řídí zahrnutí statistiky exponenciálního formuláře. Je-li toto nastavení vybráno, výstup zahrnuje statistiku exponenciálního tvaru, včetně exponenciálního efektu velikosti a intervalu spolehlivosti. Nastavení je k dispozici, je-li **Velikost efektu** uvedena jako **Poměr lichých protokolů**, **Peto Logs Odd** **Poměrnebo** **Poměr rizika protokolu**.

## Definování nastavení binárního kontrastu Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na volbu **Kontrast**.

3. Vyberte a definujte vhodná nastavení kontrastu.

4. Klepněte na tlačítko **Pokračovat**.

## Binární analýza: Bias

Dialogové okno **Bias** poskytuje nastavení pro povolení předpojatosti publikování testů na základě regresních testů pro meta-analýzu s binárními výsledky v nezpracovaných datech, které jsou poskytnuty v aktivní datové sadě pro odhad velikosti efektu.

## Testy založené na regresi

Poskytuje volby pro určení regresních testů. Lze vybrat více než jeden test.

### Eggeův test

Je-li vybráno, provede test Egger.

### Harbordův test

Je-li vybráno, provede test Harbord. Test je k dispozici, když je zadána hodnota **Efekt efektu** jako **Poměr lichých protokolů** nebo **Poměr rizika protokolu**.

### Peters ' test

Je-li vybráno, provede test pro Peters. Test je dostupný, když je zadána hodnota **Efekt efektu** jako **Poměr lichých protokolů**.

## Proměnné

Seznam poskytuje všechny dostupné proměnné datové sady.

## Kovariáta (proměnné)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako proměnné. Je povoleno více proměnných covariates.

## Faktor (y)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako faktory. Je povoleno více faktorů.

## Zahrnout zachycení v regresi

Řídí výraz zachycení v regresním testu.

## Zahrnout parametr disperze v modelu s pevnou efektům

Řídí nastavení multiplikačního modelu a zavádí multiplikační parametr pro analýzu. Nastavení je k dispozici pouze v případě, že je vybrán model s pevným efektem.

## Odhadnout statistiku na základě t-distribuce

Řídí distribuci použitou v regresních testech. Nastavení je standardně povoleno, což odhaduje statistiku na základě *t*-rozdělení. Není-li toto nastavení vybráno, statistiky se odhadují na základě normální distribuce.

## Definování nastavení binárních analýz Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na **Bias**.

3. Vyberte a definujte odpovídající nastavení předpojatosti.

4. Klepněte na tlačítko **Pokračovat**.

### Binární analýza: Trim-and-Fill

Dialogové okno **Trim-and-Fill** poskytuje nastavení pro implementaci analýzy obruby publikování a výplně pro analýzu meta-analýzy s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

#### Odhadnout počet chybějících studií

Řídí analýzu obruby-a-výplň publikování. Výběrem tohoto nastavení povolíte jiná nastavení dialogových oken.

#### Ze strany Impute Studies

Poskytuje volby pro uvedení strany trychtýřového grafu, na které jsou přisouzena chybějící studie.

#### Stanoveno svahy Eggerově testu

Výchozí nastavení určuje stranu založenou na odhadovaném sklonu testu Egger.

#### Vlevo

Impuuje levou stranu trychtýřku grafu.

#### Vpravo

Impuuje pravou stranu trychtýřového grafu.

#### Metoda

Uvádí metodu pro odhad počtu chybějících studií.

#### Lineární

Výchozí nastavení vypočítá lineární odhad.

#### Spustit

Vypočte odhad spuštění.

#### Kvadratické

Vypočte kvadratický odhad.

#### Proces iterace

Poskytuje nastavení pro určení odhadu iterace a standardní úpravy chyb.

#### Model s pevnou efecí

Je-li vybrána tato volba, je použit model s pevným efektem a nejsou k dispozici standardní volby úpravy iterace a nejsou k dispozici standardní volby úpravy chyby.

#### Model náhodných efektů

Je-li vybrána tato volba, je použit model s náhodnými efekty a jsou k dispozici následující nastavení.

#### Odhadce

Poskytuje nastavení pro určení odhadu iterace.

#### Omezená maximální pravděpodobnost (REML)

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

#### Maximální pravděpodobnost (ML)

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

#### Empirical Bayes

Použije iterativní metodu a vypočítá empirický odhad Bayes.

### **Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

### **Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

### **DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

### **Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

### **Úprava standardní chyby**

Poskytuje nastavení pro řízení, zda má být použita standardní úprava chyb Knapp-Hartung na iterace algoritmu obruby a výplně.

### **Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

### **Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

### **Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## **Definování binárních hodnot trim-and-fill metadat Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na **Oříznout-a-Vyplnit**.

3. Vyberte a definujte odpovídající nastavení obruby a výplně.

4. Klepněte na tlačítko **Pokračovat**.

## **Binární meta-analýza: tisk**

Dialogové okno **Tisk** poskytuje nastavení pro řízení výstupů tabulky pro meta-analýzu s binárními výsledky na prvotních datech, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### **Homogenita/heterogenita**

Poskytuje nastavení pro řízení homogenity a testů heterogenity.

#### **Zkouška homogenity**

Je-li tato volba vybrána, je ve výstupu uveden odpovídající test homogenity.

#### **Opatření heterogenity**

Je-li tato volba vybrána, jsou ve výstupu poskytnuta opatření heterogenity.

### **velikost účinku**

Poskytuje následující nastavení velikosti efektu.

#### **Jednotlivé studie**

Řídí zobrazení jednotlivých studií. Je-li tato volba vybrána, bude poskytnut odpovídající výstup.

#### **Velikosti kumulativního účinku**

Řídí zobrazení kumulativní analýzy. Je-li tato volba vybrána, bude poskytnut odpovídající výstup. Nastavení je k dispozici pouze v případě, že je v dialogovém okně **Analýza** vybrána proměnná

#### **Kumulativní analýza .**

#### **Interval předpovědi v modelu s náhodnými účinky**

Řídí zobrazení intervalu předpovědi. Je-li tato volba vybrána, bude poskytnut odpovídající výstup. Nastavení je k dispozici pouze v případě, že je zadán model s náhodnými účinky.

### **Zobrazit exponovanou statistiku**

Řídí zahrnutí statistiky exponenciálního formuláře. Je-li vybráno nastavení, jsou výstupem exponenciální statistiky, včetně velikosti exponenciálního efektu a mezních hodnot intervalu

spolehlivosti. Tento proces se týká jak povinných, tak i volitelných tabulek odhadu a předpovědi. Nastavení je k dispozici, je-li **Velikost efektu** uvedena jako **Poměr lichých protokolů**, **Peto Logs Odd** **Poměrnebo** **Poměr rizika protokolu**.

## Definování nastavení binárního tisku Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na tlačítko **Tisk**.

3. Vyberte a definujte odpovídající nastavení tisku.

4. Klepněte na tlačítko **Pokračovat**.

## Binární meta-analýza: uložení

Dialogové okno **Uložit** poskytuje nastavení pro uložení odhadované statistiky na aktivní datovou sadu pro meta-analýzu s binárními výsledky na nezpracovaná data, která jsou poskytnuta v aktivní datové sadě pro odhad velikosti efektu.

### Jednotlivé studie

Poskytuje volby pro uložení odhadované velikosti efektu.

#### Velikost individuálního účinku

Uloží odhadovanou velikost efektu.

#### Velikost individuálního účinku (exponovaná forma exponentu)

Uloží odhadovanou velikost efektu do exponenciální podoby.

#### Standardní chyba

Uloží odhadovanou směrodatnou chybu velikosti efektu.

#### Dolní hranice intervalu spolehlivosti

Uloží odhadovanou mez intervalu spolehlivosti dolní meze velikosti efektu.

#### Horní hranice intervalu spolehlivosti

Uloží odhadovanou horní mez intervalu spolehlivosti o velikosti efektu.

#### Dolní hranice intervalu spolehlivosti (exponovaná forma exponent)

Uloží odhad intervalu spolehlivosti nižší, než je velikost efektu ve exponenciálním tvaru.

#### Horní hranice intervalu spolehlivosti (exponovaná forma exponent)

Uloží odhadovanou horní mez intervalu spolehlivosti do velikosti exponenciálu ve exponenciálním tvaru.

#### Hodnota P

Uloží odhadovanou hodnotu  $p$ -value efektu velikosti.

#### (%)

Uloží odhadovanou váhu studie.

#### Procentní část tělesné hmotnosti

Uloží normalizovanou váhu studie jako procentní část.

## Definování nastavení binárního ukládání metadat Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Binární meta-analýza** klepněte na tlačítko **Uložit**.

3. Vyberte a definujte odpovídající odhadované nastavení pro uložení statistiky.

4. Klepněte na tlačítko **Pokračovat**.

## Binární meta-analýza: Graf

Dialogové okno **Zakreslit** poskytuje nastavení pro následující typy výkresu:

- “Výkres lesa” na stránce 73
- “Kumulativní vykreslení lesa” na stránce 74
- “Zákresový” na stránce 75
- “Trychtýř” na stránce 76
- “Galbraitní nákres” na stránce 77
- “L' Abb'e Plot” na stránce 78

## Velikost binárního efektu metaanalýzy

Procedura Metaanalýza velikosti binárního efektu provádí meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předem vypočteného efektu.

### Příklad

Několik výzkumných studií bylo provedeno v historii, aby prošetřila faddish, ale sporný lék na pomoc při léčbě diabetu typu II. Perorální přípravek byl po jídle tvrdil, že je schopen snížit hladinu glukózy v krvi. Údaje byly shromažďovány z různých výzkumných míst od roku 1979 do roku 1986.

Hlavní vyšetřovatel by rád nakreslil statistické inference o účinku perorálního léku. Vzhledem k tomu, že údaje byly získány z různých studií, navrhla myšlenku syntetizovat výsledky v rámci studií za účelem dosažení celkového pochopení účinku a zjištění těchto základních zdrojů variace ve výsledcích.

### Statistika

Konfidenční interval, poměr log Odds, Poměr rizika protokolu, Poměr rizika protokolu, rozdíl rizik, náhodné účinky, fixní efekty, inverzní rozptyl, Mantel-Haenzel, iterace, krok-zastavení, konvergence, kumulativní statistika, kumulativní velikost účinku, maximální pravděpodobnost, maximální pravděpodobnost, ML, Empirical Bayes, Hedges, Hunter-Schmidt, DerSironian-Laird, Sidik-Jonkman, Knapp-Hartung, Egger's Test, Harbord's Test, Peters ' Test, intercept in regression, dispersion parameter, homogeniety, heterogeniety, exponentiated statistics, standard error,  $p$ -value, gravait weight.

## Analýza analýzy velikosti binárního efektu Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. Vyberte proměnnou **Effect Size**, která označuje velikost efektu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
3. Vyberte buď volbu **Směrodatná odchylka**, chcete-li určit směrodatnou odchylku velikosti efektu, nebo volbu **Odchylka**, chcete-li určit odchylku velikosti efektu a poté vybrat proměnnou, která bude představovat standardní odchylku/varpro velikost efektu.
4. Volitelně můžete vybrat proměnné **ID studie** a/nebo **Štítek studie**. Zvolená proměnná **Study ID** nemůže být stejná jako vybraná proměnná **Štítek studie**.
5. Volitelně můžete vybrat nastavení **Velikost efektu**. Dostupné volby jsou **Poměr pravděpodobnosti protokolu**, **Poměr protokolu odds protokolu**, **Poměr rizika protokolu** a **Rozdíl rizik**.
6. Volitelně můžete vybrat nastavení **Model**. Je-li povolena nastavení **Oříznout a-Vyplnit**, nastavení také řídí model, který je používán fondem v analýze obruby a výplně. Je-li nastavení **Bias** povoleno, nastavení také řídí model, který používá test založený na regresi.

### Náhodné účinky

Výchozí nastavení vytváří model s náhodnými efekty.

### Fixní-účinky

Sestaví model s neměnným účinkem.

7. Volitelně můžete:

- Klepněte na **Kritéria ...** specifikujte obecná kritéria.
  - Klepnutím na volbu **Analýza** určete podskupinu a kumulativní analýzu.
  - Chcete-li určit metody odhadu, klepněte na volbu **Inference** .
  - Klepnutím na tlačítko **Kontrast** otestujte kontrast testu.
  - Klepnutím na tlačítko **Bias** získáte přístup k předpojatosti publikování tím, že provedete test na regresní testu EGGER.
  - Klepnutím na volbu **Oříznout-a-Vyplnit** implementujte analýzu obruby publikování a výplně.
  - Chcete-li řídit výstupy tabulky, klepněte na tlačítko **Tisk** .
  - Klepnutím na tlačítko **Uložit** uložíte odhadovanou statistiku do aktivní datové sady.
  - Chcete-li určit vykreslení dat ve výstupu, klepněte na volbu **Zakreslit** .
8. Klepněte na tlačítko **OK**.

## Velikost binárního efektu meta-analýzy: Kritéria

Dialog **Kritéria** poskytuje nastavení pro uvedení kritérií pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Interval spolehlivosti

Volitelné nastavení určuje úroveň důvěry. Hodnota musí být číselná hodnota mezi 0 a 100. Předvolené nastavení je 95.

### Chybějící rozsah dat

Volitelná nastavení řídí, jak procedura zpracovává chybějící data.

### Vyloučit analýzu případů podle analýzy

Výchozí nastavení zahrnuje všechny případy s dostatkem dat na proměnných použitých v každé konkrétní analýze.

### Vyloučit případy jako litwise

Zahrnuje všechny případy s dostatečnými daty pro všechny proměnné použité ve všech analýzách uvedených v rámci procedury.

### Uživatelské chybějící hodnoty

Volitelná nastavení řídí způsob zpracování uživatelských chybějících hodnot.

### Vyloučit

Výchozí nastavení považuje uživatele za platné chybějící hodnoty.

### Zahrnout

Ignoruje uživatele, kterému chybí hodnota, a zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami.

## Iterace

### Maximum iterací

Volitelné nastavení určuje maximální počet iterací v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 100. Hodnota 0 znamená, že se neprovedou žádné iterace.

### Maximální krok-snížení

Volitelné nastavení určuje maximální krok-halving v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 5. Hodnota 0 znamená, že krok-hallování není použit.

### Konvergence

Volitelné nastavení určuje toleranci konvergence. Hodnota musí být jedna kladná hodnota. Nastavení je k dispozici při použití iterativních metod. Standardní hodnota je  $1E-6$ .

## Definování binárních kritérií velikosti metadat Meta-Analysis

1. Z nabídky vyberte:

## **Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu Meta-Analysis** klepněte na volbu **Kritéria**.
3. Vyberte a definujte příslušné nastavení kritérií.
4. Klepněte na tlačítko **Pokračovat**.

## **Velikost binárního efektu metaanalýzy: analýza**

Dialogové okno **Analýza** poskytuje nastavení pro uvedení podskupiny a kumulativní analýzy pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### **Analýza podskupin**

Vyberte proměnnou, která vyvolává analýzu podskupin. Proměnná nemůže být stejná jako proměnná definovaná pro **Kumulativní analýza**.

### **Kumulativní analýza**

Vyberte proměnnou, která vyvolává kumulativní analýzu, a na níž je provedena kumulativní analýza meta-analýzy. Proměnná nemůže být stejná jako proměnná definovaná pro **Analýza podskupiny**. Je-li vybrána volba **Vzestupně**, souhrnná analýza je založena na zadané proměnné ve vzestupném pořadí. Je-li vybrána volba **Sestupně**, souhrnná analýza je založena na zadané proměnné v sestupném pořadí.

### **Kumulativní statistika**

Poskytuje volby pro uložení odhadované kumulativní celkové velikosti efektu. Nastavení jsou k dispozici pouze v případě, že je vybrána proměnná **Kumulativní analýza**.

#### **Velikost kumulativního účinku**

Uloží odhadovanou kumulativní celkovou velikost celkového účinku.

#### **Velikost kumulativního účinku (exponovaná forma exponentu)**

Uloží odhadovanou kumulativní celkovou velikost celkového účinku v exponenciální podobě.

#### **Standardní chyba**

Uloží odhadovanou standardní chybu kumulativní celkové velikosti efektu.

#### **Dolní hranice intervalu spolehlivosti**

Uloží odhadovaný interval spolehlivosti dolní hranice kumulativní celkové velikosti efektu.

#### **Horní hranice intervalu spolehlivosti**

Uloží odhadovanou horní mez intervalu spolehlivosti kumulativní celkové velikosti efektu.

#### **Dolní hranice intervalu spolehlivosti (exponovaná forma exponent)**

Uloží odhadovaný interval spolehlivosti dolní limit kumulativní celkové velikosti celkového účinku ve exponenciální podobě.

#### **Horní hranice intervalu spolehlivosti (exponovaná forma exponent)**

Uloží odhadovaný interval spolehlivosti horní limit kumulativní celkové velikosti celkového účinku v exponenciálním tvaru.

#### **Hodnota P**

Uloží odhadovanou hodnotu  $p$ -hodnota kumulativní celkové velikosti efektu.

### **Cíl**

Poskytuje volby pro uvedení souboru dat typu save file nebo datového souboru. Je-li vybrána volba **Datová sada**, můžete zadat název nové datové sady (můžete zachovat výchozí název datové sady). Je-li vybrána volba **Datový soubor**, klepněte na tlačítko **Procházet ...**, abyste vybrali název souboru typu save file a umístění.

## **Definování nastavení analýzy velikosti binárního efektu Meta-Analysis**

1. Z nabídky vyberte:

### **Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu meta-analýza** klepněte na volbu **Analýza**.

3. Vyberte a definujte odpovídající nastavení analýzy.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost binárního efektu metaanalýzy: Inference

Dialogové okno **Inference** poskytuje nastavení pro uvedení metod odhadu pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Odhadce

Poskytuje nastavení pro určení odhadu.

#### Poznámky:

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také odhad, který se používá při slučování v analýze obruby a výplně.
- Je-li zadána nastavení Bias, řídí se také odhad, který používá test založený na regresi.

#### Omezená maximální pravděpodobnost (REML)

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

#### Maximální pravděpodobnost (ML)

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

#### Empirical Bayes

Použije iterativní metodu a vypočítá empirický odhad Bayes.

#### Hrany

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

#### Hunterová-

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

#### DerSimonian-Laird

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

#### Sidik-Jonkman

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

### Úprava standardní chyby

Poskytuje nastavení, které řídí, zda má být použita standardní úprava nastavení chyb Knapp-Hartung.

#### Poznámky:

- Je-li zadána volba Oříznout a Vyplnit nastavení, řídí se také standardní úprava chyb, která je použita při slučování v analýze obruby a výplně.
- Je-li zadána hodnota Bias, bude také použita standardní úprava chyb, kterou používá test založený na regresi.

#### Žádná úprava

Výchozí nastavení tuto úpravu nepoužije.

#### Použití úpravy Knapp-Hartung

Použije se seřizovací metoda Knapp-Hartung.

#### Použití redukované úpravy Knapp-Hartung

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## Definování nastavení odvozování binárních efektů analýzy Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu metaanalýzy** klepněte na volbu **Inference**.

3. Vyberte a definujte odpovídající nastavení odvozování.



4. Klepněte na tlačítko **Pokračovat**.

## Velikost binárního efektu meta-analýzy: Kontrast

Dialogové okno **Kontrast** poskytuje nastavení pro řízení kontrastní testu pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Proměnné

Seznam zobrazí všechny dostupné proměnné datové sady. Vyberte proměnné ze seznamu a přesuňte je do seznamu **Contrast (s)**.

### Kontrast (y)

Seznam identifikuje koeficienty, které jsou uloženy jako proměnné v aktivní datové sadě. Je povoleno více proměnných. Řetězcové proměnné nejsou podporovány.

### User-Input Coefficient Values

Poskytuje nastavení pro určení koeficientů kontrastu uživatele. Povoleny jsou pouze číselné hodnoty. Chcete-li formulovat platnou kontrastní test, musí se počet uvedených hodnot shodovat s počtem platných studií.

### Zobrazit exponovanou statistiku

Řídí zahrnutí statistiky exponenciálního formuláře. Je-li toto nastavení vybráno, výstup zahrnuje statistiku exponenciálního tvaru, včetně exponenciálního efektu velikosti a intervalu spolehlivosti. Nastavení je k dispozici, je-li **Velikost efektu** uvedena jako **Poměr lichých protokolů**, **Peto Logs Odd** nebo **Poměr rizika protokolu**.

## Definování nastavení kontrastu velikosti binárního efektu Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu Meta-Analysis** klepněte na volbu **Kontrast**.

3. Vyberte a definujte vhodná nastavení kontrastu.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost binárního efektu metaanalýzy: Bias

Dialogové okno **Bias** poskytuje nastavení pro povolení předpojatosti publikování provedením regresních testů pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### Zkouška na regresní přímce Egger

Výběrem této volby povolíte předpojatost publikování provedením testu na regresní testu Eggers.

### Proměnné

Seznam poskytuje všechny dostupné proměnné datové sady.

### Kovariáta (proměnné)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako proměnné. Je povoleno více proměnných covariates.

### Faktor (y)

Proměnné vybrané ze seznamu **Proměnné** se zpracovávají jako faktory. Je povoleno více faktorů.

### Zahrnout zachycení v regresi

Řídí výraz zachycení v regresním testu.

### Zahrnout parametr disperze v modelu s pevnou efektem

Řídí nastavení multiplikačního modelu a zavádí multiplikační disperzní parametr pro analýzu. Nastavení je k dispozici pouze v případě, že je vybrán model s pevným efektem.

### **Odhadnout statistiku na základě t-distribuce**

Řídí distribuci použitou v regresních testech. Nastavení je standardně povoleno, což odhaduje statistiku na základě *t*-rozdělení. Není-li toto nastavení vybráno, statistiky se odhadují na základě normální distribuce.

## **Definování nastavení předpojatosti binárních efektů analýzy Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu Meta-Analysis** klepněte na volbu **Bias**.

3. Vyberte a definujte odpovídající nastavení předpojatosti.

4. Klepněte na tlačítko **Pokračovat**.

## **Velikost binárního efektu meta-analýzy: Trim-and-Fill**

Dialogové okno **Trim-and-Fill** poskytuje nastavení pro implementaci analýzy obruby publikování a výplně pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### **Odhadnout počet chybějících studií**

Řídí analýzu obruby-a-výplň publikování. Výběrem tohoto nastavení povolíte jiná nastavení dialogových oken.

### **Ze strany Impute Studies**

Poskytuje volby pro uvedení strany trychtýřového grafu, na které jsou přisouzena chybějící studie.

#### **Stanoveno svahy Eggerově testu**

Výchozí nastavení určuje stranu založenou na odhadovaném sklonu testu Egger.

#### **Vlevo**

Impuuje levou stranu trychtýřku grafu.

#### **Vpravo**

Impuuje pravou stranu trychtýřového grafu.

### **Metoda**

Uvádí metodu pro odhad počtu chybějících studií.

#### **Lineární**

Výchozí nastavení vypočítá lineární odhad.

#### **Spustit**

Vypočte odhad spuštění.

#### **Kvadratické**

Vypočte kvadratický odhad.

### **Proces iterace**

Poskytuje nastavení pro určení odhadu iterace a standardní úpravy chyb.

#### **Model s pevnou efecí**

Je-li vybrána tato volba, je použit model s pevným efektem a nejsou k dispozici standardní volby úpravy iterace a nejsou k dispozici standardní volby úpravy chyby.

#### **Model náhodných efektů**

Je-li vybrána tato volba, je použit model s náhodnými efekty a jsou k dispozici následující nastavení.

#### **Odhadce**

Poskytuje nastavení pro určení odhadu iterace.

#### **Omezená maximální pravděpodobnost (REML)**

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

### **Maximální pravděpodobnost (ML)**

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

### **Empirical Bayes**

Použije iterativní metodu a vypočítá empirický odhad Bayes.

### **Hrany**

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

### **Hunterová-**

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

### **DerSimonian-Laird**

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

### **Sidik-Jonkman**

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

### **Úprava standardní chyby**

Poskytuje nastavení pro řízení, zda má být použita standardní úprava chyb Knapp-Hartung na iterace algoritmu obruby a výplně.

#### **Žádná úprava**

Výchozí nastavení tuto úpravu nepoužije.

#### **Použití úpravy Knapp-Hartung**

Použije se seřizovací metoda Knapp-Hartung.

#### **Použití redukované úpravy Knapp-Hartung**

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## **Definování nastavení trim-a-fill Meta Effect Meta**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu Meta-Analysis** klepněte na volbu **Oříznout a-Vyplnit**.

3. Vyberte a definujte odpovídající nastavení obruby a výplně.

4. Klepněte na tlačítko **Pokračovat**.

## **Velikost binárního efektu metaanalýzy: Tisk**

Dialogové okno **Tisk** poskytuje nastavení pro řízení výstupů tabulky pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### **Homogenita/heterogenita**

Poskytuje nastavení pro řízení homogenity a testů heterogenity.

#### **Zkouška homogenity**

Je-li tato volba vybrána, je ve výstupu uveden odpovídající test homogenity.

#### **Opatření heterogenity**

Je-li tato volba vybrána, jsou ve výstupu poskytnuta opatření heterogenity.

### **velikost účinku**

Poskytuje následující nastavení velikosti efektu.

#### **Jednotlivé studie**

Řídí zobrazení jednotlivých studií. Je-li tato volba vybrána, bude poskytnut odpovídající výstup.

#### **Velikosti kumulativního účinku**

Řídí zobrazení kumulativní analýzy. Je-li tato volba vybrána, bude poskytnut odpovídající výstup. Nastavení je k dispozici pouze v případě, že je v dialogovém okně **Analýza** vybrána proměnná

**Kumulativní analýza** .

### **Interval předpovědi v modelu s náhodnými účinky**

Řídí zobrazení intervalu předpovědi. Je-li tato volba vybrána, bude poskytnut odpovídající výstup. Nastavení je k dispozici pouze v případě, že je zadán model s náhodnými účinky.

### **Zobrazit exponovanou statistiku**

Řídí zahrnutí statistiky exponenciálního formuláře. Je-li vybráno nastavení, jsou výstupem exponenciální statistiky, včetně velikosti exponenciálního efektu a mezních hodnot intervalu spolehlivosti. Tento proces se týká jak povinných, tak i volitelných tabulek odhadu a předpovědi. Nastavení je k dispozici, je-li **Velikost efektu** uvedena jako **Poměr lýchých protokolů**, **Peto Logs Odd** **Poměrnebo** **Poměr rizika protokolu**.

## **Definování nastavení tisku velikosti binárního efektu Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Binární velikost efektu metadat** klepněte na tlačítko **Tisk**.

3. Vyberte a definujte odpovídající nastavení tisku.

4. Klepněte na tlačítko **Pokračovat**.

## **Velikost binárního efektu meta-analýzy: Uložit**

Dialogové okno **Uložit** poskytuje nastavení pro uložení odhadované statistiky na aktivní datovou sadu pro meta-analýzu s binárními výsledky, když jsou v aktivní datové sadě poskytnuta data o velikosti předvypočteného efektu.

### **Jednotlivé studie**

Poskytuje volby pro uložení odhadované velikosti efektu.

#### **Velikost individuálního účinku (exponovaná forma exponentu)**

Uloží odhadovanou velikost efektu do exponenciální podoby.

#### **Standardní chyba**

Uloží odhadovanou směrodatnou chybu velikosti efektu.

#### **Dolní hranice intervalu spolehlivosti**

Uloží odhadovanou mez intervalu spolehlivosti dolní meze velikosti efektu.

#### **Horní hranice intervalu spolehlivosti**

Uloží odhadovanou horní mez intervalu spolehlivosti o velikosti efektu.

#### **Dolní hranice intervalu spolehlivosti (exponovaná forma exponent)**

Uloží odhad intervalu spolehlivosti nižší, než je velikost efektu ve exponenciálním tvaru.

#### **Horní hranice intervalu spolehlivosti (exponovaná forma exponent)**

Uloží odhadovanou horní mez intervalu spolehlivosti do velikosti exponenciálu ve exponenciálním tvaru.

#### **Hodnota P**

Uloží odhadovanou hodnotu  $p$ -value efektu velikosti.

#### **(%)**

Uloží odhadovanou váhu studie.

#### **Procentní část tělesné hmotnosti**

Uloží normalizovanou váhu studie jako procentní část.

## **Definování nastavení uložení velikosti binárního efektu Meta-Analysis**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Binární výsledky > Velikost předvypočteného efektu ...**

2. V dialogovém okně **Velikost binárního efektu metaanalýzy** klepněte na tlačítko **Uložit**.

3. Vyberte a definujte odpovídající odhadované nastavení pro uložení statistiky.

4. Klepněte na tlačítko **Pokračovat**.

## Velikost binárního efektu metaanalýzy: Zakreslit

Dialogové okno **Zakreslit** poskytuje nastavení pro následující typy výkresu:

- “Výkres lesa” na stránce 73
- “Kumulativní vykreslení lesa” na stránce 74
- “Zákresový” na stránce 75
- “Trychtýř” na stránce 76
- “Galbraitní nákres” na stránce 77
- “L' Abb'e Plot” na stránce 78

## Regrese meta-analýzy

Regresní procedura Meta-analýza provádí analýzu meta-regrese.

### Příklad

Několik výzkumných studií bylo provedeno v historii, aby prošetřila faddish, ale sporný lék na pomoc při léčbě diabetu typu II. Perorální přípravek byl po jídle tvrdil, že je schopen snížit hladinu glukózy v krvi. Údaje byly shromažďovány z různých výzkumných míst od roku 1979 do roku 1986.

Hlavní vyšetřovatel by rád nakreslil statistické inference o účinku perorálního léku. Vzhledem k tomu, že údaje byly získány z různých studií, navrhla myšlenku syntetizovat výsledky v rámci studií za účelem dosažení celkového pochopení účinku a zjištění těchto základních zdrojů variace ve výsledcích.

### Statistika

Úroveň spolehlivosti, iterační metoda, krok-zastavení, směrodatná odchylka, vzorek odchylky, směrodatná odchylka, odhadnutá velikost účinku, metoda odhadu, regresní test, model s náhodnou účinku, model s pevným efektem, odhad maximální pravděpodobnosti, empirický odhad-odhadce hran, kvali-Hartungův odhad chyb, zkrácený Knapp-Hartungův standard-korekční úprava, zkrácený Knapp-Hartungův standard-korekční úprava Knapp-Hartungův standard-korekční úprava, koeficienty, koeficient koeficientu chyby, exponent statistiky.

## Získání regresní analýzy Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Meta regrese**

2. Vyberte jednu závislou proměnnou **Effect Size**, která označuje velikost efektu. Zvolená proměnná musí být číselná (řetězcové proměnné nejsou podporovány).
3. Vyberte jedno z následujících nastavení a poté vyberte příslušnou jedinou číselnou proměnnou:

### Standardní chyba

Vyberte proměnnou, která určuje standardní chybu, která je převedena na váhu. Toto je výchozí nastavení.

### Odchylka

Vyberte proměnnou, která uvádí odchylku, která se převede na váhu.

### Váha

Vyberte proměnnou, která určuje váhu.

4. Volitelně přidejte proměnné faktoru do seznamu **Faktor (faktory)**. Pro každou vybranou proměnnou faktoru lze zadat volitelnou hodnotu pro každou proměnnou, aby bylo možné určit vlastní **Poslední kategorie**. Můžete klepnout na tlačítko **Resetovat**, chcete-li obnovit výchozí hodnoty **Poslední kategorie**.

**Poznámka:** Nejsou-li k dispozici žádné případy, které by odpovídaly zadaným hodnotám **Poslední kategorie**, budou naposledy provedené hodnoty považovány za poslední kategorie.

5. Volitelně můžete vybrat číselné proměnné kovariate.

6. Volitelně můžete vybrat nastavení **Model** .

#### **Náhodné účinky**

Výchozí nastavení vytváří model s náhodnými efekty.

#### **Fixní-účinky**

Sestaví model s neměnným účinkem. Volitelně můžete vybrat nastavení **Zahrnout parametr dispersion** .

7. Volitelně můžete:

- Klepněte na **Kritéria ...** specifikujte obecná kritéria.
- Chcete-li určit metody odhadu, klepněte na volbu **Inference** .
- Chcete-li řídit výstupy tabulky, klepněte na tlačítko **Tisk** .
- Klepněte na tlačítko **Uložit** , chcete-li předpovědět a uložit odhadovanou statistiku do aktivní datové sady.
- Chcete-li určit vykreslení dat ve výstupu, klepněte na volbu **Zakreslit** .

8. Klepněte na tlačítko **OK**.

## **Regrese meta-analýza: kritéria**

Dialogové okno **Kritéria** obsahuje nastavení pro určení kritérií pro regresi meta-analýzy.

#### **Interval spolehlivosti**

Volitelné nastavení určuje úroveň důvěry. Hodnota musí být číselná hodnota mezi 0 a 100. Předvolené nastavení je 95.

#### **Uživatelské chybějící hodnoty**

Volitelná nastavení řídí způsob zpracování uživatelských chybějících hodnot.

##### **Vyloučit**

Výchozí nastavení považuje uživatele za platné chybějící hodnoty.

##### **Zahrnout**

Ignoruje uživatele, kterému chybí hodnota, a zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami.

#### **Iterace**

##### **Maximum iterací**

Volitelné nastavení určuje maximální počet iterací v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 100. Hodnota 0 znamená, že se neprovedou žádné iterace.

##### **Maximální krok-snížení**

Volitelné nastavení určuje maximální krok-halving v iteračních metodách. Nastavení je k dispozici při použití iterativních metod. Hodnota musí být jediné kladné celé číslo. Výchozí hodnota je 5. Hodnota 0 znamená, že krok-hallování není použit.

##### **Konvergence**

Volitelné nastavení určuje toleranci konvergence. Hodnota musí být jedna kladná hodnota. Nastavení je k dispozici při použití iterativních metod. Standardní hodnota je  $1E-6$ .

## **Definování kritérií regrese meta-analýzy**

1. Z nabídky vyberte:

#### **Analýza > Metaanalýza > Meta regrese**

2. V dialogovém okně **Regrese v metaanalýze** klepněte na volbu **Kritéria**.

3. Vyberte a definujte příslušné nastavení kritérií.

4. Klepněte na tlačítko **Pokračovat**.

## Regrese meta-analýza: inference

Dialogové okno **Inference** poskytuje nastavení pro uvedení metod odhadu pro regresi v meta-analýze.

### Zahrnout zachycení v regresi

Nastavení je standardně povoleno.

### Odhadnout statistiku na základě t-distribuce

Řídí distribuci použitou v regresních testech. Nastavení je standardně povoleno, což odhaduje statistiku na základě  $t$ -rozdělení. Není-li toto nastavení vybráno, statistiky se odhadují na základě normální distribuce.

### Odhadce

Poskytuje nastavení pro určení odhadu.

#### Omezená maximální pravděpodobnost (REML)

Výchozí nastavení použije iterační metodu a vypočítá omezený maximální odhad pravděpodobnosti.

#### Maximální pravděpodobnost (ML)

Použije iterativní metodu a vypočítá maximální odhad pravděpodobnosti.

#### Empirical Bayes

Použije iterativní metodu a vypočítá empirický odhad Bayes.

#### Hrany

Použije neiterativní metodu a vypočítá odhad Hhrabetor.

#### Hunterová-

Použije neiterativní metodu a vypočítá odhadce Hunter-Schmidt.

#### DerSimonian-Laird

Použije neiterativní metodu a vypočítá odhad Dermonian-Laird.

#### Sidik-Jonkman

Použije neiterativní metodu a vypočítá odhad Sidik-Jonkman.

### Úprava standardní chyby

Poskytuje nastavení, které řídí, zda má být použita standardní úprava nastavení chyb Knapp-Hartung.

#### Žádná úprava

Výchozí nastavení tuto úpravu nepoužije.

#### Použití úpravy Knapp-Hartung

Použije se seřizovací metoda Knapp-Hartung.

#### Použití redukované úpravy Knapp-Hartung

Použije metodu úpravy Knapp-Hartung a zkracuje hodnotu, pokud je při odhadu matice variance-kovariance menší než 1.

## Definování nastavení regrese regresní analýzy metadat

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Meta regrese**

2. V dialogovém okně **Regrese v metaanalýze** klepněte na volbu **Inference**.

3. Vyberte a definujte odpovídající nastavení odvozování.

4. Klepněte na tlačítko **Pokračovat**.

## Regrese meta-analýza: Tisk

Dialogové okno **Tisk** poskytuje nastavení pro řízení výstupů tabulky pro regresi meta-analýzy.

### Zkouška koeficientem modelu

Řídí test koeficientu modelu. Nastavení není ve výchozím nastavení zadáno, což potlačuje test. Je-li toto nastavení určeno, je ve výstupu uveden test.

## Zobrazit exponovanou statistiku

Řídí odhady parametrů. Nastavení není standardně uvedeno, což potlačí odhady parametrů. Je-li nastaveno nastavení, jsou ve výstupu poskytnuty odhady parametrů.

## Definování nastavení regresního tisku Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Meta regrese**

2. V dialogovém okně **Regrese v metaanalýze** klepněte na tlačítko **Tisk**.

3. Vyberte a definujte odpovídající nastavení tisku.

4. Klepněte na tlačítko **Pokračovat**.

## Regrese meta-analýza: Uložit

Dialogové okno **Uložit** poskytuje nastavení pro uložení odhadované statistiky na aktivní datovou sadu pro regresi v meta-analýze.

### Jednotlivé studie

Poskytuje volby pro uložení odhadované velikosti efektu.

#### Předpovězené hodnoty

Předvídejte a uloží nainstalované hodnoty.

#### Standardní chyba předpokládaných hodnot

Předvídá a uloží standardní chybu namontování hodnot.

#### Dolní hranice intervalu spolehlivosti

Uloží odhadovaný interval spolehlivosti dolní mez namontovaných hodnot.

#### Horní hranice intervalu spolehlivosti

Uloží odhadnutý interval spolehlivosti horní hranice namontování hodnot.

#### Reziduum

Předvídá a uloží zbytkové chyby.

#### Standardní chyba zbytkových chyb

Předvídá a uloží standardní chybu zbytkových chyb.

#### Nápoje

Předvídejte a šetří to průměry.

### Kumulativní statistika

Poskytuje volby pro ukládání lineárních předpovědí.

**Poznámka:** Kumulativní nastavení statistiky se ignorují, je-li jako **Model** vybráno **Pevné efekty**.

#### Pevné lineární předpovědi

Předvídejte a uložte lineární předpověď.

#### Směrodatná chyba fixních lineárních předpovědí

Předvídá a uloží standardní chybu lineární předpovědi.

#### Nejlepší lineární nezaujaté předpovědi

Předvídejte a uloží nejlepší lineární nepředpojatou předpověď náhodných efektů.

#### Standardní chyba BLUP

Předvídejte a uloží standardní chybu v nejlepší lineární nepředpovědné předpovědi náhodných efektů.

## Definování nastavení úspory regrese Meta-Analysis

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Meta regrese**

2. V dialogovém okně **Regrese v metaanalýze** klepněte na tlačítko **Uložit**.



3. Vyberte a definujte odpovídající nastavení ukládání.

4. Klepněte na tlačítko **Pokračovat**.

## Regrese meta-analýza: Zakreslit

Dialogové okno **Zakreslit** poskytuje nastavení pro definování bublinových grafů. Více informací naleznete v části "[Zákresový](#)" na stránce 75.

## Meta-analýza-volby

Pro procedury Meta-Analysis jsou k dispozici následující volby vykreslení a jsou přístupné v dialogových oknech **Zákresové**.

### Výkres lesa

Karta **Forest plot** poskytuje nastavení pro řízení pruhových grafů, které se zobrazují ve výstupu pro meta-analýzu s průběžným a binárním výsledkem, když jsou v aktivní datové sadě poskytnuta data o raw data a data o velikosti předvypočteného efektu.

#### Lesní pozemek

Volitelné nastavení povoluje a zakazuje výstup do lesního grafu. Nastavení je ve výchozím nastavení vypnuto.

#### Sloupce pro zobrazení

Volitelná nastavení určují odhadovanou statistiku, která je tabulová.

#### Velikost efektu

Zahrnuje odhadovanou velikost efektu.

#### standardní chyba

Zahrne standardní chybu odhadované velikosti efektu.

#### Limity intervalu spolehlivosti

Zahrnuje interval spolehlivosti dolní a horní meze odhadované velikosti efektu.

#### Hodnota P

Obsahuje hodnotu  $p$ -value odhadované velikosti efektu pro každou jednotlivou studii.

#### Váha

Zahrnuje hmotnost primárních studií.

#### Zobrazit exponovanou formu

**Poznámka:** Toto nastavení platí pouze pro meta-analýzu s binárními výsledky.

Řídí sloupec zákresového grafu a zobrazení efektu velikosti a intervalu spolehlivosti. Když je vybráno nastavení, vynáší poměr pravděpodobnosti, Poměr pravděpodobnosti k poměru, poměr rizika a poměr rizik a tabuluje exponenciální statistiku. Není-li toto nastavení vybráno, vynáší se poměr pravděpodobnosti protokolu, poměr pravděpodobnosti protokolu Peto nebo poměr rizika protokolu a tabuluje statistiky transformované protokolem.

#### Proměnné

Vypisuje dostupné proměnné datové sady.

#### Další sloupce

Volitelně můžete vybrat proměnné, které se mají zobrazit jako další sloupce. Pořadí zobrazení sloupců určuje pořadí proměnných.

#### Řadit

Volitelně můžete určit proměnnou, podle které je zalesněné vykreslení seřazeno. Je-li zadána proměnná, je pralesnický graf seřazen v pořadí **Vzestupně** (výchozí nastavení). Je-li vybrána volba **Sestupně**, je pralesnický graf seřazen sestupně.

#### Pozice sloupku výkresu

Řídí umístění sloupce výkresu.

### Vpravo

Výchozí nastavení umístí sloupec do zákresového sloupce vpravo od ostatních sloupců tabulky.

### Vlevo

Umístí sloupec výkresu do levé části ostatních sloupců tabulky.

### Referenční čáry

Řídí referenční čáry, které se přidávají do zalesněné plochy. Volba **Celková velikost efektu** přidá řádek, který označuje odhadovanou celkovou velikost efektu. **Efekt efektu null** přidá řádek k označení velikosti efektu null.

### Anotace

Řídí zobrazení anotací.

### Homogenita

Vypíše statistické údaje o testu homogenity.

### Heterogenní

Vypíše statistické údaje o testu heterogenity.

### Testovat

Vytiskne test celkové velikosti efektu a testu homogenity mezi podskupinami za přítomnosti analýzy podskupin.

### Oblast oříznutí

Určuje rozsah plodin. Je-li toto nastavení vybráno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ .

Pro meta-analýzu s binárními výsledky, je-li nastavení **Velikost efektu** nastaveno na hodnotu **Rozdíl v riziku**, musí také vyhovovat  $Upper\ bound \leq 1$  a  $Lower\ bound \geq -1$ . Je-li vybráno nastavení **Zobrazit exponovanou formu**, musí také splňovat  $Lower\ bound \geq 0$ .

## Definování nastavení globálního výkresu

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Souvislé výstupy nebo binární výstup > Nezpracovaná data ...** nebo **Před vypočtenou velikostí efektu ...**

2. V dialogovém okně **Metaanalýza** klepněte na volbu **Zákresový graf** a poté klepněte na kartu **Výkres lesa**.

3. Vyberte a definujte odpovídající nastavení nákresu v lese.

4. Klepněte na tlačítko **Pokračovat**.

## Kumulativní vykreslení lesa

Ouško **Kumulativní vykreslení doménové struktury** obsahuje nastavení pro řízení kumulativních grafů zákresových grafů, které se zobrazují ve výstupu pro metaanalýzu s průběžným a binárním výsledkem, když jsou v aktivní datové sadě poskytnuta data o raw data a data o velikosti předvypočteného efektu.

### Kumulativní zalesněné plochy

Volitelné nastavení povoluje a zakazuje výstup do lesního grafu. Nastavení je ve výchozím nastavení vypnuto.

### Sloupce pro zobrazení

Volitelná nastavení určují odhadovanou statistiku, která je tabulová.

### Velikost kumulativního účinku

Zahrnuje kumulativní celkovou velikost efektu.

### standardní chyba

Zahrne standardní chybu celkové kumulativní velikosti celkového účinku.

### Limity intervalu spolehlivosti

Zahrnuje interval spolehlivosti dolní a horní meze kumulativní celkové velikosti účinku.

## Hodnota P

Obsahuje  $p$ -hodnotu kumulativní celkové velikosti účinku pro každou jednotlivou studii.

## Zobrazit exponovanou formu

**Poznámka:** Toto nastavení platí pouze pro meta-analýzu s binárními výsledky.

Řídí sloupec zákresového grafu a zobrazení efektu velikosti a intervalu spolehlivosti. Je-li toto nastavení vybráno, vykreslí poměr pravděpodobnosti, poměr pravděpodobnosti Peto, poměr rizika nebo poměr rizika a tabulátorová statistika exponenciálního tvaru. Pokud není nastavení povoleno, vynáší poměr šancí protokolu, poměr pravděpodobnosti protokolu Peto nebo poměr rizika protokolu a tabuluje statistické údaje transformované protokolem.

## Proměnné

Vypisuje dostupné proměnné datové sady.

## Další sloupce

Volitelně můžete vybrat proměnné, které se mají zobrazit jako další sloupce. Pořadí zobrazení sloupců určuje pořadí proměnných.

## Pozice sloupku výkresu

Řídí umístění sloupce výkresu.

### Vpravo

Výchozí nastavení umístí sloupek do zákresového sloupce vpravo od ostatních sloupců tabulky.

### Vlevo

Umístí sloupek výkresu do levé části ostatních sloupců tabulky.

## Oblast oříznutí

Určuje rozsah plodin. Je-li toto nastavení vybráno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ .

Pro meta-analýzu s binárními výsledky, je-li nastavení **Velikost efektu** nastaveno na hodnotu **Rozdíl v riziku**, musí také vyhovovat  $Upper\ bound \leq 1$  a  $Lower\ bound \geq -1$ . Je-li vybráno nastavení **Zobrazit exponovanou formu**, musí také splňovat  $Lower\ bound \geq 0$ .

## Definování kumulativních nastavení lesního výkresu

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Souvislé výstupy nebo binární výstup > Nezpracovaná data ...** nebo **Před vypočtenou velikostí efektu ...**

2. V dialogovém okně **Metaanalýza** klepněte na volbu **Zákresový graf** a poté na kartu **Kumulativní zákresový graf**.

3. Vyberte a definujte odpovídající kumulativní nastavení pro vykreslení forest.

4. Klepněte na tlačítko **Pokračovat**.

## Zákresový

Karta **Bublinový graf** poskytuje nastavení pro řízení bublinových zákresových grafů, které se zobrazují ve výstupu pro meta-analýzu s průběžným a binárním výsledkem, když jsou v aktivní datové sadě poskytnuta prvotní data a data o velikosti předvypočteného efektu.

**Poznámka:** Bublinový zákresový graf je jediný dostupný typ grafu pro regresní analýzu meta-analýzy.

### Zákresový

Volitelné nastavení povoluje a zakazuje výstup v bublinovém grafu. Nastavení je ve výchozím nastavení vypnuto.

**Poznámka:** Toto nastavení není dostupné pro regresní analýzu meta-analýzy.

## Proměnné

Vypisuje dostupné proměnné datové sady.

## Prediktor (y)

Určete proměnné, které jsou považovány za souvislé prediktory na ose X. Vícenásobné proměnné jsou povoleny při každém vytvoření samostatného bublinového grafu.

**Poznámka:** Pro regresní analýzu meta-analýza je toto nastavení označeno **Bublinový zákresový predikát (s)** a umožňuje vám vybrat proměnné, pro které se mají vytvořit bublinové grafy.

## Prediktor, který se středem v průměru (s)

Volitelná nastavení řídí spojitý prediktor na ose X. Nastavení je ve výchozím nastavení vypnuto, což prediktorem neuvede do středu. Je-li nastavení povoleno, prediktorem se rovná střední-střed.

## Popisek

Volitelně určete proměnnou, na které je vykreslený bublinový graf označen. Seznam **Umístění** poskytuje volby pro automatické umístění nálepek (**Automatický**) nebo na **Jasně., Vlevo, Velká a Poslední**.

## Kreslit bubliny proporcionální k váze

Řídí, jak se bubliny kreslí ve vztahu k váze. Nastavení je standardně povoleno, což nakreslí bubliny proporcionálně k váhám. Je-li nastavení zakázáno, budou všechny bubliny vykresleny do stejné velikosti.

## Zobrazit namontovaný řádek

Řídí zobrazení naložené regresní přímkou. Nastavení je standardně povoleno, což přidá k bublinovému grafu zabudovanou regresní čáru. Je-li nastavení zakázáno, nepřidaná regresní čára se nepřidá do bublinového grafu.

## Zobrazit meze intervalu spolehlivosti

Řídí zobrazení hranic intervalu spolehlivosti. Nastavení je standardně povoleno, což přidá meze intervalu spolehlivosti k bublinovému grafu. Je-li nastavení zakázáno, vazby intervalu spolehlivosti se nepřidají do bublinového grafu.

## Rozsah osy X

Určuje rozsah zákresu na ose X. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ . Všimněte si, že uvedené hodnoty platí pro všechny bublinové grafy.

## Rozsah osy y

Určuje rozsah zákresové osy Y. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ . Všimněte si, že uvedené hodnoty platí pro všechny bublinové grafy.

Pro meta-analýzu s binárními výsledky, je-li nastavení **Velikost efektu** nastaveno na hodnotu **Rozdíl v riziku**, musí také vyhovovat  $Upper\ bound \leq 1$  a  $Lower\ bound \geq -1$ .

## Definování nastavení bublinového grafu

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Souvislé výstupy nebo binární výstup > Nezpracovaná data ...** nebo **Před vypočtenou velikostí efektu ...**

nebo

**Analýza > Metaanalýza > Meta regrese**

2. V dialogovém okně **Metaanalýza** klepněte na volbu **Zakreslit** a poté klepněte na kartu **Bublinový zákres**.

3. Vyberte a definujte odpovídající nastavení bublinového grafu.

4. Klepněte na tlačítko **Pokračovat**.

## Trychtýř

Karta **Výkres trychtýře** poskytuje nastavení pro řídicí grafy trychtýře, které se zobrazí ve výstupu pro meta-analýzu s průběžným a binárním výsledkem, když jsou v aktivní datové sadě poskytnuta data o raw data a data o velikosti předvypočteného efektu.

## Trychtýř

Volitelné nastavení povoluje a zakazuje výstup zákresového grafu. Nastavení je ve výchozím nastavení vypnuto.

## Hodnoty osy Y

Řídí hodnoty osy Y.

### Standardní chyba

Vykreslí standardní chybu.

### Inverzní standardní chyba

Vykreslí inverzní standardní chybu.

### Odchylka

Spoustu rozptylu.

### Inverzní odchylka

Vykreslí inverzní odchylku.

## Rozsah osy X

Určuje rozsah zákresu na ose X. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ . Všimněte si, že uvedené hodnoty se vztahují na všechny trychtýřové grafy.

Pro meta-analýzu s binárními výsledky, je-li nastavení **Velikost efektu** nastaveno na hodnotu **Rozdíl v riziku**, musí také vyhovovat  $Upper\ bound \leq 1$  a  $Lower\ bound \geq -1$ .

## Rozsah osy y

Určuje rozsah zákresové osy Y. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound \geq 0$ . Všimněte si, že uvedené hodnoty se vztahují na všechny trychtýřové grafy.

## Zahrnout imputované studie s obrubou a vyplněním

Řídí, zda jsou imputované studie zahrnuty do grafu trychtýře, když je v dialogovém okně **Oříznout a-Vyplnit** povoleno nastavení **Odhadnout počet chybějících studií**. Nastavení je standardně vypnuto, což vylučuje imputované studie. Je-li vybrána volba **Zobrazit celkovou velikost pozorovaných studií**, bude přidána svislá referenční čára, která označuje odhadovanou celkovou velikost efektu. Toto nastavení bere v úvahu jak pozorované, tak i imputované studie.

## Proměnné

Vypisuje dostupné proměnné datové sady.

## Popisek

Volitelně můžete určit proměnnou, na které je popisek trychtýře označen. Seznam **Umístění** poskytuje volby pro automatické umístění nálepek (**Automatický**) nebo na **Jasně., Vlevo, Velkáa Poslední**.

## Definování nastavení zákresového grafu

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Souvislé výstupy nebo binární výstup > Nezpracovaná data ...** nebo **Před vypočtenou velikostí efektu ...**

2. V dialogovém okně **Metaanalýza** klepněte na volbu **Zákresový graf** a poté klepněte na kartu **Výkres trychtýř**.

3. Vyberte a definujte odpovídající nastavení zákresového grafu.

4. Klepněte na tlačítko **Pokračovat**.

## Galbraithní nákres

Ouško **Galbraith Plot** poskytuje nastavení pro řízení grafů Galbraith, které se zobrazují ve výstupu pro metaanalýzu s průběžným a binárním výsledkem, když jsou v aktivní datové sadě poskytnuta data o raw data a data o velikosti předvypočteného efektu.

### **Galbraitský pozemek**

Volitelné nastavení povoluje a zakazuje výstup v bublinovém grafu. Nastavení je ve výchozím nastavení vypnuto.

### **Zobrazit meze intervalu spolehlivosti**

Řídí zobrazení hranic intervalu spolehlivosti. Nastavení je standardně povoleno, což přidá meze intervalu spolehlivosti k bublinovému grafu. Je-li nastavení zakázáno, vazby intervalu spolehlivosti se nepřidají do bublinového grafu.

### **Rozsah osy X**

Určuje rozsah zákresu na ose X. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound \geq 0$ . Všimněte si, že uvedené hodnoty se vztahují na všechny diagramy Galbraith.

### **Rozsah osy y**

Určuje rozsah zákresové osy Y. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ . Všimněte si, že uvedené hodnoty se vztahují na všechny diagramy Galbraith.

### **Proměnné**

Vypisuje dostupné proměnné datové sady.

### **Popisek**

Volitelně určete proměnnou, na které je vykreslený bublinový graf označen. Seznam **Umístění** poskytuje volby pro automatické umístění nálepek (**Automatický**) nebo na **Jasně., Vlevo, Velkáa Poslední**.

## **Definování nastavení pro Galbraith**

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Souvislé výstupy nebo binární výstup > Nezpracovaná data ...** nebo **Před vypočtenou velikostí efektu ...**

2. V dialogovém okně **Metaanalýza** klepněte na volbu **Zakreslit** a poté klepněte na kartu **Galbraith Plot**.

3. Vyberte a definujte odpovídající nastavení výkresu programu Galbraith.

4. Klepněte na tlačítko **Pokračovat**.

## **L' Abb'e Plot**

Karta **L' Abb'e Plot** poskytuje nastavení pro řízení grafů L' Abb'e, které se zobrazují ve výstupu pro meta-analýzu s průběžným a binárním výsledkem, když jsou v aktivní datové sadě poskytnuty raw data a data o velikosti předvypočteného efektu.

### **Výkres L' Abb'e**

Volitelné nastavení zapíná a vypíná výstup výkresu L' Abb'e. Nastavení je ve výchozím nastavení vypnuto.

### **Kreslit body úměrné závažími**

Řídí, jak se kreslí tečky ve vztahu k váhám. Nastavení je ve výchozím nastavení povoleno, což teze proporcionálně proporcionálně odpovídá. Je-li nastavení zakázáno, všechny body se kreslí na stejnou velikost.

### **Referenční čáry**

#### **Celková velikost účinku**

Řídí zobrazení referenční čáry, která představuje odhadovanou celkovou velikost efektu. Nastavení je standardně vypnuto, což potlačuje referenční čáru. Referenční čára se zobrazí, když je nastavení povoleno.

#### **Nulová hodnota efektu**

Řídí zobrazení referenční čáry, která představuje žádný efekt. Nastavení je standardně povoleno, což zobrazuje referenční čáru. Referenční řádek je potlačen, když není nastavení povoleno.

### Rozsah osy X

Určuje rozsah zákresu na ose X. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ . Všimněte si, že uvedené hodnoty se vztahují na všechny diagramy Galbraith.

Pro meta-analýzu s binárními výsledky, je-li nastavení **Velikost efektu** nastaveno na hodnotu **Rozdíl v riziku**, musí také vyhovovat  $Lower\ bound \geq 0$  a  $Upper\ bound \leq 1$ . Je-li nastavení **Velikost efektu** nastaveno na hodnotu **Poměr rizika protokolu**, musí splňovat  $Upper\ bound \leq 0$ .

### Rozsah osy y

Určuje rozsah zákresové osy Y. Je-li nastavení povoleno, musí být zadány dvě číselné hodnoty a musí vyhovovat  $Upper\ bound > Lower\ bound$ . Všimněte si, že uvedené hodnoty se vztahují na všechny diagramy Galbraith.

Pro meta-analýzu s binárními výsledky, je-li nastavení **Velikost efektu** nastaveno na hodnotu **Rozdíl v riziku**, musí také vyhovovat  $Lower\ bound \geq 0$  a  $Upper\ bound \leq 1$ . Je-li nastavení **Velikost efektu** nastaveno na hodnotu **Poměr rizika protokolu**, musí splňovat  $Upper\ bound \leq 0$ .

### Proměnné

Vypisuje dostupné proměnné datové sady.

### Popisek

Volitelně určete proměnnou, na které je vykreslený bublinový graf označen. Seznam **Umístění** poskytuje volby pro automatické umístění nálepek (**Automatický**) nebo na **Jasně., Vlevo, Velkáa Poslední**.

## Definování nastavení zákresového grafu L' Abb'e

1. Z nabídky vyberte:

**Analýza > Metaanalýza > Souvislé výstupy nebo binární výstup > Nezpracovaná data ...**

2. V dialogovém okně **Metaanalýza** klepněte na volbu **Zakreslit** a poté klepněte na kartu **L' Abb'e Plot**.

3. Vyberte a definujte odpovídající nastavení výkresu L' Abb'e.

4. Klepněte na tlačítko **Pokračovat**.

## Kódová kniha

---

Kódová kniha uvádí informace o slovníku -- jako např. názvy proměnných, popisky proměnných, popisky hodnot, chybějící hodnoty -- a souhrnné statistiky pro všechny nebo zadané proměnné a více sad odpovědí v aktivní datové sadě. Pro nominální a ordinální proměnné a více sad odpovědí zahrnuje souhrnné statistiky počty a procenta. Pro proměnné stupnice jsou souhrnné statistiky střední, směrodatné odchylky a kvartily.

Poznámka: Kódová kniha ignoruje stav rozděleného souboru. Patří sem skupiny rozdělení souborů vytvořené pro více imputace chybějících hodnot (dostupné ve volbě Chybějící hodnoty pro chybějící hodnoty).

Získání kódové knihy

1. Z nabídky vyberte:

**Analyzovat > Sestavy > Kódová kniha**

2. Klepněte na kartu Proměnné.

3. Vyberte jednu nebo více proměnných a/nebo více sad odpovědí.

Volitelně můžete:

- Řídit informace o proměnné, které se zobrazí.
- Řídit statistiku, která se zobrazí (nebo vyloučí všechny souhrnné statistiky).
- Stanovte pořadí, ve kterém jsou zobrazeny proměnné a více sad odpovědí.

- Změňte úroveň měření pro každou proměnnou ve zdrojovém seznamu tak, abyste změnili souhrnné statistiky zobrazené. Další informace naleznete v tématu [“Karta Statistika kódové knihy”](#) na stránce [81](#).

Změna úrovně měření

Můžete dočasně změnit úroveň měření pro proměnné. (Úroveň měření nemůžete změnit pro více sad odpovědí. Ty jsou vždy považovány za nominální.)

1. Klepněte pravým tlačítkem myši na proměnnou ve zdrojovém seznamu.
2. Vyberte úroveň měření z rozevírací nabídky.

Tím se dočasně změní úroveň měření. V praxi to je užitečné pouze pro numerické proměnné. Úroveň měření pro řetězcové proměnné je omezena na nominální nebo pořadové číslo, které se s procedurou Codebook zachází stejně jako s procedurou kódové knihy.

## Karta Výstup z kódové knihy

Karta Výstup řídí informace o proměnné zahrnuté v každé proměnné a vícenásobné sadě odezvy, pořadí, ve kterém se zobrazují proměnné a více sad odpovědí, a obsah tabulky s informacemi o volitelném souboru.

Informace o proměnné

Tento ovládací prvek určuje informace o slovníku zobrazené pro každou proměnnou.

**Umístění.** Celé číslo, které představuje pozici proměnné v pořadí souborů. Tato volba není k dispozici pro více sad odpovědí.

**Popisek.** Popisný popisek přidružený k proměnné nebo sadě více odpovědí.

**Typ.** Základní datový typ. To je buď *Číslo*, *Řetězec*, nebo *Vícenásobná sada odpovědí*.

**Formát.** Formát zobrazení pro proměnnou, například *A4*, *F8.2* nebo *DATE11*. Tato volba není k dispozici pro více sad odpovědí.

**Úroveň měření.** Možné hodnoty jsou *Nominální*, *Ordinální*, *Měřítka* a *Neznámé*. Zobrazená hodnota je úroveň měření uložená ve slovníku a není ovlivněna žádnou dočasnou úrovní úrovně měření zadanou změnou úrovně měření ze seznamu zdrojových proměnných na kartě Proměnné. Tato volba není k dispozici pro více sad odpovědí.

Poznámka: Úroveň měření pro číselné proměnné může být "neznámá" před prvním průchodem dat, když úroveň měření nebyla explicitně nastavena, jako například data přečtená z externího zdroje nebo nově vytvořené proměnné. Další informace naleznete v tématu .

**Funkce.** Některé dialogy podporují schopnost předvýběru proměnných pro analýzu založenou na definovaných rolích.

**Popisky hodnot.** Popisné popisky přidružené ke specifickým datovým hodnotám.

- Je-li na kartě Statistika vybrán počet nebo procento, jsou do výstupu zahrnuty popisky definovaných hodnot i v případě, že zde nevyberete hodnotu Popisky hodnot.
- Pro více sad dichotomie jsou "popisky hodnot" buď popisky proměnných pro elementární proměnné v sadě, nebo popisky započítaných hodnot, v závislosti na tom, jak je definována sada. Další informace naleznete v tématu .

**Chybějící hodnoty.** Uživatelem definované chybějící hodnoty. Je-li na kartě Statistika vybrán počet nebo procento, jsou do výstupu zahrnuty popisky definovaných hodnot i v případě, že zde nevyberete chybějící hodnoty. Tato volba není k dispozici pro více sad odpovědí.

**Vlastní atributy.** Uživatelem definované atributy vlastních proměnných. Výstup obsahuje jak názvy, tak i hodnoty pro všechny vlastní proměnné atributy přidružené k jednotlivým proměnným. Další informace naleznete v tématu . Tato volba není k dispozici pro více sad odpovědí.

**Vyhrazené atributy.** Vyhrazené atributy systémových proměnných. Systémové atributy můžete zobrazit, ale neměli byste je měnit. Názvy systémových atributů začínají znakem dolaru (\$). Nezobrazované



atributy, jejichž názvy začínají řetězcem "@" nebo "\$@", nejsou zahrnuty. Výstup zahrnuje jak názvy, tak i hodnoty pro jakékoli systémové atributy přidružené k jednotlivým proměnným. Tato volba není k dispozici pro více sad odpovědí.

Informace o souboru

Volitelná tabulka informací o souboru může obsahovat kterýkoli z následujících atributů souboru:

**Název souboru.** Název datového souboru IBM SPSS Statistics . Pokud nebyla datová sada nikdy uložena ve formátu IBM SPSS Statistics , pak neexistuje žádný název datového souboru. (Pokud v pruhu titulku v okně Editor dat není zobrazen žádný název souboru, pak aktivní datová sada nemá název souboru.)

**Umístění.** Umístění adresáře (složky) datového souboru IBM SPSS Statistics . Pokud nebyla datová sada nikdy uložena ve formátu IBM SPSS Statistics , pak neexistuje žádné umístění.

**Počet případů.** Počet případů v aktivní datové sadě. Jedná se o celkový počet případů, včetně všech případů, které mohou být vyloučeny ze souhrnných statistik kvůli podmínkám filtru.

**Popisek.** Jedná se o jmenovku souboru (je-li nějaká) definovaný příkazem FILE LABEL .

**Dokumenty.** Text dokumentu datového souboru.

**Stav váhy.** Je-li váha zapnuta, zobrazí se název proměnné váhy. Další informace naleznete v tématu .

**Vlastní atributy.** Uživatelem definované atributy vlastních datových souborů. Atributy datového souboru definované pomocí příkazu DATAFILE ATTRIBUTE .

**Vyhrazené atributy.** Vyhrazené atributy datového souboru systému. Systémové atributy můžete zobrazit, ale neměli byste je měnit. Názvy systémových atributů začínají znakem dolaru (\$). Nezobrazované atributy, jejichž názvy začínají řetězcem "@" nebo "\$@", nejsou zahrnuty. Výstup obsahuje jak názvy, tak i hodnoty pro všechny atributy systémového datového souboru.

Pořadí zobrazení proměnných

Pro řízení pořadí, ve kterém jsou zobrazeny proměnné a více sad odpovědí, jsou k dispozici následující alternativy.

**Abecedně.** Abecední pořadí podle názvu proměnné.

. Pořadí, ve kterém se proměnné objeví v datové sadě (pořadí, ve kterém jsou zobrazeny v editoru dat). Ve vzestupném pořadí se po všech vybraných proměnných zobrazí více sad odpovědí za poslední.

**Úroveň měření.** Řadit podle úrovně měření. Tím se vytvoří čtyři třídící skupiny: nominální, ordinální, měřítko a neznámé. S více sadami odpovědí se zachází jako s nominální hodnotou.

Poznámka: Úroveň měření pro číselné proměnné může být "neznámá" před prvním průchodem dat, když úroveň měření nebyla explicitně nastavena, jako například data přečtená z externího zdroje nebo nově vytvořené proměnné.

**Seznam proměnných.** Pořadí, ve kterém se proměnné a více sad odpovědí zobrazí ve vybraném seznamu proměnných na kartě Proměnné.

**Název vlastního atributu.** Seznam voleb pořadí řazení také obsahuje názvy všech uživatelsky definovaných vlastních atributů proměnných. Ve vzestupném pořadí proměnné, které nemají řazení atributu na začátek, následují proměnné, které mají atribut, ale nemají definovanou hodnotu pro atribut, za nímž následují proměnné s definovanými hodnotami atributu v abecedním pořadí hodnot.

Maximální počet kategorií

Pokud výstup zahrnuje popisky hodnot, počty nebo procenta pro každou jedinečnou hodnotu, můžete tyto informace potlačit z tabulky, pokud počet hodnot překročí zadanou hodnotu. Při výchozím nastavení jsou tyto informace potlačeny, pokud počet jedinečných hodnot pro proměnnou překročí 200.

## Karta Statistika kódové knihy

Karta Statistika vám umožňuje řídit souhrnnou statistiku, která je zahrnuta ve výstupu, nebo potlačit zobrazení souhrnné statistiky zcela.

## Počty a procenta

Pro nominální a ordinální proměnné, více sad odpovědí a označených hodnot proměnných měřítka jsou k dispozici následující statistiky:

*Počet.* Počet nebo počet případů, které mají každou hodnotu (nebo rozsah hodnot) proměnné.

*Procento.* Procento případů, které mají konkrétní hodnotu.

## Střední tendence a disperze

Pro proměnné měřítka jsou k dispozici následující statistiky:

*Střední.* Měřítka centrální tendence. Aritmetický průměr, součet dělený počtem případů.

*Směrodatná odchylka.* Ukazatel disperze kolem střední hodnoty. V normálním rozložení se 68% případů propadne do jedné směrodatné odchylky střední hodnoty a 95% případů spadá do dvou směrodatných odchylek. Například, je-li střední věk 45, se směrodatnou odchylkou 10, 95% případů by mělo být mezi 25 a 65 v normálním rozdělení.

*kvartily.* Zobrazí hodnoty odpovídající hodnotám 25th, 50th a 75th percentilů.

Poznámka: Můžete dočasně změnit úroveň měření přidruženou k proměnné (a tím změnit souhrnnou statistiku zobrazenou pro tuto proměnnou) ve zdrojovém seznamu proměnných na kartě Proměnné.

## Frekvence

---

Procedura Frekvence poskytuje statistiky a grafické zobrazení, které jsou užitečné pro popis mnoha typů proměnných. Postup frekvence je dobré místo pro zahájení pohledu na vaše data.

Pro sestavu frekvence a sloupcový graf můžete uspořádat odlišné hodnoty ve vzestupném nebo sestupném pořadí, nebo můžete řadit kategorie podle jejich četnosti. Hlášení o frekvencích může být potlačeno, když má proměnná mnoho odlišných hodnot. Můžete označit grafy s frekvencemi (výchozí) nebo procentními podíly.

### Příklad

Jaká je distribuce zákazníků společnosti podle typu odvětví? Z výstupu se můžete dozvědět, že 37,5% vašich zákazníků jsou ve vládních agenturách, 24,9% jsou ve společnostech, 28,1% jsou v akademických institucích a 9,4% je ve zdravotnickém průmyslu. Pro souvislé, kvantitativní data, jako např. výnos z prodeje, můžete zjistit, že průměrný prodej produktu je \$3,576, se standardní odchylkou 1 078 USD.

### Statistika a vykreslení

Počet frekvencí, procenta, kumulativní procenta, střední, medián, režim, součet, směrodatná odchylka, odchylka, rozsah, minimální a maximální hodnoty, standardní chyba průměru, šikmosti a špičatosti (obojí se standardními chybami), kvartily, uživatelem zadané percentily, pruhové grafy, výšečové grafy a histogramy.

## Aspekty dat

### Data

Použijte číselné kódy nebo řetězce k kódům kategoriických proměnných (měření nominální nebo ordinální úrovně).

### Předpoklady

Tabulace a procenta poskytují užitečný popis pro data z libovolného rozdělení, zejména pro proměnné s objednanými nebo neobjednanými kategoriemi. Většina z volitelných souhrnných statistik, jako například střední a směrodatná odchylka, vychází z běžné teorie a je vhodná pro kvantitativní proměnné s symetrickými distribucemi. Robustní statistika, jako je medián, kvartily a percentily, je vhodná pro kvantitativní proměnné, které mohou nebo nemusí splňovat předpoklad normality.

## Získávání frekvenčních tabulek

1. Z nabídky vyberte:

## Analyzovat > Deskriptivní statistika > Frekvence ...

2. Vyberte jednu nebo více kategoriálních nebo kvantitativních proměnných.
3. Volitelně můžete výběrem volby **Vytvořit tabulky stylu APA** vytvořit výstupní tabulky, které budou vyhovovat pokynům stylu APA.
4. Volitelně můžete:
  - Pro deskriptivní statistiky pro kvantitativní proměnné klepněte na volbu **Statistika**.
  - Klepněte na volbu **Grafy** pro pruhové grafy, výsečové grafy a histogramy.
  - Klepněte na **Formát** pro pořadí, ve kterém se zobrazí výsledky.
  - Chcete-li určit podmínky pro automatické změny vlastností kontingenčních tabulek na základě specifických podmínek, klepněte na volbu **Styl**.
  - Klepněte na volbu **Zaváděcí program**, chcete-li odvodit robustní odhady standardních chyb a intervalů spolehlivosti pro odhady, jako např. střední, medián, poměr, poměr pravděpodobnosti, korelační koeficient nebo regresní koeficient. Může být také použit pro vytváření testů hypotéz.

## Statistika frekvencí

**Procentuální hodnoty.** Hodnoty kvantitativní proměnné, které rozdělují řazená data do skupin tak, že určité procento je nad a jiné procento je pod. Kvartily (25th, 50th, 75th percentily) rozdělují pozorování do čtyř skupin stejné velikosti. Chcete-li mít stejný počet jiných skupin než čtyři, vyberte volbu **Vyjmout body pro n stejné skupiny**. Můžete také uvést jednotlivé percentily (například percentil 95th, níže uvedenou hodnotu, která 95% z pozorování se sníží).

**Střední Tendence.** Statistika, která popisuje umístění distribuce, zahrnuje střední, medián, režim a součet všech hodnot.

- *Střední.* Měřítka centrální tendence. Aritmetický průměr, součet dělený počtem případů.
- *Medián.* Hodnota nad a pod níž polovina případů padá, 50th percentil. Existuje-li sudý počet případů, střední je průměr dvou středních případů, kdy jsou seřazeny ve vzestupném nebo sestupném pořadí. Medián je měřítkem centrální tendence, která není citlivá na odlehlé hodnoty (na rozdíl od střední hodnoty, která může být ovlivněna několika extrémně vysokými nebo nízkými hodnotami).
- *Režim.* Nejčastěji se vyskytující hodnota. Pokud se více hodnot sdílí s největší frekvencí výskytu, každý z nich je v režimu. Procedura Frekvence oznamuje pouze nejmenší z těchto vícenásobných režimů.
- *Součet.* Součet nebo součet hodnot ve všech případech s nechybějícími hodnotami.

**Disperze.** Statistické údaje, které měří množství změn nebo rozptylu v datech, zahrnují směrodatnou odchylku, rozsoh, minimum, maximum a směrodatnou chybu střední hodnoty.

- *Směrodat. odchylka.* Ukazatel disperze kolem střední hodnoty. V normálním rozložení se 68% případů propadne do jedné směrodatné odchylky střední hodnoty a 95% případů spadá do dvou směrodatných odchylek. Například, je-li střední věk 45, se směrodatnou odchylkou 10, 95% případů by mělo být mezi 25 a 65 v normálním rozdělení.
- *Odchylka.* Ukazatel disperze kolem střední hodnoty rovnající se součtu čtverců odchylky od průměru vyděleného o jeden menší než počet případů. Odchylka se měří v jednotkách, které jsou čtvercem těch, které mají vlastní proměnnou.
- *Rozsoh.* Rozdíl mezi největší a nejmenší hodnotou číselné proměnné, maximum minus minimum.
- *Minimální.* Nejmenší hodnota číselné proměnné.
- *Maximální.* Největší hodnota číselné proměnné.
- *S. E. střední hodnotu.* Měřítka toho, jak hodnota střední hodnoty se může lišit od vzorku na vzorek odebraného ze stejného rozdělení. Lze ji použít k hrubšímu porovnání sledovaného průměru s hypotetizovanou hodnotou (to znamená, že můžete uzavřít dvě hodnoty se liší, pokud poměr rozdílu mezi standardní chybou je menší než -2 nebo větší než + 2).

**Distribuce.** Skewness a kurtóza jsou statistiky, které popisují tvar a symetrii rozdělení. Tyto statistiky se zobrazují s jejich standardními chybami.

- **Šikmost.** Měřitko asymetrie rozdělení. Normální rozdělení je symetrické a má hodnotu šikmosti 0. Rozdělení s významnou kladnou šikmou má dlouhý ocas. Rozdělení s významnou negativní šikmou má dlouhý levý ocas. Jako vodítko se bere hodnota šikmosti více než dvojnásobek standardní chyby k označení odchylky od symetrie.
- **Kurtóza.** Měřitko rozsahu, v jakém jsou přehlé hodnoty. Pro běžnou distribuci je hodnota statistiky špičatosti rovna nule. Kladná špičatost indikuje, že data vykazují extrémnější odlehle hodnoty než normální rozdělení. Negativní špičatost indikuje, že data vykazují méně extrémní odlehle hodnoty než normální rozdělení. Definice kurtózy, která se používá, je-li hodnota 0 pro normální distribuci, je někdy označována jako nadměrná špičatost. Některé softwarové produkty mohou vykazovat špičatost tak, že hodnota je 3 pro normální distribuci.

**Hodnoty jsou středy skupiny.** Jsou-li hodnoty ve vašich datech středy skupin (například stáří všech lidí ve svých třicítce jsou kódovány jako 35), vyberte tuto volbu pro odhad mediánu a percentilů pro původní neseskupená data.

## Grafy četnosti

**Poznámka:** Grafy se ve výstupu nevytvorí, když je v dialogovém okně Zaváděcí program povoleno **Provádět bootstrapping**.

### Typ grafu

Výsečový graf zobrazuje příspěvek částí k celku. Každá výseč výsečového grafu odpovídá skupině definované jednou proměnnou seskupení. Pruhový graf zobrazuje počet pro každou jednotlivou hodnotu nebo kategorii jako samostatný pruh, což vám umožňuje vizuálně porovnat kategorie. Histogram má také pruhy, ale jsou vyneseny podél rovnoměrné stupnice intervalu. Výška každého pruhu je počet hodnot kvantitativní proměnné, která spadá do intervalu. Histogram ukazuje tvar, střed a rozložení distribuce. Normální křivka uložená na histogramu pomáhá posoudit, zda jsou data normálně distribuována.

### Hodnoty grafu

Pro pruhové grafy lze osu stupnice označit podle počtu frekvencí nebo procent.

## Formát frekvence

**Seřadit podle.** Frekvenční tabulka může být uspořádána podle skutečných hodnot v datech nebo podle počtu (frekvence výskytu) těchto hodnot a tabulka může být uspořádána ve vzestupném nebo sestupném pořadí. Pokud však požadujete histogram nebo percentily, četnost předpokládá, že proměnná je kvantitativní a zobrazí její hodnoty ve vzestupném pořadí.

**Více proměnných.** Pokud vytvoříte statistické tabulky pro více proměnných, můžete buď zobrazit všechny proměnné v jedné tabulce (**Porovnat proměnné**), nebo zobrazit samostatnou tabulku statistiky pro každou proměnnou (**Uspořádat výstup podle proměnných**).

**Potlačit tabulky s mnoha kategoriemi.** Tato volba zabraňuje zobrazení tabulek s více než určeným počtem hodnot.

## Popisovače

Procedura Descriptives zobrazuje jedinečné souhrnné statistické údaje pro několik proměnných v jedné tabulce a počítá normalizované hodnoty (skóre). Proměnné lze řadit podle velikosti jejich středních hodnot (ve vzestupném nebo sestupném pořadí), abecedně nebo podle pořadí, ve kterém vyberete proměnné (výchozí nastavení).

Když jsou skóre z ukládána, přidávají se do dat v editoru dat a jsou k dispozici pro grafy, výpisy dat a analýzy. Pokud jsou proměnné zaznamenány v různých jednotkách (například hrubý domácí produkt na osobu a procento gramotnosti), transformace skóre zumístí proměnné ve společném měřítku pro snazší vizuální porovnání.

**Příklad.** Pokud každý případ ve vašich datech obsahuje celkové denní prodeje pro každého člena prodejního personálu (například jedna položka pro Bob, jedna položka pro Kim a jedna položka pro Brian) shromážděné každý den po dobu několika měsíců, může procedura s deskriptivními ukazateli vypočítat

průměrný denní prodej každého zaměstnance a může objednávat výsledky od nejvyššího průměrného prodeje k nejnižším průměrným prodejům.

**Statistika.** Ukázková velikost, průměr, minimum, maximum, směrodatná odchylka, odchylka, rozsah, součet, směrodatná chyba střední hodnoty a kurtóza a šikmost s jejich standardními chybami.

Posouzení dat s deskriptivními ukazateli

**Data.** Numerické proměnné použijte až poté, co je vizuálně prověřujete, chcete-li zaznamenat chyby, odlehlé hodnoty a distribuční anomálie. Procedura Descriptives je velmi efektivní pro velké soubory (v tisících případech).

**Předpoklady.** Většina dostupných statistik (včetně skóre z ) je založena na normální teorii a je vhodná pro kvantitativní proměnné (měření na úrovni intervalu nebo poměru úrovní poměru) se symetrickými distribucemi. Vyhnout se proměnným s netříděnými kategoriemi nebo zkreslené distribucí. Distribuce skóre z má stejný tvar jako u původních dat. Výpočetní skóre z proto není lékem pro data problému.

Jak získat popisné statistiky

1. Z nabídky vyberte:

**Analýza > Deskriptivní statistiky > deskriptivní statistiky ...**

2. Vyberte jednu nebo více proměnných.

Volitelně můžete:

- Vyberte volbu **Uložit standardizované hodnoty jako proměnné** , chcete-li uložit skóre z jako nové proměnné.
- Klepněte na **Volby** pro volitelnou statistiku a pořadí zobrazení.

## Deskriptivní volby

**Průměr a součet.** Střední hodnota nebo aritmetický průměr se zobrazí jako výchozí.

**Disperze.** Statistika, která měří rozptyl nebo změnu dat, zahrnuje směrodatnou odchylku, odchylku, rozsah, minimum, maximum a směrodatnou chybu střední hodnoty.

- *Směrodat. odchylka.* Ukazatel disperze kolem střední hodnoty. V normálním rozložení se 68% případů propadne do jedné směrodatné odchylky střední hodnoty a 95% případů spadá do dvou směrodatných odchylek. Například, je-li střední věk 45, se směrodatnou odchylkou 10, 95% případů by mělo být mezi 25 a 65 v normálním rozdělení.
- *Odchylka.* Ukazatel disperze kolem střední hodnoty rovnající se součtu čtverců odchylky od průměru vyděleného o jeden menší než počet případů. Odchylka se měří v jednotkách, které jsou čtvercem těch, které mají vlastní proměnnou.
- *Rozsah.* Rozdíl mezi největší a nejmenší hodnotou číselné proměnné, maximum minus minimum.
- *Minimální.* Nejmenší hodnota číselné proměnné.
- *Maximální.* Největší hodnota číselné proměnné.
- *S.E. střední hodnota.* Měřítka toho, jak hodnota střední hodnoty se může lišit od vzorku na vzorek odebraného ze stejného rozdělení. Lze ji použít k hrubšímu porovnání sledovaného průměru s hypotetizovanou hodnotou (to znamená, že můžete uzavřít dvě hodnoty se liší, pokud poměr rozdílu mezi standardní chybou je menší než -2 nebo větší než + 2).

**Distribuce.** Kurtóza a šikmosti jsou statistiky, které charakterizují tvar a symetrii rozdělení. Tyto statistiky se zobrazují s jejich standardními chybami.

- *Kurtóza.* Měřítka rozsahu, v jakém jsou přelehlé hodnoty. Pro běžnou distribuci je hodnota statistiky špičatosti rovna nule. Kladná špičatost indikuje, že data vykazují extrémnější odlehlé hodnoty než normální rozdělení. Negativní špičatost indikuje, že data vykazují méně extrémní odlehlé hodnoty než normální rozdělení. Definice kurtózy, která se používá, je-li hodnota 0 pro normální distribuci, je někdy označována jako nadměrná špičatost. Některé softwarové produkty mohou vykazovat špičatost tak, že hodnota je 3 pro normální distribuci.

- **Šikmost.** Měřítka asymetrie rozdělení. Normální rozdělení je symetrické a má hodnotu šikmosti 0. Rozdělení s významnou kladnou šikmou má dlouhý ocas. Rozdělení s významnou negativní šikmou má dlouhý levý ocas. Jako vodítko se bere hodnota šikmých hodnoty více než dvojnásobek standardní chyby k označení odchylky od symetrie.

**Pořadí zobrazení.** Ve výchozím nastavení jsou proměnné zobrazeny v pořadí, ve kterém jste je vybrali. Volitelně můžete zobrazit proměnné abecedně, podle vzestupného způsobem, nebo sestupně.

## Další funkce příkazu DESCRIPTIVA

Jazyk syntaxe příkazu vám také umožňuje:

- Uložit standardizované skóre (z skóre) pro některé, ale ne všechny proměnné (s dílčím příkazem VARIABLES).
- Zadejte názvy nových proměnných, které obsahují standardizované skóre (s dílčím příkazem VARIABLES).
- Vyloučit z analytických případů s chybějícími hodnotami pro každou proměnnou (s dílčím příkazem MISSING).
- Třídít proměnné na obrazovce podle hodnoty libovolné statistiky, ne pouze průměrné hodnoty (s dílčím příkazem SORT).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## percentily

Procedura Percentily zobrazuje tabulky percentilů.

### Hodnoty percentilů

Vyberete-li volbu **Procentuální hodnoty**, při výchozím nastavení se zobrazí hodnoty pro 5th, 10th, 25th, 50th, 75th, 90th a 95th percentily. Hodnoty kvantitativní proměnné, které rozdělují řazená data do skupin tak, že určité procento je nad a jiné procento je pod.

#### kvartily

Kvartily (25th, 50th a 75th percentily) rozdělují pozorování do čtyř skupin stejné velikosti.

#### Vlastní

Když vyberete volbu **Vlastní**, zadejte alespoň jednu hodnotu pro spuštění analýzy. Vstupní hodnoty musí být číslo v rozsahu 0 až 100. Pomocí tlačítek **Přidat**, **Změnit** a **Odebrat** můžete pracovat s hodnotami v seznamu percentilů hodnot.

### Metoda percentilů

Ve výchozím nastavení je metoda HAVERAGE vybrána pro výpočet percentilů.

## Získání tabulek percentilů

1. Z nabídky vyberte:

**Analýza > Deskriptivní statistika > Percentily ..**

2. Vyberte jednu nebo více číselných proměnných a přesuňte je do pole Proměnné.

3. Chcete-li určit hodnoty pro výpočet percentilu, vyberte volbu **Procentuální hodnoty**.

4. Chcete-li vybrat metodu percentilu výpočtu, vyberte volbu **Metoda percentilů**.

5. Volitelně můžete:

- Chcete-li řídit zpracování chybějících hodnot, klepněte na tlačítko **Chybí**.
- Klepnutím na volbu **Zaváděcí program** odvodíte robustní odhady standardních chyb a intervalů spolehlivosti pro odhady percentilů.

## Percentily Chybějící hodnoty

### Chybějící hodnoty

Řídí zpracování chybějících hodnot.

### Vyloučení případů s Lisewise

Případy s chybějícími hodnotami pro kteroukoli uvedenou proměnnou jsou vyloučeny ze všech analýz. Jedná se o výchozí nastavení.

### Podle proměnné

Vyloučit případy s chybějícími hodnotami na bázi proměnné podle proměnné.

## Prozkoumat

---

Procedura Prozkoumat vytváří souhrnné statistiky a grafické zobrazení buď pro všechny vaše případy, nebo samostatně pro skupiny případů. Existuje mnoho důvodů pro použití procedury průzkumu --data screening, outlier identification, description, assuming checking a characterizing differences among subpopulations (groups of cases). Prověření dat může ukázat, že máte neobvyklé hodnoty, extrémní hodnoty, mezery v datech nebo jiných zvláštností. Prozkoumání dat vám může pomoci určit, zda jsou statistické metody zvažující pro analýzu dat vhodné. Prozkoumání může označovat, že potřebujete transformovat data, pokud tato technika vyžaduje normální distribuci. Nebo se můžete rozhodnout, že potřebujete neparametrické testy.

**Příklad.** Podívejte se na distribuci bludiště-učovací doby pro krysy pod čtyři různé výztuže plány. Pro každou ze čtyř skupin můžete zjistit, zda je distribuce časů přibližně normální a zda jsou čtyři rozptyly stejné. Můžete také identifikovat případy s pěti největšími a pěti nejkratšími časy. Krabičkové grafy a diagramy kmenových a kmenových buněk graficky shrnují rozdělení času učení pro každou skupinu.

**Statistika a grafy.** Průměr, medián, 5% zkrácená chyba, směrodatná odchylka, směrodatná odchylka, minimum, maximum, rozsah, mezikvartilní rozsah, šikmost a špičatost a jejich standardní chyby, interval spolehlivosti pro střední hodnotu (a stanovenou úroveň důvěry), percentily, Huber M-estimator, Andrews ' wave estimator, Hampelův resetizátor M-estimator, pět největších a pět nejmenších hodnot, statistiky Kolmogorov-Smirnov s úrovní významnosti Kolmogorov-Smirnov s úrovní významnosti Lilliefors pro testování normality a statistiky Shapiro-Wilk. Boxplots, stem-and-leaf plots, histogramy, normality plots, and spread-versus-level plots with Levene tests and transformations.

Aspekty zkoumání dat

**Data.** Postup prozkoumání lze použít pro kvantitativní proměnné (měření na úrovni intervalu nebo poměru míry). Proměnná faktorů (použitá pro rozdělení dat do skupin případů) by měla mít přiměřený počet odlišených hodnot (kategorie). Tyto hodnoty mohou být krátké řetězce nebo numerické. Proměnná popisku případu použitá k označení odlehklých hodnot v krabičkových grafech může být krátký řetězec, dlouhý řetězec (prvních 15 bajtů) nebo číselný.

**Předpoklady.** Distribuce vašich dat nemusí být symetrická nebo normální.

Prozkoumání vašich dat

1. Z nabídky vyberte:

**Analyzovat > Deskriptivní statistika > Prozkoumat ...**

2. Vyberte jednu nebo více závislých proměnných.

Volitelně můžete:

- Vyberte jednu nebo více proměnných faktorů, jejichž hodnoty nadefinují skupiny případů.
- Vyberte identifikační proměnnou pro případy popisku.
- Klepněte na volbu **Statistika** pro robustní odhady, přelivy, percentily a tabulky četnosti.
- Klepněte na volbu **Plot** pro histogramy, běžné hodnoty a testy pravděpodobnosti a testy rozložení na úrovni s úrovní Leveneovy statistiky.
- Klepněte na **Volby** pro zacházení s chybějícími hodnotami.

## Prozkoumat statistiku

**Popisovače.** Tato opatření centrální tendence a rozptylu se zobrazují jako výchozí. Míry centrální tendence označují umístění distribuce; zahrnují střední, střední a 5% oříznuté hodnoty. Ukazatele disperze ukazují nepodobnost hodnot; mezi ně patří standardní chyba, odchylka, směrodatná odchylka, minimum, maximum, rozsah a interkvartilní rozsah. Popisná statistika rovněž zahrnuje opatření tvaru rozdělení; zkršenost a špičatost se zobrazují s jejich standardními chybami. Interval spolehlivosti 95% pro střední hodnotu se také zobrazí; můžete uvést odlišnou úroveň důvěry.

**M-estimators.** Robustní alternativy k odběru vzorků střední a střední hodnoty pro odhad umístění. Odhadci, kteří se spočítají, se liší od vah, které se používají na případy. Huber je M-estimator, Andrews ' wave estimator, Hampel reseed-estimator M-estimator, a Tukeyův biweight odhad se zobrazí.

**Odlehlé hodnoty.** Zobrazí pět největších a pět nejmenších hodnot s popisky případů.

### percentily

Procedura Percentily zobrazuje tabulky percentilů. Vyberete-li volbu **Percentil**, při výchozím nastavení se zobrazí hodnoty pro 5th, 10th, 25th, 50th, 75th, 90th a 95th percentily. Hodnoty kvantitativní proměnné, které rozdělují řazená data do skupin tak, že určité procento je nad a jiné procento je pod.

### kvartily

Kvartily ( 25th, 50th a 75th percentily) rozdělují pozorování do čtyř skupin stejné velikosti.

### Vlastní

Když vyberete volbu **Vlastní**, zadejte alespoň jednu hodnotu pro spuštění analýzy. Vstupní hodnoty musí být číslo v rozsahu 0 až 100. Pomocí tlačítek **Přidat**, **Změnit** a **Odebrat** můžete pracovat s hodnotami v seznamu percentilů hodnot.

### Metoda

Ve výchozím nastavení je metoda HAVERAGE vybrána pro výpočet percentil.

## Prozkoumat paty

**Krabičkové grafy.** Tyto alternativy řídí zobrazení krabičkových grafů, když máte více než jednu závislou proměnnou. **Úrovně faktorů dohromady** generují oddělené zobrazení pro každou závislou proměnnou. V rámci obrazovky jsou krabičkové grafy zobrazeny pro každou ze skupin definovaných proměnnou faktoru. **Závislé na sobě** vytvoří oddělené zobrazení pro každou skupinu definovanou proměnnou faktoru. V rámci obrazovky jsou krabičkové grafy zobrazeny vedle sebe pro každou závislou proměnnou. Tato obrazovka je zvláště užitečná, když různé proměnné představují jednu charakteristiku měřenou v různých časech.

**Popisné.** Popisná skupina vám umožňuje vybrat diagramy kmenových a listů a histogramy.

**Normality grafů s testy.** Zobrazí normální pravděpodobnost a detrendovat normální hodnoty pravděpodobnosti. Zobrazí se statistika Kolmogorov-Smirnov s úrovní významnosti Lilliefors pro testování normality. Jsou-li zadány neceločíselné váhy, statistika Shapiro-Wilk se vypočítá, když se velikost váženého vzorku nachází mezi 3 a 50. Pro žádné váhy nebo celočíselné váhy se statistika vypočítá, když se vážená velikost vzorku nachází mezi 3 a 5000.

**Rozložení versus úroveň s Levenem Test.** Řídí transformaci dat pro grafy s rozpětím a úrovní. Pro všechny plochy na úrovni rozložení se zobrazí sklon regresní přímky a Levene v robustních zkouškách homogenity rozptylu. Zvolíte-li transformaci, Levenovy testy jsou založeny na transformovaných datech. Není-li vybrána žádná proměnná faktoru, nejsou vykreslené grafy s úrovní rozložení dat typu spread-s-level. **Odhad napájení** vytváří zakres přirozených logů mezikvartilového rozpětí proti přírodním protokolům mediánů pro všechny buňky, stejně jako odhad energetické přeměny pro dosažení stejných rozptylů v buňkách. Zámkový graf na úrovni rozložení pomáhá určit sílu transformace ke stabilizaci (více stejných) rozptylů ve skupinách. **Transformované** vám umožňuje vybrat jednu z možností výměny energie, možná podle doporučení ze odhadu spotřeby energie, a vytvořit grafy transformovaných dat. Je vykreslen interkvartilový rozsah a střední hodnota transformovaných dat. Volba **Netransformovaná** produkuje grafy nezpracovaných dat. To je ekvivalentní k transformaci s mocnou 1.



## Prozkoumat napájecí sestavy

Jedná se o transformace napájení pro grafy s rozpětím na úrovni. Chcete-li transformovat data, musíte vybrat sílu pro transformaci. Můžete zvolit jednu z následujících možností:

- **Přirozený protokol.** Přirozená transformace protokolu. Jedná se o výchozí nastavení.
- **1/square kořen.** Pro každou datovou hodnotu se vypočítá převrácená hodnota druhé odmocniny.
- **Recipro.** Vypočítává se reciproční hodnota jednotlivých datových hodnot.
- **Druhá odmocnina.** Vypočítává se druhá odmocnina z každé hodnoty dat.
- **Náměstí.** Jednotlivé datové hodnoty jsou čtvercové.
- **Krychle.** Každá datová hodnota je krychleová.

## Volby prozkoumání

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy jako litwise.** Případy s chybějícími hodnotami pro všechny závislé nebo faktorové proměnné jsou vyloučeny ze všech analýz. Jedná se o výchozí nastavení.
- **Vyloučit případy po dvojicích.** Případy bez chybějících hodnot pro proměnné ve skupině (buňce) jsou zahrnuty do analýzy této skupiny. Případ může mít chybějící hodnoty pro proměnné použité v jiných skupinách.
- **Hodnoty sestavy.** Chybějící hodnoty pro proměnné faktoru jsou považovány za samostatnou kategorii. Všechny výstupy jsou vytvořeny pro tuto dodatečnou kategorii. Tabulky frekvence zahrnují kategorie pro chybějící hodnoty. Chybějící hodnoty pro proměnnou faktoru jsou zahrnuty, ale označeny jako chybějící.

## Dodatečné funkce příkazu EXAMINE

Procedura Prozkoumat používá syntaxi příkazu EXAMINE . Jazyk syntaxe příkazu vám také umožňuje:

- Vyžádat celkový výstup a grafy kromě výstupu a grafů pro skupiny definované proměnnými faktoru (s dílčím příkazem TOTAL ).
- Uvedte společné měřítko pro skupinu krabčkových grafů (s dílčím příkazem SCALE ).
- Určete interakce s proměnnými faktoru (s dílčím příkazem VARIABLES ).
- Určete percentily jiné než standardní hodnoty (s dílčím příkazem PERCENTILES ).
- Vypočítat percentily podle jakékoli z pěti metod (s dílčím příkazem PERCENTILES ).
- Určete libovolnou transformaci napájení pro grafy s rozpětím a s úrovní (s dílčím příkazem PLOT ).
- Uvedte počet extrémních hodnot, které se mají zobrazit (spolu s dílčím příkazem STATISTICS ).
- Zadejte parametry pro měřiče M-estimators, robustní odhad umístění (s dílčím příkazem MESTIMATORS ).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Kontingenční tabulky

Procedura kontingenčních tabulek vytváří dvoucestné a vícecestné tabulky a poskytuje různé testy a ukazatele přidružení pro dvousměrné tabulky. Struktura tabulky a zda jsou kategorie seřazeny podle kategorií, určují, jaký test nebo ukazatel se má použít.

S výjimkou parciálních koeficientů gama se výpočty kontingenčních tabulek a ukazatelů přidružení vypočítávají zvláště pro každou dvoucestnou tabulku. Uvedete-li řádek, sloupec a faktor vrstvy (řídící proměnná), procedura Kontingenční tabulky tvoří jeden panel přidružených statistik a ukazatelů pro každou hodnotu faktoru vrstvy (nebo kombinaci hodnot pro dvě nebo více řídících proměnných). Například, pokud *pohlaví* je faktorem vrstvy pro tabulku *ženatá* (ano, ne) proti *životě* (je vzrušující, rutinní nebo matná), výsledky pro dvoucestnou tabulku pro ženy se vypočítají odděleně od těch pro muže a potisknou se jako panely za sebou.

**Příklad.** Jsou zákazníci z malých společností s vyšší pravděpodobností ziskový při prodeji služeb (například školení a konzultace) než od větších společností? Z přechodu se můžete naučit, že většina malých společností (méně než 500 zaměstnanců) přináší vysoké zisky služeb, zatímco většina velkých společností (více než 2500 zaměstnanců) přináší nízké zisky služeb.

**Statistika a opatření přidružení.** Pearson Chi-kvadrát, pravděpodobnost-poměr chi-square, lineární-in-linear association test, Fisher je přesný test, Yates 'opravené chi-náměstí, Pearson's  $\chi^2$ , Spearman's rho, Noční koeficient,  $F$ , Cramér je  $V$ , symetrický a asymetrický lambdas, Goodman a Kruskal je tau, nejistota součinitel, gama, Somers' 44%, Kendall's tau- $b$ , Kendall's tau- $k$ , cena koeficient, Cohen je kappa, relativní odhad rizika, poměr pravděpodobnosti, McNemarský test, Cochran a Mantel-Haenszel statistiky a statistiky o poměry sloupců.

Aspekty kontingenčních dat

**Data.** Chcete-li definovat kategorie každé proměnné tabulky, použijte hodnoty číselné nebo řetězcové proměnné (osm nebo méně bajtů). Například pro *gender* můžete kódovat data jako 1 a 2 nebo jako *male* a *female*.

**Předpoklady.** Některé statistiky a ukazatele předpokládají objednané kategorie (ordinální data) nebo kvantitativní hodnoty (interval nebo poměr dat), jak je probíráno v oddílu o statistice. Jiné jsou platné, pokud proměnné tabulky mají neřazené kategorie (nominální data). Pro statistiku chí-kvadrát statistiky ( $F$ , Cramér  $V$  a nouzový koeficient) by měla být data náhodného vzorku z polynomiální distribuce.

*Poznámka:* Ordinální proměnné mohou být buď číselné kódy, které představují kategorie (například 1 = nízká, 2 = střední, 3 = vysoká) nebo řetězcová hodnota. Avšak abecední pořadí řetězcových hodnot se předpokládá, aby odráželo skutečné pořadí kategorií. Například pro řetězcovou proměnnou s hodnotami *nízké, střední, vysoké* je pořadí kategorií interpretováno jako *vysoká, nízké, střední*-- což není správné pořadí. Obecně lze říci, že je spolehlivější používat číselné kódy k reprezentaci pořadových dat.

Jak získat křížové tabulky

1. Z nabídky vyberte:

**Analyzovat > Deskriptivní statistika > Kontingenční tabulky ...**

2. Vyberte jednu nebo více proměnných řádku a jednu nebo více sloupcových proměnných.

Volitelně můžete:

- Vyberte jednu nebo více řídících proměnných.
- Klepněte na tlačítko **Statistika** pro testy a ukazatele přidružení pro dvoucestné tabulky nebo dílčí tabulky.
- Klepněte na **Buňky** pro sledované a očekávané hodnoty, procentní části a zbytkové chyby.
- Chcete-li řídit pořadí kategorií, klepněte na volbu **Formát**.

## Vrstvy kontingenční tabulky

Vyberete-li jednu nebo více proměnných vrstvy, bude pro každou kategorii každé proměnné vrstvy (ovládací proměnná) vytvořena oddělená kontingenční tabulka. Máte-li například jednu proměnnou řádku, jednu proměnnou sloupcu a jednu proměnnou vrstvy se dvěma kategoriemi, získáte dvoucestnou tabulku pro každou kategorii proměnné vrstvy. Chcete-li vytvořit další vrstvu řídících proměnných, klepněte na tlačítko **Další**. Podtabulky jsou vytvořeny pro každou kombinaci kategorií pro každou proměnnou první vrstvy, každou druhou proměnnou vrstvy atd. Jsou-li požadovány statistiky a ukazatele přidružení, použijí se pouze na dvoucestné podtabulky.

## Sloupcové pruhové grafy kontingenční tabulky

**Zobrazte klastrované pruhové grafy.** Klastrovaný pruhový graf pomáhá sumarizovat vaše data pro skupiny případů. Pro každou hodnotu proměnné, kterou jste zadali v řádcích, je jeden klastr sloupců. Proměnná, která definuje pruhy v rámci každého klastru, je proměnná, kterou jste zadali v části Sloupcu. Pro každou hodnotu této proměnné je k dispozici jedna sada různých barevných nebo vzorných sloupců.

Určíte-li více než jednu proměnnou v rámci sloupců nebo řádků, bude pro každou kombinaci obou proměnných vytvořen klastrovaný pruhový graf.

## Kontingenční tabulky zobrazující proměnné vrstvy v vrstvách tabulky

**Zobrazte proměnné vrstvy v vrstvách tabulky.** Můžete zvolit zobrazení proměnných vrstvy (řídících proměnných) jako vrstev tabulky v tabulce crosstablation. To vám umožňuje vytvářet pohledy, které zobrazují celkovou statistiku pro řádková a sloupcová proměnná a umožňují také přechod o úroveň níž na kategorie proměnných vrstvy.

Níže je uveden příklad, který používá datový soubor *demo.sav* (dostupný v adresáři Ukázky instalačního adresáře) níže a byl získán takto:

1. Vyberte položku *Kategorie příjmu v tisících (incctat)* jako proměnnou řádku, *Vlastní PDA (ownpda)* jako proměnnou sloupce a *Úroveň vzdělání (ed)* jako proměnné vrstvy.
2. Vyberte volbu **Zobrazit proměnné vrstvy v vrstvách tabulky**.
3. V dílčím dialogovém okně Zobrazení buňky vyberte volbu **Sloupec**.
4. Spusťte proceduru kontingenčních tabulek, poklepejte na tabulku crosstablation a vyberte položku **College degree** z rozevíracího seznamu Úroveň vzdělání.

Vybraný pohled na tabulku crosstablation zobrazuje statistiky pro respondenty, kteří mají vysokoškolský titul.

## Statistika kontingenčních tabulek

**Chi-náměstí.** Pro tabulky se dvěma řádky a dvěma sloupci vyberte **Chi-square** pro výpočet Pearsonova chí-kvadrát, poměr pravděpodobnosti chí kvadrát, Fisherova přesná zkouška a Yates ' corrected chi-square (continuity correction). Pro  $2 \times 2$  tabulky je Fisherův přesný test vypočítán, pokud tabulka, která není výsledkem chybějících řádků nebo sloupců ve větší tabulce, má buňku s očekávanou frekvencí menší než 5. Yates ' corrected chi-square is computed for all other  $2 \times 2$  tables. Pro tabulky s libovolným počtem řádků a sloupců vyberte **Chi-square** pro výpočet Pearsonova chí-kvadrát a poměr pravděpodobnosti chí kvadrát. Jsou-li obě proměnné tabulky kvantitativní, **chí-kvadrát** předá lineární test přidružení.

**Korelace.** Pro tabulky, v nichž řádky i sloupce obsahují seřazené hodnoty, **Korelace** vrátí korelační koeficient Spearman, rho (pouze číselná data). Spearman's rho je měřítkem asociace mezi řadových objednávek. Jsou-li obě proměnné tabulky (faktory) kvantitativní, funkce **Korelace** vrátí Pearsonův korelační koeficient  $r$ , měřítko lineárního přidružení mezi proměnnými.

**Jmenovitá hodnota.** Pro nominální data (žádná vnitřní objednávka, jako je katolická, protestantská a židovská), můžete vybrat **Koeficient nepředvídaných událostí, Pi** (koeficient) a **Cramér's V, Lambda** (symetrický a asymetrický lambdas a Goodman a Kruskal's tau) a **Koeficient nejistoty**.

- *Součinitel nepředvídaných událostí.* Měřítka přidružení založené na chí-kvadrát. Hodnota se pohybuje mezi 0 a 1, přičemž hodnota 0 označuje, že žádné přidružení mezi řádkem a sloupci proměnných a hodnotami blízkými 1 označuje vysoký stupeň přidružení mezi proměnnými. Maximální možná hodnota závisí na počtu řádků a sloupců v tabulce.
- *Phi a Cramer's V.* Phi je měřítkem přidružení založeného na kvadratu chí kvadrát, které zahrnuje dělení statistiky chí kvadrát vzorkovou velikostí a použití druhé odmocniny výsledku. Cramer V je opatření sdružení založené na chí kvadrát.
- *Lambda.* Míra přidružení, která odráží poměrné snížení chyby, když se hodnoty nezávislé proměnné používají k předpovědi hodnot závislé proměnné. Hodnota 1 znamená, že nezávislá proměnná dokonale předpovídá závislou proměnnou. Hodnota 0 znamená, že nezávislá proměnná nemá žádnou pomoc při předvídání závislé proměnné.
- *Koeficient nejistoty.* Míra přidružení, která označuje poměrné snížení chyby, když se hodnoty jedné proměnné používají k předpovídání hodnot jiné proměnné. Například hodnota 0.83 označuje, že znalost jedné proměnné snižuje chybu v předvídání hodnot jiné proměnné o 83%. Program vypočítá symetrické i asymetrické verze koeficientu nejistoty.

**Pořadový.** Pro tabulky, v nichž řádky i sloupce obsahují seřazené hodnoty, vyberte **Gamma** (nula-order pro dvoucestné tabulky a podmíněné pro 3-cestné 3 cesty), **Kendall's tau-b** **Kendall's tau-c**. Pro odhad kategorií sloupců z kategorií řádků vyberte volbu **Somers ' d**.

- *Gama*. Symetrický ukazatel přidružení mezi dvěma pořadovými proměnnými, který se pohybuje mezi -1 a 1. Hodnoty blízké 1 označují silný vztah mezi dvěma proměnnými. Hodnoty close to 0 označují málo nebo žádný vztah. Pro dvoucestné tabulky se zobrazí gammas s nulovými obkladovými příkazy. Pro trojcestné tabulky se zobrazí podmíněné tabulky gammas.
- *Somery ' d*. Měřitko přidružení mezi dvěma pořadovými proměnnými, které se pohybují od -1 do 1. Hodnoty blízké 1 označují silný vztah mezi dvěma proměnnými a hodnoty blízké 0 označují malý nebo žádný vztah mezi proměnnými. Somers ' d je asymetrické rozšíření gama, které se liší pouze v tom, aby se počet párů neváže na nezávislé proměnné. Vypočítá se také symetrická verze této statistiky.
- *Kendall's tau-b*. Neparametrické měřítko korelace pro ordinální nebo ohodnocené proměnné, které berou vazby na účet. Znaménko koeficientu určuje směr vztahu a jeho absolutní hodnota udává sílu, přičemž větší absolutní hodnoty označují silnější vztahy. Možné hodnoty jsou v rozsahu -1 až 1, ale hodnota -1 nebo + 1 lze získat pouze z čtvercových tabulek.
- *Kendall's tau-c*. Neparametrický ukazatel přidružení pro ordinální proměnné, které ignoruje vazby. Znaménko koeficientu určuje směr vztahu a jeho absolutní hodnota udává sílu, přičemž větší absolutní hodnoty označují silnější vztahy. Možné hodnoty jsou v rozsahu -1 až 1, ale hodnota -1 nebo + 1 lze získat pouze z čtvercových tabulek.

**Nominální za interval.** Je-li jedna proměnná kategoriální a druhá je kvantitativní, vyberte volbu **Eta**. Kategoriální proměnná musí být kódována numericky.

- *Eta*. Rozsah přidružení, který se pohybuje od 0 do 1, přičemž 0 označuje žádné přidružení mezi řádkovou a sloupcovou proměnnými a hodnotami blízkými 1 indikující vysokou míru přidružení. Eta je vhodná pro závislou proměnnou měřenou na stupnici intervalu (například příjem) a samostatnou proměnnou s omezeným počtem kategorií (např. gender). Vypočítávají se dvě hodnoty éta: jeden považuje řádkovou proměnnou jako proměnnou intervalu a druhou se zachází s proměnnou sloupce jako proměnnou intervalu.

*Kappa*. Cohen's kappa měří dohodu mezi hodnoceními dvou hodnotitelů, když oba hodnotí stejný objekt. Hodnota 1 označuje dokonalou shodu. Hodnota 0 indikuje, že dohoda není lepší než šance. Kappa je založen na čtvercové tabulce, ve které hodnoty řádku a sloupce představují stejné měřítko. Každá buňka, která má pozorované hodnoty pro jednu proměnnou, ale ne druhou, má přiřazen počet 0. Kappa se nevypočítá, jestliže datový typ úložiště (řetězec nebo číslo) není stejný pro obě proměnné. V případě řetězcové proměnné musí mít obě proměnné stejnou definovanou délku.

*Riziko*. Pro tabulky 2 x 2 je to měřítko síly přidružení mezi přítomností faktoru a výskytem události. Pokud interval spolehlivosti pro statistiku zahrnuje hodnotu 1, nemůžete předpokládat, že je faktor přidružen k události. Poměr pravděpodobnosti může být použit jako odhad nebo relativní riziko, je-li výskyt faktoru vzácný.

*McNemar*. Neparametrický test pro dvě související dichomomolární proměnné. Testuje změny v odezvách pomocí distribuce chí kvadrát. Užitečné pro zjišťování změn v odezvách v důsledku experimentálního zásahu v návrzích "před-a-po". Pro větší čtvercové tabulky je nahlášen test McNemar-Bower test symetrie.

*Statistiky Cochran's a Mantel-Haenszel*. Statistiky Cochran's a Mantel-Haenszel lze použít k testování nezávislosti mezi proměnnou dichotomoni faktoru a proměnnou odezvy dichotomové, která je podmíněná proměnnými vrstvami definovanými jednou nebo více vrstvami (ovládání). Všimněte si, že zatímco ostatní statistiky jsou vypočítány vrstvy podle vrstvy, statistiky Cochran a Mantel-Haenszel se vypočítají jednou pro všechny vrstvy.

## Zobrazení kontingenční tabulky

Aby se vám v datech, která přispívají k významnému testu chí kvadrát, pomohla odhalit vzory, zobrazí procedura kontingenční tabulky očekávané frekvence a tři typy zbytkových chyb (deviates), které měří

rozdíl mezi pozorovanými a očekávanými frekvencemi. Každá buňka tabulky může obsahovat libovolnou kombinaci počtů, procentních částí a zbytkových chyb.

**Počty.** Počet skutečně zaznamenaných případů a počet očekávaných případů, pokud jsou proměnné řádku a sloupce nezávislé na sobě. Můžete se rozhodnout skrýt počty, které jsou menší než zadané celé číslo. Skryté hodnoty se zobrazí jako  $< N$ , kde  $N$  je zadané celé číslo. Uvedené celé číslo musí být větší než nebo rovno 2, ačkoli hodnota 0 je povolena a uvádí, že žádné počty nejsou skryté.

**Porovnejte proporce sloupců.** Tato volba vypočítá porovnávání po dvojicích sloupců a udává, které páry sloupců (pro daný řádek) jsou výrazně odlišné. Významné rozdíly jsou označeny v tabulce crosstablation s formátováním ve stylu APA pomocí dolních písmen a jsou počítány na úrovni významnosti 0,05.

*Poznámka:* Je-li tato volba zadána bez výběru pozorovaných počtů nebo procentní části sloupce, pak jsou pozorované počty zahrnuty do tabulky crosstablation, s dolní indexy ve stylu APA, které označují výsledky testů vzájemných poměrů sloupců.

- **Seřídít hodnoty p (Bonferroniho metoda).** Porovnávání sloupcových proporcí využívá korekci Bonferroniho, která upravuje pozorovanou úroveň významnosti pro to, že se provádí více porovnání.

**Procenta.** Procentní část může přidat přes řádky nebo dolů sloupce. Procentuální podíl celkového počtu případů představovaných v tabulce (jedna vrstva) je také k dispozici.

**Poznámka:** Je-li ve skupině počtů vybrána volba **Skrýt malé počty**, budou skryty také procentní části přidružené ke skrytému počtu.

**Residuals.** Prvotní nestandardizované zbytkové chyby poskytují rozdíl mezi pozorovanými a očekávanými hodnotami. Standardizované a upravené standardizované zbytkové chyby jsou také k dispozici.

- *Nestandardizováno.* Rozdíl mezi pozorovanou hodnotou a očekávanou hodnotou. Očekávaná hodnota je počet případů, které byste očekávali v buňce, pokud mezi těmito dvěma proměnnými nebyl žádný vztah. Kladné reziduum označuje, že v buňce existuje více případů než v případě, že by proměnné řádku a sloupce byly nezávislé.
- *Standardizováno.* Zbytek vydělený odhadem jeho směrodatné odchylky. Standardizované zbytkové chyby, které jsou také známé jako Pearsonové zbytkové chyby, mají střední hodnotu 0 a směrodatnou odchylku 1.
- *Upravené standardizované.* Zbylá hodnota buňky (pozorovaná minus očekávaná hodnota) dělená odhadem její standardní chyby. Výsledná standardizovaná zbytková hodnota je vyjádřena ve standardních jednotkách odchylky nad nebo pod průměrem.

**Vytvořte tabulku stylu APA.** Vytvoří výstupní tabulky, které odpovídají pokynům stylu APA.

**Poznámka:** Volby **Pozorováno, Očekávaný, Řádek, Sloupeca Celkem** nejsou k dispozici, když je vytvořen APA tabulka stylů je vybrána.

**Neceločíselné váhy.** Počty buněk jsou obvykle celočíselné hodnoty, protože představují počet případů v každé buňce. Je-li však datový soubor momentálně vážený hodnotou váhy s zlomkovými hodnotami (např. 1,25), mohou být počty buněk také zlomkové hodnoty. Můžete zkrátit nebo zaokrouhlit buď před nebo po výpočtu počtu buněk, nebo použít frakční počet buněk pro zobrazení tabulky i pro statistické výpočty.

- *Počty zaokrouhlených buněk.* Váhy případů se používají jako, ale akumulovaná váha v buňkách je zaokrouhlena před výpočtem jakékoli statistiky.
- *Oříznutí počtů buněk.* Jsou použity váhy případu, ale akumulovaná váha v buňkách je před výpočtem statistiky oříznuta.
- *Váhy zaokrouhlených velikostí.* váhy případu jsou před použitím zaokrouhleny.
- *Oříznout váhy případu.* Váhy případů jsou před použitím zkráceny.
- *Žádné úpravy.* Váhy případů se používají jako jsou a jsou použity zlomkové počty buněk. Pokud jsou však požadovány přesné statistiky (dostupné pouze pro vzorkování a testování), akumulovaná váha v buňkách je buď oříznuta, nebo zaokrouhlena před výpočtem přesné statistiky testu.

## Formát tabulky kontingenčních tabulek

Můžete uspořádat řádky ve vzestupném nebo sestupném pořadí hodnot proměnné řádku.

## Shrnout

---

Procedura Shrnout vypočítá statistiky podskupin pro proměnné v rámci kategorií jedné nebo více seskupujících proměnných. Všechny úrovně proměnné seskupení jsou crossstabulated. Můžete zvolit pořadí, ve kterém se statistika zobrazí. Zobrazeny jsou také souhrnné statistiky pro každou proměnnou ve všech kategoriích. Datové hodnoty v každé kategorii mohou být zobrazeny nebo potlačeny. Pomocí velkých datových sad můžete zvolit, že chcete vypsat pouze prvních  $n$  případů.

**Příklad.** Jaký je průměrný objem prodeje produktů podle regionu a odvětví zákazníků? Můžete zjistit, že průměrná prodejní částka je o něco vyšší v západní oblasti než v jiných regionech, přičemž firemní zákazníci v západní oblasti mají nejvyšší průměrnou prodejní částku.

**Statistika.** Součet, počet případů, střední hodnota, seskupený medián, směrodatná odchylka střední hodnoty, minimální, maximální, rozsah, variabilní hodnota první kategorie proměnné seskupení, hodnota proměnné poslední kategorie proměnné seskupení, směrodatná odchylka, odchylka, špičatost, směrodatná chyba špičatosti, šikmost, směrodatná odchylka šiklosti, procentní podíl z celkové částky, procento z celkového počtu  $N$ , procento z celkové hodnoty  $N$  vyjádřené v procentech, v procentech, geometrická střední hodnota a harmonický průměr.

Aspekty souhrnu dat

**Data.** Seskupení proměnných jsou kategorické proměnné, jejichž hodnoty mohou být číselné nebo řetězcové. Počet kategorií by měl být přiměřeně malý. Ostatní proměnné by měly být řazeny podle důležitosti.

**Předpoklady.** Některé z volitelných statistik podskupin, jako je střední a směrodatná odchylka, jsou založeny na normální teorii a jsou vhodné pro kvantitativní proměnné s symetrickými distribucemi. Robustní statistiky, jako je medián a rozsah, jsou vhodné pro kvantitativní proměnné, které mohou nebo nemusí splňovat předpoklad normality.

Získání souhrnů případů

1. Z nabídky vyberte:

**Analyzovat > Sestavy > Souhrny případů ...**

2. Vyberte jednu nebo více proměnných.

Volitelně můžete:

- Vyberte jednu nebo více seskupujících proměnných k rozdělení dat do podskupin.
- Chcete-li změnit název výstupu, klepněte na tlačítko **Volby**, přidejte popisec pod výstup nebo vynechte případy s chybějícími hodnotami.
- Klepněte na **Statistika** pro volitelnou statistiku.
- Vyberte **Zobrazit případy**, chcete-li zobrazit seznam případů v každé podskupině. Při výchozím nastavení systém vypíše pouze prvních 100 případů ve vašem souboru. Můžete zvýšit nebo snížit hodnotu pro **Omezit případy na první  $n$**  nebo zrušit výběr této položky, chcete-li vypsat všechny případy.

## Volby sumarizace

Shrnout vám umožňuje změnit titulek vašeho výstupu nebo přidat popisec, který se objeví pod výstupní tabulkou. Můžete řídit zalamování řádků v titulcích a titulcích zadáním \n všude, kam chcete vložit zalomení řádku do textu.

Můžete také zvolit zobrazení nebo potlačení podpoložek pro celkové součty a zahrnutí nebo vyloučení případů s chybějícími hodnotami pro jakoukoli z proměnných použitých v některé z analýz. Často je žádoucí označovat chybějící případy ve výstupu s tečkou nebo hvězdičkou. Zadejte znak, frázi nebo kód,

kteřý se má zobrazít, když chybí hodnota; v opačném případě se žádné speciální zacházení nepoužije na chybějící případy ve výstupu.

## Shrnout statistiku

Můžete zvolit jednu nebo více z následujících statistických údajů podskupiny pro proměnné v každé kategorii každé proměnné seskupení: součet, počet případů, střední, medián, seskupený medián, standardní chyba průměru, minimum, maximum, rozsah, proměnná hodnota první kategorie proměnné seskupení, hodnota proměnné poslední kategorie proměnné seskupení, směrodatná odchylka, odchylka, špičatost, směrodatná chyba špičatosti, šikmost, směrodatná chyba šikmosti, procentní podíl z celkové sumy, procento celkového součtu  $N$ , celkové procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu: procento součtu v procentní části  $N$  v geometrickém průměru, harmonické průměr. Pořadí, ve kterém se statistika objeví v seznamu Statistika buňky, je pořadí, ve kterém budou zobrazeny ve výstupu. Pro každou proměnnou ve všech kategoriích jsou také zobrazeny souhrnné statistiky.

*První.* Zobrazí první datovou hodnotu zjištěnou v datovém souboru.

*Geometrický průměr.* Hodnota  $n$ th produktu datových hodnot, kde  $n$  představuje počet případů.

*Seskupený medián.* Medián, který je vypočítán pro data, která jsou zakódována do skupin. Například s věkovými daty, je-li každá hodnota ve 30s kódována 35, každá hodnota ve 40s je kódována 45 a tak dále, seskupený medián je medián vypočtený z kódovaných dat.

*Harmonická střední hodnota.* Používá se k odhadu průměrné velikosti skupiny, když se velikosti vzorku ve skupinách nerovnaj. Harmonický průměr je celkový počet vzorků vydělený součtem recipročních velikostí vzorku velikosti vzorku.

*Kurtóza.* Měřitko rozsahu, v jakém jsou přelehle hodnoty. Pro běžnou distribuci je hodnota statistiky špičatosti rovna nule. Kladná špičatost indikuje, že data vykazují extrémnější odlehle hodnoty než normální rozdělení. Negativní špičatost indikuje, že data vykazují méně extrémní odlehle hodnoty než normální rozdělení. Definice kurtózy, která se používá, je-li hodnota 0 pro normální distribuci, je někdy označována jako nadměrná špičatost. Některé softwarové produkty mohou vykazovat špičatost tak, že hodnota je 3 pro normální distribuci.

*Last.* Zobrazí poslední datovou hodnotu zjištěnou v datovém souboru.

*Maximální.* Největší hodnota číselné proměnné.

*Střední.* Měřitko centrální tendence. Aritmetický průměr, součet dělený počtem případů.

*Medián.* Hodnota nad a pod níž polovina případů padá, 50th percentil. Existuje-li sudý počet případů, střední je průměr dvou středních případů, kdy jsou seřazeny ve vzestupném nebo sestupném pořadí. Medián je měřítkem centrální tendence, která není citlivá na odlehle hodnoty (na rozdíl od střední hodnoty, která může být ovlivněna několika extrémně vysokými nebo nízkými hodnotami).

*Minimální.* Nejmenší hodnota číselné proměnné.

*N.* Počet případů (pozorování nebo záznamy).

*Procento z celku N.* Procentní část celkového počtu případů v každé kategorii.

*Procenta z celkového součtu.* Procentní část celkové částky v každé kategorii.

*Rozsah.* Rozdíl mezi největší a nejmenší hodnotou číselné proměnné, maximum minus minimum.

*Šikmost.* Měřitko asymetrie rozdělení. Normální rozdělení je symetrické a má hodnotu šikmosti 0. Rozdělení s významnou kladnou šikmou šikmý ocas má dlouhý ocas. Rozdělení s významnou negativní šikmou má dlouhý levý ocas. Jako vodítko se bere hodnota šikmý šikmý hodnoty více než dvojnásobek standardní chyby k označení odchylky od symetrie.

*Směrodatná odchylka.* Ukazatel disperze kolem střední hodnoty. V normálním rozložení se 68% případů propadne do jedné směrodatné odchylky střední hodnoty a 95% případů spadá do dvou směrodatných

odchylek. Například, je-li střední věk 45, se směrodatnou odchylkou 10, 95% případů by mělo být mezi 25 a 65 v normálním rozdělení.

*Standardní chyba Kurtózy.* Poměr špičatosti k její standardní chybě může být použit jako test normality (to znamená, že můžete odmítnout normalnost, je-li poměr menší než -2 nebo větší než + 2). Velká kladná hodnota pro špičatost znamená, že ocásky distribuce jsou delší než u normálního rozložení; záporná hodnota pro špičatost indikuje kratší ocasy (které se stávají jako standardní rozložení v bokovníku).

*Standardní chyba střední hodnoty.* Měřítka toho, jak hodnota střední hodnoty se může lišit od vzorku na vzorek odebraného ze stejného rozdělení. Lze ji použít k hrubšímu porovnání sledovaného průměru s hypotetizovanou hodnotou (to znamená, že můžete uzavřít dvě hodnoty se liší, pokud poměr rozdílu mezi standardní chybou je menší než -2 nebo větší než + 2).

*Standardní chyba Skewness.* Poměr skewness k jeho standardní chybě může být použit jako test normality (to znamená, že můžete odmítnout normalnost, je-li poměr menší než -2 nebo větší než + 2). Velká kladná hodnota šikmost označuje dlouhý pravý konec; extrémní záporná hodnota označuje dlouhý levý ocas.

*Součet.* Součet nebo součet hodnot ve všech případech s nechybějícími hodnotami.

*Odchylka.* Ukazatel disperze kolem střední hodnoty rovnající se součtu čtverců odchylky od průměru vyděleného o jeden menší než počet případů. Odchylka se měří v jednotkách, které jsou čtvercem těch, které mají vlastní proměnnou.

## Prostředky

Procedura Způsob využití vypočítá podskupinu a související statistiky univariate pro závislé proměnné v rámci kategorií jedné nebo více nezávislých proměnných. Volitelně můžete získat jednocestnou analýzu rozptylu, éta a testů na linearitu.

**Příklad.** Změřte průměrnou velikost tuku absorbovaného třemi různými druhy oleje na vaření a proveďte jednosměrnou analýzu rozptylu, abyste zjistili, zda se to může lišit.

**Statistika.** Součet, počet případů, střední hodnota, seskupený medián, směrodatná odchylka střední hodnoty, minimální, maximální, rozsah, variabilní hodnota první kategorie proměnné seskupení, hodnota proměnné poslední kategorie proměnné seskupení, směrodatná odchylka, odchylka, špičatost, směrodatná chyba špičatosti, šikmost, směrodatná odchylka šiklosti, procentní podíl z celkové částky, procento z celkového počtu  $N$ , procento z celkové hodnoty  $N$  vyjádřené v procentech, v procentech, geometrická střední hodnota a harmonický průměr. Volby zahrnují analýzu rozptylu, éta, eta druhou a testy pro linearitu  $R$  a  $R^2$ .

Aspekty týkající se dat

**Data.** Závislé proměnné jsou kvantitativní a nezávislé proměnné jsou kategorické. Hodnoty kategorických proměnných mohou být číselné nebo řetězcové.

**Předpoklady.** Některé z volitelných statistik podskupin, jako je střední a směrodatná odchylka, jsou založeny na normální teorii a jsou vhodné pro kvantitativní proměnné s symetrickými distribucemi. Robustní statistiky, jako je medián, jsou vhodné pro kvantitativní proměnné, které mohou nebo nemusí splňovat předpoklad normality. Analýza rozptylu je odolná vůči odchylkám od normality, ale data v každé buňce by měla být symetrická. Analýza rozptylu také předpokládá, že skupiny pocházejí z populací se stejnými odchylkami. Chcete-li otestovat tento předpoklad, použijte test homogenity rozptylu Levene, který je k dispozici v režimu ANOVA jednofaktorového ANOVA.

Jak získat podskupinu znamená

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Prostředky ...**

2. Vyberte jednu nebo více závislých proměnných.

3. Chcete-li vybrat kategoriální nezávislé proměnné, použijte jednu z následujících metod:

- Vyberte jednu nebo více nezávislých proměnných. Pro každou nezávislou proměnnou se zobrazí samostatné výsledky.



- Vyberte jednu nebo více vrstev nezávislých proměnných. Každá vrstva dále dělí vzorek. Máte-li jednu nezávislou proměnnou ve vrstvě 1 a jednu nezávislou proměnnou ve vrstvě 2, výsledky se zobrazí v jedné přeškrtnuté tabulce, na rozdíl od samostatných tabulek pro každou nezávislou proměnnou.
4. Volitelně klepněte na volbu **Volby** pro volitelnou statistiku, analýzu tabulky odchylek, eta, eta druhou,  $Ra R^2$ .

## Možnosti

Můžete zvolit jednu nebo více z následujících statistických údajů podskupiny pro proměnné v každé kategorii každé proměnné seskupení: součet, počet případů, střední, medián, seskupený medián, standardní chyba průměru, minimum, maximum, rozsah, proměnná hodnota první kategorie proměnné seskupení, hodnota proměnné poslední kategorie proměnné seskupení, směrodatná odchylka, odchylka, špičatost, směrodatná chyba špičatosti, šikmost, směrodatná chyba šikmosti, procentní podíl z celkové sumy, procento celkového součtu  $N$ , celkové procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento celkového součtu, procento součtu, procentní části  $N$ , geometrické střední hodnoty a harmonické průměrné hodnoty. Můžete změnit pořadí, ve kterém se objeví statistika podskupiny. Pořadí, ve kterém se statistika objeví v seznamu Statistika buňky, je pořadí, ve kterém jsou zobrazeny ve výstupu. Pro každou proměnnou ve všech kategoriích jsou také zobrazeny souhrnné statistiky.

*První.* Zobrazí první datovou hodnotu zjištěnou v datovém souboru.

*Geometrický průměr.* Hodnota  $n$ th produktu datových hodnot, kde  $n$  představuje počet případů.

*Seskupený medián.* Medián, který je vypočítán pro data, která jsou zakódována do skupin. Například s věkovými daty, je-li každá hodnota ve 30s kódována 35, každá hodnota ve 40s je kódována 45 a tak dále, seskupený medián je medián vypočtený z kódovaných dat.

*Harmonická střední hodnota.* Používá se k odhadu průměrné velikosti skupiny, když se velikosti vzorku ve skupinách nerovnjí. Harmonický průměr je celkový počet vzorků vydělený součtem recipročních velikostí vzorku velikosti vzorku.

*Kurtóza.* Měřitko rozsahu, v jakém jsou přehlé hodnoty. Pro běžnou distribuci je hodnota statistiky špičatosti rovna nule. Kladná špičatost indikuje, že data vykazují extrémnější odlehlé hodnoty než normální rozdělení. Negativní špičatost indikuje, že data vykazují méně extrémní odlehlé hodnoty než normální rozdělení. Definice kurtózy, která se používá, je-li hodnota 0 pro normální distribuci, je někdy označována jako nadměrná špičatost. Některé softwarové produkty mohou vykazovat špičatost tak, že hodnota je 3 pro normální distribuci.

*Last.* Zobrazí poslední datovou hodnotu zjištěnou v datovém souboru.

*Maximální.* Největší hodnota číselné proměnné.

*Střední.* Měřitko centrální tendence. Aritmetický průměr, součet dělený počtem případů.

*Medián.* Hodnota nad a pod níž polovina případů padá, 50th percentil. Existuje-li sudý počet případů, střední je průměr dvou středních případů, kdy jsou seřazeny ve vzestupném nebo sestupném pořadí. Medián je měřítkem centrální tendence, která není citlivá na odlehlé hodnoty (na rozdíl od střední hodnoty, která může být ovlivněna několika extrémně vysokými nebo nízkými hodnotami).

*Minimální.* Nejmenší hodnota číselné proměnné.

*N.* Počet případů (pozorování nebo záznamy).

*Procento z celkem N.* Procentní část celkového počtu případů v každé kategorii.

*Procento celkové sumy.* Procentní část celkové částky v každé kategorii.

*Rozsah.* Rozdíl mezi největší a nejmenší hodnotou číselné proměnné, maximum minus minimum.

*Šikmost.* Měřitko asymetrie rozdělení. Normální rozdělení je symetrické a má hodnotu šikmosti 0. Rozdělení s významnou kladnou šikmou šikmou ocas má dlouhý ocas. Rozdělení s významnou negativní

šikmosti má dlouhý levý ocas. Jako vodítko se bere hodnota šikmé šikmé hodnoty více než dvojnásobek standardní chyby k označení odchylky od symetrie.

*Směrodatná odchylka.* Ukazatel disperze kolem střední hodnoty. V normálním rozložení se 68% případů propadne do jedné směrodatné odchylky střední hodnoty a 95% případů spadá do dvou směrodatných odchylek. Například, je-li střední věk 45, se směrodatnou odchylkou 10, 95% případů by mělo být mezi 25 a 65 v normálním rozdělení.

*Standardní chyba Kurtózy.* Poměr špičatosti k její standardní chybě může být použit jako test normality (to znamená, že můžete odmítnout normalnost, je-li poměr menší než -2 nebo větší než + 2). Velká kladná hodnota pro špičatost znamená, že ocásky distribuce jsou delší než u normálního rozložení; záporná hodnota pro špičatost indikuje kratší ocasy (které se stávají jako standardní rozložení v bokovníku).

*Standardní chyba střední hodnoty.* Měřítka toho, jak hodnota střední hodnoty se může lišit od vzorku na vzorek odebraného ze stejného rozdělení. Lze ji použít k hrubšímu porovnání sledovaného průměru s hypotetizovanou hodnotou (to znamená, že můžete uzavřít dvě hodnoty se liší, pokud poměr rozdílu mezi standardní chybou je menší než -2 nebo větší než + 2).

*Standardní chyba Skewness.* Poměr skewness k jeho standardní chybě může být použit jako test normality (to znamená, že můžete odmítnout normalnost, je-li poměr menší než -2 nebo větší než + 2). Velká kladná hodnota šikmost označuje dlouhý pravý konec; extrémní záporná hodnota označuje dlouhý levý ocas.

*Součet.* Součet nebo součet hodnot ve všech případech s nechybějícími hodnotami.

*Odchylka.* Ukazatel disperze kolem střední hodnoty rovnající se součtu čtverců odchylky od průměru vyděleného o jeden menší než počet případů. Odchylka se měří v jednotkách, které jsou čtvercem těch, které mají vlastní proměnnou.

Statistika pro první vrstvu

*Tabulka Anova a ta.* Zobrazí tabulku jednofaktorová analýza rozptylu a vypočítá hodnotu éta a éta (ukazatele přidružení) pro každou nezávislou proměnnou v první vrstvě.

*Test na linearitu.* Vypočte součet čtverců, stupňů volnosti a střední kvadratická hodnota spojená s lineárními a nelineárními součástmi, jakož i s poměrem F, R a R na druhou. Linearita se nevypočítá, je-li nezávislá proměnná krátký řetězec.

## Krychle OLAP

---

Procedura OLAP (Online Analytical Processing) Cubes vypočítá součty, průměry a další univariantní statistiky pro souvislé souhrnné proměnné v rámci kategorií jedné nebo více proměnných kategorických seskupení. Pro každou kategorii každé proměnné seskupení je vytvořena samostatná vrstva v tabulce.

**Příklad.** Celkové a průměrné tržby pro různé regiony a produktové řady v rámci regionů.

**Statistika.** Součet, počet případů, střední hodnota, seskupený medián, směrodatná odchylka střední hodnoty, minimální, maximální, rozsah, variabilní hodnota první kategorie proměnné skupiny, hodnota proměnné poslední kategorie proměnné seskupení, směrodatná odchylka, odchylka, špičatost, směrodatná chyba špičatosti, šikmost, směrodatná odchylka šikmosti, procentní podíl z celkového počtu případů, procento celkového součtu, procento celkového počtu případů v rámci seskupení proměnných, procento celkové sumy v seskupeních proměnných, geometrický průměr a harmonické průměr.

Aspekty týkající se dat krychlí OLAP

**Data.** Souhrnné proměnné jsou kvantitativní (spojité proměnné měřené na stupnici intervalu nebo poměru) a proměnné seskupení jsou kategorické. Hodnoty kategorických proměnných mohou být číselné nebo řetězcové.

**Předpoklady.** Některé z volitelných statistik podskupin, jako je střední a směrodatná odchylka, jsou založeny na normální teorii a jsou vhodné pro kvantitativní proměnné s symetrickými distribucemi. Robustní statistiky, jako je medián a rozsah, jsou vhodné pro kvantitativní proměnné, které mohou nebo nemusí splňovat předpoklad normalnosti.

Získání krychlí OLAP

1. Z nabídky vyberte:

**Analyzovat > Sestavy > Krychle OLAP ...**

2. Vyberte jednu nebo více spojitých součtových proměnných.

3. Vyberte jednu nebo více proměnných kategorických seskupení.

Volitelně:

- Vyberte jinou souhrnnou statistiku (klepněte na volbu **Statistika**). Než budete moci vybrat souhrnnou statistiku, musíte vybrat jednu nebo více proměnných seskupení.
- Vypočítejte rozdíly mezi dvojicemi proměnných a dvojicemi skupin, které jsou definovány proměnnou seskupení (klepněte na volbu **Rozdíly**).
- Vytvořte vlastní titulky tabulky (klepněte na volbu **Nadpis**).
- Skrýt počty, které jsou menší než uvedené celé číslo. Skryté hodnoty se zobrazí jako **< N**, kde **N** je zadané celé číslo. Uvedené celé číslo musí být větší než nebo rovno 2.

## Statistika datových krychlí OLAP

Můžete zvolit jednu nebo více z následujících statistických údajů podskupiny pro každou kategorii každé proměnné seskupení: součet, počet případů, střední, medián, seskupený medián, standardní chyba průměru, minimum, maximum, rozsah, hodnota proměnné první kategorie proměnné seskupení, hodnota proměnné poslední kategorie proměnné seskupení, směrodatná odchylka, odchylka, špičatost, směrodatná chyba fluktuace, šikmost, procento celkového součtu, procento celkového součtu, procentní část celkových případů v seskupeních proměnných, procento celku součet v seskupeních proměnných, geometrickém průměru, a harmonické střední hodnoty.

Můžete změnit pořadí, ve kterém se objeví statistika podskupiny. Pořadí, ve kterém se statistika objeví v seznamu Statistika buňky, je pořadí, ve kterém jsou zobrazeny ve výstupu. Pro každou proměnnou ve všech kategoriích jsou také zobrazeny souhrnné statistiky.

*První.* Zobrazí první datovou hodnotu zjištěnou v datovém souboru.

*Geometrický průměr.* Hodnota nth produktu datových hodnot, kde n představuje počet případů.

*Seskupený medián.* Medián, který je vypočítán pro data, která jsou zakódována do skupin. Například s věkovými daty, je-li každá hodnota ve 30s kódována 35, každá hodnota ve 40s je kódována 45 a tak dále, seskupený medián je medián vypočtený z kódovaných dat.

*Harmonická střední hodnota.* Používá se k odhadu průměrné velikosti skupiny, když se velikosti vzorku ve skupinách nerovnájí. Harmonický průměr je celkový počet vzorků vydělený součtem recipročních velikostí vzorku velikosti vzorku.

*Kurtóza.* Měřitko rozsahu, v jakém jsou přelehle hodnoty. Pro běžnou distribuci je hodnota statistiky špičatosti rovna nule. Kladná špičatost indikuje, že data vykazují extrémnější odlehlé hodnoty než normální rozdělení. Negativní špičatost indikuje, že data vykazují méně extrémní odlehlé hodnoty než normální rozdělení. Definice kurtózy, která se používá, je-li hodnota 0 pro normální distribuci, je někdy označována jako nadměrná špičatost. Některé softwarové produkty mohou vykazovat špičatost tak, že hodnota je 3 pro normální distribuci.

*Last.* Zobrazí poslední datovou hodnotu zjištěnou v datovém souboru.

*Maximální.* Největší hodnota číselné proměnné.

*Střední.* Měřitko centrální tendence. Aritmetický průměr, součet dělený počtem případů.

*Medián.* Hodnota nad a pod níž polovina případů padá, 50th percentil. Existuje-li sudý počet případů, střední je průměr dvou středních případů, kdy jsou seřazeny ve vzestupném nebo sestupném pořadí. Medián je měřítkem centrální tendence, která není citlivá na odlehlé hodnoty (na rozdíl od střední hodnoty, která může být ovlivněna několika extrémně vysokými nebo nízkými hodnotami).

*Minimální.* Nejmenší hodnota číselné proměnné.

*N.* Počet případů (pozorování nebo záznamy).

*Procento z N v.* Procentní část počtu případů pro uvedenou proměnnou seskupení v rámci kategorií jiných seskupujících proměnných. Máte-li pouze jednu proměnnou seskupení, tato hodnota je identická s procentním podílem z celkového počtu případů.

*Procenta součtu v.* Procentní část součtu pro uvedenou proměnnou seskupení v rámci kategorií jiných proměnných seskupení. Pokud máte pouze jednu proměnnou seskupení, tato hodnota je identická s procentem z celkové sumy.

*Procento z celku N.* Procentní část celkového počtu případů v každé kategorii.

*Procenta z celkového součtu.* Procentní část celkové částky v každé kategorii.

*Rozsah.* Rozdíl mezi největší a nejmenší hodnotou číselné proměnné, maximum minus minimum.

*Šikmost.* Měřitko asymetrie rozdělení. Normální rozdělení je symetrické a má hodnotu šikmosti 0. Rozdělení s významnou kladnou šikmou má dlouhý ocas. Rozdělení s významnou negativní šikmou má dlouhý levý ocas. Jako vodítko se bere hodnota šikmé šikmé hodnoty více než dvojnásobek standardní chyby k označení odchylky od symetrie.

*Směrodatná odchylka.* Ukazatel disperze kolem střední hodnoty. V normálním rozložení se 68% případů propadne do jedné směrodatné odchylky střední hodnoty a 95% případů spadá do dvou směrodatných odchylek. Například, je-li střední věk 45, se směrodatnou odchylkou 10, 95% případů by mělo být mezi 25 a 65 v normálním rozdělení.

*Standardní chyba Kurtózy.* Poměr špičatosti k její standardní chybě může být použit jako test normality (to znamená, že můžete odmítnout normalnost, je-li poměr menší než -2 nebo větší než + 2). Velká kladná hodnota pro špičatost znamená, že ocásky distribuce jsou delší než u normálního rozložení; záporná hodnota pro špičatost indikuje kratší ocasy (které se stávají jako standardní rozložení v bokovníku).

*Standardní chyba střední hodnoty.* Měřitko toho, jak hodnota střední hodnoty se může lišit od vzorku na vzorek odebraného ze stejného rozdělení. Lze ji použít k hrubšímu porovnání sledovaného průměru s hypotetizovanou hodnotou (to znamená, že můžete uzavřít dvě hodnoty se liší, pokud poměr rozdílu mezi standardní chybou je menší než -2 nebo větší než + 2).

*Standardní chyba Skewness.* Poměr skewness k jeho standardní chybě může být použit jako test normality (to znamená, že můžete odmítnout normalnost, je-li poměr menší než -2 nebo větší než + 2). Velká kladná hodnota šikmost označuje dlouhý pravý konec; extrémní záporná hodnota označuje dlouhý levý ocas.

*Součet.* Součet nebo součet hodnot ve všech případech s nechybějícími hodnotami.

*Odchylka.* Ukazatel disperze kolem střední hodnoty rovnající se součtu čtverců odchylky od průměru vyděleného o jeden menší než počet případů. Odchylka se měří v jednotkách, které jsou čtvercem těch, které mají vlastní proměnnou.

## Rozdíly krychlí OLAP

Toto dialogové okno vám umožňuje vypočítat procentní část a aritmetické rozdíly mezi souhrnnými proměnnými nebo mezi skupinami, které jsou definovány proměnnou seskupení. Rozdíly jsou vypočítány pro všechna měřítka, která jsou vybrána v dialogovém okně Statistika datových krychlí OLAP.

**Rozdíly mezi proměnnými.** Vypočte rozdíly mezi dvojicemi proměnných. Hodnoty statistiky souhrnu pro druhou proměnnou (proměnná minus) v každém páru jsou odečteny ze souhrnných statistických hodnot pro první proměnnou v páru. Pro procentní rozdíly se jako jmenovatel použije hodnota souhrnné proměnné pro proměnnou Minus. Musíte vybrat alespoň dvě souhrnné proměnné v hlavním dialogovém okně před tím, než můžete uvést rozdíly mezi proměnnými.

**Rozdíly mezi skupinami případů.** Vypočte rozdíly mezi dvojicemi skupin, které jsou definovány proměnnou seskupení. Souhrnné statistické hodnoty pro druhou kategorii v každém páru (kategorie minus) se odečtou ze souhrnných statistických hodnot pro první kategorii v páru. Rozdíly v procentech používají hodnotu souhrnné statistiky pro kategorii Minus jako jmenovatel. Musíte vybrat jednu nebo více proměnných seskupení v hlavním dialogovém okně před tím, než můžete uvést rozdíly mezi skupinami.

## Nadpis datových krychlí OLAP

Můžete změnit název vašeho výstupu nebo přidat popisek, který se zobrazí pod výstupní tabulkou. Můžete také řídit zalamování řádků titulků a titulků tak, že napíšete \n všude, kam chcete vložit zalomení řádku do textu.

## rozměry

---

### Úvod do porcí

Proporce vypočítá testy a intervaly spolehlivosti pro binomické proporce nebo rozdíly v proporcích. Statistika je k dispozici pro jeden vzorek podílů (testovaný proti zadané hodnotě), párové vzorky (různé proměnné) nebo nezávislé ukázky (různé skupiny případů). Jsou nabízeny různé možnosti zkoušek typu zkoušky a intervaly spolehlivosti. Další procedury, které poskytují některé ze stejných funkcí, zahrnují **CROSSTABS**, **NPAR TESTS** a **NPTESTS**.

#### Jednovzorkový proporce

Jeden vzorek testů a intervaly spolehlivosti. Produkce zahrnuje pozorovaný podíl, odhad rozdílu mezi poměrnou částí populace a hypotetickou základní populací, asymptotické směrodatné odchytky pod nulovými a alternativními hypotézami, specifikovanou zkušební statistikou s dvoustrannými hodnotami pravděpodobnosti a intervaly spolehlivosti s ohledem na proporce.

#### Paired-rozměry-porce

Paired-vzorky testů a intervaly spolehlivosti pro rozdíly mezi proporcemi. Produkce zahrnuje pozorované proporce, odhady rozdílu v populačních proporcích, asymptotické běžné chyby v populační rozdíly pod hodnotou null a alternativní hypotézy, stanovené zkušební statistiky s oboustrannými pravděpodobnostmi a intervaly spolehlivosti pro rozdíly v proporcích.

#### Nezávisle-části vzorků

Nezávislé-vzorky testů a intervaly spolehlivosti. Produkce zahrnuje pozorované proporce, odhady rozdílu v populačních proporcích, asymptotické běžné chyby v populační rozdíly pod hodnotou null a alternativní hypotézy, stanovené zkušební statistiky s oboustrannými pravděpodobnostmi a intervaly spolehlivosti pro rozdíly v proporcích.

### Jednovzorkový proporce

Procedura One-Sample Proportions poskytuje testy a intervaly spolehlivosti pro jednotlivé binomické proporce. Předpokládá se, že data jsou z jednoduchého náhodného vzorku a každá hypotéza test nebo interval spolehlivosti je samostatný test nebo jednotlivý interval na základě binomického podílu. Produkce zahrnuje pozorovaný podíl, odhad rozdílu mezi poměrnou částí populace a hypotetickou základní populací, asymptotické směrodatné odchytky pod nulovými a alternativními hypotézami, specifikovanou zkušební statistikou s dvoustrannými hodnotami pravděpodobnosti a intervaly spolehlivosti s ohledem na proporce.

#### Příklad

##### Statistika

Agresti-Coull, Anscombe, Clopper-Pearson (Exact), Jeffreys, Logit, Wald, Wald (continuity corrected), Wilson Score, Wilson Score (continuity corrected), Exact Binomial, Mid-p Adjusted Binomial, Score, Score (continuity corrected).

### Aspekty dat

#### Data

Procedura zobrazí požadovanou statistiku testu a oboustrannou pravděpodobnost, intervaly spolehlivosti pro rozdíly v proporcích, stejně jako proporce, standardní chyby a počty pro každou skupinu nebo proměnnou. Postup je omezen na maximálně jednu testovací hodnotu.

#### Předpoklady

## Získání jednovzorných testů proporcí

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Jedna-ukázková-části ...**

2. Vyberte jednu nebo více kvantitativních proměnných testu.

3. Volitelně můžete:

- Nastavení kritérií úspěchu vyberte v sekci **Definovat úspěch** :

### **Poslední hodnota**

Používá se poslední nebo nejvyšší hodnota mezi tříděnými odlišenými hodnotami v datech. To platí pro číselné nebo řetězcové proměnné. Toto je výchozí nastavení.

### **První hodnota**

Používá se první nebo nejnižší hodnota mezi tříděnými odlišenými hodnotami v datech. To platí pro číselné nebo řetězcové proměnné.

### **Hodnoty**

Jedna nebo více specifických hodnot v závorkách. Více hodnot musí být odděleno mezerou. To platí pro číselné nebo řetězcové proměnné. Hodnoty řetězcových proměnných by měly být uzavřeny v jednoduchých uvozovkách.

### **Středový bod**

Hodnoty ve středu nebo nad středem rozsahu pozorovaných hodnot v datech. To platí pouze pro číselná data.

### **Řez bodu**

Hodnoty nad nebo nad určenou hodnotou. To platí pouze pro číselná data.

- Klepněte na volbu **Intervaly důvěry ...** chcete-li určit, které typy intervalů spolehlivosti jsou zobrazeny, nebo k potlačení všech intervalů spolehlivosti.
- Klepněte na tlačítko **Testy ...** chcete-li určit, které typy statistiky testu se mají zobrazit, nebo potlačit všechny testy.
- Klepněte na volbu **Chybějící hodnoty ...** aby bylo možné kontrolovat nakládání s chybějícími údaji.
- Klepněte na **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako je průměr, medián, poměr, kurzový poměr, korelační koeficient nebo regresní koeficient.

4. Klepněte na tlačítko **OK**.

## jedna-vzorková proporce: Intervaly spolehlivosti

Dialogové okno Intervaly důvěry poskytuje volby pro uvedení úrovně pokrytí a pro výběr toho, které typy intervalů spolehlivosti se zobrazí.

### **Úroveň krytí**

Určuje procentní část intervalu spolehlivosti. Musí být uvedena číselná hodnota v rozsahu (0,100). Výchozí nastavení je 95.

### **Typ intervalu (y)**

Poskytuje volby pro určení, které typy intervalů spolehlivosti se zobrazí. Dostupné volby zahrnují:

- Agresti-Coull
- Ancombe
- Cloupper-Pearson (Přesná)
- Jeffreydae
- Logit
- Wald
- Wald (Vodivost nápravného zařízení)
- Skóre Wilson

- Wilsonovo skóre (Vodilosti opravené)

## Stanovení intervalů spolehlivosti pro jednovzorkovou dílčí část

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Jedna-ukázková-části ...**

2. Klepněte na **Intervaly důvěry**, abyste určili, které typy intervalů spolehlivosti se zobrazí, nebo abyste potlačili všechny intervaly spolehlivosti.

## Jedna-ukázková část: Testy

Dialogové okno Testy poskytuje volby pro určení, které typy testů statistiky se zobrazí.

### ALL

Ve výstupu se zobrazí všechny statistiky testu.

### Žádný

Ve výstupu se nezobrazují žádné statistiky testu.

### Přesná Binomické

Zobrazí přesně binomické pravděpodobnosti.

### Upravené Binomické střední hodnoty

Zobrazí střední hodnoty binomického pravděpodobnosti v polovině období. Jedná se o výchozí nastavení.

### Skóre

Zobrazí statistiku testu Zhodnocení Z. Jedná se o výchozí nastavení.

### Skóre (Vodilosti opravené)

Zobrazuje statistiku testu korigovaná kontinuita Z.

### Wald

Zobrazí statistiku testu Wald Z.

### Wald (Vodivost nápravného zařízení)

Zobrazí statistiku testu Wald Z korigovaná kontinuita.

### Testovací hodnota

Určuje testovací hodnotu v rozsahu 0 až 1. Výchozí hodnota je 0.5.

## Získání jednovzorných testů proporcí

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Jedna-ukázková-části ...**

2. V dialogovém okně One-Sample Proportions klepněte na tlačítko **Zkoušky**.

3. Vyberte jeden nebo více dostupných testů.

## Jedna-ukázková proporcí: Chybějící hodnoty

Dialogové okno Chybějící hodnoty poskytuje volby pro práci s chybějícími hodnotami.

### Chybějící rozsah dat

#### Vyloučit analýzu případů podle analýzy

Označuje zahrnutí všech případů s dostatečnými daty na proměnné použité v každé konkrétní analýze. Toto je výchozí nastavení.

#### Vyloučit případy jako litwise

Označuje zahrnutí všech případů s dostatečnými daty na všechny proměnné použité ve všech analýzách.

## Uživatelské chybějící hodnoty

Volba **Vyloučit** zachází s uživatelem s chybějícími hodnotami jako chybějící. **Zahrnout** ignoruje uživatele s chybějícími hodnotami a zachází s uživatelem, které chybějící hodnoty jsou platné.

## Definování nastavení chybějících hodnot pro vlastnost One-Sample Proportions

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Jedna-ukázková-části ...**

2. V dialogovém okně s jednou ukázkou proportions klepněte na volbu **Chybějící hodnoty**.

3. Vyberte požadovaná nastavení chybějících hodnot.

## Paired-rozměry-porce

Postup zkoušky ve dvojicích poskytuje testy a intervaly spolehlivosti pro rozdíl ve dvou souvisejících nebo párovaných binomiálních poměrech. Předpokládá se, že data jsou z jednoduchého náhodného vzorku a každý test hypotézy nebo interval spolehlivosti je samostatný test nebo jednotlivý interval. Produkce zahrnuje pozorované proporce, odhady rozdílů v populačních proporcích, asymptotické běžné chyby v populační rozdíly pod hodnotou null a alternativní hypotézy, stanovené zkušební statistiky s oboustrannými pravděpodobnostmi a intervaly spolehlivosti pro rozdíly v proporcích.

### Příklad

#### Statistika

Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuity corrected), Exact Binomial, Mid-p Adjusted Binomial, McNemar, McNemar (continuity corrected).

## Aspekty dat

### Data

- Požaduje se seznam proměnných obsahující alespoň dvě proměnné.
- Je-li zadán jeden seznam proměnných, každý člen seznamu je spárován s každým dalším členem seznamu.

### Předpoklady

- Pokud jsou dva seznamy proměnných odděleny **WITH** bez klíčového slova (**PAIRED**), každý člen prvního seznamu je spárován s každým členem druhého seznamu.
- Pokud jsou dva seznamy proměnných odděleny **WITH** a druhý seznam je následován (**PAIRED**), členy obou seznamů v pořadí jsou spárovány: první člen prvního seznamu je spárován s prvním členem druhého seznamu, druhé členy každého seznamu jsou spárovány, atd. Neodpovídající proměnné jsou ignorovány a je vydána varovná zpráva.

## Vydávání vzorků-zkoušky proporcí vzorků

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Paired-části ukázek ...**

2. Vyberte jednu nebo více kvantitativních proměnných testů.

3. Volitelně můžete:

- Nastavení kritérií úspěchu vyberte v sekci **Definovat úspěch** :

#### Poslední hodnota

Používá se poslední nebo nejvyšší hodnota mezi tříděnými odlišenými hodnotami v datech. To platí pro číselné nebo řetězcové proměnné. Toto je výchozí nastavení.

#### První hodnota

Používá se první nebo nejnižší hodnota mezi tříděnými odlišenými hodnotami v datech. To platí pro číselné nebo řetězcové proměnné.



### Hodnoty

Jedna nebo více specifických hodnot v závorkách. Více hodnot musí být odděleno mezerou. To platí pro číselné nebo řetězcové proměnné. Hodnoty řetězcových proměnných by měly být uzavřeny v jednoduchých uvozovkách.

### Středový bod

Hodnoty ve středu nebo nad středem rozsahu pozorovaných hodnot v datech. To platí pouze pro číselná data.

### Řez bodu

Hodnoty nad nebo nad určenou hodnotou. To platí pouze pro číselná data.

- Klepněte na volbu **Intervaly důvěry ...** chcete-li určit, které typy intervalů spolehlivosti jsou zobrazeny, nebo k potlačení všech intervalů spolehlivosti.
- Klepněte na tlačítko **Testy ...** chcete-li určit, které typy statistiky testu se mají zobrazit, nebo potlačit všechny testy.
- Klepněte na volbu **Chybějící hodnoty ...** aby bylo možné kontrolovat nakládání s chybějícími údaji.
- Klepněte na **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako je průměr, medián, poměr, kurzový poměr, korelační koeficient nebo regresní koeficient.

4. Klepněte na tlačítko **OK**.

## Paired-rozměry vzorků: Intervaly spolehlivosti

Dialogové okno Intervaly důvěry poskytuje volby pro uvedení úrovně pokrytí a pro výběr toho, které typy intervalů spolehlivosti se zobrazí.

### Úroveň krytí

Určuje procentní část intervalu spolehlivosti. Musí být uvedena číselná hodnota v rozsahu (0,100). Výchozí nastavení je 95.

### Typ intervalu (y)

Poskytuje volby pro určení, které typy intervalů spolehlivosti se zobrazí. Dostupné volby zahrnují:

- Agresti-Min
- Bonett-cena
- Novocombe
- Wald
- Wald (Vodivost nápravného zařízení)

## Určení intervalů spolehlivosti pro paired-rozměry vzorků

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Paired-části ukázek ...**

2. Klepněte na **Intervaly důvěry**, abyste určili, které typy intervalů spolehlivosti se zobrazí, nebo abyste potlačili všechny intervaly spolehlivosti.

## Paired-Ukázky-části: testy

Dialogové okno Testy poskytuje volby pro určení, které typy testů statistiky se zobrazí.

### ALL

Ve výstupu se zobrazí všechny statistiky testu.

### Žádný

Ve výstupu se nezobrazují žádné statistiky testu.

### Přesná Binomické

Zobrazí přesně binomické pravděpodobnosti.

### **Upravené Binomické střední hodnoty**

Zobrazí střední hodnoty binomického pravděpodobnosti v polovině období. Jedná se o výchozí nastavení.

### **McNemar**

Zobrazí statistiku testu McNemar Z. Jedná se o výchozí nastavení.

### **McNemar (Kontinuita opravena)**

Zobrazuje statistiku testu McNemar Z McNemar Z.

### **Wald**

Zobrazí statistiku testu Wald Z.

### **Wald (Vodivost nápravného zařízení)**

Zobrazí statistiku testu Wald Z korigovaná kontinuita.

## **Vydávání vzorků-zkoušky proporcí vzorků**

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Paired-části ukázek ...**

2. V dialogovém okně Paired-Samples Proportions klepněte na volbu **Testy**.

3. Vyberte jeden nebo více dostupných testů.

## **Paired-Ukázky-porce: chybějící hodnoty**

Dialogové okno Chybějící hodnoty poskytuje volby pro práci s chybějícími hodnotami.

### **Chybějící rozsah dat**

#### **Vyloučit analýzu případů podle analýzy**

Označuje zahrnutí všech případů s dostatečnými daty na proměnné použité v každé konkrétní analýze. Toto je výchozí nastavení.

#### **Vyloučit případy jako litwise**

Označuje zahrnutí všech případů s dostatečnými daty na všechny proměnné použité ve všech analýzách.

### **Uživatelské chybějící hodnoty**

Volba **Vyloučit** zachází s uživatelem s chybějícími hodnotami jako chybějící. **Zahrnout** ignoruje uživatele s chybějícími hodnotami a zachází s uživatelem, které chybějící hodnoty jsou platné.

## **Definování nastavení chybějících hodnot pro Paired-portions**

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Paired-části ukázek ...**

2. V dialogovém okně Paired-Samples Proportions klepněte na volbu **Chybějící hodnoty**.

3. Vyberte požadovaná nastavení chybějících hodnot.

## **Nezávisle-části vzorků**

Postup pro porce nezávislých vzorků poskytuje testy a intervaly spolehlivosti pro rozdíl ve dvou nezávislých binomiálních proporcích. Předpokládá se, že data jsou z jednoduchého náhodného vzorku a každý test hypotézy nebo interval spolehlivosti je samostatný test nebo jednotlivý interval. Produkce zahrnuje pozorované proporce, odhady rozdílů v populačních proporcích, asymptotické běžné chyby v populační rozdíly pod hodnotou null a alternativní hypotézy, stanovené zkušební statistiky s oboustrannými pravděpodobnostmi a intervaly spolehlivosti pro rozdíly v proporcích.

### **Příklad**

#### **Statistika**

Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuity corrected), Exact Binomial, Mid-p Adjusted Binomial, McNemar, McNemar (continuity corrected).

## Aspekty dat

### Data

- Je požadována alespoň jedna závislá proměnná a jedna proměnná k identifikaci dvou skupin, které mají být porovnány.
- Proměnná seskupení může být buď číselná, nebo řetězcová.

### Předpoklady

## Získávání údajů o porcích nezávislých vzorků

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Nezávislí-části ukázek ...**

2. Vyberte jednu nebo více kvantitativních proměnných testů.

3. Vyberte jednu proměnnou **Grouping Variable**, která identifikuje dvě skupiny, které mají být porovnány.

4. Volitelně zadejte nastavení pro vybranou **Grouping Variable**.

- Je-li vybrána volba **Hodnota (y)**, můžete pro hodnoty, které mají být porovnány, zadat dvě číselné nebo řetězcové hodnoty v závorkách. Řetězcové hodnoty by měly být uzavřeny v jednoduchých uvozovkách. Případy s ostatními hodnotami jsou ignorovány.
- Hodnota **Midpoint** platí pouze pro číselné proměnné. Případy ve nebo nad střed na střed rozdělení proměnné seskupení jsou přiřazeny druhé skupině, případy pod středním bodem jsou přiřazeny první skupině.
- Volba **Vyjmout bod** použije pouze číselné proměnné a povoluje specifikaci se závorkami jediné číselné hodnoty. Případy nad bodem přerušení nebo nad bodem přerušení u proměnné seskupení jsou přiřazeny druhé skupině, případy pod bodem přerušení jsou přiřazeny k první skupině.

5. Volitelně můžete:

- Nastavení kritérií úspěchu vyberte v sekci **Definovat úspěch** :

#### Poslední hodnota

Používá se poslední nebo nejvyšší hodnota mezi tříděnými odlišenými hodnotami v datech. To platí pro číselné nebo řetězcové proměnné. Toto je výchozí nastavení.

#### První hodnota

Používá se první nebo nejnižší hodnota mezi tříděnými odlišenými hodnotami v datech. To platí pro číselné nebo řetězcové proměnné.

#### Hodnoty

Jedna nebo více specifických hodnot v závorkách. Více hodnot musí být odděleno mezerou. To platí pro číselné nebo řetězcové proměnné. Hodnoty řetězcových proměnných by měly být uzavřeny v jednoduchých uvozovkách.

#### Středový bod

Hodnoty ve středu nebo nad středem rozsahu pozorovaných hodnot v datech. To platí pouze pro číselná data.

#### Řez bodu

Hodnoty nad nebo nad určenou hodnotou. To platí pouze pro číselná data.

- Klepněte na volbu **Intervaly důvěry ...** chcete-li určit, které typy intervalů spolehlivosti jsou zobrazeny, nebo k potlačení všech intervalů spolehlivosti.
- Klepněte na tlačítko **Testy ...** chcete-li určit, které typy statistiky testu se mají zobrazit, nebo potlačit všechny testy.
- Klepněte na volbu **Chybějící hodnoty ...** aby bylo možné kontrolovat nakládání s chybějícími údaji.
- Klepněte na **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako je průměr, medián, poměr, kurzový poměr, korelační koeficient nebo regresní koeficient.

6. Klepněte na tlačítko **OK**.

## Nezávislí-Vzorky na porci: Intervaly spolehlivosti

Dialogové okno Intervaly důvěry poskytuje volby pro uvedení úrovně pokrytí a pro výběr toho, které typy intervalů spolehlivosti se zobrazí.

### Úroveň krytí

Určuje procentní část intervalu spolehlivosti. Musí být uvedena číselná hodnota v rozsahu (0,100).  
Výchozí nastavení je 95.

### Typ intervalu (y)

Poskytuje volby pro určení, které typy intervalů spolehlivosti se zobrazí. Dostupné volby zahrnují:

- Agresti-Caffo
- Brown-Li-Jeffreys
- Hauck-Anderson
- Novocombe
- Newcombe (Kontinuita opravena)
- Wald
- Wald (Vodivost nápravného zařízení)

## Určení intervalů spolehlivosti pro nezávislé-části-ukázky

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Nezávislí-části ukázek ...**

2. Klepněte na **Intervaly důvěry**, abyste určili, které typy intervalů spolehlivosti se zobrazí, nebo abyste potlačili všechny intervaly spolehlivosti.

## Nezávislí-rozměry vzorků: Zkoušky

Dialogové okno Testy poskytuje volby pro určení, které typy testů statistiky se zobrazí.

### ALL

Ve výstupu se zobrazí všechny statistiky testu.

### Žádný

Ve výstupu se nezobrazují žádné statistiky testu.

### Hauck-Anderson

Zobrazí statistiku testu Hauck-Anderson Z.

### Wald

Zobrazí statistiku testu Wald Z.

### Wald (Vodivost nápravného zařízení)

Zobrazí statistiku testu Wald Z korigovaná kontinuita.

### Wald H<sub>0</sub>

Zobrazí statistiku testu Wald Z s použitím odhadů rozptylu pod H<sub>0</sub>.

### Wald H<sub>0</sub> (Kontinuita opravena)

Zobrazuje statistiku testu Wald Z s opravenou nepřetržitost za použití odhadů rozptylu pod H<sub>0</sub>.

## Získávání údajů o porcích nezávislých vzorků

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Nezávislí-části ukázek ...**

2. V dialogovém okně Nezávisle-části s ukázkami klepněte na tlačítko **Testy**.

3. Vyberte jeden nebo více dostupných testů.

## Nezávislí-Proporce vzorků: chybějící hodnoty

Dialogové okno Chybějící hodnoty poskytuje volby pro práci s chybějícími hodnotami.

### Chybějící rozsah dat

#### Vyloučit analýzu případů podle analýzy

Označuje zahrnutí všech případů s dostatečnými daty na proměnné použité v každé konkrétní analýze. Toto je výchozí nastavení.

#### Vyloučit případy jako litwise

Označuje zahrnutí všech případů s dostatečnými daty na všechny proměnné použité ve všech analýzách.

### Uživatelské chybějící hodnoty

Volba **Vyloučit** zachází s uživatelem s chybějícími hodnotami jako chybějící. **Zahrnout** ignoruje uživatele s chybějícími hodnotami a zachází s uživatelem, které chybějící hodnoty jsou platné.

## Definování chybějících nastavení hodnot pro části Nezávisle-ukázky

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Nezávislí-části ukázek ...**

2. V dialogovém okně Nezávisle-části s ukázkami klepněte na volbu **Chybějící hodnoty**.

3. Vyberte požadovaná nastavení chybějících hodnot.

## Testy T

---

### Testy T

K dispozici jsou tři typy testů  $t$  :

**Nezávislý-testovací test  $t$ -test (dvouvýběrový  $t$  test).** Porovnává prostředky jedné proměnné pro dvě skupiny případů. K dispozici jsou popisné statistiky pro každou skupinu a Leveneův test na rovnost rozptylů, jakož i hodnoty rovného a nerovného rozptylu  $t$  a 95% interval spolehlivosti pro rozdíl ve středních hodnotách.

**Paired-samples  $t$  test (dependent  $t$  test).** Porovnává dva proměnné pro jednu skupinu. Tento test je také určen pro párové páry nebo návrhy případové studie. Výstup obsahuje deskriptivní statistiky pro testovací proměnné, korelaci mezi proměnnými, deskriptivní statistiky pro párové rozdíly, test  $t$  a interval spolehlivosti 95%.

**Jednovzorkový test  $t$ .** Porovná střední hodnotu jedné proměnné se známou nebo hypotetizovanou hodnotou. Společně s testem  $t$  se zobrazují popisné statistiky pro testovací proměnné. 95% interval spolehlivosti pro rozdíl mezi střední hodnotou testovací proměnné a hypotetizovanou testovací hodnotou je část výchozího výstupu.

## Nezávislý-Test č. T

Testovací procedura T-ukázky T porovnává prostředky pro dvě skupiny případů a automatizuje výpočet velikosti testovacího efektu  $t$ . V ideálním případě by měly být předměty náhodně vybrány do dvou skupin, takže případný rozdíl v odezvě je způsoben léčbou (nebo nedostatkem léčby) a jinými faktory. To není případ, kdy porovnáváte průměrný příjem mužů a žen. Osoba není náhodně přiřazena jako muž nebo žena. V takových situacích byste měli zajistit, aby rozdíly v jiných faktorech nemaskovali nebo zvyšovali významný rozdíl v prostředcích. Rozdíly v průměrném příjmu mohou být ovlivněny faktory, jako je vzdělání (a ne pouze sex).

### Příklad

Pacienti s vysokým krevním tlakem jsou náhodně zařazeni do skupiny s placebem a léčenými skupinami. Pacienti s placebem dostávali neúčinný prášek a pacienti léčení dostávali nový lék, který se očekává snížení krevního tlaku. Po dvouměsíční léčbě se u jedinců používá dvouvýběrový test

rk porovnání průměrných krevního tlaku pro skupinu léčenou placebem a skupinu léčené. Každý pacient je měřen jednou a patří do jedné skupiny.

### Statistika

Pro každou proměnnou: velikost vzorku, střední, směrodatná odchylka směrodatné odchylky střední hodnoty a odhad velikosti efektu pro  $t$ -test. Pro rozdíl v prostředcích: střední, standardní chyba a interval spolehlivosti (můžete uvést úroveň důvěry). Zkoušky: Levenův test na rovnost rozptylů a obou sdružených rozptylů a samostatných rozptylů  $t$  testuje rovnost středních hodnot.

### Aspekty dat

#### Data

Hodnoty kvantitativní proměnné zájmu se nacházejí v jednom sloupci v datovém souboru. Procedura používá proměnnou seskupení se dvěma hodnotami pro oddělení případů do dvou skupin. Proměnná seskupení může být číselná (hodnoty jako 1 a 2 nebo 6.25 a 12.5) nebo krátký řetězec (například *yes* a *no*). Jako alternativu můžete použít kvantitativní proměnnou, jako např. *age*, abyste rozdělili případy do dvou skupin uvedením omezení (řezného bodu 21 rozdělí *age* do skupiny *under-21* a *21-a-over group*).

#### Předpoklady

Pro test *equal-variance t* by pozorování měly být nezávislé, náhodné vzorky z běžných distribucí se stejnou odchylkou počtu obyvatel. Pokud jde o test nerovnoměrného rozptylu  $t$ , měly by být pozorování nezávislé, náhodné vzorky z běžných distribucí. Dvouvýběrový test  $t$  je poměrně robustní, aby odklonil z normality. Při kontrole distribucí graficky, podívejte se na to, že jsou symetrické a nemají žádné outliers.

### Získání testu T-ukázka T

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Nezávislý-Test T ...**

2. Vyberte jednu nebo více kvantitativních proměnných testu. Pro každou proměnnou se vypočte samostatný test  $t$ .

3. Vyberte jednu proměnnou seskupení a poté klepněte na volbu **Definovat skupiny**, abyste určili dva kódy pro skupiny, které chcete porovnat.

4. Volitelně můžete:

- Vyberte volbu **Estimate effect sizes**, chcete-li řídit odhad velikosti testovacího efektu  $t$ -testu.
- Klepněte na tlačítko **Volby**, chcete-li řídit zpracování chybějících dat a úroveň intervalu spolehlivosti.
- Klepněte na tlačítko **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako jsou například střední hodnoty, medián, poměr, poměr pravděpodobnosti, korelační koeficient nebo regresní koeficient.

### Nezávislý-Vzorky T-testu skupin

Pro číselné proměnné seskupení definujte dvě skupiny pro  $t$ -test zadáním dvou hodnot nebo z bodu řezů:

- **Použijte zadané hodnoty.** Zadejte hodnotu pro skupinu 1 a další hodnotu pro skupinu 2. Případy s jakýmkoli jinými hodnotami jsou vyloučeny z analýzy. Čísla nemusí být celá čísla (např. 6.25 a 12.5 jsou platné).
- **Místo uzavření objektu.** Zadejte číslo, které rozdělí hodnoty proměnné seskupení do dvou sad. Všechny případy s hodnotami, které jsou nižší než jedna skupina, a případy s hodnotami, které jsou větší než nebo rovny ve výstřiku, tvoří druhou skupinu.

U řetězcových proměnných seskupení zadejte řetězec pro skupinu 1 a další hodnotu pro skupinu 2, jako je *ano* a *ne*. Případy s ostatními řetězci jsou z analýzy vyloučeny.

## Nezávisle-Volby testu T

**Interval spolehlivosti.** Ve výchozím nastavení se zobrazí 95% interval spolehlivosti pro rozdíl v prostředcích. Zadejte hodnotu mezi 1 a 99, abyste požádali o odlišnou úroveň důvěry.

**Chybějící hodnoty.** Když testujete několik proměnných a data chybí pro jednu nebo více proměnných, můžete říci procedury, které případy mají zahrnout (nebo vyloučit).

- **Vyloučit analýzu případů podle analýzy.** Každý test  $t$  používá všechny případy, které mají platná data pro testované proměnné. Velikost vzorku se může lišit od zkoušky k testování.
- **Vyloučit případy jako litwise.** Každý test  $t$  používá pouze ty případy, které mají platná data pro všechny proměnné, které se používají v požadovaných testech  $t$ . Velikost vzorku je v testech konstantní.

## Paired-test s ukázkami T

Postup zkoušky Paired-Samples T porovnává střední hodnoty dvou proměnných pro jednotlivou skupinu. Procedura vypočítá rozdíly mezi hodnotami obou proměnných pro každý případ a otestuje, zda se průměr liší od 0. Procedura také automatizuje výpočet velikosti testovacího efektu  $t$ -testu.

### Příklad

Ve studii s vysokým krevním tlakem se na začátku studie měří všichni pacienti, a to s ohledem na léčbu a znovu naměřeny. Každý předmět má tedy dvě měřítka, která se často nazývají *před* a *po*. Alternativním návrhem, pro který se tento test používá, je srovnávací studie nebo případová studie, v níž každý záznam v datovém souboru obsahuje odpověď pro pacienta a také pro jeho nebo její odpovídající kontrolní subjekt. Ve studii krevního tlaku mohou být pacienti a kontroly porovnávány podle věku (75-letého pacienta s 75-ti letým členem kontrolní skupiny).

### Statistika

Pro každou proměnnou: střední, velikost vzorku, směrodatná odchylka a směrodatná odchylka střední hodnoty. Pro každou dvojici proměnných: korelace, průměrný rozdíl ve středních hodnotách,  $t$  test, interval spolehlivosti pro průměrný rozdíl (můžete uvést úroveň důvěry) a odhad velikosti efektu pro  $t$ -test. směrodatná odchylka a směrodatná odchylka střední hodnoty rozdílu.

## Aspekty dat

### Data

Pro každý spárovaný test zadejte dvě kvantitativní proměnné (úroveň intervalu měření nebo úroveň poměru měření). U porovnávaných párů nebo případových studií musí být odezva pro každý předmět testu a jeho odpovídající kontrolní subjekt ve stejném případě v datovém souboru.

### Předpoklady

Pozorování pro každou dvojici by mělo být provedeno za stejných podmínek. Střední rozdíly by měly být normálně distribuovány. Varianty každé proměnné mohou být stejné nebo nerovné.

## Vydávání zkušebního vysílání T-test vzorků T

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Paired-test T Test ...**

2. Vyberte jednu nebo více dvojic proměnných.

3. Volitelně můžete změnit/vybrat volbu **Odhadnout velikost efektů**. Nastavení řídí, jak se počítá standardizace při odhadování korekce Cohenů  $d$  a Hhrages pro každou proměnnou páru.

### směrodatná odchylka rozdílu

Jmenovatel použitý při odhadu velikosti účinku. Cohen  $d$  používá vzorkovou směrodatnou odchylku střední hodnoty rozdílu. Korekce hamínu používá vzorkovou směrodatnou odchylku střední hodnoty rozdílu upravenou korekční součinitelem.

### Korigovaná směrodatná odchylka rozdílu

Jmenovatel použitý při odhadu velikosti účinku. Cohen  $d$  používá vzorkovou směrodatnou odchylku průměru rozdílu, který je upraven souvztázností mezi opatřeními. Korekce hamínu používá vzorkovou směrodatnou odchylku střední hodnoty rozdílu mezi opatřeními a korekční faktor.

## Průměr rozptylů

Jmenovatel použitý při odhadu velikosti účinku. Cohenova  $d$  používá druhou odmocninu průměrné odchylky ukazatelů. Korekce za okraj používá druhou odmocninu průměrné odchylky ukazatelů plus korekčního faktoru.

4. Volitelně můžete:

- Vyberte volbu **Estimate effect sizes**, chcete-li řídit odhad velikosti testovacího efektu  $t$ -testu. Když je nastavení vybráno, můžete dále kontrolovat, jak se standardizace vypočítá při odhadování korekce  $d$  a Hesh Cohen pro každou proměnnou páru.
- Klepněte na tlačítko **Volby**, chcete-li řídit zpracování chybějících dat a úroveň intervalu spolehlivosti.
- Klepněte na tlačítko **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako jsou například střední hodnoty, medián, poměr, poměr pravděpodobnosti, korelační koeficient nebo regresní koeficient.

## Vyvěrováno-Volby testu T

**Interval spolehlivosti.** Ve výchozím nastavení se zobrazí 95% interval spolehlivosti pro rozdíl v prostředcích. Zadejte hodnotu mezi 1 a 99, abyste požádali o odlišnou úroveň důvěry.

**Chybějící hodnoty.** Když testujete několik proměnných a chybí data pro jednu nebo více proměnných, můžete říci procedury, které případy mají zahrnout (nebo vyloučit):

- **Vyloučit analýzu případů podle analýzy.** Každý test  $t$  používá všechny případy, které mají platná data pro testovanou dvojici proměnných. Velikost vzorku se může lišit od zkoušky k testování.
- **Vyloučit případy jako litwise.** Každý test  $t$  používá pouze ty případy, které mají platná data pro všechny dvojice testovaných proměnných. Velikost vzorku je v testech konstantní.

## Dodatečné funkce příkazu T TEST

Jazyk syntaxe příkazu vám také umožňuje:

- Pomocí jednoho příkazu vytvořte testy jednoho vzorku i nezávislé ukázky.
- Testujte proměnnou na každé proměnné v seznamu ve spárovaných testu  $t$  (s dílčím příkazem PAIRS).
- Řídit odhad velikosti efektu  $t$ -testu (s dílčím příkazem ES).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Jednovzorkový test T

Procedura jednovzorkového testu T testuje, zda se střední hodnota jedné proměnné liší od určené konstanty a automatizuje výpočet velikosti výsledku testu  $t$ -testu.

### Příklady

Výzkumník by mohl chtít otestovat, zda se průměrný IQ skóre pro skupinu studentů liší od 100. Nebo výrobce obilovin může vzít vzorek beden z výrobní linky a ověřit, zda střední hmotnost vzorků se liší od 1,3 kilo na úrovni 95% spolehlivosti.

### Statistika

Pro každou proměnnou testu: střední, směrodatná odchylka, směrodatná odchylka střední hodnoty a odhad velikosti efektu pro  $t$ -test. Průměrný rozdíl mezi každou hodnotou dat a hypotetizovanou testovací hodnotou, test  $t$ , který testuje, že tento rozdíl je 0, a interval spolehlivosti pro tento rozdíl (můžete uvést úroveň důvěry).

## Aspekty dat

### Data

Chcete-li testovat hodnoty kvantitativní proměnné vůči hypotetické testovací hodnotě, zvolte kvantitativní proměnnou a zadejte hypotetizovanou testovací hodnotu.



## Předpoklady

Tento test předpokládá, že data jsou normálně distribuována; tento test je však poměrně robustní pro odchylky od normality.

## Získání jednovzorkového testu T

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > První-ukázka T testu ...**

2. Vyberte jednu nebo více proměnných, které mají být testovány proti stejné hypotetické hodnotě.

3. Zadejte číselnou hodnotu testu, proti níž má být porovnávána každá střední hodnota vzorku.

4. Volitelně můžete:

- Vyberte volbu **Estimate effect sizes**, chcete-li řídit odhad velikosti testovacího efektu  $t$ -testu.
- Klepněte na tlačítko **Volby**, chcete-li řídit zpracování chybějících dat a úroveň intervalu spolehlivosti.

## Jedna-ukázková T-testovací volby

**Interval spolehlivosti.** Standardně se zobrazí 95% interval spolehlivosti pro rozdíl mezi střední hodnotou a hypotetizovanou testovací hodnotou. Zadejte hodnotu mezi 1 a 99, abyste požádali o odlišnou úroveň důvěry.

**Chybějící hodnoty.** Když testujete několik proměnných a data chybí pro jednu nebo více proměnných, můžete říci procedury, které případy mají zahrnout (nebo vyloučit).

- **Vyloučit analýzu případů podle analýzy.** Každý test  $t$  používá všechny případy, které mají platná data pro testovanou proměnnou. Velikost vzorku se může lišit od zkoušky k testování.
- **Vyloučit případy jako listwise.** Každý test  $t$  používá pouze ty případy, které mají platná data pro všechny proměnné, které se používají v libovolném z požadovaných testů  $t$ . Velikost vzorku je v testech konstantní.

## Dodatečné funkce příkazu T TEST

Jazyk syntaxe příkazu vám také umožňuje:

- Pomocí jednoho příkazu vytvořte testy jednoho vzorku i nezávislé ukázky.
- Testujte proměnnou na každé proměnné v seznamu ve spárovaných testu  $t$  (s dílčím příkazem PAIRS).
- Řídit odhad velikosti efektu  $t$ -testu (s dílčím příkazem ES).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Dodatečné funkce příkazu T TEST

Jazyk syntaxe příkazu vám také umožňuje:

- Pomocí jednoho příkazu vytvořte testy jednoho vzorku i nezávislé ukázky.
- Testujte proměnnou na každé proměnné v seznamu ve spárovaných testu  $t$  (s dílčím příkazem PAIRS).
- Řídit odhad velikosti efektu  $t$ -testu (s dílčím příkazem ES).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Jednosměrný test ANOVA

Procedura One-Way ANOVA vytváří jednocestnou analýzu rozptylu pro kvantitativní závislou proměnnou jedinou proměnnou faktoru (nezávislá) a odhadne velikost účinku v jednofaktorové ANOVA. Analýza rozptylu se používá k testování hypotézy, že několik prostředků je stejné. Tato technika je rozšířením dvouvzorového  $t$  testu.

Kromě určení, že existují rozdíly mezi prostředky, můžete chtít vědět, které prostředky se liší. Existují dva typy testů pro porovnání středních hodnot: a priori kontrasty a post hoc testy. Kontrastní testy jsou nastaveny na hodnotu *před* spuštěním experimentu a post hoc testy jsou spuštěny *po* provedení experimentu. Můžete také testovat trendy napříč kategoriemi.

### Příklad

Doubské ořechy absorbují tuk v různých množstvích, když jsou vařené. Experiment se skládá ze tří druhů tuku: burákového oleje, kukuřičného oleje a sádla. Burákový olej a kukuřičný olej jsou nenasycené tuky a sádlo je nasycený tuk. Spolu s určením, zda množství absorbovaného tuku závisí na druhu použitého tuku, můžete stanovit a priori kontrast určit, zda se množství absorpce tuku liší u nasycených a nenasycených tuků.

### Statistika

Pro každou skupinu: počet případů, střední směrodatná odchylka, směrodatná odchylka střední hodnoty, minimální, maximální, 95% interval spolehlivosti pro střední hodnotu a odhad velikosti účinku v jednofaktorové ANOVA. Levene testy pro homogenitu rozptylu, analýza rozptylového stolu a robustní zkoušky rovnosti prostředků pro každou závislou proměnnou, uživatel-určil a priori rozpětí testy a mnohonásobné srovnání: Bonferroni, Sidak, Tukey je čestně významný rozdíl, Hochberg je GT2, Gabriel, Dunnett, Ryan-Einot-Gabriel-Welsch *E F* test (R-E-G-W *E F*), Ryan-Einot-Gabriel-Welsch range test (R-E-G-W *OTÁZKA*), Tamhaneův T2, Dunnett's T3, hry-Howelll, Dunnett's S, Dunnettův test s vícenásobným rozsahem, Student-Newman-Keuls (S-N-K), Tukey's *b*, Waller-Duncan, Scheffé, a least-significant difference.

## Aspekty dat

### Data

Závislá proměnná by měla být kvantitativní (úroveň intervalu měření).

### Předpoklady

Každá skupina je nezávislým namátkovým vzorkem z běžné populace. Analýza rozptylu je odolná vůči odklonům z normality, ačkoli by data měla být symetrická. Skupiny by měly pocházet z populací se stejnými odchylkami. Chcete-li otestovat tento předpoklad, použijte test Leveneova homogenity rozptylu.

## Získání jednofaktorové analýzy rozptylu

1. Z nabídky vyberte:

**Analyzovat > Porovnat prostředky > Jednofaktorová ANOVA ...**

2. Vyberte jednu nebo více závislých proměnných.
3. Vyberte jednu nezávislou proměnnou faktoru.

Volitelně můžete:

- Vyberte volbu **Odhadnout velikost efektu pro celkové testy**, abyste mohli řídit výpočet velikosti efektu pro celkový test. Je-li tato volba vybrána, tabulka "Velikosti efektů ANOVA" se zobrazí ve výstupu.
- Klepnutím na volbu **Kontrast** rozdělují mezi seskupeními čtverců součty čtverců do komponent trendů nebo určete a priori kontrasty.
- Klepněte na **Post hoc**, chcete-li použít post hoc testy rozsahu a porovnávání po dvojicích, abyste určili, které prostředky se liší.
- Klepněte na tlačítko **Volby**, chcete-li řídit zpracování chybějících dat a úroveň intervalu spolehlivosti.
- Klepněte na tlačítko **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako jsou například střední hodnoty, medián, poměr, poměr pravděpodobnosti, korelační koeficient nebo regresní koeficient.

## Kontrasty ANOVA One-Way

Do komponent trendů můžete rozdělit součty čtverců mezi seskupeními nebo určit a priori kontrastové.

## Polynomní

Oddíly mezi-seskupí součty čtverců do komponent trendů. Můžete testovat tendenci závislé proměnné napříč řazenou úrovní proměnné faktoru. Například, můžete testovat lineární trend (zvýšení nebo snížení) v platu na objednané úrovni nejvyššího stupně vydělané.

- **Stupeň.** Můžete zvolit 1st, 2nd, 3rd, 4th nebo 5th stupeň polynomiální funkce.

## Koeficienty

Uživatel-určil a priori kontrastuje být testován pomocí statistiky  $t$ . Zadejte koeficient pro každou skupinu (kategorii) proměnné faktoru a klepněte na tlačítko **Přidat** za každou položku. Každá nová hodnota se přidá na konec seznamu koeficientů. Chcete-li určit další sady kontrastů, klepněte na tlačítko **Další**. Použijte **Další** a **Předchozí** pro přesun mezi sadami protikladů.

## Odhad velikosti efektu pro kontrasty

Řídí výpočet velikosti efektu pro celkový test. Je-li toto nastavení povoleno, musí být vybrána alespoň jedna z následujících voleb pro výpočet velikostí efektu. Toto nastavení je povoleno, je-li zadán alespoň jeden kontrast a výsledkem je tabulka velikostí efektů ANOVA ve výstupu.

### Použití směšné směrodatné odchylky pro všechny skupiny jako standardizace

Používá se standardní směrodatné odchylky pro všechny skupiny jako standardizér při odhadu velikosti efektu. Jedná se o výchozí nastavení a je k dispozici, je-li vybrána volba **Odhadnout velikost efektu pro kontrasty**.

### Použití směšné směrodatné odchylky u těchto skupin působících na rozdíl od standardizace

Použití směšného standardního úchytku pro skupiny, které jsou zapojeny do kontrastu jako standardizace. Nastavení je k dispozici, je-li vybrána volba **Odhadnout velikost efektu pro kontrasty**.

Pořadí koeficientů je důležité, protože odpovídá vzestupnému pořadí hodnot kategorie proměnné faktoru. První koeficient na seznamu odpovídá nejnižší hodnotě skupiny faktoru proměnné a poslední koeficient odpovídá nejvyšší hodnotě. Pokud například existuje šest kategorií proměnné faktoru, koeficienty -1, 0, 0, 0, 0,5 a 0,5 kontrastují první skupinu s pátou a šestou skupinou. U většiny aplikací by koeficienty měly být součtem 0. Lze také použít sady, které se nesují na 0, ale zobrazí se varovná zpráva.

## Jednorázové testy ANOVA jednofaktorové

Jakmile zjistíte, že rozdíly existují mezi prostředky, testy po jednotlivých rozsah a porovnávání po dvojicích mohou určit, které prostředky se liší. Rozsahové testy identifikují homogenní podmnožiny prostředků, které se od sebe vzájemně neliší. Při vícenásobném srovnávání se testuje rozdíl mezi jednotlivými páry středních hodnot středních hodnot a výsledkem je matice, kde hvězdičky označují významně odlišné skupiny, a to na úrovni alfa 0,05.

## Předpokládané odchylky ve stejné variantě

Tukey je opravdu významný rozdíl test, Hochberg je GT2, Gabriel, a Scheffé jsou více porovnávacích testů a range testy. Další dostupné testy rozsahu jsou Tukey's  $b$ , S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W  $F$  (Ryan-Einot-Gabriel-Welsch  $F$  test), R-E-G-W  $Q$  (Ryan-Einot-Gabriel-Welsch range test) a Waller-Duncan. K dispozici je více porovnávacích testů jsou Bonferroniho, Tukey je opravdu významný rozdíl test, Sidak, Gabriel, Hochberg, Dunnett, Scheffé, a LSD (nejméně významný rozdíl).

- **LSD.** Používá  $t$  testy k provedení všech porovnávání po dvojicích mezi skupinou. Pro více porovnání se neprovedla žádná úprava pro četnost chyb.
- **Bonferroni.** používá testy  $t$  k provedení porovnávání po dvojicích mezi skupinou prostředků, ale řídí celkovou chybovost tím, že nastavuje četnost chyb pro každou zkoušku na experimentující chybu s chybami vydělenou celkovým počtem testů. Z tohoto důvodu je pozorovaná úroveň významnosti upravena tak, aby byla provedena vícenásobná porovnání.
- **Sidak.** Vícenásobný porovnávací test založený na  $t$  statistickém období. Sidak upravuje úroveň významnosti pro více porovnání a poskytuje přísnější meze než Bonferroni.
- **Scheffe.** Provádí souběžné porovnávání po dvojicích pro všechny možné kombinace středních hodnot po dvojicích. Používá distribuci vzorkování  $F$ . Lze použít k prozkoumání všech možných lineárních kombinací středních hodnot skupiny, nikoli pouze porovnávání po dvojicích.

- *R-E-G-W F*. Ryan-Einot-Gabriel-Welsch více postupný postup založený na F testu.
- *R-E-G-W Q*. Ryan-Einot-Gabriel-Welsch více postupný postup založený na Studentizovaném rozsahu.
- *S-N-K*. Provede všechna porovnání po dvojicích mezi prostředky pomocí distribuce Studentized range. Se stejnou velikostí vzorku, to také porovnává dvojice prostředků v rámci homogenní podmnožiny, pomocí krokového postupu. Prostředky jsou seřazeny od nejvyšší po nejnižší a extrémní rozdíly jsou testovány jako první.
- *Tukey*. používá statistiku Studentizovaného rozsahu, aby provedla všechna porovnání po dvojicích mezi skupinami. Nastaví četnost chyb experimentů při chybovém poměru pro kolekci pro všechna porovnání po dvojicích.
- *Tukey's b*. používá rozdělení Studentizovaného rozsahu k vytvoření porovnání po dvojicích mezi skupinami. Klíčovou hodnotou je průměr odpovídající hodnoty pro zkoušku Tukey s poctivým významným rozdílem a Student-Newman-Keuls.
- *Duncan*. Provede porovnávání po dvojicích pomocí krokového pořadí porovnání identických s pořadím použitým testem Student-Newman-Keuls, ale nastavuje úroveň ochrany pro četnost chyb pro kolekci testů, spíše než četnost chyb pro jednotlivé testy. Používá statistiku Studentizovaného rozsahu.
- *Hochbergův GT2*. Vícenásobný test porovnání a rozsah, který používá Studentizovaný maximální modul. Podobně jako v Tukey je opravdu významný rozdíl test.
- *Gabriel*. Test porovnání po dvojicích, který použil maximální modulový modul Studentized a je obecně výkonnější než Hochbergův GT2, když jsou velikosti buněk nestejně. Gabrielův test se může stát liberálním, když se velikost buněk velmi liší.
- *Waller-Duncan*. Vícenásobný porovnávací test založený na statistice  $t$ ; používá Bayesovský přístup.
- *Dunnnett*. Test vícenásobného porovnání  $t$ , který porovnává sadu procedur s jedním řídicím průměrem. Poslední kategorií je výchozí řídicí kategorie. Alternativně můžete zvolit první kategorii. **Dvoustranné** testy, že střední hodnota na libovolné úrovni (kromě řídicí kategorie) faktoru se nerovná úrovni řídicí kategorie. **< Control** testuje, zda je průměr na libovolné úrovni faktoru menší než průměr řídicí kategorie. **> Řízení** testů, je-li průměr na jakékoli úrovni faktoru větší než průměr řídicí kategorie.

## Rovné rozdíly, které nebyly uvozeny

Více porovnávacích testů, které nepředpokládají stejné rozptyly jsou Tamhaneův T2, Dunnnettův T3, Hry-Howellův a Dunnnettův C.

- *Tamhaneův T2*. Konzervativní porovnávání po dvojicích test založený na  $t$  testu. Tento test je vhodný, když se liší rozptyly.
- *Dunnnettův T3*. Test porovnání po dvojicích založený na maximálním modulu Studentized. Tento test je vhodný, když se liší rozptyly.
- *Hry-Howell*. porovnávací test po dvojicích, který je někdy liberální. Tento test je vhodný, když se liší rozptyly.
- *Dunnnett's C*. Test porovnání po dvojicích založený na rozsahu Studentized. Tento test je vhodný, když se liší rozptyly.

**Poznámka:** You may find it easier to interpret the output from post hoc tests if you deselect **Skrýt prázdné řádky a sloupce** in the Table Properties dialog box (in an activated pivot table, choose **Vlastnosti tabulky** from the Format menu).

## Zkouška hypotéza s hodnotou null

Uvádí, jak se zachází s úrovní významnosti (alfa) pro post hoc test.

### Použit stejnou úroveň významnosti (alfa) jako nastavení ve volbách

Je-li tato volba vybrána, použije stejné nastavení, které je určeno v dialogovém okně Volby.

### Určete úroveň významnosti (alfa) pro post hoc test

Je-li tato volba vybrána, můžete v poli **Úroveň** určit úroveň významnosti (alfa).

## Získání post-hoc testů pro jednofaktorovou ANOVA

### Možnosti jednofaktorové ANOVA

#### Statistika

Vyberte jednu nebo více z následujících možností:

#### Popisné

Vypočítá počet případů, střední směrodatnou odchylku, směrodatnou chybu střední, minimální, maximální a 95% intervaly spolehlivosti pro každou závislou proměnnou pro každou skupinu.

#### Pevné a náhodné účinky

Zobrazí směrodatnou odchylku, standardní chybový výstup a 95% interval spolehlivosti pro model s pevným efektem a směrodatnou chybu, 95% interval spolehlivosti a odhad rozptylu mezi komponentami pro model s náhodným efektem.

#### Homogenita testů rozptylu

Vypočte statistiku Levene, která má být testována pro rovnost skupinových rozdíků. Tato zkouška není závislá na předpokladu normálnosti.

#### Brown-Forsythe

Vypočítá statistickou zkoušku pro dosažení rovnosti středních hodnot Brown-Forsythe. Tato statistika je vhodnější než statistika  $F$ , když se předpokládá, že se liší rozptyl.

#### Svaz

Vypočítává Svařovací statistiku ke zkoušce pro rovnost středních hodnot. Tato statistika je vhodnější než statistika  $F$ , když se předpokládá, že se liší rozptyl.

#### Chybějící hodnoty

Řídí zpracování chybějících hodnot.

#### Vyloučit analýzu případů podle analýzy

Případ s chybějící hodnotou buď pro závislou, nebo proměnnou faktoru pro danou analýzu není v této analýze použit. Také případ mimo rozsah zadaný pro proměnnou faktoru není použit.

#### Vyloučit případy jako litwise

Případy s chybějícími hodnotami pro proměnnou faktoru nebo pro jakoukoli závislou proměnnou zahrnutou na závislém seznamu v hlavním dialogovém okně jsou vyloučeny ze všech analýz. Pokud jste nezadal více závislých proměnných, nebude to mít žádný vliv.

#### Interval spolehlivosti

Standardně se zobrazí 95% interval spolehlivosti pro rozdíl mezi střední hodnotou a hypotetizovanou testovací hodnotou. Zadejte hodnotu mezi 1 a 99, abyste požádali o odlišnou úroveň důvěry.

#### Zápletka

Zobrazí graf, který zakresluje podskupinu prostředků (prostředky pro každou skupinu definovanou hodnotami proměnné faktoru).

### Určení voleb pro jednofaktorovou ANOVA

### Další funkce příkazu ONEWAY

Jazyk syntaxe příkazu vám také umožňuje:

- Získat statistiku fixních a náhodných efektů. Standardní odchylka, směrodatná chyba průměru a 95% intervaly spolehlivosti pro model s pevnými účinky. Standardní chybový výstup, 95% interval spolehlivosti a odhad rozptylu mezi komponentami pro model s náhodnými účinky (pomocí produktu STATISTICS=EFFECTS).
- Uvedte úroveň alfa pro rozdíl rozdílů významnosti, Bonferroi, Duncan a Scheffé více porovnávacích testů (s dílčím příkazem RANGES).
- Zapište matici středních hodnot, směrodatné odchylky a frekvence, nebo si přečtěte matici prostředků, frekvencí, směsných odchylek a stupňů volnosti pro sloučené rozptyly. Tyto matrice lze použít místo prvotních dat k získání jednosměrné analýzy rozptylu (s dílčím příkazem MATRIX).

## Analýza proměnné GLM Univariate

Procedura GLM Univariate poskytuje regresní analýzu a analýzu rozptylu pro jednu závislou proměnnou jedním nebo více faktory a/nebo proměnnými. Proměnné faktorů rozdělují populaci do skupin. Při použití této procedury obecného lineárního modelu můžete testovat hypotézy s hodnotou null o účincích jiných proměnných na prostředky různých seskupení jedné závislé proměnné. Můžete vyšetřit interakce mezi faktory stejně jako účinky jednotlivých faktorů, některé z nich mohou být náhodné. Kromě toho mohou být zahrnuty účinky kovariátů a kovariátové interakce s faktory. V případě regresní analýzy jsou nezávislé proměnné (prediktory) uvedeny jako proměnné.

Vyzkoušet lze jak vyvážené, tak nevyvážené modely. Návrh se vyváží, pokud každá buňka v modelu obsahuje stejný počet případů. Kromě testování hypotéz, GLM Univariate produkuje odhady parametrů.

Běžně k dispozici jsou a priori kontrasty jsou dostupné k provedení testování hypotéz. Navíc po zobrazení celkového testu  $F$  můžete použít post hoc testy k vyhodnocení rozdílů mezi specifickými prostředky. Odhadované marginální prostředky poskytují odhady předpovídaných průměrných hodnot pro buňky v modelu a profilové grafy (grafy interakce) z těchto prostředků umožňují snadno vizualizovat některé vztahy.

Residuals, predicted values, Cook's distance, and páku values can be saved as new new variables in your data file for checking assumptions.

Váha WLS umožňuje určit proměnnou použitou k vyjádření různých vah pro analýzu nejmenších čtverců (WLS), která by mohla být kompenzována jinou přesností měření.

**Příklad.** Data se shromažďují pro jednotlivé běžce v Chicagu maratonu na několik let. Čas, ve kterém každý spouštěč končí, je závislá proměnná. Mezi další faktory patří počasí (chladný, příjemný, nebo horký), počet měsíců tréninku, počet předchozích maratonů a pohlaví. Věk je považován za kovariát. Možná zjistíte, že pohlaví je významný vliv a že interakce mezi pohlavími a počasím je významná.

**Metody.** Typ I, Typ II, Typ III a součet IV součty čtverců lze použít k vyhodnocení různých hypotéz. Typ III je předvolba.

**Statistika.** Testy post hoc rozsahu a vícenásobná porovnání: nejméně významný rozdíl, Bonferroniho, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch Multiple  $E F$ , Ryan-Einot-Gabriel-Welsch Multiple range, Student-Newman-Keuls, Tukey's tonely significant difference, Tukey's  $b$ , Duncan, Hochberg's  $GT2$ , Gabriel, Waller-Duncan  $r$  test, Dunnett (jednostranný a dvoustranný), Tamhaneův  $T2$ , Dunnett je  $T3$ , hry-Howelll a Dunnett je  $S$ . Deskriptivní statistiky: pozorované prostředky, směrodatné odchylky a počty pro všechny závislé proměnné ve všech buňkách. Levenová zkouška na homogenitu rozptylu.

**Plot.** Rozložení-versus úroveň, zbytková hodnota a profil (interakce).

Posouzení GLM Univariate Data

**Data.** Závislá proměnná je kvantitativní. Faktory jsou kategorické. Mohou mít číselné hodnoty nebo hodnoty řetězce až osm znaků. Kovarioty jsou kvantitativní proměnné, které souvisejí se závislou proměnnou.

**Předpoklady.** Údaje jsou náhodným vzorkem z běžné populace; v populaci jsou všechny buněčné rozptyly stejné. Analýza rozptylu je odolná vůči odklonům z normality, ačkoli by data měla být symetrická. Chcete-li zkontrolovat hypotézy, můžete použít homogenitu testů rozptylů a grafů na úrovni rozložení. Můžete také přezkoumat zbytkové chyby a zbytkové grafy.

Získání tabulek GLM Univariate

1. Z nabídky vyberte:

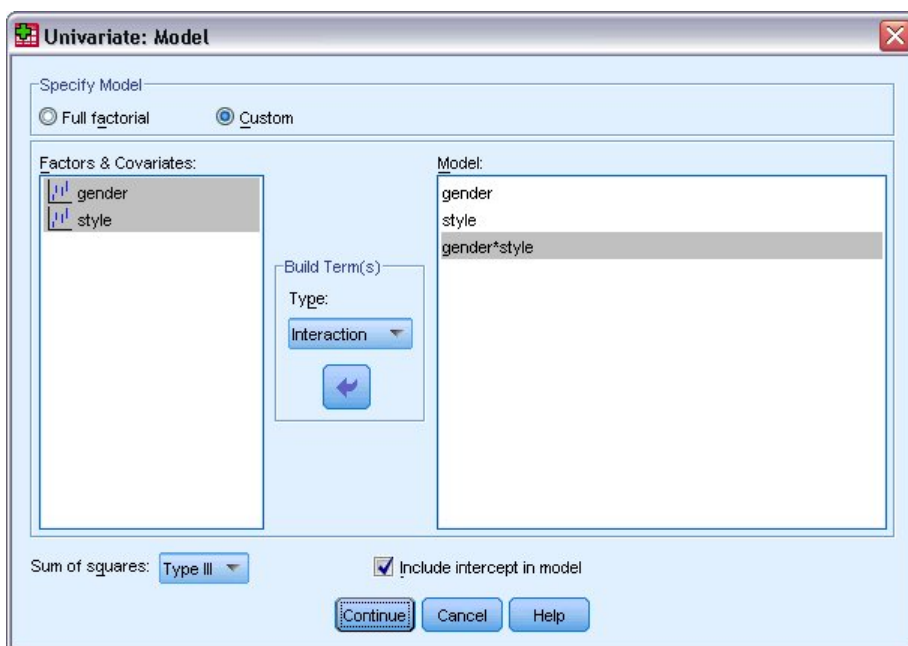
**Analyzovat > Obecný lineární model > Univariate ...**

2. Vyberte závislou proměnnou.

3. Vyberte proměnné pro pevný faktor (y), Náhodný (é) faktor (y) a Kovariate (s), podle vhodnosti pro vaše data.

4. Volitelně můžete použít váhu WLS k určení proměnné váhy pro analýzu nejmenších čtverců (weighted least-squares). Je-li hodnota váhové proměnné nula, záporná nebo chybějící, případ je vyloučen z analýzy. Proměnná, která je již použita v modelu, nemůže být použita jako váhová proměnná.

## Model GLM



Obrázek 1. Dialogové okno Univariate Model

**Určete model.** Úplný faktoriální model obsahuje všechny hlavní efekty všech faktorů, všechny hlavní účinky a všechny interakce mezi faktory a faktory. Neobsahuje interakce covariate. Vyberte volbu **Vlastní**, chcete-li určit pouze část interakcí, nebo určete interakce faktoru-podle-kovariate. Je třeba označit všechny podmínky, které mají být zahrnuty do modelu.

**Faktory a proměnné.** Jsou zde uvedeny faktory a kovariány.

**Model.** Model závisí na povaze vašich dat. Po výběru volby **Vlastní** můžete vybrat hlavní efekty a interakce, které mají zájem na analýze.

**Součet čtverců.** Metoda výpočtu součtů čtverců. U vyvážených nebo nevyvážených modelů bez chybějících buněk se nejčastěji používá metoda typu III-metodou čtverců.

**Zahrnout zachycení do modelu.** Zachycení je obvykle zahrnuto v modelu. Pokud můžete předpokládat, že data procházejí zdrojem dat, můžete toto zachycení vyloučit.

## Sestavit podmínky a vlastní podmínky

### Podmínky sestavení

Tuto volbu použijte, chcete-li zahrnout nevnořené podmínky určitého typu (například hlavní efekty) pro všechny kombinace vybrané sady faktorů a kovariány.

### Sestavit vlastní výrazy

Tuto volbu použijte, chcete-li zahrnout vnořené výrazy nebo pokud chcete explicitně sestavit libovolnou proměnnou term pomocí proměnné. Sestavení vnořeného termínu zahrnuje následující kroky:

## Součet čtverců

U modelu si můžete vybrat typ součtů čtverců. Typ III je nejčastěji používaný a je předvolený.

**Typ I.** Tato metoda je také známá jako hierarchická dekompozice metody sum-of-squares. Každý termín je upraven pouze pro výraz, který je před ním uveden v modelu. Typ I částky ze čtverců se běžně používají pro:

- Vyvážený model ANOVA, v němž jsou všechny hlavní efekty určeny před jakýmkoli prvokrevným efektem interakce, jsou všechny účinky interakce first-order zadány před účinky interakce sekund-order, atd.
- Polynomiální regresní model, ve kterém jsou před případnými vzestupnými termíny uvedeny všechny termíny s nízkou objednávkou.
- Vnořený model, ve kterém je první zadaný efekt vnořen v rámci druhého určeného efektu, je druhý určený efekt vnořen do třetí a tak dále. (Tato forma vnoření může být zadána pouze pomocí syntaxe.)

**Typ II.** Tato metoda vypočítá součet druhých mocnin efektu v modelu upraveném pro všechny ostatní "vhodné" účinky. Vhodný účinek je takový, který odpovídá všem účinkům, které neobsahují zkoumaný účinek. Typ II metodou součtu čtverců se obvykle používá pro:

- Vyvážený model ANOVA.
- Jakýkoli model, který má vliv pouze na hlavní faktor.
- Libovolný regresní model.
- Čistě vnořený návrh. (Tento tvar vnoření lze zadat pomocí syntaxe.)

**Typ III.** Předvolba. Tato metoda vypočítá součty čtverců efektu v návrhu jako součty čtverců, upravené pro všechny ostatní efekty, které neobsahují efekt, a ortogonální vůči jakýmkoli účinkům (pokud existují), které tento efekt obsahují. Typ III částky čtverců mají jednu velkou výhodu v tom, že jsou invariantní s ohledem na buněčných frekvencí tak dlouho, jak je obecná forma odhadu zůstává konstantní. Proto je tento typ částek čtverců často považován za užitečný pro nevyvážený model bez chybějících buněk. V faktoriálním provedení bez chybějících buněk se tato metoda rovná metodě Yates' weighted-squares-of-means. Typ III metodou součtu čtverců se běžně používá pro:

- Všechny modely uvedené v typu I a II.
- Jakýkoli vyvážený nebo nevyvážený model bez prázdných buněk.

**Typ IV.** Tato metoda je navržena pro situaci, ve které chybí buňky. V případě jakéhokoli efektu  $F$  v návrhu, pokud  $F$  není obsažen v žádném jiném efektu, pak typ IV = Typ III = Typ II. Je-li  $F$  obsaženo v jiných účincích, typ IV distribuuje rozdíly mezi parametry v  $F$  na všechny vysoce vysoké účinky na úrovni, které jsou vyrovnané. Typ IV součet čtverců se běžně používá pro:

- Všechny modely uvedené v typu I a II.
- Jakýkoli vyvážený model nebo nevyvážený model s prázdnými buňkami.

## Kontrasty GLM

Kontrast se používá k testování rozdílů mezi úrovněmi faktoru. Můžete určit kontrast pro každý faktor v modelu (v modelu s opakovanými měřeními, pro každý faktor mezi předměty). Kontrast představuje lineární kombinace parametrů.

**GLM Univariate.** Testování hypotéz je založeno na nulové hypotéze  $\mathbf{LB} = 0$ , kde  $\mathbf{L}$  je matice kontrastu matic a  $\mathbf{B}$  je vektor parametru. Je-li zadán kontrast, vytvoří se matice  $\mathbf{L}$ . Sloupce matice  $\mathbf{L}$  odpovídající faktoru se shodují s faktorem. Zbývající sloupce jsou seřazeny tak, aby matice  $\mathbf{L}$  byla odhadovatelná.

Výstup obsahuje statistiku  $F$  pro každou sadu kontrastů. Také se zobrazují pro kontrastní rozdíly jsou Bonferroniho typ simultánní intervaly spolehlivosti založené na Studentovo  $t$  distribuci.

Dostupné kontrasty

Dostupné kontrasty jsou odchylky, jednoduché, rozdíl, Helmert, opakující se a polynom. Pro odchylky kontrastů a jednoduchých kontrastů si můžete zvolit, zda je kategorie odkazů poslední nebo první kategorií.



## Typy kontrastu

**Odchylka.** Porovnává střední hodnotu každé úrovně (kromě referenční kategorie) se střední hodnotou všech úrovní (průměrná střední hodnota). Úrovně faktoru mohou být v libovolném pořadí.

**Jednoduchý.** Porovnává střední hodnotu každé úrovně se střední hodnotou uvedené úrovně. Tento typ kontrastu je užitečný v případě, že je zde kontrolní skupina. Jako referenci můžete zvolit první nebo poslední kategorii.

**Rozdíl.** Porovnává střední hodnotu každé úrovně (kromě první) na střední hodnotu z předchozích úrovní. (Někdy se nazývá Reverse Helmert kontrasty.)

**Helmert.** Porovnává střední hodnotu každé úrovně faktoru (kromě poslední) na střední hodnotu následujících úrovní.

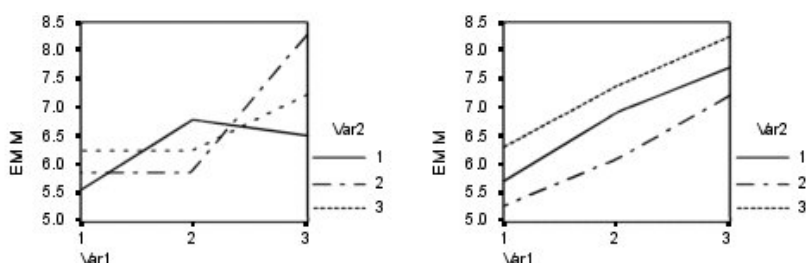
**Opakováno.** Porovnává střední hodnotu každé úrovně (kromě poslední) na střední hodnotu na následující úrovní.

**Polynomiální.** Porovná lineární efekt, kvadratický efekt, krychlový efekt a tak dále. První stupeň svobody obsahuje lineární účinek napříč všemi kategoriemi; druhý stupeň svobody, kvadratický efekt a tak dále. Tyto kontrasty se často používají k odhadu polynomiálních trendů.

## Plosy profilu GLM

Profilové grafy (grafy interakce) jsou užitečné k porovnání mezních hodnot ve vašem modelu. Vykreslení profilu je řádkový graf, ve kterém každý bod označuje odhadovanou okrajovou střední hodnotu závislé proměnné (upravené pro každé proměnné) na jedné úrovni faktoru. Úrovně druhého faktoru lze použít k vytvoření samostatných řádků. Každá úroveň ve třetím faktoru může být použita k vytvoření samostatného výkresu. Všechny pevné a náhodné faktory, jsou-li nějaké, jsou k dispozici pro pozemky. U vícerozměrných analýz jsou vytvořeny grafy profilů pro každou závislou proměnnou. V rámci opakované analýzy ukazatelů lze v profilových zkusných plochách použít faktory mezi faktory a faktory v rámci jednotlivých subjektů. GGLM Multivariate a GLM Repeated Measures jsou k dispozici pouze tehdy, máte-li nainstalovanou volbu Rozšířená statistika.

Vykreslení profilu jednoho faktoru ukazuje, zda se odhadované mezní hodnoty zvyšují nebo snižují napříč úrovněmi. U dvou nebo více faktorů ukazují paralelní čáry, že mezi faktory žádná interakce neexistuje, což znamená, že můžete vyšetřit úrovně pouze jednoho faktoru. Neparalelní čáry označují interakci.



Obrázek 2. Neparalelní vykreslení (vlevo) a paralelní zkusné plochy (vpravo)

Poté, co je zakreslení určeno výběrem faktorů pro vodorovnou osu a volitelně faktory pro samostatné řádky a samostatné zkusné plochy, musí být zakreslení přidáno do seznamu Ploty.

### Typ grafu

Graf může být čárový graf nebo sloupcový graf.

### Chybové úsečky

Můžete zahrnout chybové úsečky, které představují interval spolehlivosti, nebo počet standardních chyb. Interval spolehlivosti je založen na úrovni významnosti zadané v dialogovém okně Volby.

### Zahrnout referenční čáru pro střední hodnotu průměru

Obsahuje referenční čáru, která představuje celkovou střední hodnotu.

## Osa Y začíná na 0

Pro spojnicové grafy se všemi pozitivními nebo všemi zápornými hodnotami vynutí, aby osa Y začínaly hodnotou 0. Sloupcové grafy vždy začínají (nebo jsou zahrnuty) 0.

## Volby GLM

Volitelné statistiky jsou k dispozici z tohoto dialogového okna. Statistika se vypočítává pomocí modelu s pevným efektem.

**Obrazovka.** Vyberte volbu **Deskriptivní statistika** k vytvoření pozorovaných středních hodnot, směrodatných odchylek a počtů pro všechny závislé proměnné ve všech buňkách. **Odhady velikosti efektu** poskytují pro každý efekt a každý odhad hodnoty dílčí hodnoty éta na druhou. Statistika na druhou mocninu popisuje část celkové variability přiřaditelné faktoru. Vyberte **Sledovaný výkon**, chcete-li získat výkon testu, když je alternativní hypotéza založena na zjištěné hodnotě. Vyberte volbu **Odhady parametrů**, chcete-li vytvořit odhady parametrů, standardní chyby, testy  $t$ , intervaly spolehlivosti a pozorovanou sílu pro každý test. Chcete-li získat matici **L**, vyberte volbu **Variační koeficient kontrastu**.

**Testy homogenity** vytváří Levene testy homogenity rozptylu pro každou závislou proměnnou napříč všemi kombinacemi úrovně mezi faktory mezi faktory, pouze pro faktory mezi předměty. Volby rozložení dat na úrovni a zbytkové hodnoty jsou užitečné při kontrole předpokladů o datech. Tato položka je zakázána, pokud zde nejsou žádné faktory. Vyberte volbu **Zbytkový graf**, chcete-li pro každou závislou proměnnou vytvořit pozorovaný zbytkový graf se standardizovaným reziduem. Tyto zkusné plochy jsou užitečné při zkoumání předpokladu stejné odchylky. Chcete-li zkontrolovat, zda vztah mezi závislou proměnnou a nezávislými proměnnými může být odpovídajícím způsobem popsán modelem, vyberte volbu **Chybějící přízpůsobení**. **General estimable function (s)** vám umožňuje vytvořit vlastní testy hypotézy založené na obecné odhadnutelné funkci (ech). Rows in any contrast coefficient matrix are linear combinations of the general estimable function (s).

**Heteroskedastické testy** jsou dostupné pro testování, zda odchylky chyb (pro každou závislou proměnnou) závisí na hodnotách nezávislých proměnných. Pro test **Breusch-Pagan, Modified Breusch-Pagan testa F test** můžete zadat model, na kterém je test založen. Model se standardně skládá z konstantního termínu, z výrazu, který je lineární v předpovězých hodnotách, z výrazu, který je kvadratický v předpovězenou hodnotu, a chybový výraz.

**Odhady parametrů s robustními standardními chybami** zobrazují tabulku odhadů parametrů spolu s robustními nebo heteroskopickými směrodatnými chybami (HC) a  $t$  statistikami, hodnotami významnosti a intervaly spolehlivosti, které používají robustní standardní chyby. Pro robustní odhad matic kovariance je k dispozici pět různých metod.

### HC0

Na základě původních asymptotických nebo velkých vzorků robustního, empirického nebo "sendvičového" odhadu kovarianční matice hodnot parametrů. Střední část sendviče obsahuje čtvercový OLS (ordinální nejmenších čtverců) nebo čtvercový Vážené WLS (vážené nejmenších čtverců) zbytkové chyby.

### HC1

Konečnou ukázkou modifikace HC0 vynásobením  $N/(N-p)$ , kde  $N$  je velikost vzorku a  $p$  je počet neredundantních parametrů v modelu.

### HC2

Modifikace HC0, která zahrnuje vydělení zbytkové zbytkové hodnoty o  $1-h$ , kde  $h$  je pákový efekt pro daný případ.

### HC3

Modifikace HC0, která se blíží odhadci jacksového odhadu. Druhá mocnina se dělí čtvercem o  $1-h$ .

### HC4

Modifikace HC0, která dělí čtvercový zbytek o  $1-h$  na moc, která se liší podle  $h$ ,  $N$  a  $p$ , s horním limitem 4.

**Hladina významnosti.** Možná budete chtít upravit úroveň významnosti použitou v post hoc testech a úroveň spolehlivosti užité pro vytváření intervalů spolehlivosti. Uvedená hodnota se také použije pro

výpočet pozorovaného výkonu pro test. Určíte-li úroveň významnosti, v dialogovém okně se zobrazí přidružená úroveň intervalů spolehlivosti.

## Další funkce příkazu UNIANOVA

Jazyk syntaxe příkazu vám také umožňuje:

- Určete vnořené efekty v návrhu (pomocí dílčího příkazu DESIGN ).
- Určete testy efektů oproti lineární kombinaci efektů nebo hodnoty (pomocí dílčího příkazu TEST ).
- Určete více kontrastů (pomocí dílčího příkazu CONTRAST ).
- Zahrnout uživatele-chybějící hodnoty (pomocí dílčího příkazu MISSING ).
- Zadejte kritéria EPS (pomocí dílčího příkazu CRITERIA ).
- Vytvořte vlastní matici **L** , matici **M** nebo matici **K** (s použitím dílčích příkazů LMATRIX, MMATRIX a KMATRIX ).
- Pro odchylky nebo jednoduché kontrasty uveďte intermediační referenční kategorii (pomocí dílčího příkazu CONTRAST ).
- Uveďte metriky pro polynomiální kontrasty (pomocí dílčího příkazu CONTRAST ).
- Určete chybové termíny pro post hoc porovnání (pomocí dílčího příkazu POSTHOC ).
- Výpočet odhadovaných mezních hodnot pro jakýkoli součinitel nebo součinitel interakcí mezi faktory v seznamu faktorů (pomocí dílčího příkazu EMMEANS ).
- Zadejte názvy pro dočasné proměnné (pomocí dílčího příkazu SAVE ).
- Vytvořit datový soubor matice korelace (pomocí dílčího příkazu OUTFILE ).
- Vytvořte datový soubor matice, který obsahuje statistické údaje z tabulky mezi subjekty ANOVA (pomocí dílčího příkazu OUTFILE ).
- Uložte matici návrhu na nový datový soubor (pomocí dílčího příkazu OUTFILE ).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Porovnání post hoc GLM

**Jednorázové více porovnávacích testů.** Jakmile zjistíte, že rozdíly existují mezi prostředky, testy po jednotlivých rozsah a porovnávání po dvojicích mohou určit, které prostředky se liší. Porovnání se provádí na neupravených hodnotách. Tyto testy se používají pouze pro stanovení mezi jednotlivými faktory. V případě opakovaných měřítek GLM nejsou tyto testy k dispozici, pokud neexistují žádné faktory mezi předměty a post hoc více porovnávacích testů se provádí pro průměr napříč úrovněmi vnitropředmětnými faktory. Pro GLM Multivariate jsou prováděny post hoc testy pro každou závislou proměnnou zvlášť. GGLM Multivariate a GLM Repeated Measures jsou k dispozici pouze tehdy, máte-li nainstalovanou volbu Rozšířená statistika.

Významné porovnávací testy Bonferroniho a Tukeyho se běžně používají pro více porovnávacích testů.

**Test Bonferroniho** na základě statistiky *t* Studenta upravuje pozorovanou úroveň významnosti pro skutečnost, že bylo provedeno více porovnání. **Test Sidak t** také upravuje úroveň významnosti a poskytuje přísnější meze než test Bonferroniho. **Tupkey's abely significant difference test** používá statistiku Studentized range *k* pro provedení všech porovnání po dvojicích skupin a nastavuje chybu experimentálního testu na četnost chyb pro kolekci pro všechny porovnání po dvojicích. Při testování velkého počtu párů prostředků je Tukey poctivě významný rozdíl testu je mnohem silnější než Bonferroniho test. Pro malý počet párů je Bonferroni silnější.

**Hochbergův GT2** je podobný testu s poctivým rozdílem Tukey, ale používá se maximální modulo Studentized (Studentized). Obvykle je Tukeyův test mnohem silnější. **Test porovnání Gabriela po dvojicích** používá také Studentizovaný maximální modulo a je obecně výkonnější než Hochbergův GT2 , když jsou velikosti buněk nestejně. Gabrielův test se může stát liberálním, když se velikosti buněk velmi liší.

**Multiple Dunnett's pairwise multiple comparison t test** porovnává sadu zpracování proti jediné řídicí skupině. Kategorie poslední kategorie je výchozí řídicí kategorie. Alternativně můžete zvolit první kategorii.

Můžete si také vybrat oboustranný nebo jednostranný test. Chcete-li otestovat, že střední hodnota na jakékoli úrovni (kromě kontrolní kategorie) faktoru se nerovná úrovni kontrolní kategorie, použijte dvoustrannou zkoušku. Chcete-li otestovat, zda je střední hodnota na jakékoli úrovni faktoru menší než velikost řídicí kategorie, vyberte volbu **< Control**. Podobně pro testování, zda je průměr na libovolné úrovni faktoru větší než průměr řídicí kategorie, vyberte volbu **> Řízení**.

Ryane, Einot, Gabriel a Welsch (R-E-G-W) vyvinuly dva vícenásobné zkoušky nevlastním rozpětím. Více postupných procedur nejprve testují, zda jsou všechny prostředky shodné. Pokud všechny prostředky nejsou stejné, podmnožiny středních hodnot jsou testovány na rovnost. **R-E-G-W F** je založen na testu *F* a **R-E-G-W Q** je založen na rozsahu Studentized. Tyto testy jsou výkonnější než Duncanův mnohonásobný test a Student-Newman-Keuls (který je také několikanásobně postupnými postupy), ale nejsou doporučovány pro nestejně velikosti buněk.

Když jsou rozptyly nestejně, použijte **Tamhaneovo T2** (konzervativní porovnávání po dvojicích na základě testu *r*), **Dunnettův T3** (test porovnávání po dvojicích založený na maximálním modulu Studentized), **Hry-Howellův porovnávací test po dvojicích** (někdy liberálně) nebo **Dunnettův C** (test porovnávání po dvojicích založený na rozsahu Studentized). Všimněte si, že tyto testy nejsou platné a nebudou vyprodukovány, pokud je v modelu více faktorů.

**Duncan's multiple range test**, Student-Newman-Keuls (**S-N-K**) a **Tukeyův b** jsou rozsahy testů, které skupina disků rank znamená a vypočítá hodnotu rozsahu. Tyto testy se nepoužívají tak často, jak se dříve diskutovalo o testech.

The **Waller-Duncan t test** používá Bayesův přístup. Tato zkouška rozsahu používá harmonický průměr velikosti vzorku, když velikosti vzorku jsou nestejně.

Úroveň významnosti testu **Scheffé** je navržena tak, aby umožňovala testování všech možných lineárních kombinací skupin, nikoli pouze po porovnávání po dvojicích, které jsou k dispozici v této funkci. Výsledkem je, že test Scheffé je často konzervativnější než jiné testy, což znamená, že významnost vyžaduje větší rozdíl mezi prostředky.

Nejméně významný rozdíl (**LSD**) vícenásobný porovnávací test je ekvivalentní s více individuálními testy *t* mezi všemi dvojicemi skupin. Nevýhodou tohoto testu je, že není proveden žádný pokus upravit pozorovanou úroveň významnosti pro více porovnání.

**Provedené testy.** Porovnávání po dvojicích jsou poskytovány pro LSD, Sidak, Bonferronii, Games-Howell, Tamhaneův T2 a T3, Dunnett's Ca Dunnett je T3. Homogenní podmnožiny pro testy rozsahu jsou poskytovány pro S-N-K, Tukeyův *b*, Duncan, R-E-G-W *F*, R-E-G-W *Q*a Waller. Tukey je opravdu významný rozdíl test, Hochberg je GT2, Gabriel test, a Scheffé test je jak více porovnávacích testů a range testů.

## Volby GLM

Volitelné statistiky jsou k dispozici z tohoto dialogového okna. Statistika se vypočítává pomocí modelu s pevným efektem.

**Obrazovka.** Vyberte volbu **Deskriptivní statistika** k vytvoření pozorovaných středních hodnot, směrodatných odchylek a počtů pro všechny závislé proměnné ve všech buňkách. **Odhady velikosti efektu** poskytují pro každý efekt a každý odhad hodnoty dílčí hodnoty éta na druhou. Statistika na druhou mocninu popisuje část celkové variability přiřaditelné faktoru. Vyberte **Sledovaný výkon**, chcete-li získat výkon testu, když je alternativní hypotéza založena na zjištěné hodnotě. Vyberte volbu **Odhady parametrů**, chcete-li vytvořit odhady parametrů, standardní chyby, testy *t*, intervaly spolehlivosti a pozorovanou sílu pro každý test. Chcete-li získat matici **L**, vyberte volbu **Variační koeficient kontrastu**.

**Testy homogenity** vytváří Levene testy homogenity rozptylu pro každou závislou proměnnou napříč všemi kombinacemi úrovně mezi faktory mezi faktory, pouze pro faktory mezi předměty. Volby rozložení dat na úrovni a zbytkové hodnoty jsou užitečné při kontrole předpokladů o datech. Tato položka je zakázána, pokud zde nejsou žádné faktory. Vyberte volbu **Zbytkový graf**, chcete-li pro každou závislou proměnnou vytvořit pozorovaný zbytkový graf se standardizovaným reziduem. Tyto zkusné plochy jsou užitečné při zkoumání předpokladu stejné odchylky. Chcete-li zkontrolovat, zda vztah mezi závislou proměnnou a nezávislými proměnnými může být odpovídajícím způsobem popsán modelem, vyberte volbu **Chybějící přízpusobení**. **General estimable function (s)** vám umožňuje vytvořit vlastní testy hypotézy založené na

obecné odhadnutelné funkci (ech). Rows in any contrast coefficient matrix are linear combinations of the general estimable function (s).

**Heteroskedastické testy** jsou dostupné pro testování, zda odchylky chyb (pro každou závislou proměnnou) závisí na hodnotách nezávislých proměnných. Pro test **Breusch-Pagan, Modified Breusch-Pagan testu F test** můžete zadat model, na kterém je test založen. Model se standardně skládá z konstantního termínu, z výrazu, který je lineární v předpovězných hodnotách, z výrazu, který je kvadratický v předpovězenou hodnotu, a chybový výraz.

**Odhady parametrů s robustními standardními chybami** zobrazují tabulku odhadů parametrů spolu s robustními nebo heteroskopickými směrodatnými chybami (HC) a *t* statistikami, hodnotami významnosti a intervaly spolehlivosti, které používají robustní standardní chyby. Pro robustní odhad matic kovariance je k dispozici pět různých metod.

#### HC0

Na základě původních asymptotických nebo velkých vzorků robustního, empirického nebo "sendvičového" odhadu kovarianční matice hodnot parametrů. Střední část sendviče obsahuje čtvercový OLS (ordinální nejmenších čtverců) nebo čtvercový Vážené WLS (vážené nejmenších čtverců) zbytkové chyby.

#### HC1

Konečnou ukázkou modifikace HC0 vynásobením  $N/(N-p)$ , kde  $N$  je velikost vzorku a  $p$  je počet neredundantních parametrů v modelu.

#### HC2

Modifikace HC0, která zahrnuje vydělení zbytkové zbytkové hodnoty o  $1-h$ , kde  $h$  je pákový efekt pro daný případ.

#### HC3

Modifikace HC0, která se blíží odhadci jacksového odhadu. Druhá mocnina se dělí čtvercem o  $1-h$ .

#### HC4

Modifikace HC0, která dělí čtvercový zbytek o  $1-h$  na moc, která se liší podle  $h$ ,  $N$  a  $p$ , s horním limitem 4.

**Hladina významnosti.** Možná budete chtít upravit úroveň významnosti použitou v post hoc testech a úroveň spolehlivosti užitou pro vytváření intervalů spolehlivosti. Uvedená hodnota se také použije pro výpočet pozorovaného výkonu pro test. Určíte-li úroveň významnosti, v dialogovém okně se zobrazí přidružená úroveň intervalů spolehlivosti.

## Další funkce příkazu UNIANOVA

Jazyk syntaxe příkazu vám také umožňuje:

- Určete vnořené efekty v návrhu (pomocí dílčího příkazu DESIGN).
- Určete testy efektů oproti lineární kombinaci efektů nebo hodnoty (pomocí dílčího příkazu TEST).
- Určete více kontrastů (pomocí dílčího příkazu CONTRAST).
- Zahrnout uživatele-chybějící hodnoty (pomocí dílčího příkazu MISSING).
- Zadejte kritéria EPS (pomocí dílčího příkazu CRITERIA).
- Vytvořte vlastní matici **L**, matici **M** nebo matici **K** (s použitím dílčích příkazů LMATRIX, MMATRIX a KMATRIX).
- Pro odchylky nebo jednoduché kontrasty uveďte intermediační referenční kategorii (pomocí dílčího příkazu CONTRAST).
- Uveďte metriky pro polynomiální kontrasty (pomocí dílčího příkazu CONTRAST).
- Určete chybové termíny pro post hoc porovnání (pomocí dílčího příkazu POSTHOC).
- Výpočet odhadovaných mezních hodnot pro jakýkoli součinitel nebo součinitel interakcí mezi faktory v seznamu faktorů (pomocí dílčího příkazu EMMEANS).
- Zadejte názvy pro dočasné proměnné (pomocí dílčího příkazu SAVE).
- Vytvořit datový soubor matice korelace (pomocí dílčího příkazu OUTFILE).

- Vytvořte datový soubor matice, který obsahuje statistické údaje z tabulky mezi subjekty ANOVA (pomocí dílčího příkazu `OUTFILE`).
- Uložte matici návrhu na nový datový soubor (pomocí dílčího příkazu `OUTFILE`).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Uložení GLM

Můžete uložit hodnoty, které model předpověděl, zbytkové chyby a související ukazatele jako nové proměnné v editoru dat. Mnohé z těchto proměnných lze použít k přezkoumání předpokladů o datech. Chcete-li uložit hodnoty pro použití v jiné relaci IBM SPSS Statistics, musíte uložit aktuální datový soubor.

**Předpovězené hodnoty.** Hodnoty, které model předpovídá pro každý případ.

- *Nestandardizováno.* Hodnota, kterou model předpovídá pro závislou proměnnou.
- *Vážené.* Vážené nestandardizované předpovězené hodnoty. Tato volba je k dispozici pouze v případě, že byla dříve vybrána proměnná WLS.
- *Standardní chyba.* Odhad směrodatné odchylky průměrné hodnoty závislé proměnné pro případy, které mají stejné hodnoty u nezávislých proměnných.

**Diagnostika.** Opatření k identifikaci případů s neobvyklými kombinacemi hodnot pro nezávislé proměnné a případy, které mohou mít velký dopad na model.

- *Kuchařova vzdálenost.* Měřítka toho, kolik zbytkových chyb ze všech případů se změní, pokud byl konkrétní případ vyloučen z výpočtu regresních koeficientů. Velké Cook D označuje, že vyloučení případu z výpočtu regresní statistiky mění koeficienty podstatně.
- *Efektivní využití hodnot.* Nesoustředěné pákové hodnoty. Relativní vliv každého pozorování na přizpůsobení modelu.

**Residuals.** Nestandardizovaná zbytková hodnota je skutečnou hodnotou závislé proměnné minus hodnota předpovězená modelem. Standardizované, Studentizované a odstraněné zbytkové chyby jsou také k dispozici. Byla-li zvolena proměnná WLS, jsou k dispozici vážené nestandardizované zbytkové chyby.

- *Nestandardizováno.* Rozdíl mezi pozorovanou hodnotou a hodnotou předpovězenou modelem.
- *Vážené.* Vážené nestandardizované zbytkové chyby. Tato volba je k dispozici pouze v případě, že byla dříve vybrána proměnná WLS.
- *Standardizováno.* Zbytek vydělený odhadem jeho směrodatné odchylky. Standardizované zbytkové chyby, které jsou také známy jako Pearsonové zbytkové chyby, mají střední hodnotu 0 a směrodatnou odchylku 1.
- *Studentizovaný.* Zbytek vydělený odhadem jeho směrodatné odchylky, která se liší případ od případu, v závislosti na vzdálenosti hodnot jednotlivých případů na nezávislých proměnných od středních hodnot nezávislých proměnných. Někdy se odkazuje jako na interně stomatozované zbytkové chyby.
- *Odstraněno.* Zbytková hodnota pro případ, kdy je tento případ vyloučen z výpočtu regresních koeficientů. Je to rozdíl mezi hodnotou závislé proměnné a upravenou předpokládanou hodnotou.

**Coefficient Statistics.** Zapisuje matici kovariance-kovarianční matici parametrů v modelu do nové datové sady v aktuální relaci nebo externího datového souboru IBM SPSS Statistics. Pro každou závislou proměnnou bude navíc existovat řádek odhadů parametrů, řádek standardních chyb odhadů parametru, řádek hodnot významnosti pro statistiku  $t$  odpovídající odhadům parametrů a řádek zbytkových stupňů volnosti. U vícerozměrného modelu existují podobné řádky pro každou závislou proměnnou. Když je vybrána funkce Heteroskedasticity-konzistentní statistika (k dispozici pouze pro modely univariate), je matice kovariance-kovariance vypočítána pomocí robustního odhadu, řádek standardních chyb zobrazuje robustní standardní chyby a hodnoty významnosti odrážejí robustní chyby. Tento maticový soubor můžete použít v jiných procedurách, které čtou soubory matice.

## Odhadované mezní hodnoty GLM

Vyberte faktory a interakce, pro které chcete, aby odhady okrajových hodnot populace byly v buňkách. Tyto prostředky jsou upraveny pro proměnné, pokud existují.

### Porovnání hlavních efektů

Poskytuje nekorigovaná porovnávání po dvojicích mezi odhadovanými marginálními prostředky pro jakýkoli hlavní účinek v modelu, a to jak mezi faktory mezi-a uvnitř-předměty. Tato položka je k dispozici pouze v případě, že jsou vybrány hlavní efekty pod seznamem Zobrazit prostředky pro seznam.

### Porovnání jednoduchých hlavních efektů

Nastavení je povoleno vždy, když cílový seznam obsahuje jeden nebo více efektů produktu nebo interakce (například  $A \times B$ ,  $A \times B \times C$ ). Nastavení podporuje specifikaci porovnání mezi jednoduchými hlavními efekty, které jsou hlavními účinky vnořenými v úrovních jiných faktorů.

### Úprava intervalu spolehlivosti

Vyberte nejméně významný rozdíl (LSD), Bonferroniho nebo nastavení hodnoty Sidak na intervaly spolehlivosti a významnost. Tato položka je k dispozici pouze v případě, že je vybrána volba **Porovnat hlavní efekty** a/nebo **Porovnat jednoduché hlavní efekty**.

## Určení odhadnutých marginální středních hodnot

1. V nabídce vyberte jednu z procedur dostupných v nabídce > **Analyzovat** > **Obecný lineární model**.
2. V hlavním dialogovém okně klepněte na volbu **EM prostředky**.

## Volby GLM

Volitelné statistiky jsou k dispozici z tohoto dialogového okna. Statistika se vypočítává pomocí modelu s pevným efektem.

**Obrazovka.** Vyberte volbu **Deskriptivní statistika** k vytvoření pozorovaných středních hodnot, směrodatných odchylek a počtů pro všechny závislé proměnné ve všech buňkách. **Odhady velikosti efektu** poskytují pro každý efekt a každý odhad hodnoty dílčí hodnotu éta na druhou. Statistika na druhou mocninu popisuje část celkové variability přiřaditelné faktoru. Vyberte **Sledovaný výkon**, chcete-li získat výkon testu, když je alternativní hypotéza založena na zjištěné hodnotě. Vyberte volbu **Odhady parametrů**, chcete-li vytvořit odhady parametrů, standardní chyby, testy  $t$ , intervaly spolehlivosti a pozorovanou sílu pro každý test. Chcete-li získat matici **L**, vyberte volbu **Variační koeficient kontrastu**.

**Testy homogenity** vytváří Levene testy homogenity rozptylu pro každou závislou proměnnou napříč všemi kombinacemi úrovně mezi faktory mezi faktory, pouze pro faktory mezi předměty. Volby rozložení dat na úrovni a zbytkové hodnoty jsou užitečné při kontrole předpokladů o datech. Tato položka je zakázána, pokud zde nejsou žádné faktory. Vyberte volbu **Zbytkový graf**, chcete-li pro každou závislou proměnnou vytvořit pozorovaný zbytkový graf se standardizovaným reziduem. Tyto zkusné plochy jsou užitečné při zkoumání předpokladu stejné odchylky. Chcete-li zkontrolovat, zda vztah mezi závislou proměnnou a nezávislými proměnnými může být odpovídajícím způsobem popsán modelem, vyberte volbu **Chybějící přizpůsobení**. **General estimable function (s)** vám umožňuje vytvořit vlastní testy hypotézy založené na obecné odhadnutelné funkci (ech). Rows in any contrast coefficient matrix are linear combinations of the general estimable function (s).

**Heteroskedastické testy** jsou dostupné pro testování, zda odchylky chyb (pro každou závislou proměnnou) závisí na hodnotách nezávislých proměnných. Pro test **Breusch-Pagan**, **Modified Breusch-Pagan** testu **F test** můžete zadat model, na kterém je test založen. Model se standardně skládá z konstantního termínu, z výrazu, který je lineární v předpovězných hodnotách, z výrazu, který je kvadratický v předpovězenou hodnotu, a chybový výraz.

**Odhady parametrů s robustními standardními chybami** zobrazují tabulku odhadů parametrů spolu s robustními nebo heteroskopickými směrodatnými chybami (HC) a  $t$  statistikami, hodnotami významnosti a intervaly spolehlivosti, které používají robustní standardní chyby. Pro robustní odhad matic kovariance je k dispozici pět různých metod.

## HC0

Na základě původních asymptotických nebo velkých vzorků robustního, empirického nebo "sendvičového" odhadu kovarianční matice hodnot parametrů. Střední část sendviče obsahuje čtvercový OLS (ordinální nejmenších čtverců) nebo čtvercový Vážené WLS (vážené nejmenších čtverců) zbytkové chyby.

## HC1

Konečnou ukázkou modifikace HC0 vynásobením  $N/(N-p)$ , kde  $N$  je velikost vzorku a  $p$  je počet neredundantních parametrů v modelu.

## HC2

Modifikace HC0, která zahrnuje vydělení zbytkové zbytkové hodnoty o  $1-h$ , kde  $h$  je pákový efekt pro daný případ.

## HC3

Modifikace HC0, která se blíží odhadci jacksového odhadu. Druhá mocnina se dělí čtvercem o  $1-h$ .

## HC4

Modifikace HC0, která dělí čtvercový zbytek o  $1-h$  na moc, která se liší podle  $h$ ,  $N$  a  $p$ , s horním limitem 4.

**Hladina významnosti.** Možná budete chtít upravit úroveň významnosti použitou v post hoc testech a úroveň spolehlivosti užitou pro vytváření intervalů spolehlivosti. Uvedená hodnota se také použije pro výpočet pozorovaného výkonu pro test. Určíte-li úroveň významnosti, v dialogovém okně se zobrazí přidružená úroveň intervalů spolehlivosti.

## Pomocný regresní model GLM

Dialogové okno Pomocný regresní model uvádí model, který se používá pro testování heteroskedasticity.

### Použití předpovězené hodnoty

Používá model, který se skládá z konstantního výrazu, z výrazu, který je lineární v předpovězných hodnotách, z výrazu, který je kvadratický v předpovězných hodnotách, a chybový výraz.

### Použití jednorozměrný model

Používá model, který je zadán v dílčím dialogovém okně Model. Výraz zachycení je zahrnut, pokud uvedený model neobsahuje jeden.

### Vlastní model

Používá model, který jste explicitně zadali.

### Podmínky sestavení

Tuto volbu použijte, chcete-li zahrnout nevnořené podmínky určitého typu (například hlavní efekty) pro všechny kombinace vybrané sady faktorů a kovariany.

### Sestavit vlastní výrazy

Tuto volbu použijte, chcete-li zahrnout vnořené výrazy nebo pokud chcete explicitně sestavit libovolnou proměnnou term pomocí proměnné. Sestavení vnořené termínu zahrnuje následující kroky:

## Další funkce příkazu UNIANOVA

Jazyk syntaxe příkazu vám také umožňuje:

- Určete vnořené efekty v návrhu (pomocí dílčího příkazu DESIGN).
- Určete testy efektů oproti lineární kombinaci efektů nebo hodnoty (pomocí dílčího příkazu TEST).
- Určete více kontrastů (pomocí dílčího příkazu CONTRAST).
- Zahrnout uživatele-chybějící hodnoty (pomocí dílčího příkazu MISSING).
- Zadejte kritéria EPS (pomocí dílčího příkazu CRITERIA).
- Vytvořte vlastní matici **L**, matici **M** nebo matici **K** (s použitím dílčích příkazů LMATRIX, MMATRIX a KMATRIX).
- Pro odchylky nebo jednoduché kontrasty uveďte intermediační referenční kategorii (pomocí dílčího příkazu CONTRAST).



- Uved'te metriky pro polynomiální kontrasty (pomocí dílčího příkazu CONTRAST ).
- Určete chybové termíny pro post hoc porovnání (pomocí dílčího příkazu POSTHOC ).
- Výpočet odhadovaných mezních hodnot pro jakýkoli součinitel nebo součinitel interakcí mezi faktory v seznamu faktorů (pomocí dílčího příkazu EMMEANS ).
- Zadejte názvy pro dočasné proměnné (pomocí dílčího příkazu SAVE ).
- Vytvořit datový soubor matice korelace (pomocí dílčího příkazu OUTFILE ).
- Vytvořte datový soubor matice, který obsahuje statistické údaje z tabulky mezi subjekty ANOVA (pomocí dílčího příkazu OUTFILE ).
- Uložte matici návrhu na nový datový soubor (pomocí dílčího příkazu OUTFILE ).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Bivariantní koráže

Procedura Bivariate Correlations vypočítá Pearsonův korelační koeficient, Spearmanův rho a Kendall's tau-*b* se svými úrovněmi významnosti. Korelace měří, jak jsou související proměnné nebo řadové objednávky. Před výpočtem korelačního koeficientu, obrazovka your data for outliers (which can cause misleading results) and evidence of a linear relationship. Pearson je korelační koeficient je míra lineární asociace. Dvě proměnné mohou být dokonale vázané, ale pokud je vztah není lineární, Pearson je korelační koeficient není vhodnou statistikou pro měření jejich přidružení.

Nastavení intervalu spolehlivosti jsou k dispozici pro Pearson a Spearman.

### Příklad

Je počet her vyhrál basketbalový tým koreloval s průměrným počtem bodů skóroval za hru? Scatterplot označuje, že existuje lineární vztah. Analýza dat z období 1994-1995 NBA přináší, že Pearson je korelační koeficient (0.581) je významný na úrovni 0.01. Můžete mít podezření, že čím více her vyhrál za sezónu, tím méně bodů soupeři skóroval. Tyto proměnné jsou negativně korelovány (-0,401) a korelace je významná na úrovni 0,05.

### Statistika

Pro každou proměnnou: počet případů s nechybějícími hodnotami, střední hodnotou a směrodatnou odchylkou. Pro každou dvojici proměnných: Pearson je korelační koeficient, Spearman's rho, Kendall's tau-*b*, cross-product of deviations, and covariance.

## Aspekty dat

### Data

Pomocí symetrických kvantitativních proměnných pro Pearsonův korelační koeficient a kvantitativní proměnné nebo proměnné s objednanými kategoriemi pro Spearman's rho a Kendall's tau-*b*.

### Předpoklady

Pearsonův korelační koeficient předpokládá, že každý pár proměnných je bivariate normální.

## Získání korelace dat Bivariate

Z nabídky vyberte:

### Analyzovat > Korelovat > Bivariate ...

1. Vyberte dvě nebo více číselných proměnných.

K dispozici jsou také následující volby:

#### Korelační koeficienty

V případě kvantitativních, obvykle distribuovaných proměnných, vyberte korelační koeficient **Pearson** . Pokud vaše data nejsou normálně distribuována nebo máte objednané kategorie, vyberte **Kendall's tau-b** nebo **Spearman**, které měří přidružení mezi řadovými objednávkami. Korelační koeficienty se pohybují v hodnotách od -1 (perfektní negativní vztah) a + 1 (perfektní pozitivní vztah). Hodnota 0 označuje, že žádný lineární vztah neexistuje. Při interpretaci výsledků

buďte opatrní, abyste z důvodu závažné korelace nekreslili žádné závěry vyvolané a neúčinnými účinky.

### Zkouška významnosti

Můžete vybrat oboustranné nebo jednostranné pravděpodobnosti. Je-li směr přidružení znám předem, vyberte volbu **Jednostranný**. V opačném případě vyberte volbu **Dvojstranný**.

### Označit významné korelace

Korelační koeficienty significant at the 0.05 level are identified with a single asterisk, and those significant at the 0.01 level are identified with two asterisks.

### Zobrazit pouze dolní trojúhelník

Je-li tato volba vybrána, ve výstupu se zobrazí pouze dolní trojúhelník tabulky matice korelace. Není-li tato volba vybrána, bude ve výstupu uvedena úplná tabulka matice korelace. Nastavení povoluje výstup tabulky tak, aby odpovídal pokynům stylu APA.

### Zobrazit diagonální

Je-li tato volba vybrána, dolní trojúhelník tabulky matice korelace spolu s diagonálními hodnotami se zobrazí ve výstupu. Nastavení povoluje výstup tabulky tak, aby odpovídal pokynům stylu APA.

2. Volitelně můžete vybrat následující:

- Klepněte na tlačítko **Volby ...** pro uvedení Pearsonovy korelační statistiky a chybějících nastavení hodnot.
- Klepněte na volbu **Styl ...** určit podmínky pro automatické změny vlastností kontingenčních tabulek na základě specifických podmínek.
- Klepněte na **Zaváděcí program** pro odvození robustních odhadů standardních chyb a intervalů spolehlivosti pro odhady, jako je průměr, medián, poměr, kurzový poměr, korelační koeficient nebo regresní koeficient.
- Klepněte na tlačítko **Interval důvěry ...** nastavit volby pro odhad intervalů spolehlivosti.

## Volby relace korelace bivariate

### Statistika

Pro Pearsonovy korelace si můžete vybrat jednu nebo obě z následujících možností:

#### Prostředky a směrodatné odchylky

Zobrazí se pro každou proměnnou. Zobrazí se také počet případů s nechybějícími hodnotami. Chybějící hodnoty jsou zpracovány na základě proměnné podle proměnné bez ohledu na nastavení chybějících hodnot.

#### Odchylky od jednotlivých produktů a kovariance

Zobrazí se pro každou dvojici proměnných. Křížový výrobek odchylek se rovná součtu výrobků střední hodnoty korigovaných proměnných. Jedná se o čitatele Pearsonova korelačního koeficientu. Kovariance je nestandardizovaný ukazatel vztahu mezi dvěma proměnnými, který se rovná rozdílu mezi produktem děleno  $N-1$ .

### Chybějící hodnoty

Můžete zvolit jednu z následujících možností:

#### Vyloučit případy po dvojicích

Případy s chybějícími hodnotami pro jeden nebo oba dvojice proměnných pro korelační koeficient jsou vyloučeny z analýzy. Vzhledem k tomu, že každý koeficient je založen na všech případech, které mají platné kódy na daném páru proměnných, jsou maximální dostupné informace použity v každém výpočtu. To může vést k sadě koeficientů založených na různém počtu případů.

#### Vyloučit případy jako litwise

Případy s chybějícími hodnotami pro jakoukoli proměnnou jsou vyloučeny ze všech korelací.

## Interval spolehlivosti Bivariate Correlations

Dialog Interval spolehlivosti poskytuje volby pro odhad intervalů spolehlivosti. Dialogové okno je k dispozici, je-li v dialogovém okně Bivariate Correlations vybrána volba **Pearson**, **Kendall's tau-b** nebo **Spearman**.

## Odhadnout interval spolehlivosti parametru korelace bivariate

Řídí odhad intervalu spolehlivosti parametru korelace bivariate. Je-li označeno, dojde k odhadu intervalu spolehlivosti.

### Interval spolehlivosti (%)

Určuje úroveň důvěry pro všechny produkované intervaly spolehlivosti. Uvedte číselnou hodnotu mezi 0 a 100. Výchozí hodnota je 95.

### Pearsonová korelace

Nastavení **Použití nastavení předpojatosti** řídí, zda se použije úprava předpojatosti. Ve výchozím nastavení není toto nastavení vybráno, což nebere v úvahu termín zkreslení. Je-li tato volba vybrána, použije se změna předpojatosti k odhadu limitů spolehlivosti. Nastavení je k dispozici, je-li v dialogovém okně Bivariate Correlations vybrána volba **Pearson**.

### Spearmanova korelace

Nastavení je dostupné, když je vybráno **Spearman** v dialogovém okně Bivariate Correlations a poskytuje volby pro odhad odchylky korelace Spearman pomocí následujících metod:

- **Fieller, Hartley a Pearson**
- **Bonett a Wright**
- **Koruso a Cliff**

## Další funkce CORVZTAS a NONPAR CORR

Jazyk syntaxe příkazu vám také umožňuje:

- Zapište korelační matici pro Pearsonovy korelace, které lze použít místo prvotních dat k získání jiných analýz, jako je analýza faktoru (s dílčím příkazem MATRIX).
- Získejte korelace každé proměnné na seznamu s každou proměnnou na druhém seznamu (pomocí klíčového slova WITH na dílčím příkazu VARIABLES).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Částečné korelace

Procedura Dílčí korelace vypočítá dílčí korelační koeficienty, které popisují lineární vztah mezi dvěma proměnnými při řízení pro účinky jedné nebo více dalších proměnných. Korelace jsou ukazatele lineárního přidružení. Dvě proměnné mohou být dokonale vázané, ale pokud vztah není lineární, korelační koeficient není vhodnou statistikou pro měření jejich přidružení.

### Příklad

Existuje vztah mezi financováním zdravotní péče a porodními onemocněními? Ačkoli můžete očekávat, že jakýkoli takový vztah bude negativní, studie vykazuje významný *pozitivní* korelace: jak se zvyšuje financování zdravotní péče, zdá se, že se míra onemocnění zvyšuje. Kontrola míry návštěv u poskytovatelů zdravotní péče však prakticky eliminuje pozorovanou pozitivní korelaci. Zdá se, že více lidí má přístup ke zdravotní péči, když se zvyšuje financování, což vede k dalším hlášeným chorobám lékařů a nemocnic, a to pouze proto, že se zvyšuje zdravotní péče a míra onemocnění.

### Statistika

Pro každou proměnnou: počet případů s nechybějícími hodnotami, střední hodnotou a směrodatnou odchylkou. Částečné a nulové korelační matice, se stupni volnosti a významností.

## Aspekty dat

### Data

Použijte symetrické, kvantitativní proměnné.

### Předpoklady

Procedura Partial Correlations předpokládá, že každá dvojice proměnných je bivariantní normální.

## Získání částečných korelací

1. Z nabídky vyberte:

**Analyzovat > Korelovat > Dílčí ...**

2. Vyberte dvě nebo více numerických proměnných, pro které se mají vypočítat dílčí korelace.

3. Vyberte jednu nebo více číselných řídicích proměnných.

K dispozici jsou také následující volby:

### Zkouška významnosti

Můžete vybrat oboustranné nebo jednostranné pravděpodobnosti. Je-li směr přidružení znám předem, vyberte volbu **Jednostranný**. V opačném případě vyberte volbu **Dvojstranný**.

### Zobrazit aktuální úroveň významnosti

Při výchozím nastavení se pro každý korelační koeficient zobrazí pravděpodobnost a stupně volnosti. Pokud zrušíte výběr této položky, budou koeficienty významné na úrovni 0,05 identifikovány jednou hvězdičkou, koeficienty významné na úrovni 0.01 jsou identifikovány dvojitou hvězdičkou a stupně volnosti jsou potlačeny. Toto nastavení ovlivňuje jak dílčí, tak i matkové matice s pořadím.

## Volby dílčích korelací

**Statistika.** Můžete si vybrat jednu nebo obě z následujících možností:

- **Střední hodnoty a směrodatné odchylky.** Zobrazí se pro každou proměnnou. Zobrazí se také počet případů s nechybějícími hodnotami.
- **Korelace nulového pořadí.** Zobrazí se matice jednoduchých korelací mezi všemi proměnnými, včetně řídicích proměnných.

**Chybějící hodnoty.** Můžete zvolit jednu z následujících možností:

- **Vyloučit případy jako litwise.** Případy, které mají chybějící hodnoty pro každou proměnnou, včetně řídicí proměnné, jsou vyloučeny ze všech výpočtů.
- **Vyloučit případy po dvojicích.** Pro výpočet korelací s nulovým pořadím, na kterých jsou dílčí korelace založeny, se nepoužívá případ s chybějícími hodnotami pro oba nebo jeden z páru proměnných. Odstranění pairwise využívá co nejvíce dat. Počet případů se však může lišit napříč koeficienty. Je-li odstranění po dvojicích v platnosti, jsou stupně volnosti pro určitý dílčí koeficient založeny na nejmenším počtu případů použitých při výpočtu všech korelací nulového pořadí.

## Další funkce příkazu PARTIAL CORR

Jazyk syntaxe příkazu vám také umožňuje:

- Přečtete si korelační matici s nulovým pořadím nebo zapište částečnou korelační matici (s dílčím příkazem **MATRIX**).
- Získejte částečné korelace mezi dvěma seznamy proměnných (pomocí klíčového slova **WITH** na dílčím příkazu **VARIABLES**).
- Získejte více analýz (s více podpříkazy **VARIABLES**).
- Pokud máte dvě řídicí proměnné (s dílčím příkazem **VARIABLES**), uveďte hodnoty pořadí k požadavku (například jak první, tak i dílčí korelace s pořadovým číslem).
- Potlačí redundantní koeficienty (s dílčím příkazem **FORMAT**).
- Zobrazit matici jednoduchých korelací, když některé koeficienty nelze vypočítat (s dílčím příkazem **STATISTICS**).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Vzdálenosti

Tato procedura vypočítá jakoukoli z široké škály statistik měřících buď podobnosti, nebo nepodobnosti (vzdálenosti) mezi dvojicemi proměnných nebo mezi dvojicemi případů. Tato podobnost nebo vzdálenost

může být použita s dalšími procedurami, jako je analýza faktorů, analýza klastrů nebo vícerozměrné škálování, které pomáhají analyzovat komplexní datové sady.

**Příklad.** Je možné měřit podobnosti mezi páry automobilů na základě určitých charakteristik, jako je velikost motoru, MPG, a koňská síla? Díky výpočetní podobnosti mezi autos můžete získat představu o tom, které autok jsou podobné navzájem a které jsou odlišné od sebe navzájem. Pro formálnější analýzu můžete zvážit použití analýzy hierarchického klastru nebo vícerozměrné změny měřítka pro podobnost se zkoumáním základní struktury.

**Statistika.** Rozdílnost (vzdálenost) opatření pro interval údaje jsou euklidovský vzdálenost, čtvercový euklidovský odstup, Chebychev, blok, Minkowského, nebo přizpůsobené; pro počítání dat, chi-square nebo fi-square; pro binární data, euklidovský vzdálenost, čtvercový euklidovský vzdálenost, velikost rozdíl, vzorec rozdílů, odchylka, tvar, nebo Lance a Williams. Podobnost měřítek pro interval data jsou Pearson correlation nebo cosinus; pro binární data, Russel a Rao, jednoduché párování, Jaccard, kostky, Rogers a Tanimoto, Sokal a Snest 2, Sokal a Snest 3, Kulczynski 1, Kulczynski 2, Sokal a Snest 4, Hamann, Lambda, Anderberg's *D*, Yule's *Y*, Yule's *Q*, Ochiai, Sokal a Sneen 5, Fí 4 bodová korelace nebo rozptyl.

Chcete-li získat zásady vzdálenosti

1. Z nabídky vyberte:

**Analyzovat > Korelovat > Vzdálenosti ...**

2. Vyberte alespoň jednu číselnou proměnnou pro výpočet vzdáleností mezi případy, nebo vyberte alespoň dvě číselné proměnné pro výpočet vzdáleností mezi proměnnými.

3. Vyberte alternativu ve skupině Compute Distances pro výpočet proximit buď mezi případy, nebo mezi proměnnými.

## Měřítka přidružení k dispodobnosti

Ze skupiny ukazatelů vyberte alternativu, která odpovídá vašemu typu dat (interval, počet nebo binární); potom z rozevíracího seznamu vyberte jedno z ukazatelů, které odpovídají tomuto typu dat. Dostupná opatření, podle datového typu, jsou:

- **Interval dat.** euklidovský vzdálenost, na druhou euklidovský odstup, Čebychev, blok, Minkowského, nebo přizpůsobené.
- **Count dat.** Chi-kvadrát ukazatel nebo měřítko Fí-náměstí.
- **Binární data.** euklidovský odstup, čtvercový euklidovský vzdálenost, velikost rozdíl, vzorec rozdíl, odchylka, tvar, nebo Lance a Williams. (Zadejte hodnoty pro prezentaci a nepřítomnost, abyste určili, které dvě hodnoty jsou smysluplné; Vzdálenosti budou ignorovat všechny ostatní hodnoty.)

Skupina Hodnoty transformace vám umožňuje standardizovat datové hodnoty buď pro případy, nebo proměnné před výpočetními proximity. Tyto transformace nejsou použitelné pro binární data. Dostupné standardizační metody jsou z skóre, rozsah -1 až 1, rozsah 0 až 1, maximální velikost 1, střední hodnota 1 nebo směrodatná odchylka 1.

Skupina Ukazatele transformace Transform vám umožňuje transformovat hodnoty generované ukazatelem vzdálenosti. Jsou použity po vypočtených vzdálenoměrných opatřeních. Dostupné volby jsou absolutní hodnoty, změna znaménka a změna měřítka na 0-1 rozsah.

## Měřítka podobnosti odpojení

Ze skupiny měřítek vyberte alternativu, která odpovídá vašemu typu dat (interval nebo binární soubor); potom z rozevíracího seznamu vyberte jedno z opatření, které odpovídá tomuto typu dat. Dostupná opatření, podle datového typu, jsou:

- **Interval dat.** Pearsonův souvztažnost nebo
- **Binární data.** Russell a Rao, jednoduché párování, Jaccard, Dice, Rogers a Tanimoto, Sokal a Snest 2, Sokal a Snest 3, Kulczynski 1, Kulczynski 2, Sokal a Snest 4, Hamann, Lambda, Anderberg's *D*, Yule's *Y*, Yule's *Q*, Ochiai, Sokal a Sneath 5, Fí 4 bodová korelace, nebo disperze. (Zadejte hodnoty pro prezentaci a nepřítomnost, abyste určili, které dvě hodnoty jsou smysluplné; Vzdálenosti budou ignorovat všechny ostatní hodnoty.)

Skupina Hodnoty transformace vám umožňuje standardizovat datové hodnoty buď pro případy, nebo proměnné před výpočtem proximities. Tyto transformace nejsou použitelné pro binární data. Dostupné standardizační metody jsou z skóre, rozsah -1 až 1, rozsah 0 až 1, maximální velikost 1, střední hodnota 1 a směrodatná odchylka 1.

Skupina Ukazatele transformace Transform vám umožňuje transformovat hodnoty generované ukazatelem vzdálenosti. Jsou použity po vypočtených vzdálenoměrných opatřeních. Dostupné volby jsou absolutní hodnoty, změna znaménka a změna měřítka na 0-1 rozsah.

## Další funkce příkazu PROXIMITIES

Procedura Dices používá syntaxi příkazu PROXIMITIES . Jazyk syntaxe příkazu vám také umožňuje:

- Uvedte libovolné celé číslo jako napájení pro měřítka Minkowského vzdálenosti.
- Uvedte jakákoli celá čísla jako mocninu a kořen pro upravené měřítka vzdálenosti.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Lineární modely

---

Lineární modely předpovídají souvislý cíl založený na lineárních vztazích mezi cílem a jedním nebo více prediktory.

Lineární modely jsou relativně jednoduché a poskytují snadno interpretovaný matematický vzorec pro bodování. Vlastnosti těchto modelů jsou dobře pochopeny a mohou být obvykle postaveny velmi rychle ve srovnání s jinými typy modelů (jako jsou neuronové sítě nebo rozhodovací stromy) na stejné datové sadě.

**Příklad.** Pojišťovna s omezenými prostředky na prošetření pojistných nároků majitelů domů chce vytvořit model pro odhad nákladů na pojistné události. Nasazením tohoto modelu do servisních center mohou zástupci zadávat informace o nárocích a okamžitě získat "očekávané" náklady na nárok na základě minulých dat.

**Požadavky na pole.** Musí existovat cíl a alespoň jeden vstup. Ve výchozím nastavení nejsou pole s předdefinovanými rolemi Obojí nebo Žádná použita. Cíl musí být spojitý (scale). Pro prediktory (vstupy) nejsou žádná omezení na úrovni měření; kategoriální (nominal a ordinal) se používají jako faktory v modelu a souvislá pole se používají jako proměnné.

**Poznámka:** Má-li kategoričné pole více než 1000 kategorič, procedura se nespustí a nebude sestaven žádný model.

## Získání lineárního modelu

Tato funkce vyžaduje volbu Statistika základu.

Z nabídky vyberte:

**Analyzovat > Regrese > Automatický lineární model ...**

1. Ujistěte se, že je zde alespoň jeden cíl a jeden vstup.
2. Chcete-li určit volitelné nastavení sestavení a modelu, klepněte na volbu **Volby sestavení** .
3. Klepněte na tlačítko **Volby modelu** , abyste uložili skóre do aktivní datové sady a exportovali model do externího souboru.
4. Klepnutím na tlačítko **Spustit** spustíte proceduru a vytvoříte objekty modelu.

## Cíle

**Jaký je váš hlavní cíl?** Vyberte příslušný cíl.

- **Vytvořte standardní model.** Metoda sestaví jeden model, aby předpověděl cíl pomocí prediktorů. Obecně řečeno, standardní modely se snáze interpretují a mohou být rychlejší než zesílené, baggované nebo velké datové sady.

- **Zvýšit přesnost modelu (posílení).** Metoda sestaví model kompletu používající zesílení, který generuje posloupnost modelů k získání přesnějších předpovědí. Komplety mohou trvat déle než sestavení a skóre než standardní model.

Posilování produkuje dědění "modelů komponent", z nichž každá je sestavena na celé datové sadě. Před sestavením každého následného modelu komponenty jsou záznamy váženy na základě zbytkových chyb předchozího modelu komponenty. Případy s velkými rezidui mají relativně vyšší váhu analýzy, takže se následující model komponenty zaměří na předpovídání těchto záznamů dobře. Tyto modely komponent společně tvoří model kompletu. Model modelu sepisuje nové záznamy pomocí kombinačního pravidla. Dostupná pravidla závisí na úrovni měření cíle.

- **Zlepšete stabilitu modelu (blokování).** Metoda sestaví model kompletu pomocí přetažení (zaváděcího programu agregací), který generuje více modelů k získání spolehlivějších předpovědí. Komplety mohou trvat déle než sestavení a skóre než standardní model.

Agregace samozavěření (blokování) produkuje repliky cvikujícího souboru dat pomocí vzorkování s náhradou z původní datové sady. Tím se vytvoří vzorky pro samozavedení stejné velikosti původní datové sady. Pak je na každé replice vytvořen "model komponent". Tyto modely komponent společně tvoří model kompletu. Model modelu sepisuje nové záznamy pomocí kombinačního pravidla. Dostupná pravidla závisí na úrovni měření cíle.

- **Vytvořte model pro velmi rozsáhlé datové sady (vyžaduje server IBM SPSS Statistics).** Metoda sestaví model kompletu rozdělením datové sady do samostatných datových bloků. Vyberte tuto volbu, pokud je vaše datová sada příliš velká na sestavení libovolného z výše uvedených modelů, nebo pro přírůstkové sestavení modelu. Tato volba může trvat méně času k sestavení, ale může trvat déle než skóre než standardní model. Tato volba vyžaduje konektivitu produktu IBM SPSS Statistics Server.

Nastavení související s nastavením zvyšování, přetahování pomocí myši a velmi velké datové sady zvyšování a přetahování pomocí myšiviz ["Komplety"](#) na stránce 136.

## Základní

**Automaticky připravit data.** Tato volba umožňuje interně transformovat cíl a prediktory za účelem maximalizace predikční síly modelu; všechny transformace se uloží s modelem a použijí se na nová data pro přidělení skóre. Původní verze transformovaných polí jsou vyloučeny z modelu. Standardně se provádí následující automatická příprava dat.

- **Práce s datem a časem.** Každý prediktor data je transformován do nového prediktoru předpovědi, který obsahuje uplynulý čas od referenčního data (1970-01-01). Každý časový prediktor se transformuje do nového souvislého prediktoru, který obsahuje čas uplynulý od referenčního času (00:00:00).
- **Upravit úroveň měření.** Souvislé prediktory s méně než 5 odlišnými hodnotami jsou přepracované jako ordinální prediktory. Ordinální prediktory s více než 10 odlišnými hodnotami jsou přepracovány jako průběžné prediktory.
- **Ošetření odlehlých hodnot.** Hodnoty průběžných prediktorů, které leží mimo hodnotu uzavření (3 směrodatné odchylky od střední hodnoty), jsou nastaveny na hodnotu uzavření objektu.
- **Zpracování chybějících hodnot.** Chybějící hodnoty nominálního prediktorů jsou nahrazeny režimem tréninkového oddílu. Chybějící hodnoty pořadových prediktorů jsou nahrazeny mediánem tréninkového oddílu. Chybějící hodnoty průběžných prediktorů jsou nahrazeny střední částí tréninkového oddílu.
- **slučovaná sloučení.** Tento model je více komplikovanější tím, že se sníží počet polí, která mají být zpracována ve spojení s cílem. Podobné kategorie jsou identifikovány na základě vztahu mezi vstupem a cílem. Kategorie, které nejsou výrazně odlišné (tj. jejichž hodnota  $p$  je větší než 0,1), jsou sloučeny. Jsou-li všechny kategorie sloučeny do jedné, budou původní a odvozené verze pole vyloučeny z modelu, protože nemají žádnou hodnotu jako prediktor.

**Úroveň důvěry.** Jedná se o úroveň důvěry použitou k výpočtu odhadu intervalů podle koeficientů modelu v pohledu [Koeficienty](#). Zadejte hodnotu větší než 0 a menší než 100. Předvolba je 95.

## Výběr modelu

**Metoda výběru modelu.** Vyberte jednu z metod výběru modelu (podrobnosti viz níže) nebo **Zahrnout všechny prediktory**, které jednoduše vstoupí do všech dostupných prediktorů jako hlavní efekty modelu efektů. Ve výchozím nastavení je použita volba **Předat po krocích**.

**Postoupit po výběru.** To začíná bez efektů v modelu a přidá a odebere jeden krok v čase, dokud nebude možné přidávat nebo odebrat další prvky podle nevlastních kritérií.

- **Kritéria pro vstup/odebrání.** Toto je statistika použitá k určení, zda má být nějaký efekt přidán do modelu nebo z něj odebrán. **Informační kritérium (AICC)** je založeno na pravděpodobnosti sady školení vzhledem k modelu a je upraveno tak, aby penalizovalo příliš složité modely. **F Statistics** je založena na statistickém testu zlepšení chyby modelu. **Upravená R-druhou** je založena na vhodnosti sady školení a je přizpůsobena tak, aby penalizovala příliš složité modely. **Kritérium pro prevenci proložení (ASE)** je založeno na vhodnosti (průměrná kvadratická chyba nebo ASE) pro přednastavenou preventivní sadu. Předvstupní prevence je náhodný dílčí vzorek přibližně 30% původní datové sady, která se nepoužívá k vycvičování modelu.

Pokud je zvoleno jiné kritérium než **Statistika F**, pak se v každém kroku přidá do modelu efekt, který odpovídá největšímu kladnému zvýšení v daném kritériu. Všechny účinky v modelu, které odpovídají snížení v kritériu, se odstraní.

Je-li jako kritérium vybrána volba **Statistika F**, je do modelu přidán efekt, který má nejmenší hodnotu *p* menší než určená prahová hodnota, **Zahrnout efekty s *p*-hodnotami nižšími než**. Výchozí hodnota je 0,05. Všechny efekty v modelu s hodnotou *p* vyšší, než je uvedená prahová hodnota, **Odebrat efekty s *p*-hodnotami vyššími než**, budou odebrány. Výchozí hodnota je 0,10.

- **Upravte maximální počet efektů v konečném modelu.** Ve výchozím nastavení lze do modelu zadat všechny dostupné efekty. Případně, pokud algoritmus krokového algoritmu ukončí krok s uvedeným maximálním počtem efektů, algoritmus se zastaví s aktuální sadou efektů.
- **Upravte maximální počet kroků.** Po určitém počtu kroků se algoritmus po krocích zastaví. Ve výchozím nastavení se jedná o trojnásobek počtu dostupných efektů. Případně uveďte kladné celé číslo, které je maximální počet kroků.

**Výběr nejlepších dílčích sad.** To kontroluje "všechny možné" modely, nebo alespoň větší část z možných modelů než dopředu po krocích, vybrat nejlepší podle nejlepší podmnožiny kritérium. **Informační kritérium (AICC)** je založeno na pravděpodobnosti nastavené sady školení vzhledem k modelu a je upraveno tak, aby penalizovalo příliš složité modely. **Upravená R-druhou** je založena na vhodnosti sady školení a je přizpůsobena tak, aby penalizovala příliš složité modely. **Kritérium pro prevenci proložení (ASE)** je založeno na vhodnosti (průměrná kvadratická chyba nebo ASE) pro přednastavenou preventivní sadu. Předvstupní prevence je náhodný dílčí vzorek přibližně 30% původní datové sady, která se nepoužívá k vycvičování modelu.

Model s největší hodnotou kritéria se volí jako nejlepší model.

**Poznámka:** Nejlepší podmnožiny výběru je více výpočetně intenzivní než dopředu po kroku výběru. Když se provádí nejlepší podmnožiny ve spojení se zvýšením, vytažením nebo velmi velkými datovými sadami, může to trvat podstatně déle, než standardní model vytvořený s využitím dopředného krokování.

## Komplety

Tato nastavení určují chování ensembleru, ke kterému dochází při požadavku na zvyšování, přetahování nebo velmi velké datové sady v cílech. Volby, které se nepoužijí na vybraný záměr, budou ignorovány.

**Barevné práce a velmi velké datové sady.** Při hodnocení kompletu se jedná o pravidlo použité ke sloučení předpokládaných hodnot ze základních modelů pro výpočet hodnoty skóre kompletu.

- **Výchozí kombinační pravidlo pro souvislé cíle.** Předpovídané kompletování hodnot pro souvislé cíle lze kombinovat pomocí střední hodnoty nebo mediánu předpokládaných hodnot ze základních modelů.

Všimněte si, že když cílem je zvýšit přesnost modelu, výběry kombinování pravidel se budou ignorovat. Posilování vždy používá vážené většinové hlasy pro skóre kategoriálních cílů a vážený medián pro skóre průběžných cílů.



**Bootig and Bagging.** Určete počet základních modelů, které se mají sestavit, je-li cílem zvýšit přesnost a stabilitu modelu; pro přetažení se jedná o počet vzorků samozavedení. Mělo by se jednat o kladné celé číslo.

## Rozšířené

**Proveďte replikaci výsledků.** Nastavení náhodného náhodného řetězce vám umožní replikovat analýzy. Generátor náhodných čísel se používá k výběru toho, které záznamy jsou v preventivní sadě pro nadměrné přizpůsobení. Zadejte celé číslo nebo klepněte na tlačítko **Generovat**, které vytvoří pseudo-náhodné celé číslo v rozsahu 1 až 2147483647 včetně. Výchozí hodnota je 54752075.

## Volby modelu

**Ukládat předpovězené hodnoty do datové sady.** Výchozí název proměnné je *PredictedValue*.

**Model exportu.** Tento zápis zapíše model do externího souboru .zip . Tento modelový soubor můžete použít k použití informací modelu na jiné datové soubory pro účely hodnocení. Uveďte jedinečný, platný název souboru. Pokud se specifikace souboru odkazuje na existující soubor, pak se soubor přepíše.

## Souhrn modelu

Zobrazení Souhrn modelu je snímek, souhrn modelu a jeho přizpůsobení.

**tabulka:** Tabulka identifikuje některá nastavení modelu vyšší úrovně, včetně:

- Název cíle uvedeného na kartě Pole ,
- Zda byla provedena automatická příprava dat, jak je uvedeno v nastavení Základy,
- Výběrová metoda modelu a kritérium výběru zadané v nastavení Výběr modelu. Také se zobrazí hodnota kritéria výběru pro konečný model a je prezentována v menším formátu, který má lepší formát.

**Graf.** Graf zobrazuje přesnost finálního modelu, který je prezentován ve větším formátu, je lepší formát. Hodnota je  $100 \times$  upravená  $R^2$  pro konečný model.

## Automatická příprava dat

Tento pohled zobrazuje informace o tom, která pole byla vyloučena a jak byla transformovaná pole odvozena v kroku automatizovaného zpracování dat (ADP). Pro každé pole, které bylo transformováno nebo vyloučeno, tabulka uvádí název pole, jeho roli v analýze a akce provedené krokem ADP. Pole jsou řazena vzestupně podle abecedního pořadí názvů polí. Možné akce provedené pro každé pole zahrnují:

- **Odvození trvání: měsíce** počítá uplynulou dobu v měsících od hodnot v poli obsahujícím data na aktuální systémové datum.
- **Odvození trvání: hodiny** počítá uplynulou dobu v hodinách z hodnot v poli obsahujících časy na aktuální systémový čas.
- **Změnit úroveň měření z průběžných na pořadové** přeloží souvislá pole s méně než 5 jedinečnými hodnotami jako ordinální pole.
- **Změnit úroveň měření z pořadového na trvalý** předřadí pole s více než 10 jedinečnými hodnotami jako souvislá pole.
- **Oříznutí odlehklých hodnot** nastavuje hodnoty spojitých predikátů, které leží mimo hodnotu uzavření (3 směrodatné odchylky od střední hodnoty) k hodnotě uzavření objektu.
- Volba **Nahradit chybějící hodnoty** nahradí chybějící hodnoty nominálních polí s režimem, pořadovým číslem s mediánem a spojitými poli se střední hodnotou.
- Volba **Sloučit kategorie pro maximalizaci přidružení s cílem** identifikuje "podobné" kategorie prediktorů založené na vztahu mezi vstupem a cílem. Kategorie, které nejsou výrazně odlišné (tj. mající  $p$ -hodnotu větší než 0,05), se sloučí.
- Volba **Vyloučit prediktor konstant/po slučování odlehklých hodnot/po sloučení kategorií** odstraní prediktory, které mají jedinou hodnotu, případně po provedení dalších akcí ADP.

## Důležitost produktu Predictor

Obvykle se budete chtít zaměřit na vaše modelovací úsilí na polích prediktorů, která se nejvíce zabývají, a zvažte zrušení nebo ignorování těch, které jsou nejméně. Graf důležitosti prediktorů vám pomáhá to tak, že indikujete relativní důležitost každého prediktoru při odhadování modelu. Vzhledem k tomu, že hodnoty jsou relativní, součet hodnot všech predikátů na obrazovce je 1.0. Důležitost předpovědí se netýkala přesnosti modelu. Vztahuje se pouze na důležitost každého prediktoru při vytváření predikce, nikoli zda je prognóza přesná.

## Predicted By Observed

Tím se zobrazí rozptýlení bodového grafu předpovídaných hodnot na vertikální ose podle pozorovaných hodnot na vodorovné ose. V ideálním případě by body měly ležet na přímce 45 °; tento pohled vám může sdělit, zda jsou některé záznamy předpovídané obzvláště špatně podle modelu.

## Residuals

Zobrazí se diagnostický graf zbytkových chyb modelu.

**Styly grafů.** Existují různé styly zobrazení, které jsou přístupné z rozevíracího seznamu **Styl**.

- **Histogram**. Tento je binned histogram studentizovaných zbytkových chyb s překryvem normální distribuce. Lineární modely předpokládají, že zbytkové chyby mají normální distribuci, takže histogram by měl ideálně úzce sblížit plynulý řádek.
- **P-P Plot**. Toto je binned pravděpodobnostní graf, který porovnává studentizované zbytkové chyby s normální distribucí. Je-li sklon zakreslených bodů méně strmé než normální linie, zbytkové chyby vykazují větší variabilitu než normální rozdělení; je-li sklon strmější, zbytkové chyby vykazují menší variabilitu než normální rozdělení. Mají-li zakreslené body křivku ve tvaru S, rozložení zbytkových chyb je posunutě.

## Odlehlé hodnoty

Tato tabulka obsahuje seznam záznamů, které mají na model nežádoucí vliv, a zobrazuje ID záznamu (je-li určeno na kartě Pole), cílovou hodnotu a vzdálenost Cook. Cook's distance is a measure of how much the residuals of all records would change if a particular record were excluded from the calculation of the model coefficients. Velká Cook vzdálenost označuje, že vyloučení záznamu od změn koeficientů podstatně, a měly by být považovány za vlivné.

Vlivné záznamy by měly být pečlivě zkoumány, aby bylo možné určit, zda můžete dát jim menší váhu při odhadu modelu, nebo zkrátit odlehlé hodnoty na přijatelnou prahovou hodnotu, nebo odstranit vlivné záznamy zcela.

## Efekty

Tento pohled zobrazuje velikost každého efektu v modelu.

**Styly.** Existují různé styly zobrazení, které jsou přístupné z rozevíracího seznamu **Styl**.

- **Diagram**. Toto je graf, ve kterém jsou účinky seřazeny od shora dolů sestupně podle důležitosti prediktorů. Spojovací čáry v diagramu jsou vážené na základě významnosti efektu a větší šířku čáry odpovídá výraznějším účinkům (menší hodnota *p*-values). Při přejetí přes připojovanou čáru se zobrazí popis, který ukazuje *p*-hodnotu a důležitost efektu. Jedná se o výchozí nastavení.
- **tabulka:** Jedná se o tabulku ANOVA pro celkový model a pro jednotlivé efekty modelu. Jednotlivé účinky jsou seřazeny od shora dolů tím, že snižují prediktor důležitost. Všimněte si, že při výchozím nastavení je tabulka sbalena tak, aby zobrazovala pouze výsledky pro celý model. Chcete-li zobrazit výsledky pro jednotlivé efekty modelu, klepněte na buňku **Opravený model** v tabulce.

**Důležitost prediktoru.** Posuvný ovladač důležitosti predikátů, který řídí, které prediktory se zobrazí v pohledu. Tím se model nezmění, ale jednoduše vám umožní soustředit se na nejdůležitější prediktory. Při výchozím nastavení se zobrazí 10 nejlepších efektů.

**Parametr významnosti.** Je zde posuvný ovladač významnosti, který dále kontroluje, které efekty jsou zobrazeny v pohledu, kromě těch, které jsou zobrazeny na základě důležitosti prediktoru. Efekty s hodnotou významnosti větší než hodnota posuvného ovladače jsou skryté. Tím se model nezmění, ale jednoduše vám umožní zaměřit se na nejdůležitější efekty. Standardně je hodnota 1.00, takže žádné účinky nejsou filtrovány na základě významnosti.

## Koeficienty

Tento pohled zobrazuje hodnotu každého koeficientu v modelu. Všimněte si, že faktory (kategoriální prediktory) jsou v modelu označeny indikátorem, takže **efekty** obsahující faktory obvykle mají více přidružených **koeficientů**; jedna pro každou kategorii kromě kategorie odpovídající redundantnímu (referenčnímu) parametru.

**Styly.** Existují různé styly zobrazení, které jsou přístupné z rozevíracího seznamu **Styl**.

- **Diagram.** Toto je graf, který zobrazuje nejprve zachycení a pak seřadí účinky shora dolů tím, že se sníží důležitost predikce. V rámci efektů obsahujících faktory jsou koeficienty řazeny vzestupně podle hodnot dat. Spojovací čáry v diagramu jsou obarveny na základě znaku koeficientu (viz klíč diagramu) a vážené na základě koeficientu významnosti, přičemž větší šířka řádku odpovídá významnějším koeficientům (menší hodnota  $p$ -values). Přetivání přes připojovanou čáru odhaluje popis, který zobrazuje hodnotu koeficientu, jeho  $p$ -hodnotu a důležitost efektu, se kterým je parametr přidružen. Toto je výchozí styl.
- **tabulka:** To ukazuje hodnoty, testy významnosti a intervaly spolehlivosti pro jednotlivé koeficienty modelu. Po zachycení jsou účinky seřazeny od shora dolů sestupně podle důležitosti prediktoru. V rámci efektů obsahujících faktory jsou koeficienty řazeny vzestupně podle hodnot dat. Všimněte si, že při výchozím nastavení je tabulka sbalena tak, aby ukazovala pouze koeficient, významnost a důležitost každého parametru modelu. Chcete-li zobrazit standardní chybu,  $t$  statistiku a interval spolehlivosti, klepněte na buňku **Spolupracující** v tabulce. Při přejetí na název parametru modelu v tabulce se zobrazí popis, který zobrazuje název parametru, efekt, který je přidružený k parametru, a (pro kategoriální prediktory), popisky hodnot přidružené k parametru modelu. To může být užitečné zejména při zobrazení nových kategorií vytvořených při automatické přípravě dat sloučí podobné kategorie kategorického prediktoru.

**Důležitost prediktoru.** Posuvný ovladač důležitosti predikátu, který řídí, které prediktory se zobrazí v pohledu. Tím se model nezmění, ale jednoduše vám umožní soustředit se na nejdůležitější prediktory. Při výchozím nastavení se zobrazí 10 nejlepších efektů.

**Parametr významnosti.** Je zde posuvný ovladač významnosti, který dále řídí, které koeficienty se v pohledu zobrazují, kromě těch, které jsou zobrazeny na základě důležitosti prediktoru. Koeficienty s hodnotami významnosti větší než hodnota posuvného ovladače jsou skryté. To nemění model, ale jednoduše vám umožní zaměřit se na nejdůležitější koeficienty. Standardně je hodnota 1.00, takže žádné koeficienty nejsou filtrovány na základě významnosti.

## Odhadované prostředky

Jedná se o grafy zobrazené pro významné prediktory. Graf zobrazuje model-odhadovanou hodnotu cíle na svislé ose pro každou hodnotu prediktoru na vodorovné ose a drží všechny ostatní prediktory prediktorů. Poskytuje užitečnou vizualizaci účinků každého prediktoru koeficientů na cíli.

*Poznámka:* Nejsou-li žádné prediktory významné, nevytvorí se žádné odhadované prostředky.

## Souhrn budovy modelu

Je-li vybrán jiný algoritmus výběru modelu než **Žádný** v nastavení výběru modelu, poskytuje se zde některé podrobnosti o procesu sestavení modelu.

**Předat krokové.** Je-li dopředným krokovým algoritmem algoritmus výběru, v tabulce se zobrazí posledních 10 kroků v rámci algoritmu krokové metody. U každého kroku se zobrazí hodnota kritéria výběru a efekty v modelu v daném kroku. To vám dává smysl, jak každý krok přispívá k modelu. Každý sloupec vám umožňuje seřadit řádky tak, abyste mohli snadněji vidět, které efekty jsou v modelu v daném kroku.

**Nejlepší podmnožiny.** Je-li nejlepší podmnožinou algoritmus výběru, zobrazí se v tabulce prvních 10 modelů. Pro každý model se zobrazí hodnota kritéria výběru a efekty v modelu. To vám dává smysl pro stabilitu špičkových modelů; pokud mají tendenci mít mnoho podobných efektů s několika rozdíly, pak můžete být docela jistý v "top" modelu; pokud mají tendenci mít velmi odlišné účinky, pak některé účinky mohou být příliš podobné a měly by být sloučeny (nebo jeden odstraněn). Každý sloupec vám umožňuje seřadit řádky tak, abyste mohli snadněji vidět, které efekty jsou v modelu v daném kroku.

## Lineární regrese

Lineární regrese odhaduje koeficienty lineární rovnice zahrnující jednu nebo více nezávislých proměnných, které nejlépe předpovídají hodnotu závislé proměnné. Můžete se například pokusit odhadnout celkové roční tržby prodejce (závislá proměnná) od nezávislých proměnných, jako je věk, vzdělání a let zkušeností.

**Příklad.** Je počet her vyhrál basketbalový tým v sezóně vztahující se k průměrnému počtu bodů na tým skóre za hru? Bodový graf označuje, že tyto proměnné jsou lineárně související. Počet her vyhrál a průměrný počet bodů gólů soupeřem jsou také lineárně spojené. Tyto proměnné mají záporný vztah. Jak se počet her vyhrál zvyšuje, průměrný počet bodů skóroval soupeřem snižuje. Pomocí lineární regrese můžete modelovat vztah těchto proměnných. Dobrý model může být použit pro předpověď, kolik her týmy vyhraje.

**Statistika.** Pro každou proměnnou: počet platných případů, střední hodnoty a směrodatné odchylky. Pro každý model: regresní koeficienty, korelační matici, část a částečné korelace, více  $P$ ,  $P^2$ , upravené  $P^2$ , změny v  $P^2$ , směrodatná chyba odhadu, tabulky odchylky, předpokládané hodnoty a zbytkové chyby. Také, 95%-confidenční intervaly pro každý regresní koeficient, variance-kovarianční matice, inflační faktor odchylky, tolerance, Durbin-Watson test, vzdálenost opatření (Mahalanobis, Cook, a pákové hodnoty), DfBeta, DfFit, predikční intervaly a diagnostické informace o casewise. Plot: bodový graf, parciální grafy, histogramy a normální grafy pravděpodobnosti.

Posouzení lineárních regresních dat

**Data.** Závislé a nezávislé proměnné by měly být kvantitativní. Kategorické proměnné, jako např. náboženství, významné oblasti studia nebo region trvalého pobytu, je třeba rekódovat do binárních (fiktivních) proměnných nebo jiných typů kontrastních proměnných.

**Předpoklady.** Pro každou hodnotu nezávislé proměnné musí být distribuce závislé proměnné normální. Rozdíl v distribuci závislé proměnné by měl být konstantní pro všechny hodnoty nezávislé proměnné. Vztah mezi závislou proměnnou a každou nezávislou proměnnou by měl být lineární a všechny pozorování by měly být nezávislé.

Jak získat lineární regresní analýzu

1. Z nabídky vyberte:

**Analyzovat > Regrese > Lineární ...**

2. V dialogovém okně Lineární regrese vyberte numerickou závislou proměnnou.
3. Vyberte jednu nebo více číselných nezávislých proměnných.

Volitelně můžete:

- Seskupte nezávislé proměnné do bloků a určete různé vstupní metody pro různé podmnožiny proměnných.
- Vyberte proměnnou výběru, chcete-li omezit analýzu na podmnožinu případů s konkrétní hodnotou této proměnné.
- Vyberte identifikační proměnnou případu pro identifikaci bodů na zkusných plochách.
- Vyberte numerickou proměnnou váhy WLS pro analýzu vážených nejmenších čtverců.

**WLS.** Umožňuje získat vážený model nejmenších čtverců. Datové body jsou váženy vzájemnou odchylností jejich rozptylů. To znamená, že pozorování s velkými rozptyly mají menší dopad na analýzu než pozorování spojená s malými odchylkami. Je-li hodnota váhové proměnné nula, záporná nebo chybějící, je případ vyloučen z analýzy.

## Metody výběru lineární regrese regresní proměnné

Výběr metody vám umožňuje uvést, jak se nezávislé proměnné zadávají do analýzy. Pomocí různých metod je možné vytvořit různé regresní modely ze stejné sady proměnných.

- *Enter (Regrese)*. Procedura pro výběr proměnné, v níž jsou všechny proměnné v bloku zadány v jednom kroku.
- *Kroková*. V každém kroku nezávislá proměnná není v rovnici, která má nejmenší pravděpodobnost F, pokud je pravděpodobnost, že je dostatečně malá. Proměnné již v regresní rovnici jsou odstraněny, pokud jejich pravděpodobnost F bude dostatečně velká. Tato metoda se ukončí, když nejsou způsobilé žádné další proměnné pro zahrnutí nebo odebrání.
- *Odebrat*. Procedura pro výběr proměnných, v níž jsou všechny proměnné v bloku odstraněny v jednom kroku.
- *Zpětná eliminace*. Výběrové řízení proměnné, ve kterém jsou všechny proměnné zadány do rovnice a pak postupně odebrány. Proměnná s nejmenší dílčí korelací se závislou proměnnou je považována za první za účelem odebrání. Pokud splňuje kritérium pro odstranění, je odstraněna. Po odebrání první proměnné je proměnná zbývající v rovnici s nejmenší dílčí souvztažnou hodnotou považována za další. Procedura se zastaví, když v rovnici, která splňuje kritéria odebrání, nejsou žádné proměnné.
- *Postoupit výběr*. Výběrové řízení proměnné, ve kterém jsou proměnné postupně zadávány do modelu. První proměnná uvažovaná pro vstup do rovnice je ta s největší pozitivní nebo negativní souvztažností se závislou proměnnou. Tato proměnná se zadává do rovnice pouze tehdy, splňuje-li kritérium pro vstup. Je-li zadána první proměnná, nezávislá proměnná není v rovnici, která má největší dílčí korelacii, je považována za další. Procedura se zastaví, když neexistují žádné proměnné, které splňují kritérium položky.

Hodnoty významnosti ve vašem výstupu jsou založeny na montáži jednoho modelu. Proto jsou hodnoty významnosti obecně neplatné, když se použije kroková metoda (kroková, dopředná nebo zpětná).

Všechny proměnné musí projít kritériem tolerance, které mají být zadány do rovnice, bez ohledu na určenou vstupní metodu. Výchozí úroveň tolerance je 0,0001. Také proměnná není zadána, pokud by způsobila, že tolerance jiné proměnné již v modelu klesne pod kritérium tolerance.

Všechny vybrané nezávislé proměnné se přidávají do jednoho regresního modelu. Můžete však určit různé vstupní metody pro různé podmnožiny proměnných. Např. můžete zadat jeden blok proměnných do regresního modelu pomocí krokového výběru a druhého bloku pomocí volby předání. Chcete-li přidat druhý blok proměnných do regresního modelu, klepněte na tlačítko **Další**.

## Pravidlo lineární regresní sady

Případy definované pravidlem výběru jsou zahrnuty do analýzy. Pokud například vyberete proměnnou, vyberte volbu **rovná se** pro hodnotu zadejte hodnotu 5, budou do analýzy zahrnuty pouze případy, pro které má zvolená proměnná hodnotu 5. Hodnota řetězce je také povolena.

## Lineární regresní paty

Plot může pomoci při validaci předpokladů normality, linearity a rovnosti odchylek. Pónky jsou také užitečné pro zjišťování outliers, neobvyklé pozorování a vlivné případy. Po jejich uložení jako nové proměnné, předpokládané hodnoty, zbytkové chyby a další diagnostické informace jsou k dispozici v editoru dat pro sestavení grafů s nezávislými proměnnými. K dispozici jsou následující grafy:

**Bodové grafy**. Můžete zakreslit kteroukoli dvě z následujících: závislá proměnná, standardizované predikované hodnoty, standardizované zbytkové chyby, odstraněné zbytkové chyby, upravené předpokládané hodnoty, Studentiované zbytkové chyby nebo Studentiované odstraněné zbytkové chyby. Vykreslete standardizované zbytkové chyby vůči standardizovaným předpovězeným hodnotám, abyste zkontrolovaly linearitu a rovnost odchylek.

*Seznam zdrojových proměnných*. Uvádí závislou proměnnou (DEPENDNT) a následující předpokládané a zbytkové proměnné: Standardizované predikované hodnoty (\*ZPRED), Standardizované zbytkové chyby

(\*ZRESID), Odstraněné zbytkové chyby (\*DRESID), Upravené předpokládané hodnoty (\*ADJPRED), Stentizované zbytkové chyby (\*SRESID), Studentizované odstraněné zbytkové chyby (\*SDRESID).

**Vytvořte všechny dílčí grafy.** Zobrazí bodový graf zbytkových chyb každé nezávislé proměnné a zbytkových chyb závislé proměnné, jsou-li obě proměnné znovu nastaveny odděleně od ostatních nezávislých proměnných. Alespoň dvě nezávislé proměnné musí být v rovnici pro částečný výkres, který má být vyroben.

**Standardizováno reziduální Plots.** Histogram standardizovaných zbytkových chyb a běžných pravděpodobnostních grafů lze získat porovnáním distribuce standardizovaných zbytkových chyb na normální distribuci.

Jsou-li požadovány nějaké grafy, souhrnné statistiky se zobrazí pro standardizované předpokládané hodnoty a standardizované zbytkové chyby (\*ZPRED a \*ZRESID).

## Lineární regrese: Uložení nových proměnných

Můžete uložit předpokládané hodnoty, zbytkové chyby a další statistiky užitečné pro diagnostické informace. Každý výběr přidá jednu nebo více nových proměnných do vašeho aktivního datového souboru.

**Předpovězené hodnoty.** Hodnoty, které regresní model předpovídá pro každý případ.

- *Nestandardizováno.* Hodnota, kterou model předpovídá pro závislou proměnnou.
- *Standardizováno.* Transformace každé predikované hodnoty do své standardizované podoby. To znamená, že střední předpovězená hodnota se odečte od předpokládané hodnoty a rozdíl se vydělí směrodatnou odchylkou předpokládaných hodnot. Standardizované predikované hodnoty mají střední hodnotu 0 a směrodatnou odchylku 1.
- *Upraveno.* Předpovídána hodnota případu, je-li tento případ vyloučen z výpočtu regresních koeficientů.
- *S.E. of mean predictions.* Standardní chyby předpokládaných hodnot. Odhad směrodatné odchylky průměrné hodnoty závislé proměnné pro případy, které mají stejné hodnoty u nezávislých proměnných.

**Vzdálenosti.** Opatření k identifikaci případů s neobvyklými kombinacemi hodnot pro nezávislé proměnné a případy, které mohou mít velký dopad na regresní model.

- *Mahalanobis.* Míra, v jaké míře se hodnoty jednotlivých případů na nezávislých proměnných liší od průměru všech případů. Velká Mahalanobis vzdálenost identifikuje případ, kdy mají extrémní hodnoty na jedné nebo více nezávislých proměnných.
- *Cook's.* Měřítka toho, kolik zbytkových chyb ze všech případů se změní, pokud byl konkrétní případ vyloučen z výpočtu regresních koeficientů. Velké Cook D označuje, že vyloučení případu z výpočtu regresní statistiky mění koeficienty podstatně.
- *Efektivní využití hodnot.* Měří vliv bodu na proložení regrese. Vystředěný pákový efekt se pohybuje od 0 (žádný vliv na přizpůsobení) na  $(N-1)/N$ .

**Intervaly předpovědi.** Horní a dolní meze pro střední a individuální predikční intervaly.

- *Střední.* Dolní a horní meze (dvě proměnné) pro interval předpovědi průměrné předpovězené odezvy.
- *Individuální.* Dolní a horní meze (dvě proměnné) pro interval předpovědi závislé proměnné pro jeden případ.
- *Interval spolehlivosti.* Zadejte hodnotu mezi 1 a 99.99, chcete-li určit úroveň důvěry pro dva intervaly předpovědi. Před zadáním této hodnoty musí být vybrána hodnota Střední nebo individuální. Typické hodnoty intervalu spolehlivosti jsou 90, 95 a 99.

**Residuals.** Skutečná hodnota závislé proměnné minus hodnota předpovězená regresní rovnicí.

- *Nestandardizováno.* Rozdíl mezi pozorovanou hodnotou a hodnotou předpovězenou modelem.
- *Standardizováno.* Zbytek vydělený odhadem jeho směrodatné odchylky. Standardizované zbytkové chyby, které jsou také známé jako Pearsonové zbytkové chyby, mají střední hodnotu 0 a směrodatnou odchylku 1.

- *Studentizovaný*. Zbytek vydělený odhadem jeho směrodatné odchylky, která se liší případ od případu, v závislosti na vzdálenosti hodnot jednotlivých případů na nezávislých proměnných od středních hodnot nezávislých proměnných. Někdy se odkazuje jako na interně stomatozované zbytkové chyby.
- *Odstraněno*. Zbytková hodnota pro případ, kdy je tento případ vyloučen z výpočtu regresních koeficientů. Je to rozdíl mezi hodnotou závislé proměnné a upravenou předpokládanou hodnotou.
- *Studentizovaný odstraněný*. Odstraněná zbytková velikost pro případ vydělená její standardní chybou. Rozdíl mezi zjištěnými odstraněnými rezidui a jeho přidruženými reziduálními rezidui označuje, jak velký rozdíl odstranění případu dělá z jeho vlastní předpovědi. Někdy se odkazuje jako na externí studentizované zbytkové chyby.

**Ovlivňující statistiky.** Změna regresních koeficientů ( $DfBeta$  [s]) a předpokládaných hodnot ( $DfFit$ ), které jsou výsledkem vyloučení konkrétního případu. Standardizované hodnoty  $DfBetas$  a  $DfFit$  jsou k dispozici také spolu s poměrem kovariance.

- *DfBetas*. Rozdíl v hodnotě beta je změna regresního koeficientu, která vyplývá z vyloučení konkrétního případu. Hodnota se vypočítává pro každý termín v modelu, včetně konstanty.
- *Standardizováno DfBeta*. Standardizovaný rozdíl v beta hodnotě. Změna v regresním koeficientu, která je výsledkem vyloučení konkrétního případu. Můžete chtít zkontrolovat případy s absolutními hodnotami vyššími než 2 děleno druhou odmocninou z  $N$ , kde  $N$  je počet případů. Hodnota se vypočítává pro každý termín v modelu, včetně konstanty.
- *DfFit*. Rozdíl v hodnotě přizpůsobení je změna v předpovězenou hodnotu, která vyplývá z vyloučení konkrétního případu.
- *Standardizováno DfFit*. Standardizovaný rozdíl v hodnotě přizpůsobení. Změna v předpovězenou hodnotu, která vyplývá z vyloučení konkrétního případu. Můžete chtít zkontrolovat standardizované hodnoty, které v absolutní hodnotě překračují 2násobek druhé odmocniny  $p/N$ , kde  $p$  je počet parametrů v modelu a  $N$  je počet případů.
- *Poměr rozptylu*. Poměr determinantu matice kovariance s konkrétním případem vylučujícím z výpočtu regresních koeficientů na determinantu kovarianční matice se všemi zahrnutí případy. Je-li poměr blízký k 1, případ významně nezmění kovarianční matici.

**Coefficient Statistics.** Uloží regresní koeficienty do datové sady nebo do datového souboru. Datové sady jsou k dispozici pro další použití ve stejné relaci, ale nejsou uloženy jako soubory, dokud nejsou explicitně uloženy před koncem relace. Názvy datových sad musí být v souladu s pravidly pojmenování proměnných.

**Exportovat informace o modelu do souboru XML.** Odhady parametrů a (volitelně) jejich kovariance se exportují do zadaného souboru ve formátu XML (PMML). Tento modelový soubor můžete použít k použití informací modelu na jiné datové soubory pro účely hodnocení.

## Lineární regresní statistika

K dispozici jsou následující statistiky:

**Regresní koeficienty- Odhady** zobrazují regresní koeficient  $B$ , standardní chybu  $B$ , standardizovaný koeficient beta,  $t$  hodnotu pro  $B$  a dvoustrannou úroveň významnosti  $t$ . **Intervaly spolehlivosti** zobrazují intervaly spolehlivosti se zadanou úrovní spolehlivosti pro každý regresní koeficient nebo kovarianční matici. **Covarianční matice** zobrazuje varianci-kovarianční matice regresních koeficientů s kovariancemi mimo diagonální a rozptyly na diagonále. Zobrazí se také korelační matice.

**Přizpůsobit model-** Jsou vypsány proměnné, které jsou zadány a odebrány z modelu, a zobrazí se následující 'statistika dobrých výsledků': více  $R$ ,  $R^2$  a upravené  $R^2$ , směrodatná chyba odhadu a tabulka analýzy rozptylu.

**R na druhou změnu-** Změna ve statistice  $R^2$ , která je vytvořena přidáním nebo odstraněním nezávislé proměnné. Je-li změna  $R^2$  přidružená k proměnné velká, znamená to, že proměnná je dobrým prediktorem závislé proměnné.

**Popisné-** Poskytuje počet platných případů, střední hodnota a směrodatné odchylky každé proměnné v analýze. Zobrazí se také korelační matice s jednou sledovanou úrovní významnosti a počtem případů pro každou korelaci.

**Korelace dílu.** Korelace mezi závislou proměnnou a nezávislou proměnnou, když byly z nezávislé proměnné odebrány lineární účinky ostatních nezávislých proměnných v modelu. Souvisí se změnou v  $R$  na druhou, je-li proměnná přidána do rovnice. Někdy se nazývá semičástečná korelace.

**Dílčí korelace.** Korelace, která zůstane mezi dvěma proměnnými po odebrání korelace, která je dána vzájemným přidružením k ostatním proměnným. Korelace mezi závislou proměnnou a nezávislou proměnnou, když byly z obou odebrány lineární efekty ostatních nezávislých proměnných v modelu.

**Diagnostika kolineárnosti-** Kolineárnost (nebo vícekolineárnost) je nežádoucí situace, kdy jedna nezávislá proměnná je lineární funkcí jiných nezávislých proměnných. Eigenvalues of the scaled and uncentered cross-products matrix, condition indices, and variance-dekompozice proportions are displayed along with variance inflation factors (VIF) and tolerances for individual variables.

**Kritéria výběru-** Zahrnuje informační kritérium Akaike (AIC), kritérium předpovědi Ameniya (PC), umožňuje podmíněné střední kvadratická chyba kritéria předpovědi ( $C_p$ ) a kritéria Schwarz Bayesovo (SBC). Statistika se zobrazí v tabulce Souhrn modelu.

**Residuals-** Můžete vybrat '**STISK-Statistic**', která má být použita jako statistika křížového ověření k porovnání různých modelů. Zobrazí se také test '**Durban-Watson**' pro sériovou korelaci zbytkových chyb. Vyberte informace '**Diagnostické informace o případu**' pro případy, které splňují kritérium výběru (odlehle hodnoty nad  $n$  směrodatnými odchylkami).

## Volby lineární regrese

K dispozici jsou následující volby:

**Kritéria krokové metody.** Tyto volby se používají, když je uvedena metoda výběru proměnné dopředu, dozadu nebo postupná proměnná. Proměnné lze zadat nebo odebrat z modelu závisí na významnosti (pravděpodobnosti) hodnoty  $F$  nebo na samotné hodnotě  $F$ .

- **Pravděpodobnost použití  $F$ .** Proměnná je zadána do modelu, pokud je úroveň významnosti její hodnoty  $F$  menší než hodnota Entry a je odebrána, je-li úroveň významnosti větší než hodnota Odebrání. Položka musí být menší než odebrání, přičemž obě hodnoty musí být kladné. Chcete-li do modelu zadat více proměnných, zvyšte hodnotu Zadání. Chcete-li odebrat více proměnných z modelu, snižte hodnotu Odebrání.
- **Použití hodnotu  $F$ .** Proměnná je zadána do modelu, je-li její hodnota  $F$  větší než hodnota Entry a je odebrána, pokud je hodnota  $F$  menší než hodnota odebrání. Položka musí být větší než odebrání, a obě hodnoty musí být kladné. Chcete-li do modelu zadat více proměnných, snižte hodnotu položky Entry. Chcete-li odebrat více proměnných z modelu, zvyšte hodnotu Odebrání.

**Tolerance.** Při výchozím nastavení je hodnota 0,0001. Tolerance je podíl rozptylu proměnné v rovnici, která není zahrnuta do jiných nezávislých proměnných v rovnici. Minimální tolerance pro každou proměnnou v rovnici, je-li tato proměnná zahrnuta do analýzy, je minimální tolerance proměnné, která není zahrnuta do rovnice. Proměnné musí projít testy tolerance a minimální tolerance, aby mohly vstoupit a zůstat v regresní rovnici. Pokud proměnná splňuje kritéria tolerance, je způsobilá pro zařazení na základě metody, která je v platnosti.

**Zahrnout konstantu do rovnice.** Ve výchozím nastavení obsahuje regresní model konstantní termín. Zrušení výběru této volby vynutí regresi prostřednictvím původu, což je zřídka udělána. Některé výsledky regrese původem nejsou srovnatelné s výsledky regrese, které obsahují konstantu. Např.  $R^2$  nelze interpretovat obvyklým způsobem.

**Chybějící hodnoty.** Můžete zvolit jednu z následujících možností:

- **Vyloučit případy jako litwise.** Do analýz jsou zahrnuty pouze případy s platnými hodnotami pro všechny proměnné.
- **Vyloučit případy po dvojicích.** Případy s úplnými údaji pro dvojici proměnných, které jsou souvztažné, se používají k výpočtu koeficientu korelace, na kterém je založena regresní analýza. Stupňů volnosti jsou založeny na minimálním párování  $N$ .
- **Nahradit střední hodnotou.** Všechny případy se používají pro výpočty, se střední hodnotou proměnné nahrazené za chybějící pozorování.



## Další funkce příkazu REGRESSION

Jazyk syntaxe příkazu vám také umožňuje:

- Zapište korelační matici nebo si přečtete matici v místě nezpracovaných dat pro získání regresní analýzy (s dílčím příkazem MATRIX).
- Uvedte úroveň tolerance (s dílčím příkazem CRITERIA).
- Získejte více modelů pro stejné nebo různé závislé proměnné (s dílčími příkazy METHOD a DEPENDENT).
- Získejte další statistiky (s dílčími příkazy DESCRIPTIVES a STATISTICS).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Ordinální regrese

Ordinální regrese vám umožňuje modelovat závislost polytomové ordinální odpovědi na souboru prediktorů, které mohou být faktory nebo kovariáty. Návrh Ordinální regrese je založen na metodice McCullagh (1980, 1998) a postup je označován jako PLUM v syntaxi.

Standardní lineární regresní analýza zahrnuje minimalizaci rozdílů mezi odezvou (závislým) a váženou kombinací prediktorů (nezávislých) proměnných odezvy. Odhadnuté koeficienty odrážejí, jak změny v prediktorů ovlivňují odezvu. Předpokládá se, že odezva je číselná, v tom smyslu, že změny úrovně odezvy jsou ekvivalentní v celém rozsahu odezvy. Například rozdíl ve výšce mezi osobou, která je 150 cm vysoká, a osobou, která je 140 cm vysoká, je 10 cm, což má stejný význam jako rozdíl ve výšce mezi osobou, která je 210 cm vysoká a osobou, která je 200 cm vysoká. Tyto vztahy nemusí nutně vést pro ordinální proměnné, ve kterých může být volba a počet kategorií odpovědí zcela libovolné.

**Příklad.** Ordinální regrese by mohla být použita ke studiu reakce pacienta na dávkování léku. Možné reakce mohou být klasifikovány jako *žádné*, *mírné*, *střední* nebo *závažné*. Rozdíl mezi mírnou a středně závažnou reakcí je obtížné nebo nemožné kvantifikovat a je založen na vnímání. Navíc rozdíl mezi mírnou a střední odpovědí může být větší nebo menší než rozdíl mezi střední a silnou odezvou.

**Statistika a grafy.** Observed and expected frequencies and cumulative frequencies, Pearson residuals for frequencies and cumulative frequencies, observed and expected probabilities, observed and expected cumulative probabilities of each response category by covariate pattern, asymptotic correlation and covariance matrices of parameter estimates, Pearson's chi-square and likelihood-ratio chi-square, goodness-of-fit statistics, iteration history, test of parallel lines assumption, parameter estimates, standard errors, confidence intervals, and Cox and Snell's, Nagelkerke's, and McFadden's  $P^2$  statistics.

Pokyny pro ordinální regresní data

**Data.** Předpokládá se, že závislá proměnná má pořadové číslo a může být číselná nebo řetězcová. Řazení se určuje tak, že se seřadí hodnoty závislé proměnné ve vzestupném pořadí. Nejnižší hodnota definuje první kategorii. Proměnné faktoru se považují za kategorické. Proměnné Covariate musí být číselné. Všimněte si, že použití více než jedné souvislé kovariance může snadno vést k vytvoření velmi velké tabulky pravděpodobností buněk.

**Předpoklady.** Je povolena pouze jedna proměnná odezvy, a musí být zadána. Pro každý odlišný vzorec hodnot v rámci nezávislých proměnných se předpokládá, že odpovědi jsou nezávislé mnohohromické proměnné.

**Související postupy.** Nominální logistická regrese používá podobné modely pro nominální závislé proměnné.

Získání ordinální regrese

1. Z nabídky vyberte:

**Analyzovat > Regrese > Pořadový ...**

2. Vyberte jednu závislou proměnnou.
3. Klepněte na tlačítko **OK**.

## Ordinální regresní volby

Dialogové okno Volby umožňuje upravit parametry použité v algoritmu iterativního odhadu, zvolit úroveň důvěry pro odhady parametrů a vybrat funkci odkazu.

**Iterace.** Iterativní algoritmus si můžete upravit.

- **Maximum iterací.** Uvedte nezáporné celé číslo. Je-li zadána hodnota 0, procedura vrátí počáteční odhad.
- **Maximální krok-halving.** Uvedte kladné celé číslo.
- **Sbližování pravděpodobnosti protokolu.** Algoritmus se zastaví, pokud je absolutní nebo relativní změna v protokolu-pravděpodobnost menší než tato hodnota. Kritérium se nepoužije, je-li zadána hodnota 0.
- **Sbližování parametrů.** Algoritmus se zastaví, pokud absolutní nebo relativní změna v každém odhadu parametru je menší než tato hodnota. Kritérium se nepoužije, je-li zadána hodnota 0.

**Interval spolehlivosti.** Uvedte hodnotu větší než nebo rovnou 0 a menší než 100.

**Rozdílová data.** Hodnota se přidala k nulové frekvenci buněk. Uvedte nezápornou hodnotu menší než 1.

**Tolerance tolerance.** Používá se pro kontrolu prediktorů s vysokou závislostí. Vyberte hodnotu ze seznamu voleb.

**Funkce propojení.** Spojovací funkce je transformace kumulativních pravděpodobností, která umožňuje odhad modelu. K dispozici je následující pět funkcí propojení.

- **Logit.**  $f(x) = \log(x/(1-x))$ . Obvykle se používá pro rovnoměrně rozdělené kategorie.
- **Complementary log-log.**  $f(x) = \log(-\log(1-x))$ . Obvykle se používá, když jsou vyšší kategorie pravděpodobnější.
- **Negativní protokol protokolu.**  $f(x) = -\log(-\log(x))$ . Obvykle se používá, když jsou nižší kategorie pravděpodobnější.
- **Probit.**  $f(x) = \Phi^{-1}(x)$ . Obvykle se používá tehdy, když je latentní proměnná normálně distribuována.
- **Cauchit (inverzní Cauchyova).**  $f(x) = \tan(\pi(x-0.5))$ . Obvykle se používá tehdy, má-li latentní proměnná mnoho extrémních hodnot.

## Ordinální regresní výstup

Dialogové okno Výstup vám umožňuje vytvořit tabulky pro zobrazení v prohlížeči a uložit proměnné do pracovního souboru.

**Obrazovka.** Vytvoří tabulky pro:

- **Tisknout historii iterace pro každý n (é) krok (y).** Hodnoty protokolu pravděpodobnosti a odhadu parametrů se vytisknou pro uvedenou frekvenci iterace tisku. První a poslední iterace se vždy vytisknou.
- **Nevhodná statistika.** Pearson a pravděpodobnosti-poměr chi-square statistiky. Vypočítávají se na základě klasifikace uvedené v seznamu proměnných.
- **Souhrnné statistiky.** Cox a Snell's, Nagelkerke's, a McFadden je  $R^2$  statistiky.
- **Odhady parametrů.** Odhady parametrů, standardní chyby a intervaly spolehlivosti.
- **Asymptotická korelace odhadů parametrů.** Matice odhadu korelací parametrů.
- **Asymptotické kovariance odhadů parametrů.** Matice parametrů odhadu kovariances.
- **Informace o buňce.** Pozorované a očekávané frekvence a kumulativní frekvence, Pearsonové zbytkové chyby pro frekvence a kumulativní frekvence, pozorované a očekávané pravděpodobnosti a pozorované a očekávané kumulativní pravděpodobnosti každé kategorie odpovědí podle vzorce kovariate. Všimněte si, že u modelů s mnoha vzory kovariate (například modely s kontinuálním kovariaty) může tato volba generovat velmi rozsáhlou a nevládnitelnou tabulku.
- **Test paralelních čar.** Test hypotézy, že parametry umístění jsou ekvivalentní k úrovním závislé proměnné. Tato volba je k dispozici pouze pro model pouze pro umístění.

**Uložené proměnné.** Uloží následující proměnné do pracovního souboru:

- **Odhadnuté pravděpodobnosti odezvy.** Model-odhad pravděpodobnosti klasifikace faktoru/kovarianát do kategorií odpovědí. Existuje mnoho pravděpodobností jako počet kategorií odezvy.
- **Kategorie Predicted.** Kategorie odezvy, která má maximální odhadovanou pravděpodobnost pro vzor součinitel/kovariát.
- **Předpokládaná pravděpodobnost kategorie.** Odhadnutá pravděpodobnost zařazení faktoru/kovarianát do předpokládané kategorie do predikované kategorie. Tato pravděpodobnost je také maximální z odhadovaných pravděpodobností faktoru/kovarianát vzor.
- **Pravděpodobnost skutečné kategorie.** Odhadnutá pravděpodobnost zařazení faktoru/kovariativního vzorku do skutečné kategorie.

**Tisknout protokol-Pravděpodobnost.** Řídí zobrazení pravděpodobnosti protokolu. **Včetně mnohomické konstanty** vám dává plnou hodnotu pravděpodobnosti. Chcete-li porovnat své výsledky napříč produkty, které neobsahují konstantu, můžete ji vyloučit.

## Model ordinální regrese

Dialogové okno Umístění vám umožňuje uvést model umístění pro svou analýzu.

**Určete model.** Model s hlavními účinky obsahuje kovariantní a hlavní účinky, ale bez interakčních efektů. Můžete vytvořit vlastní model pro uvedení dílčích sad interakcí faktoru nebo kovarianát interakcí.

**Činitele/kovariany.** Jsou vypsány faktory a kovariany.

**Model umístění.** Model závisí na hlavních účincích a účincích interakce, které vyberete.

Pro vybrané faktory a kovariáty:

### Interakce

Vytvoří funkční období interakce nejvyšší úrovně všech vybraných proměnných. Jedná se o výchozí nastavení.

### Hlavní účinky

Vytvoří výraz main-effects pro každou vybranou proměnnou.

### Všechny dvoucestné

Vytvoří všechny možné dvousměrné interakce mezi vybranými proměnnými.

### Všechny 3-cesty

Vytvoří všechny možné trojcestné interakce mezi vybranými proměnnými.

### Všechny čtyřcestné

Vytvoří všechny možné čtyřcestné interakce mezi vybranými proměnnými.

### Všechny 5-way

Vytvoří všechny možné interakce pěti cest u vybraných proměnných.

## Sestavit podmínky a vlastní podmínky

### Podmínky sestavení

Tuto volbu použijte, chcete-li zahrnout nevnořené podmínky určitého typu (například hlavní efekty) pro všechny kombinace vybrané sady faktorů a kovariany.

### Sestavit vlastní výrazy

Tuto volbu použijte, chcete-li zahrnout vnořené výrazy nebo pokud chcete explicitně sestavit libovolnou proměnnou term pomocí proměnné. Sestavení vnořeného termínu zahrnuje následující kroky:

## Model Ordinální regresní stupnice

Dialogové okno Měřítka vám umožňuje uvést model stupnice pro vaši analýzu.

**Činitele/kovariany.** Jsou vypsány faktory a kovariany.

**Roztáhnout model.** Model závisí na hlavních a interakcích, které vyberete.

Pro vybrané faktory a kovariáty:

#### **Interakce**

Vytvoří funkční období interakce nejvyšší úrovně všech vybraných proměnných. Jedná se o výchozí nastavení.

#### **Hlavní účinky**

Vytvoří výraz main-effects pro každou vybranou proměnnou.

#### **Všechny dvoucestné**

Vytvoří všechny možné dvousměrné interakce mezi vybranými proměnnými.

#### **Všechny 3-cesty**

Vytvoří všechny možné trojcestné interakce mezi vybranými proměnnými.

#### **Všechny čtyřcestné**

Vytvoří všechny možné čtyřcestné interakce mezi vybranými proměnnými.

#### **Všechny 5-way**

Vytvoří všechny možné interakce pěti cest u vybraných proměnných.

## **Sestavit podmínky a vlastní podmínky**

#### **Podmínky sestavení**

Tuto volbu použijte, chcete-li zahrnout nevnořené podmínky určitého typu (například hlavní efekty) pro všechny kombinace vybrané sady faktorů a kovariány.

#### **Sestavit vlastní výrazy**

Tuto volbu použijte, chcete-li zahrnout vnořené výrazy nebo pokud chcete explicitně sestavit libovolnou proměnnou term pomocí proměnné. Sestavení vnořeného termínu zahrnuje následující kroky:

## **Další funkce příkazu PLUM**

Ordinální regresi můžete přizpůsobit, pokud vkládáte své výběry do okna syntaxe a upravíte výslednou syntaxi příkazu PLUM . Jazyk syntaxe příkazu vám také umožňuje:

- Vytvářejte přizpůsobené testy hypotézy zadáním hypotéz s hodnotou null jako lineární kombinace parametrů.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## **Lineární regrese elastické sítě**

Lineární Elastická síť používá třídu Python sklearn.linear\_model.ElasticNet k odhadu nalativých lineárních regresních modelů pro závislou proměnnou na jedné nebo více nezávislých proměnných. Regularizace kombinuje pokuty L1 (Lasso) a L2 (Ridge). Přípona zahrnuje volitelné režimy pro zobrazení trasovacích grafů pro různé hodnoty alfa pro daný poměr L1 a k výběru hodnoty L1 a hodnot alfa hyperparametru na základě vzájemné validace. Je-li pro výběr poměru k pokutě a/nebo alfa použit jeden model nebo je použita křížová validace, lze k odhadu výkonu výstupu z výběrového souboru použít rozdělení dat z odsunutí na vzorku.

Kromě montáže modelu s uvedenými hodnotami poměru L1 pokutového a regularizačního parametru alfa, lineární elastická síť může zobrazit stopové zkusné hodnoty koeficientu hodnot pro rozsah hodnot alfa pro daný poměr, nebo usnadnit výběr hodnoty hyperparations přes k-násobné mezivalidaci u určených mřížek hodnot. Je-li instalován jednotlivý model nebo je-li proveden poměr a/nebo výběr alfa prostřednictvím křížové validace, lze použít konečný model pro zadrhovací data, která jsou vytvořena oblastí vstupních dat za účelem získání platného odhadu výkonu out-of-of-ukázky modelu.

## **Získání lineární regrese regresní regrese**

1. Z nabídky vyberte:

## Analyzovat > Regrese > Lineární prvky OLS Alternativy > Elastická síť

Dialogové okno vám umožňuje uvést proměnnou, která přiřadí každý případ v aktivní datové sadě ke cvičicím nebo holdout ukázce.

2. Vyberte číselnou cílovou proměnnou. Pro spuštění analýzy je vyžadována pouze jedna cílová proměnná.
3. Uveďte numerickou závislost.
4. Určete alespoň jednu proměnnou kategorického faktoru nebo číselnou proměnnou kovariate.

Volitelně **Oblast** poskytuje způsob vytvoření díze nebo zkušební podmnožiny vstupních dat pro odhad výkonu zadaného nebo vybraného modelu. U všech případů s neplatnými daty pro každou proměnnou použitou procedurou se provádí dělení na oblasti po odstranění všech případů. Všimněte si, že pro křížové ověření se vytvoří záhyby nebo oblasti dat o školení v Python. Data o holdoutu vytvářená oblastí nejsou použita v rámci odhadu, bez ohledu na režim, který je v platnosti.

Oblast lze definovat zadáním poměru případů, které jsou náhodně přiřazovány jednotlivým vzorům (v části **Školení a zadržení oblastí**), nebo proměnnou, která každému případu přiřazuje školení nebo vzorek z výplně. Nemůžete uvést jak školení, tak proměnné. Není-li oblast zadána, vytvoří se vzorek pro pouzdro, který se vytvoří přibližně z 30% vstupních dat.

Hodnota **Školení%** uvádí relativní počet případů v aktivní datové sadě náhodně přiřazeného vzorku školení. Výchozí odborná příprava je 70%.

## Lineární regrese elastické sítě: Volby

Karta Volby poskytuje volby pro:

### Režim

Tento výběr poskytuje volby pro uvedení jednoho z následujících režimů:

#### Přizpůsobit s uvedeným poměrem L1 a alfa

Vyberete-li tento režim, bude do tréninkových dat instalován jediný model pomocí uvedených hodnot L1 a hodnot regularizace alfa. Jedná se o výchozí nastavení. Je-li zadána logická oblast, použije se jediný nebo poslední model, který je instalován, na zadrhovací testovací data k odhadu výkonu na vzorku.

V části **Zákresový graf** lze vybrat vykreslení pozorovaných a/nebo zbytkových chyb oproti předpovězeným hodnotám.

Pod volbou **Uložit** můžete určit předpokládané hodnoty a zbytkové chyby, které chcete uložit.

#### Trasovací graf

Vyberete-li tento režim, zobrazí se tři grafy pro data o školení jako funkce alfa pro určenou sadu hodnot alfa:

- Trasovací graf regresních koeficientů.
- Zákresový graf  $R^2$ .
- Vykreslení střední kvadratická chyba (MSE).

Ačkoli je oblast počtena, nejsou poskytnuty žádné výsledky pro zadrhovací testovací data, protože z tohoto režimu nejsou výsledkem žádné konečné výsledky modelu.

#### Poměr L1 a/nebo výběr alfa prostřednictvím křížového ověření

Když vyberete tento režim, provede se vyhledávání v mřížce s křížovým ověřením pro vyhodnocení modelů a nejlepší poměr a hodnoty alfa jsou zvoleny na základě nejlepšího průměru  $R^2$  přes záhyby validace. Pole **Počet křížových ověření platnosti** lze použít ke změně výchozí hodnoty pěti dělení nebo záhybů pro vzájemné ověření platnosti. Je-li zadána logická oblast, použije se jediný nebo poslední model, který je instalován, na zadrhovací testovací data k odhadu výkonu na vzorku.

V části **Zobraziti** můžete vybrat zobrazení základních informací o modelu se zvolenou hodnotou poměru a alfa (**Nejlepší**), základní informace o všech porovnávaných modelech (**Compare models**)

nebo úplné informace o všech rozděleních nebo záhybech pro všechny modely (**Porovnat modely a záhyby**). **Nejlepší** je výchozí nastavení.

V části **Zákresový graf** lze vybrat grafy střední hodnoty  $R^2$  a/nebo MSE nad validací. Lze vybrat také hodně pozorovaných a/nebo zbytkových chyb ve srovnání s předpovězenou hodnotou.

Pod volbou **Uložit** můžete určit předpokládané hodnoty a zbytkové chyby, které chcete uložit.

### Určete jednotlivé poměry L1

Když vyberete tento režim pro režim **Přizpůsobit s uvedeným poměrem L1 a alfa** nebo **Trasovací graf**, můžete zadat jedinou hodnotu poměru L1. Je-li vybrána volba **L1 a/nebo výběr alfa v režimu křížového ověření**, můžete zadat více hodnot.

#### Uvést mřížku hodnot alfa

When you select this mode for the **L1 poměr a/nebo alfa výběr přes křížové ověření** mode, a grid of unique alpha values can be specified from a **Začátek** value (value1) to an **Ukončit** value (value2) with the increment of **Podle** (value3). If specified, only one valid set of [value1 TO value2 BY value3] is allowed. Musí se ujistit, že  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . V případech, kdy  $\text{value1} = \text{value2}$ , je ekvivalentní zadání jednoho value1, bez ohledu na value3.

Vykreslí jsou zobrazeny pomocí zadané metriky pro vodorovné osy X hodnot alfa.

### Určete jednotlivé alphas

Vyberete-li tento režim pro režim **Přizpůsobit s uvedeným poměrem L1 a alfa**, můžete zadat jednu hodnotu regularizace alfa. Pokud jste vybrali volbu **Zákresový graf** nebo **L1 a/nebo výběr alfa v režimu křížového ověření**, můžete zadat více hodnot.

#### Uvést mřížku hodnot alfa

Vyberete-li tento režim pro režim **Trasovací graf** nebo **L1 poměr a/nebo alfa výběr přes křížové ověření**, lze z hodnoty **Začátek** (value1) do hodnoty **Ukončit** (value2) s přírůstkem **Podle** (value3) zadat mřížku s jedinečnými hodnotami alfa. Je-li zadán, je povolen pouze jeden platný soubor [value1 TO value2 BY value3]. Musí se ujistit, že  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . V případech, kdy  $\text{value1} = \text{value2}$ , je ekvivalentní zadání jednoho value1, bez ohledu na value3.

**Metrika Alfa** pro rozsahy hodnot může být buď **Lineární**, nebo **Base 10 logarithmic** (10 se umocní na hodnoty zadané hodnoty).

Vykreslí jsou zobrazeny pomocí zadané metriky pro vodorovné osy X hodnot alfa.

## Podmínky

Řídí analýzy.

### Zahrnout zachycení

Toto kritérium zahrnuje zachycení v namontovaném modelu (modelu). Všimněte si, že procedura rozšíření necentrování nebo standardizace závislé proměnné a zachycení není během odhadu penalizováno.

### Standardizace prediktorů

Standardizuje všechny nezávislé proměnné.

### Počet překřížení mezi ověřeními

Počet dělení nebo záhyby pro vyhodnocení křížového ověření modelů. Musí se jednat o kladnou celočíselnou hodnotu větší než 1. Výchozí nastavení je 5.

### Náhodný stav Python

Hodnota nastavení random\_state v parametru Python se používá při provádění vzájemného ověření platnosti modelů. Umožňuje replikaci výsledků, které zahrnují pseudo-náhodná čísla. Hodnota musí být celé číslo v rozsahu 0 až  $2^{32}-1$ . Výchozí nastavení je 0.

### Časový limit (minuty)

Počet minut povolených pro spuštění modelových výpočtů. Určíte-li hodnotu 0, časovač se vypne. Výchozí hodnota je 5.

## Zobrazit

Tato volba uvádí množství výstupu, které se má zobrazit pro poměr **L1 a/nebo výběr alfa přes validační ověření**.

### Nejlepší

Zobrazí pouze základní výsledky pro vybraný nejlepší model. Toto je nastaveno jako výchozí.

### Porovnat modely

Zobrazí základní výsledky pro všechny vyhodnocené modely.

### Porovnat modely a záhyby

Zobrazí úplně podrobné výsledky pro každé rozdělení nebo rozdělení pro každý vyhodnocený model.

## Graf

Tato volba určuje vykreslení pozorovaných nebo zbytkových hodnot oproti předpovězených hodnot a s křížovým validací specifikací zakreslených grafů průměrné střední kvadratické chyby (MSE) a/nebo průměru  $R^2$  přes hodnoty křížových ověření platnosti a hodnoty alfa.

### Průměrná střední kvadratická chyba (MSE) versus alfa

Pro atribut **L1 a/nebo výběr alfa prostřednictvím režimu křížového ověření** se zobrazí čára průměru průměrných hodnot MSE nad křížovými validací pro zadané nebo vybrané nejlepší hodnoty poměru L1 alfa. Pro režim **Trasovací graf** je automaticky vytvořen podobný graf na základě úplných dat o školení.

### Průměrné cross-validace $R^2$ Čtverec versus alfa

Pro atribut **L1 a/nebo výběr alfa prostřednictvím režimu křížového ověření** se zobrazí čára s průměrem  $R^2$  přes převod mezi validací a alfa pro zadanou nebo vybranou nejlepší hodnotu poměru L1. Pro režim **Trasovací graf** je automaticky vytvořen podobný graf na základě úplných dat o školení.

### Pozorovaný versus předpovězený

Zobrazí bodový graf pozorovaných versus předpokládaných hodnot pro uvedený nebo nejlepší model.

### Reziduum versus predicted

Zobrazí bodový graf zbytkových chyb oproti předpovězeným hodnotám pro uvedený nebo nejlepší model.

## Uložit

Uvádí proměnné, které se mají uložit do aktivní datové sady.

### Předpovězené hodnoty

Ukládat předpovězené hodnoty z uvedeného nebo nejlepšího modelu do aktivní datové sady. Můžete také uvést **Vlastní název proměnné**.

### Reziduum

Uložit zbytkové chyby z uvedených nebo nejlepších predikcí modelu do aktivní datové sady. Můžete také uvést **Vlastní název proměnné**.

## Lineární greso Regression

---

Linear Lasso používá třídu Python sklearn.linear\_model.Lasso k odhadu ztrát L1 regularizovaných lineárních regresních modelů pro závislou proměnnou na jedné nebo více nezávislých proměnných a zahrnuje volitelné režimy pro zobrazení trasovacích grafů a pro výběr hodnoty parametru alfa hyperparametru na základě vzájemné validace. Je-li k výběru příkazu alfa použit jeden model nebo je-li použita volba crossvalidation, lze k odhadu výkonu výstupu z dat použít oddíl obsahující data holdout.

Kromě přizpůsobení modelu se zadanou hodnotou parametru regulace alfa, lineární laso může zobrazit trasovací graf hodnot koeficientu pro rozsah hodnot alfa nebo usnadnit výběr hodnoty hyperparametru k n-násobnému vzájemnému ověření hodnot zadaných mřížek hodnot. Je-li k dispozici jediný model nebo je proveden výběr alfa prostřednictvím křížové validace, lze použít konečný model pro zadrhovací data vytvořená oblastí vstupních dat za účelem získání platného odhadu výkonu out-of-of-ukázky modelu.

## Získání analýzy lineární regrese laso

1. Z nabídky vyberte:

**Analyzovat > Regrese > Lineární prvky OLS Alternativy > Lasso**

Dialogové okno vám umožňuje uvést proměnnou, která přiřadí každý případ v aktivní datové sadě ke cvičicím nebo holdout ukázce.

2. Vyberte číselnou cílovou proměnnou. Pro spuštění analýzy je vyžadována pouze jedna cílová proměnná.

3. Uveďte numerickou závislost.

4. Určete alespoň jednu proměnnou kategorického faktoru nebo číselnou proměnnou kovariate.

Volitelně **Oblast** poskytuje způsob vytvoření díze nebo zkušební podmnožiny vstupních dat pro odhad výkonu zadaného nebo vybraného modelu. U všech případů s neplatnými daty pro každou proměnnou použitou procedurou se provádí dělení na oblasti po odstranění všech případů. Všimněte si, že pro křížové ověření se vytvoří záhyby nebo oblasti dat o školení v Python. Data holdout vytvořená oblastí nejsou použita v odhadu, bez ohledu na režim v platnosti.

Oblast lze definovat zadáním poměru případů náhodně vybraných do jednotlivých ukázek (v části **Školení a zadržení oblastí**) nebo prostřednictvím proměnné, která každému případu přiřadí školení nebo ukázkou zadržení. Nemůžete uvést jak školení, tak proměnné. Není-li oblast zadána, vytvoří se vzorek pro pouzdro, který se vytvoří přibližně z 30% vstupních dat.

Hodnota **Školení%** uvádí relativní počet případů v aktivní datové sadě náhodně přiřazeného vzorku školení. Výchozí odborná příprava je 70%.

## Lineární Lasso Regrese: Volby

Karta Volby poskytuje volby pro:

### Režim

Tento výběr poskytuje volby pro uvedení jednoho z následujících režimů:

#### Přizpůsobit se zadaným alfanumerickým

Vyberete-li tento režim, je do tréninkových dat instalován jediný model s použitím pouze jedné hodnoty regulování alfa. Toto je nastaveno jako výchozí. Je-li zadána logická oblast, použije se jediný nebo poslední model, který je instalován, na zadrhovací testovací data k odhadu výkonu na vzorku.

V části **Zákresový graf** lze vybrat vykreslení pozorovaných a/nebo zbytkových chyb oproti předpovězeným hodnotám.

Pod volbou **Uložit** můžete určit předpokládané hodnoty a zbytkové chyby, které chcete uložit.

#### Trasovací graf

Vyberete-li tento režim, zobrazí se tři grafy pro data o školení jako funkce alfa pro určenou sadu hodnot alfa:

- Trasovací graf regresních koeficientů.
- Zákresový graf  $R^2$ .
- Vykreslení střední kvadratická chyba (MSE).

Ačkoli je oblast počtena, nejsou poskytnuty žádné výsledky pro zadrhovací testovací data, protože z tohoto režimu nejsou výsledkem žádné konečné výsledky modelu.

#### Výběr Alfa přes křížovou validaci

Když vyberete vyhledávání v mřížce s křížovým ověřením pro vyhodnocení modelů, a zvolte nejlepší alfa na základě nejlepšího průměru,  $R^2$  nad přehybovými záhyby. Pole **Počet křížových ověření platnosti** lze použít ke změně výchozí hodnoty pěti dělení nebo záhybů pro vzájemné ověření platnosti. Je-li zadána logická oblast, použije se jediný nebo poslední model, který je instalován, na zadrhovací testovací data k odhadu výkonu na vzorku.



V části **Zobrazit** můžete vybrat zobrazení základních informací o modelu s vybranou hodnotou alfa (**Nejlepší**), základní informace o všech modelech ve srovnání (**Porovnat modely**) nebo úplné informace o všech rozděleních nebo záhybech pro všechny modely (**Porovnat modely a záhyby**). **Nejlepší** je výchozí nastavení.

V části **Zákresový graf** lze vybrat grafy střední hodnoty  $R^2$  a/nebo MSE nad validací. Lze vybrat také hodně pozorovaných a/nebo zbytkových chyb ve srovnání s předpovězenou hodnotou.

Pod volbou **Uložit** můžete určit předpokládané hodnoty a zbytkové chyby, které chcete uložit.

### Určete jednotlivé alphas

Vyberete-li režim **Přízpusobit se zadaným alfa**, můžete zadat jedinou hodnotu regularizace alfa. Když vyberete režim **Trasovací graf** nebo **Výběr Alfa přes režim křížového ověření**, můžete uvést více hodnot.

### Hodnoty

Uveďte jednu nebo více kladných regulizačních hodnot alfa. Vícenásobné hodnoty mohou být specifikovány jednotlivě nebo jako rozsahy. Výchozí nastavení je 1.

### Uvést mřížku hodnot alfa

Když vyberete režim **Trace plot** nebo **Alpha selection via cross validation**, může být mřížka jedinečných hodnot alfa zadána z hodnoty **Start** (value1) na hodnotu **End** (value2) s přírůstkem **Podle** (value3). Je-li zadán, je povolen pouze jeden platný soubor [value1 TO value2 BY value3]. Musí se ujistit, že  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . V případech, kdy  $\text{value1} = \text{value2}$ , je ekvivalentní zadání jednoho value1, bez ohledu na value3.

**Metrika Alfa** pro rozsahy hodnot může být buď **Lineární**, nebo **Base 10 logarithmic** (10 se umocní na hodnoty zadané hodnoty).

Vykreslí jsou zobrazeny pomocí zadané metriky pro vodorovné osy X hodnot alfa.

## Podmínky

Řídí analýzy.

### Zahrnout zachycení

Zahrnuje zachycení v namontovaných (ých) modelu (ech). Všimněte si, že procedura rozšíření necentrování nebo standardizace závislé proměnné a zachycení není během odhadu penalizováno.

### Standardizace prediktorů

Standardizuje všechny nezávislé proměnné.

### Počet překřížení mezi ověřením

Počet dělení nebo záhyby pro vyhodnocení křížového ověření modelů. Musí se jednat o kladnou celočíselnou hodnotu větší než 1. Výchozí nastavení je 5.

### Náhodný stav Python

Hodnota nastavení atributu random\_state v Python se používá při provádění vyhodnocení křížového ověření modelů. Umožňuje replikaci výsledků, které zahrnují pseudonáhodná čísla. Musí být celé číslo v rozsahu 0 až  $2^{32}-1$ . Výchozí nastavení je 0.

### Časový limit (minuty)

Počet minut povolených pro spuštění modelových výpočtů. Určíte-li hodnotu 0, časovač se vypne. Výchozí hodnota je 5.

## Zobrazit

Určuje množství výstupu, které se má zobrazit pro volbu **Výběr Alfa přes režim křížového ověření**.

### Nejlepší

Zobrazí pouze základní výsledky pro vybraný nejlepší model. Jedná se o výchozí nastavení.

### Porovnat modely

Zobrazí základní výsledky pro všechny vyhodnocené modely.

## Porovnat modely a záhyby

Zobrazí úplné podrobné výsledky pro každé rozdělení nebo rozdělení pro každý vyhodnocený model.

## Graf

Určuje vykreslení pozorovaných nebo zbytkových hodnot oproti předpovězených hodnotách a s křížovým validací, specifikací zakreslených buněk průměrné střední kvadratické chyby (MSE) a/nebo průměru  $R^2$  přes hodnoty křížových ověření platnosti versus alfa hodnoty.

### Průměrná hodnota střední kvadratické chyby (MSE) versus alfa

Pro režim **Alfa výběr přes cross-validation** se zobrazí čárový graf průměrného MSE přes cross-validation folds versus alfa. Pro režim **Trasovací graf** je automaticky vytvořen podobný graf na základě úplných dat o školení.

### Průměrné cross-validace $R^2$ Čtverec versus alfa

Pro režim **Alfa výběr přes režim křížového ověření** zobrazí spojnicový graf průměrného  $R^2$  přes přeškrťovací záhyby versus alfa. Pro režim **Trasovací graf** je automaticky vytvořen podobný graf na základě úplných dat o školení.

### Pozorovaný versus předpovězený

Zobrazí bodový graf pozorovaných versus předpokládaných hodnot pro uvedený nebo nejlepší model.

### Reziduum versus predicted

Zobrazí bodový graf zbytkových chyb oproti předpovězeným hodnotám pro uvedený nebo nejlepší model.

## Uložit

Uvádí proměnné, které se mají uložit do aktivní datové sady.

### Předpovězené hodnoty

Ukládat předpovězené hodnoty z uvedeného nebo nejlepšího modelu do aktivní datové sady. Můžete také uvést **Vlastní název proměnné**.

### Reziduum

Uložit zbytkové chyby z uvedených nebo nejlepších predikcí modelu do aktivní datové sady. Můžete také uvést **Vlastní název proměnné**.

## Regrese lineárního Ridge

---

Lineární Ridge používá třídu Python sklearn.linear\_model.Ridge k odhadu L2 nebo čtvercové ztráty regularizované lineární regresní modely pro závislou proměnnou na jedné nebo více nezávislých proměnných a zahrnuje volitelné režimy pro zobrazení trasovacích grafů a pro výběr hodnoty alfa hyperparametru na základě křížového ověření. Je-li k výběru příkazu alfa použit jeden model nebo je-li použita volba crossvalidation, lze k odhadu výkonu výstupu z dat použít oddíl obsahující data holdout.

Kromě instalace modelu se zadanou hodnotou parametru alfa regularizace může lineární ridge zobrazit trasovací graf hřebenové hodnoty součinitele hodnot pro rozsah hodnot alfa nebo usnadnit výběr hodnoty hyperparametru k n-násobnému vzájemnému ověření hodnot zadaných mřížek hodnot. Je-li k dispozici jediný model nebo je proveden výběr alfa prostřednictvím křížové validace, lze použít konečný model pro zadrhovací data vytvořená oblastí vstupních dat za účelem získání platného odhadu výkonu out-of-of-ukázky modelu.

## Získání regresní analýzy lineárního Ridge

1. Z nabídky vyberte:

**Analyzovat > Regrese > Lineární prvky Alternativy OLS > Ridge**

Dialogové okno vám umožňuje uvést proměnnou, která přiřadí každý případ v aktivní datové sadě ke cvičicím nebo holdout ukázce.

2. Vyberte číselnou cílovou proměnnou. Pro spuštění analýzy je vyžadována pouze jedna cílová proměnná.
3. Uveďte numerickou závislost.
4. Určete alespoň jednu proměnnou kategorického faktoru nebo číselnou proměnnou kovariate.

Volitelně **Oblast** poskytuje způsob vytvoření díze nebo zkušební podmnožiny vstupních dat pro odhad výkonu zadaného nebo vybraného modelu. U všech případů s neplatnými daty pro každou proměnnou použitou procedurou se provádí dělení na oblasti po odstranění všech případů. Všimněte si, že pro křížové ověření se vytvoří záhyby nebo oblasti dat o školení v Python. Data o holdoutu vytvářená oblastí nejsou použita v rámci odhadu, bez ohledu na režim, který je v platnosti.

Oblast lze definovat zadáním poměru případů, které jsou náhodně přiřazovány jednotlivým vzorům (v části **Školení a zadržení oblastí**), nebo proměnnou, která každému případu přiřazuje školení nebo vzorek z výplně. Nemůžete uvést jak školení, tak proměnné. Není-li oblast zadána, vytvoří se vzorek pro pouzdro, který se vytvoří přibližně z 30% vstupních dat.

Hodnota **Školení%** uvádí relativní počet případů v aktivní datové sadě náhodně přiřazeného vzorku školení. Výchozí odborná příprava je 70%.

## Lineární regrese Ridge: Volby

Karta Volby poskytuje volby pro:

### Režim

Tento výběr poskytuje volby pro uvedení jednoho z následujících režimů:

#### Přízpusobit se zadaným alfanumerickým

Vyberete-li tuto volbu, bude k údajům o školení, který používá pouze jednu hodnotu regularizace alfa, namontovat jeden model. Toto je nastaveno jako výchozí. Je-li zadána logická oblast, použije se jediný nebo poslední model, který je instalován, na zadrhovací testovací data k odhadu výkonu na vzorku.

V části **Zákresový graf** lze vybrat vykreslení pozorovaných a/nebo zbytkových chyb oproti předpovězeným hodnotám.

Pod volbou **Uložit** můžete určit předpokládané hodnoty a zbytkové chyby, které chcete uložit.

#### Trasovací graf

Vyberete-li tuto volbu, budou zobrazeny tři grafy pro data o školení jako funkce alfa pro určenou sadu hodnot alfa:

- Zákresový graf křivky regresních koeficientů.
- Zákresový graf  $R^2$ .
- Vykreslení střední kvadratická chyba (MSE).

Ačkoli je oblast počtena, nejsou poskytnuty žádné výsledky pro zadrhovací testovací data, protože z tohoto režimu nejsou výsledkem žádné konečné výsledky modelu.

#### Výběr Alfa přes křížovou validaci

Když vyberete tuto volbu, provede se prohledání mřížky s křížovým ověřením pro vyhodnocení modelů, a nejlepší alfa je vybrán na základě nejlepšího průměru  $R^2$  přes záhyby validace. Pole **Počet křížových ověření platnosti** lze použít ke změně výchozí hodnoty pěti dělení nebo záhybů pro vzájemné ověření platnosti. Je-li zadána logická oblast, použije se jediný nebo poslední model, který je instalován, na zadrhovací testovací data k odhadu výkonu na vzorku.

V části **Zobrazit** můžete vybrat zobrazení základních informací o modelu s vybranou hodnotou alfa (**Nejlepší**), základní informace o všech modelech ve srovnání (**Porovnat modely**) nebo úplné informace o všech rozdělenech nebo záhybech pro všechny modely (**Porovnat modely a záhyby**). **Nejlepší** je výchozí nastavení.

V části **Zákresový graf** lze vybrat grafy střední hodnoty  $R^2$  a/nebo MSE nad validací. Lze vybrat také hodně pozorovaných a/nebo zbytkových chyb ve srovnání s předpovězenou hodnotou.

Pod volbou **Uložit** můžete určit předpokládané hodnoty a zbytkové chyby, které chcete uložit.

### Určete jednotlivé alphas

Vyberete-li režim **Přízpusobit se zadaným alfa**, můžete zadat jedinou hodnotu regularizace alfa. Když vyberete režim **Trasovací graf** nebo **Výběr Alfa přes režim křížového ověření**, můžete uvést více hodnot.

### Hodnoty

Uveďte jednu nebo více kladných regulizačních hodnot alfa. Můžete uvést více hodnot jednotlivě, nebo jako rozsahy. Výchozí nastavení je 1.

### Uvést mřížku hodnot alfa

Když vyberete režim **Trace plot** nebo **Alpha selection via cross validation**, může být mřížka jedinečných hodnot alfa zadána z hodnoty **Start** (value1) na hodnotu **End** (value2) s přírůstkem **Podle** (value3). Je-li zadán, je povolen pouze jeden platný soubor [value1 TO value2 BY value3]. Musí se ujistit, že  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . V případech, kdy  $\text{value1} = \text{value2}$ , je ekvivalentní zadání jednoho value1, bez ohledu na value3.

**Metrika Alfa** pro rozsahy hodnot může být buď **Lineární**, nebo **Base 10 logarithmic** (10 se umocní na hodnoty zadané hodnoty).

Vykreslí jsou zobrazeny pomocí zadané metriky pro vodorovné osy X hodnot alfa.

## Podmínky

Řídí analýzy.

### Zahrnout zachycení

Toto kritérium zahrnuje zachycení v jednom nebo více namontovaných modelech. Všimněte si, že procedura rozšíření necentrování nebo standardizace závislé proměnné a zachycení není během odhadu penalizováno.

### Standardizace prediktorů

Toto kritérium standardizuje všechny nezávislé proměnné.

### Počet překřížení mezi ověřeními

Použijte tato kritéria k nastavení počtu rozdělení nebo přehybů pro vyhodnocení křížového ověření modelů. Počet musí být kladné celé číslo větší než 1. Výchozí nastavení je 5.

### Náhodný stav Python

Hodnota nastavení random\_state v prostředí Python se používá při provádění vzájemného ověření platnosti modelů. Umožňuje replikaci výsledků, které zahrnují pseudonáhodná čísla. Musí být celé číslo v rozsahu 0 až  $2^{32}-1$ . Výchozí nastavení je 0.

### Časový limit (minuty)

Počet minut povolených pro spuštění modelových výpočtů. Určíte-li hodnotu 0, časovač se vypne. Výchozí hodnota je 5.

## Zobrazit

Tato sekce uvádí množství výstupu, které se má zobrazit pro **Výběr Alfa přes režim křížového ověření**.

### Nejlepší

Zobrazí pouze základní výsledky pro vybraný nejlepší model. Toto je nastaveno jako výchozí.

### Porovnat modely

Také zobrazí základní výsledky pro všechny vyhodnocené modely.

### Porovnat modely a záhyby

Nakonec se zobrazí úplné podrobné výsledky pro každé rozdělení nebo rozdělení pro každý hodnocený model.

## Graf

Určuje vykreslení pozorovaných nebo zbytkových hodnot oproti předpovězených hodnot s křížovým validací, specifikací zakreslených buněk průměrné střední kvadratické chyby (MSE) a/nebo průměrného  $R^2$  přes hodnoty křížových ověření ve srovnání s hodnotami alfa.

### Průměrná hodnota střední kvadratické chyby (MSE) versus alfa

Volba **Výběr Alfa přes režim křížového ověření** zobrazuje spojnicový graf průměrného maximálního počtu objektů MSE nad křížovými validací pro záhyby oproti alfa. Pro režim **Trasovací graf** je automaticky vytvořen podobný graf na základě úplných dat o školení.

### Průměrné cross-validace $R^2$ Čtverec versus alfa

Volba **Výběr Alfa přes režim křížového ověření** zobrazuje spojnicový graf průměrné hodnoty  $R^2$  nad křížovými validací pro záhyby oproti alfa. Pro režim **Trasovací graf** je automaticky vytvořen podobný graf na základě úplných dat o školení.

### Pozorovaný versus předpovězený

Zobrazí bodový graf pozorovaných versus předpokládaných hodnot pro uvedený nebo nejlepší model.

### Reziduum versus predicted

Také se zobrazí bodový graf zbytkových chyb oproti předpovězeným hodnotám pro uvedený nebo nejlepší model.

## Uložit

Uvádí proměnné, které se mají uložit do aktivní datové sady.

### Předpovězené hodnoty

Ukládat předpovězené hodnoty z uvedeného nebo nejlepšího modelu do aktivní datové sady. Jinak můžete zadat **Vlastní název proměnné**.

### Reziduum

Uložit zbytkové chyby z uvedených nebo nejlepších predikcí modelu do aktivní datové sady. Jinak můžete zadat **Vlastní název proměnné**.

## Odhad křivky

Procedura odhadu křivky vytvoří regresní statistiku odhadu křivky a související grafy pro 11 různých regresních modelů pro odhad křivky. Pro každou závislou proměnnou je vytvořen samostatný model. Předpokládané hodnoty, zbytkové chyby a intervaly předpovědi můžete také uložit jako nové proměnné.

**Příklad.** Poskytovatel služeb sítě Internet sleduje v průběhu času procento nakažených virů elektronické pošty ve svých sítích. Scatterplot ukazuje, že vztah je nelineární. Do dat můžete umístit kvadratický model nebo kvadratický model a zkontrolujete platnost předpokladů a dobrotu, která je pro daný model vhodná.

**Statistika.** Pro každý model: regresní koeficienty, násobné  $R$ ,  $R^2$ , upravené  $R^2$ , směrodatná chyba tabulky odhadu, analýzy rozptylu, předpokládané hodnoty, zbytkové chyby a intervalů předpovědí. Modely: lineární, logaritmická, inverzní, kvadratická, krychlová, energie, sloučenina, S-křivka, logistika, růst a exponenciální.

Pokyny pro data odhadu křivky

**Data.** Závislé a nezávislé proměnné by měly být kvantitativní. Pokud vyberete volbu **Čas** z aktivní datové sady jako nezávislé proměnné (místo výběru proměnné), procedura odhadu křivky vygeneruje časovou proměnnou, kde je délka mezi případy uniformní. Je-li vybrána volba **Čas**, měla by být závislá proměnná měřítkem časové řady. Analýza časových řad vyžaduje strukturu datového souboru, ve které každý případ (řádek) představuje sadu pozorování v jiném čase a doba mezi případy je uniformní.

**Předpoklady.** Zobrazením dat graficky určete, jak nezávislé a závislé proměnné souvisejí (lineárně, exponenciálně, atd.). Celkové zbytkové chyby dobrého modelu by měly být náhodně rozděleny a normální. Je-li použit lineární model, měly by být splněny následující předpoklady: Pro každou hodnotu nezávislé proměnné musí být distribuce závislé proměnné normální. Rozdíl v distribuci závislé proměnné by měl být konstantní pro všechny hodnoty nezávislé proměnné. Vztah mezi závislou proměnnou a nezávislou proměnnou by měl být lineární a všechny pozorování by měly být nezávislé.

Získání odhadu křivky

1. Z nabídky vyberte:

**Analyzovat > Regrese > Odhad křivky ...**

2. Vyberte jednu nebo více závislých proměnných. Pro každou závislou proměnnou je vytvořen samostatný model.

3. Vyberte nezávislou proměnnou (buď vyberte proměnnou v aktivní datové sadě, nebo vyberte volbu **Čas**).

4. Volitelně:

- Vyberte proměnnou pro označování případů ve scatterplots. Pro každý bod v bodovém grafu můžete použít nástroj pro výběr bodů k zobrazení hodnoty proměnné Popisek případu.
- Klepnutím na tlačítko **Uložit** uložíte předpokládané hodnoty, zbytkové chyby a intervaly předpovědi jako nové proměnné.

K dispozici jsou také následující volby:

- **Zahrnout konstantu do rovnice.** Odhaduje konstantní termín v regresní rovnici. Konstanta je standardně zahrnuta.
- **Zákresové modely.** Vykreslí hodnoty závislé proměnné a každého vybraného modelu proti nezávislé proměnné. Pro každou závislou proměnnou se vytvoří samostatný graf.
- **Zobrazit tabulku ANOVA.** Zobrazí tabulku rozptylu souhrnu pro každý vybraný model.

## Modely odhadu křivky

Můžete zvolit jeden nebo více regresních modelů pro odhad křivky. Chcete-li určit, který model použít, zakreslejte svá data. Pokud se vaše proměnné jeví jako související lineárně, použijte jednoduchý lineární regresní model. Když se vaše proměnné nelineárně vztahují, zkuste transformovat vaše data. Když transformace nepomůže, možná budete potřebovat komplikovanější model. Zobrazit bodový graf vašich dat; pokud se zákres podobá matematické funkci, kterou rozpoznáváte, propadá vaše data na tento typ modelu. Pokud se například data podobají exponenciální funkci, použijte exponenciální model.

*Lineární.* Model, jehož rovnice je  $Y = b_0 + (b_1 * t)$ . Hodnoty řady jsou modelovány jako lineární funkce času.

*Logaritmický.* Model, jehož rovnice je  $Y = b_0 + (b_1 * \ln(t))$ .

*Inverzní.* Model, jehož rovnice má tvar  $Y = b_0 + (b_1 / t)$ .

*Kvadratický.* Model, jehož rovnice má tvar  $Y = b_0 + (b_1 * t) + (b_2 * t^{**2})$ . Kvadratický model lze použít k modelování řady, která "odsunuje" nebo sérii, která tlumí.

*Kubická.* Model, který je definován rovnicí  $Y = b_0 + (b_1 * t) + (b_2 * t^{**2}) + (b_3 * t^{**3})$ .

*Napájení.* Model, jehož rovnice je  $Y = b_0 * (t^{**b_1})$  nebo  $\ln(Y) = \ln(b_0) + (b_1 * \ln(t))$ .

*Složené.* Model, jehož rovnice je  $Y = b_0 * (b_1^{**t})$  nebo  $\ln(Y) = \ln(b_0) + (\ln(b_1) * t)$ .

*S-křivka.* Model, jehož rovnice má tvar  $Y = e^{**} (b_0 + (b_1/t))$  nebo  $\ln(Y) = b_0 + (b_1/t)$ .

*Logistický.* Model, jehož rovnice je  $Y = 1/(1/u + (b_0 * (b_1^{**t})))$  nebo  $\ln(1/y-1/u) = \ln(b_0) + (\ln(b_1) * t)$  kde u je hodnota horního ohraničení. Po výběru Logistický určete hodnotu horního ohraničení, která se má použít v regresní rovnici. Hodnota musí být kladné číslo, které je větší než největší hodnota závislé proměnné.

*Nárůst.* Model, jehož rovnice je  $Y = e^{**} (b_0 + (b_1 * t))$  nebo  $\ln(Y) = b_0 + (b_1 * t)$ .

*Exponenciální.* Model, jehož rovnice je  $Y = b_0 * (e^{**} (b_1 * t))$  nebo  $\ln(Y) = \ln(b_0) + (b_1 * t)$ .

## Uložit odhad křivky

**Uložit proměnné.** Pro každý vybraný model můžete uložit předpokládané hodnoty, zbytkové chyby (sledovaná hodnota závislé proměnné minus predikovaná hodnota modelu) a intervaly předpovědi (horní a dolní meze). Názvy nových proměnných a popisné popisky se zobrazí v tabulce ve výstupním okně.

**Predikční případy.** Pokud v aktivní datové sadě vyberete volbu **Čas** místo proměnné jako nezávislé proměnné, můžete zadat období prognózy za konec časové řady. Můžete zvolit jednu z následujících možností:

- **Předpovídat od období odhadu až do posledního případu.** Předvídá hodnoty pro všechny případy v souboru, založené na případech v období odhadu. Období odhadu, zobrazené v dolní části dialogového okna, je definováno s poddialogovým oknem Rozsah u volby Výběr případů v nabídce Data. Není-li žádné období odhadu definováno, použijí se všechny případy k předpovědi hodnot.
- **Předpověz.** Předvídá hodnoty prostřednictvím zadaného data, času nebo čísla sledování na základě případů v období odhadu. Tuto funkci lze použít k prognóze hodnot za posledním případem v časové řadě. Momentálně definované proměnné data určují, která textová pole jsou k dispozici pro určení konce období předpovědi. Pokud zde nejsou žádné definované proměnné data, můžete uvést koncové pozorovací číslo (velikost písmen).

Použijte volbu Definovat data v nabídce Data k vytvoření proměnných data.

## Regrese částečných nejmenších čtverců

Procedura regrese částečných nejmenších čtverců odhaduje, že regresní modely s částečnými alespoň čtverci (PLS, také známé jako "projekce na latentní strukturu") regresní modely. PLS je prediktivní metoda, která představuje alternativu k regresi typu OLS (ordinary least square), kanonické korelaci nebo modelování strukturálních rovnic, a je zvláště užitečná, když jsou proměnné prediktorů vysoce korelované nebo když počet prediktorů překračuje povolený počet případů.

PLS kombinuje funkce analýzy základních komponent a vícenásobnou regresi. Nejprve vyjme soubor latentních faktorů, které vysvětlují co nejvíce kovariance mezi nezávislými a závislými proměnnými. Pak regresní krok předpovídá hodnoty závislých proměnných pomocí dekompozice nezávislých proměnných.

### Tabulky

Část rozptylu rozptylu (latentní faktor), latentní váhové váhy, latentní zatížení faktoru, důležitost nezávislé proměnné v projekci (VIP) a regresní odhady parametrů (podle závislé proměnné) jsou ve výchozím nastavení vytvořeny.

### Grafy

Proměnná důležitost ve projekci (VIP), skóre faktoru, váhová váha pro první tři latentní faktory a vzdálenost k modelu se všechny vytvoří z karty [Volby](#) .

## Aspekty dat

### Úroveň měření

Závislé a nezávislé (prediktory) proměnné mohou být stupnice, nominální nebo ordinální. Tato procedura předpokládá, že byla přiřazena vhodná úroveň měření ke všem proměnným, ačkoli můžete dočasně změnit úroveň měření proměnné tak, že pravým tlačítkem myši klepnete na proměnnou ze seznamu zdrojových proměnných a vyberete úroveň měření z rozevírací nabídky. Kategorické (nominální nebo ordinální) proměnné se v rámci postupu chovají rovnocenně.

### Kódování kategoriálního proměnné

Procedura dočasně přeprogramuje kategoriální závislé proměnné s použitím jednoho kódování po dobu trvání procedury. Pokud existuje  $c$  kategorií proměnné, pak je proměnná uložena jako vektory  $c$  , přičemž první kategorie je označena  $(1, 0, \dots, 0)$  , další kategorie  $(0, 1, 0, \dots, 0)$  , ..., a konečné kategorie  $(0, 0, \dots, 0, 1)$  . Kategorické závislé proměnné jsou reprezentovány pomocí fiktivního kódování; to znamená, že jednoduše vynechává indikátor odpovídající kategorii odkazu.

### Váhy četnosti

Hodnoty váhy jsou zaokrouhleny na nejbližší celé číslo před použitím. Případy s chybějícími váhami nebo se závažími nižšími než 0,5 se v analýzách nepoužijí.

## Chybějící hodnoty

Hodnoty typu uživatel-a systémem-chybí hodnoty jsou považovány za neplatné.

## Reskování

Všechny proměnné modelu jsou vycentrovány na střed a standardizovány, včetně proměnných indikátorů reprezentujících kategoriální proměnné.

## Získání částečné regrese nejmenších čtverců

Z nabídky vyberte:

### Analyzovat > Regrese > Dílčí nejmenších čtverců ...

1. Vyberte alespoň jednu závislou proměnnou.
2. Vyberte alespoň jednu nezávislou proměnnou.

Volitelně můžete:

- Uvedte referenční kategorii pro kategoriální (nominální nebo pořadové) závislé proměnné.
- Zadejte proměnnou, která má být použita jako jedinečný identifikátor pro výstup a uložené datové sady s použitím malých a velkých písmen.
- Uvedte horní limit počtu latentních faktorů, které mají být extrahovány.

## Předpoklady

Procedura regrese částečných nejmenších čtverců je příkaz rozšíření Python a vyžaduje funkčnost Python, která je součástí vašeho produktu IBM SPSS Statistics. Vyžaduje také knihovny NumPy a SciPy Python, které jsou volně dostupné.

**Poznámka:** Pro uživatele pracující v režimu distribuované analýzy (vyžaduje server IBM SPSS Statistics), NumPy a SciPy musí být na serveru nainstalovány. Problém řešte ve spolupráci s administrátorem systému.

### Uživatelé systému Windows a Mac

Pro systémy Windows a Mac musí být NumPy a SciPy instalovány na samostatnou verzi produktu Python 3.10 z verze instalované s produktem IBM SPSS Statistics. Nemáte-li samostatnou verzi produktu Python 3.10, můžete ji stáhnout z <http://www.python.org>. Pak nainstalujte NumPy a SciPy pro Python verze 3.10. Instalační programy jsou k dispozici na webu <http://www.scipy.org/Download>.

Chcete-li povolit použití NumPy a SciPy, musíte nastavit umístění Python na verzi produktu Python 3.10, kde jste nainstalovali NumPy a SciPy. Umístění Python je nastaveno na kartě Umístění souboru v dialogovém okně Volby (Upravit > Volby).

### Uživatelé systému Linux

Doporučujeme vám stáhnout zdroj a sestavení NumPy a SciPy sami. Zdroj je k dispozici v produktu <http://www.scipy.org/Download>. Můžete nainstalovat NumPy a SciPy na verzi produktu Python 3.10, která je nainstalována s produktem IBM SPSS Statistics. Nachází se v adresáři Python v umístění, kde je nainstalován produkt IBM SPSS Statistics.

Pokud se rozhodnete instalovat NumPy a SciPy na verzi produktu Python 3.10 jinou než je verze, která je nainstalována s produktem IBM SPSS Statistics, pak musíte nastavit umístění Python tak, aby ukazovaly na tuto verzi. Umístění Python je nastaveno z karty Umístění souboru v dialogovém okně Volby (**Upravit > Volby**).

### Windows a Unix Server

Na serveru musí být na serveru nainstalována NumPy a SciPy na samostatnou verzi produktu Python 3.10 od verze, která je nainstalována s produktem IBM SPSS Statistics. Pokud na serveru není samostatná verze Python 3.10, pak je možné ji stáhnout z <http://www.python.org>. Funkce NumPy a SciPy pro Python 3.10 jsou k dispozici v produktu <http://www.scipy.org/Download>. Chcete-li povolit použití NumPy a SciPy, umístění Python pro server musí být nastaveno na verzi Python 3.10, kde



jsou nainstalovány NumPy a SciPy. Umístění Python je nastaveno z IBM SPSS Statistics Administration Console.

## Model

**Uved'te efekty modelu.** Model s hlavní účinky obsahuje všechny hlavní účinky přípravku a kovariuje hlavní účinky. Chcete-li určit interakce, vyberte volbu **Vlastní**. Je třeba označit všechny podmínky, které mají být zahrnuty do modelu.

**Faktory a proměnné.** Jsou zde uvedeny faktory a kovariány.

**Model.** Model závisí na povaze vašich dat. Po výběru volby **Vlastní** můžete vybrat hlavní efekty a interakce, které mají zájem na analýze.

Podmínky sestavení

Pro vybrané faktory a kovariáty:

**Interakce.** Vytvoří funkční období interakce nejvyšší úrovně všech vybraných proměnných. Jedná se o výchozí nastavení.

**Hlavní účinky.** Vytvoří výraz main-effects pro každou vybranou proměnnou.

**Všechny 2-way.** Vytvoří všechny možné dvousměrné interakce mezi vybranými proměnnými.

**Všechny 3-way.** Vytvoří všechny možné trojcestné interakce mezi vybranými proměnnými.

**Všechny 4-way.** Vytvoří všechny možné čtyřcestné interakce mezi vybranými proměnnými.

**Všechny 5-way.** Vytvoří všechny možné interakce pěti cest u vybraných proměnných.

## Volby

Karta Volby umožňuje uživateli uložit a vykreslit odhad modelu pro jednotlivé případy, latentní faktory a prediktory.

Pro každý typ dat zadejte název datové sady. Názvy datových sad musí být jedinečné. Pokud uvedete název existující datové sady, její obsah bude nahrazen jiným, jinak se vytvoří nová datová sada.

- **Ukládat odhady pro jednotlivé případy.** Uloží následující odhady modelu případu: předpokládané hodnoty, zbytkové chyby, vzdálenost k modelu latentního faktoru a skóre latentního faktoru. Poukazuje také na skóre latentního faktoru.
- **Uložte odhady pro latentní faktory.** Šetří latentní nastavení faktoru a váhy latentního faktoru. Pokrývá také latentní váhy faktoru.
- **Ukládat odhady pro nezávislé proměnné.** Uloží odhady parametrů regrese a proměnnou důležitost pro projekci (VIP). Pokrývá také VIP podle latentního faktoru.

## Analýza nejbližšího souseda

Analýza nejbližšího souseda je metoda pro klasifikaci případů na základě jejich podobnosti s ostatními případy. V počítačově učení bylo vyvinuto jako způsob, jak rozpoznat vzorky dat bez požadavku na přesnou shodu se všemi uloženými vzory nebo případy. Podobné případy jsou blízko sebe a nepodobné případy jsou vzdálené od sebe navzájem. Vzdálenost mezi dvěma případy je tedy měřítkem jejich nepodobnosti.

Mezi věci, které jsou blízko sebe, se říká, že jsou "sousedé". Při předkládání nového případu ("Holdout") se vypočítává jeho vzdálenost od každého z případů v modelu. Klasifikace nejpodobnějších případů-nejbližší sousedy-se prolhaní a nový případ se umístí do kategorie, která obsahuje největší počet nejbližších sousedů.

Můžete uvést počet nejbližších sousedů, které se mají prozkoumat; tato hodnota se nazývá  $k$ .

Analýza nejbližšího souseda může být také použita k výpočtu hodnot pro souvislý cíl. V této situaci se průměrná nebo střední cílová hodnota nejbližších sousedů používá k získání předpovězené hodnoty pro nový případ.












**Cíl a funkce.** Cíl a funkce mohou být:

- **Nominální.** Proměnná může být považována za nominální, když její hodnoty představují kategorie bez vlastního hodnocení (například oddělení společnosti, v níž zaměstnanec pracuje). Příklady nominálních proměnných zahrnují region, poštovní směrovací číslo a náboženskou příslušnost.
- **Pořadové.** Proměnnou lze považovat za ordinální, když její hodnoty představují kategorie s nějakým vlastním hodnocením (například úroveň spokojenosti služeb od vysoce nespokojené s vysoce spokojené). Příklady ordinálních proměnných zahrnují skóre přístupu představující stupeň spokojenosti nebo důvěryhodnost a skóre hodnocení preferencí.
- **Měřítka.** Proměnná může být považována za měřítka (souvislá), když její hodnoty reprezentují uspořádané kategorie se smysluplnou metrikou, takže porovnání vzdálenosti mezi hodnotami je vhodné. Příklady proměnných měřítka zahrnují věk v rocích a příjem v tisících dolarů.

Jmenovité a Ordinální proměnné jsou považovány za ekvivalentně z nejbližšího souseda analýzy souseda. Procedura předpokládá, že byla přiřazena příslušná úroveň měření ke každé proměnné; nicméně můžete dočasně změnit úroveň měření proměnné tak, že pravým tlačítkem myši klepnete na proměnnou ve zdrojovém seznamu proměnných a vyberete úroveň měření z rozevírací nabídky.

Ikona vedle každé proměnné v seznamu proměnných identifikuje úroveň měření a datový typ:

*Tabulka 1. Ikony úrovně měření*

	Číselné	Řetězec	Datum	Čas
Měřítka (spojitá)		není k dispozici		
Pořadové				
Nominální				

**Kategorické kódování proměnných.** Procedura dočasně přeprogramuje kategoriální prediktory a závislé proměnné pomocí kódu one-of-c kódu po dobu trvání procedury. Pokud existuje c kategorií proměnné, pak je proměnná uložena jako c vektory, s první kategorií označenou (1,0, ..., 0), další kategorie (0,1,0, ..., 0), ..., a konečné kategorie (0,0, ..., 0, 1).

Toto schéma kódování zvyšuje dimenzionalitu prostoru funkce. Celkový počet dimenzí je zejména počet prediktorů měřítka a počet kategorií napříč všemi kategoriálními prediktory. Výsledkem je, že tento kódovací program může vést k pomalejšímu školení. Pokud váš nejbližší trénink probíhá velmi pomalu, můžete se pokusit snížit počet kategorií ve vašich kategoriálních prediktorů kombinací podobných kategorií nebo vyzarování případů, které mají před spuštěním procedury velmi vzácné kategorie.

Všechny kódy jsou založeny na datech odborné přípravy, a to i v případě, že je definován vzorek Holdout (viz "Oblasti" na stránce 164). Pokud tedy vzorek z nedostatku obsahuje případy s prediktory, které nejsou přítomny v údajích o výcviku, pak tyto případy nevyhodují. Pokud vzorek obluďného pouzdra obsahuje případy se závislými kategoriemi proměnných, které nejsou obsaženy v datech odborné přípravy, pak se tyto případy skóruje.

**Rescaling.** Funkce měřítka jsou standardně normalizovány. Veškeré opakované skenování se provádí na základě údajů o školení, a to i v případě, že je definován vzorek Holdout (viz "Oblasti" na stránce 164). Určíte-li proměnnou pro definování logických oblastí, je důležité, aby funkce měly podobné distribuce v rámci odborné přípravy a ve vzorcích pouzdro z umístění. Použijte například proceduru Prozkoumat, chcete-li zkontrolovat rozdělení mezi oblastmi.

**Frekvence frekvencí.** Tato procedura se ignoruje váhami frekvence.

**Replikace výsledků.** Procedura používá náhodné generování čísel během náhodného přiřazení oblastí a přehybů mezi validacemi. Chcete-li replikovat výsledky přesně, kromě použití stejných nastavení

procedury, nastavte počáteční hodnotu pro vlastnost Mersenne Twister (viz "Oblasti" na stránce 164) nebo použijte proměnné k definování oblastí a křížových ověření platnosti.

Chcete-li získat nejbližší analýzu souseda,

Z nabídky vyberte:

### **Analyzovat > Klasifikovat > Nejbližší soused ...**

1. Uveďte jednu nebo více funkcí, které mohou být považovány za nezávislé proměnné nebo prediktory, pokud existuje cíl.

**Cíl (volitelné).** Není-li zadán žádný cíl (závislá proměnná nebo odezva), pak procedura najde  $k$  nejbližších sousedů-bez klasifikace nebo předpovědi.

**Normalizovat funkce měřítka.** Normalizované funkce mají stejný rozsah hodnot, které mohou zlepšit výkonnost algoritmu odhadu. Používá se upravená normalizace,  $[2 * (x - \min) / (\max - \min)] - 1$ . Upravené normalizované hodnoty spadají mezi  $-1$  až  $1$ .

**Identifikátor produktu Focal Case (volitelné).** To vám umožní označit případy zvláštního zájmu. Například výzkumník chce určit, zda výsledky zkoušek z jednoho školského okrsku-kontaktní osoby jsou srovnatelné s výsledky zkoušek z podobných školských obvodů. Používá nejbližší analýzu souseda k nalezení školních okrsků, které jsou nejvíce podobné s ohledem na danou sadu funkcí. Pak srovnává výsledky zkoušek z ústředního školského okrsku k těm, které jsou od nejbližších sousedů.

Ohnisková pouzdra by mohla být také použita v klinických studiích k výběru kontrolních případů, které jsou podobné klinickým případům. Aplikace Focal Point se zobrazují v  $k$  nejbližších sousedech a vzdálenostech tabulce, v grafu prostorových dat, grafu typu peer a v kvadrantové mapě. Informace o ústředních případech se ukládají do souborů uvedených na kartě Výstup.

Případy s kladnou hodnotou u zadané proměnné jsou považovány za závažné případy. Je neplatné zadat proměnnou bez kladných hodnot.

**Popisek případu (volitelný).** Případy jsou označeny pomocí těchto hodnot ve sloupcovém grafu funkcí, grafu typu peer a kvadrantové mapě.

Pole s neznámou úrovní měření

Výstraha na úrovni měření se zobrazí, když je úroveň měření pro jednu nebo více proměnných (polí) v datové sadě neznámá. Jelikož úroveň měření ovlivňuje výpočet výsledků pro tuto proceduru, všechny proměnné musí mít definovanou úroveň měření.

**Data skenování.** Přečte data v aktivní datové sadě a přiřadí výchozí úroveň měření k jakýmkoli polím s momentálně neznámou úrovní měření. Je-li datová sada velká, může to nějakou dobu trvat.

**Přiřadit ručně.** Otevře dialogové okno se seznamem všech polí s neznámou úrovní měření. Toto dialogové okno můžete použít k přiřazení úrovně měření k těmto polím. Úroveň měření můžete také přiřadit v pohledu Proměnné v editoru dat.

Vzhledem k tomu, že úroveň měření je pro tuto proceduru důležitá, nemůžete přistupovat k dialogovému oknu pro spuštění této procedury, dokud nebude mít všechna pole definovanou úroveň měření.

## **Sousedé**

**Počet nejbližším sousedů (k).** Uveďte počet nejbližších sousedů. Všimněte si, že použití většího počtu sousedů nemusí nutně vést k přesnějšímu modelu.

Je-li na kartě Proměnné zadán cíl, můžete alternativně určit rozsah hodnot a povolit, aby procedura vybrala "nejlepší" počet sousedů v daném rozsahu. Metoda pro určení počtu nejbližších sousedů závisí na tom, zda je na kartě Součásti požadován výběr funkcí.

- Je-li výběr funkcí v platnosti, pak je výběr funkcí proveden pro každou hodnotu  $k$  v požadovaném rozsahu a  $k$ a doprovodná sada funkcí s nejnižší chybovost (nebo nejmenší chyba součtu čtverců, je-li cílová hodnota měřítka) vybrána.
- Pokud výběr funkcí není v platnosti, pak  $V$ -fold cross-validation se používá k výběru "nejlepšího" počtu sousedů. Chcete-li ovládat přiřazení přehybů, prohlédněte si kartu Oblast.

**Vypočítávání vzdálenosti.** Jedná se o metriku použitou k určení metriky vzdálenosti používané k měření podobnosti případů.

- **euklidovská metrika.** Vzdálenost mezi dvěma případy,  $x$  a  $y$ , je druhou odmocninou součtu všech rozměrů a čtvercových rozdílů mezi hodnotami pro případy.
- **Metrika bloku města.** Vzdálenost mezi dvěma případy je součtem absolutních rozdílů mezi hodnotami pro dané případy ve všech dimenzích. Také se jmenuje Manhattan.

Volitelně, je-li cíl zadán na kartě Proměnné, můžete zvolit váhu funkcí podle jejich normalizovaného významu při výpočtu vzdáleností. Význam ukazatele pro prediktor je vypočítán poměrem chyby nebo součtu chyb modelu s prediktorem odebraným z modelu k chybovému poměru nebo k chybě součtu druhých mocnin pro úplný model. Normalizovaný význam je vypočítán převážením hodnot důležitosti funkce tak, aby jejich součet byl nastaven na 1.

**Predikce pro cíl měřítka.** Je-li na kartě Proměnné uveden cíl měřítka, určuje, zda se předpovídaná hodnota vypočítá na základě střední hodnoty nebo střední hodnoty nejbližších sousedů.

## Funkce

Karta Součásti vám umožňuje požadovat a uvést volby pro výběr funkcí, když je cíl uveden na kartě Proměnné. Standardně jsou všechny funkce zvažované pro výběr funkcí, ale volitelně můžete vybrat podmnožinu funkcí, které se mají vynutit v modelu.

**Zastavení kritéria.** V každém kroku je funkce, jejíž přidání do modelu má za následek nejmenší chybu (vypočtenou jako četnost chyb pro kategorický cíl a chybu součtu čtverců pro cílové měřítko), považována za zahrnutí do sady modelů. Výběr postoupení pokračuje, dokud není splněna uvedená podmínka.

- **Zadaný počet funkcí.** Algoritmus navíc přidává do modelu pevný počet funkcí navíc k těm, které jsou k dispozici. Uveďte kladné celé číslo. Klesající hodnoty počtu pro výběr vytvoří více lichého modelu, na riziko chybějících důležitých funkcí. Zvýšení hodnot počtu pro výběr bude zachycovat všechny důležité funkce, a to s rizikem eventuálně přidání funkcí, které ve skutečnosti zvyšují chybu modelu.
- **Minimální poměr změn v absolutním chybovém poměru.** Algoritmus se zastaví, když se změna v absolutním chybovém poměru označuje, že model nelze dále zlepšit přidáním dalších funkcí. Uveďte kladné číslo. Klesající hodnoty minimální změny budou mít tendenci zahrnovat více funkcí, s rizikem zahrnutí funkcí, které do modelu nepřidávají mnoho hodnot. Zvýšení hodnoty minimální změny bude mít tendenci vyloučit více funkcí, s rizikem ztráty funkcí, které jsou důležité pro model. Hodnota "optimální" minimální změny bude záviset na vašich datech a aplikacích. Podívejte se do protokolu chyb výběru funkcí ve výstupu, který vám pomůže posoudit, které funkce jsou nejdůležitější. Další informace naleznete v tématu ["Protokol chyb výběru funkcí"](#) na stránce 168.

## Oblasti

Ouško Oddíly vám umožňuje rozdělit datovou sadu na trénovací a pouzdřované sady a v případě potřeby přiřadit případy do převaděčů napříč validacemi

**Výcvik a zadržení oddílů.** Tato skupina uvádí metodu rozdělení aktivního datového souboru na oblasti do výcviku a pouzdřovo ven. **Školící vzorek** se skládá z datových záznamů používaných k vyškolení nejbližšího modelu sousedního zařízení; některé procento případů v datové sadě musí být přiděleno vzorku výcviku, aby bylo možné získat model. **Výpočetní ukázka** je nezávislá sada datových záznamů používaných k odhadu konečného modelu; chyba pro ukázkou Holdout poskytuje "čestný" odhad schopnosti prediktivní schopnosti modelu, protože pouzdřovo pouzdřovo nebylo použito k sestavení modelu.

- **Náhodně přiřazujte případy k oblastem.** Určete procentní část případů, které mají být přiřazeny ke vzorku školení. Zbytek je přiřazen k souboru holdout.
- **Použijte proměnnou k přiřazení případů.** Uveďte numerickou proměnnou, která přiřadí každý případ v aktivní datové sadě ke cvičicím nebo holdovému vzorku. Případy s kladnou hodnotou na proměnné jsou přiřazeny ke vzorku školení, případy s hodnotou 0 nebo záporné hodnoty do vzorku "holdout". Případy se systémem-chybějící hodnota jsou z analýzy vyloučeny. Jakýkoli uživatel-chybějící hodnoty proměnné logické oblasti je vždy považován za platný.

**Flds-Validation.** V-fold cross-validation se používá k určení "nejlepšího" počtu sousedů. Ve spojení s výběrem funkcí z důvodu výkonu není k dispozici.

Křížové ověření rozděluje vzorek na několik dílčích vzorků nebo záhyby. Poté jsou generovány nejdražší sousední modely, s vyloučením dat z jednotlivých dílčích vzorků postupně. První model je založen na všech případech s výjimkou těch, které se nacházejí v prvním vzorku, druhý model je založen na všech případech s výjimkou těch, které jsou uvedeny v druhém vzorku a tak dále. Pro každý model se chyba odhaduje použitím modelu na dílčí ukázkou, která má být generována při generování. "Nejlepší" počet nejbližších sousedů je ten, který produkuje nejnižší chybu napříč záhyby.

- **Náhodně přiřadit případy k přehybům.** Uvedte počet přehybů, které by měly být použity pro křížové ověření platnosti. Procedura náhodně přiřadí případy k přehybům, očíslovaných od 1 do  $V$ , počet záhyby.
- **Použijte proměnnou k přiřazení případů.** Uvedte číselnou proměnnou, která přiřadí každý případ v aktivní datové sadě fold. Proměnná musí být číselná a musí mít hodnoty od 1 do  $V$ . Pokud všechny hodnoty v tomto rozsahu chybí, a na každém rozdělení, pokud jsou rozdělené soubory v platnosti, způsobí to chybu.

**Set seed for Mersenne Twister.** Nastavení objektu typu seed vám umožní replikovat analýzy. Použití tohoto ovládacího prvku je podobné nastavení aktivního generátoru Mersenne Twister a uvedení pevného počátečního bodu v dialogovém okně Generátory náhodného čísla s důležitým rozdílem, že nastavení počátečního bodu v tomto dialogovém okně zachová aktuální stav generátoru náhodných čísel a obnoví tento stav po dokončení analýzy.

## Uložit

**Názvy uložených proměnných.** Automatické generování názvu zajišťuje, že zachováte veškerou svou práci. Vlastní názvy vám umožňují vyřadit/nahradiť výsledky z předchozích spuštění bez předchozího odstranění uložených proměnných v Editoru dat.

Proměnné pro uložení

- **Předpokládaná hodnota nebo kategorie.** Tím uložíte předpovězenou hodnotu pro cíl měřítka nebo předpovědanou kategorii pro kategorický cíl.
- **Předpokládaná pravděpodobnost.** Tím se ukládají predikované pravděpodobnosti pro kategorický cíl. Pro každou z prvních kategorií  $n$  je uložena samostatná proměnná, kde  $n$  je určen v ovládacím prvku **Maximální kategorie pro uložení kategoriálního cíle**.
- **Proměnné oddílu Školení/Zadržené logické oblasti.** Jsou-li případy náhodně přiřazeny ke vzorkům školení a pouzdro na kartě Oblasti, uloží se hodnota oddílu (školení nebo pouzdra), ke kterému byl případ přiřazen.
- **Proměnná násobného ověření.** Jsou-li případy náhodně přiřazeny ke křížovým ověřením na kartě Oblasti, uloží se tato hodnota počtu, do kterého byl případ přiřazen.

## Výstup

Výstup programu Viewer

- **Souhrn zpracování případu.** Zobrazí souhrnnou tabulku zpracování případu, která shrnuje počet zahrnutých a vyloučených případů v analýze, celkem a ve formě školení a výsuvených vzorků.
- **Grafy a tabulky.** Zobrazí výstup související s modelem, včetně tabulek a grafů. Tabulky v zobrazení modelu zahrnují  $k$  nejbližší sousedy a vzdálenosti pro cílové případy, klasifikaci kategoriálních proměnných odpovědí a souhrn chyb. Grafický výstup v zobrazení modelu obsahuje protokol chyb výběru, graf důležitosti funkcí, graf prostoru funkcí, graf typu peer a kvadrantovou mapu. Další informace naleznete v tématu ["Zobrazení modelu"](#) na stránce 166.

Soubory

- **Exportovat model do XML.** Tento modelový soubor můžete použít k použití informací modelu na jiné datové soubory pro účely hodnocení. Tato volba není k dispozici, pokud byly definovány rozdělené soubory.

- **Vyexportujte vzdálenosti mezi ohniskovými případy a k nejbližším sousedům.** Pro každý ústřední případ se vytvoří samostatná proměnná pro nejbližší sousední  $k$  nejbližší sousední země (z ukázky školení) a odpovídající  $k$  nejbližší vzdálenosti.

## Možnosti

**Uživatel-Chybějící hodnoty.** Kategorické proměnné musí mít platné hodnoty pro případ, který má být zahrnut do analýzy. Tyto ovládací prvky umožňují rozhodnout o tom, zda se s uživatelem chybějícími hodnotami zachází jako s kategoriálními proměnnými jako s platnými hodnotami.

Systémem-chybějící hodnoty a chybějící hodnoty pro proměnné měřítka se vždy považují za neplatné.

## Zobrazení modelu

Vyberete-li volbu **Grafy a tabulky** na kartě Výstup, vytvoří procedura v prohlížeči nejbližší objekt modelu Neighbor Sousední objekt. Aktivováním (poklepáním) na tento objekt získáte interaktivní pohled na model. Pohled modelu má okno se 2 panely:

- Na prvním panelu je zobrazen přehled modelu s názvem hlavního pohledu.
- Druhý panel zobrazuje jeden ze dvou typů zobrazení:

Pomocný model zobrazení ukazuje více informací o modelu, ale není zaměřen na model samotný.

Propojený pohled je pohled, který zobrazuje podrobnosti o jedné funkci modelu, když uživatel provede přechod na část hlavního pohledu.

První panel standardně zobrazuje prostor funkcí a druhý panel ukazuje graf důležitosti proměnných. Není-li graf důležitosti proměnné k dispozici, tj. **Funkce váhy podle důležitosti** nebyla vybrána na kartě Funkce, zobrazí se první dostupné zobrazení v rozevírací nabídce Zobrazit.

Když pohled nemá k dispozici žádné informace, je její text položky v rozevírací nabídce Zobrazit zakázán.

## Prostor funkce

Graf prostoru funkcí je interaktivní graf prostoru funkcí (nebo podprostor, pokud je více než 3 funkce). Každá osa představuje funkci v modelu a umístění bodů v grafu zobrazuje hodnoty těchto funkcí pro případy ve školících a holdingových oblastech.

**Klávesy.** Kromě hodnot vlastností ukazují body v grafu další informace.

- Tvar označuje oddíl, do kterého bod patří, buď Training, nebo Holdout.
- Barva/stínování bodu označuje hodnotu cíle pro daný případ; s odlišnými hodnotami barev rovnými kategoriálním cílem a odstínům označujícím rozsah hodnot spojitého cíle. Označená hodnota pro oblast školení je sledovaná hodnota; pro oblast Holdout se jedná o předpovězenou hodnotu. Není-li zadán žádný cíl, tento klíč se nezobrazí.
- Těžší obrysy ukazují, že případ je ohniskem. Ústřední případy se zobrazují s nejbližšími sousedy  $k$ .

**Ovládací prvky a interaktivita.** Počet ovládacích prvků v grafu umožňuje prozkoumat prostor funkcí.

- Můžete vybrat podmnožinu funkcí, které se mají zobrazit v grafu, a změnit funkce, které jsou reprezentovány v dimenzích.
- "Focal cases" are simply points selected in the Feature Space chart. Pokud jste zadali proměnnou cílového případu, budou nejprve vybrány body představující cílové případy. Nicméně, jakýkoli bod se může dočasně stát ústředním případem, pokud jej vyberete. "Obvyklé" ovládací prvky pro výběr bodů platí; klepání na bod vybírá tento bod a zruší výběr všech ostatních; Řízení-klepnutí na bod přidá tuto hodnotu do sady vybraných bodů. Propojené pohledy, jako např. Peers Chart, se automaticky aktualizují na základě vybraných případů v prostoru funkce.
- Můžete změnit počet nejbližších sousedů ( $k$ ), které se zobrazí pro cílové případy.
- Přetivání nad bodem v grafu zobrazuje popis s hodnotou popisku případu, nebo číslo případu, pokud nejsou popisky případů definovány, a zjištěné a předpovězené cílové hodnoty.

- Tlačítko "Reset" vám umožňuje vrátit prostor funkcí do jeho původního stavu.

## **Přidání a odebrání poli/variables**

Do prostoru funkcí můžete přidat nová pole/var, nebo odstranit ty, které jsou momentálně zobrazeny.

Paleta proměnných

Před přidáním a odebráním proměnných musí být zobrazena paleta proměnných. Chcete-li zobrazit paletu Proměnné, prohlížeč modelů musí být v režimu úprav a v prostoru funkcí musí být vybrán případ.

1. Chcete-li prohlížeč modelů umístit do režimu úprav, vyberte z nabídky následující volby:

### **Zobrazit > Režim úprav**

2. Jednou v režimu úprav klepněte na libovolný případ v prostoru funkcí.
3. Chcete-li zobrazit paletu Proměnné, vyberte z nabídky:

### **Zobrazení > Palety > Proměnné**

Na paletě Proměnné jsou uvedeny všechny proměnné v prostoru funkcí. Ikona vedle názvu proměnné označuje úroveň měření proměnné.

4. Chcete-li dočasně změnit úroveň měření proměnné, klepněte pravým tlačítkem myši na proměnnou v paletě proměnných a vyberte volbu.

Zóny proměnných

Proměnné jsou přidány do "zón" v prostoru funkcí. Chcete-li zobrazit zóny, spusťte přetahování proměnné z palety Proměnné nebo vyberte volbu **Zobrazit zóny**.

Prostor funkcí má zóny pro osy  $x$ ,  $y$  a  $z$ .

Přesun proměnných do zón

Zde jsou některá obecná pravidla a tipy pro přesouvání proměnných do zón:

- Chcete-li přesunout proměnnou do zóny, klepněte na ni a přetáhněte ji z palety Proměnné a umístěte ji do dané zóny. Vyberete-li volbu **Zobrazit zóny**, můžete také klepnout pravým tlačítkem myši na zónu a vybrat proměnnou, kterou chcete přidat do zóny.
- Pokud přetáhnete proměnnou z palety Proměnné do zóny, která je již obsazena jinou proměnnou, bude původní proměnná nahrazena novou.
- Přetahujete-li proměnnou z jedné zóny do zóny, která je již obsazena jinou proměnnou, proměnné přehození proměnných.
- Klepnutím na  $X$  v zóně odeberete proměnnou z této zóny.
- Je-li ve vizualizaci více grafických prvků, každý grafický prvek může mít své vlastní proměnné zóny. Nejprve vyberte grafický prvek.

## **Důležitost proměnné**

Obvykle se budete chtít zaměřit na modelovací úsilí na proměnné, které jsou nejvíce důležité, a zvažte možnost zrušení nebo ignorování těch, které jsou nejméně. Graf důležitosti proměnných vám pomáhá to, že indikuje relativní důležitost každé proměnné při odhadu modelu. Vzhledem k tomu, že hodnoty jsou relativní, součet hodnot všech proměnných na obrazovce je 1.0. Proměnná důležitost se nevztahuje na přesnost modelu. Vztahuje se pouze na důležitost každé proměnné při vytváření predikce, nikoli na to, zda je prognóza přesná nebo ne.

## **Peers**

Tento graf zobrazuje cílové případy a jejich  $k$  nejbližší sousedy na každé funkci a na cíli. Je k dispozici, je-li v prostoru funkce vybrán ústřední případ.

**Chování propojení.** Graf Peers je propojen s prostorem funkcí dvěma způsoby.

- Vybrané případy (focal) v prostoru funkcí se zobrazí v grafu Peers spolu s nejbližšími sousedy  $k$ .

- Hodnota  $k$  vybraná v prostoru funkcí se používá v grafu Rovnocenní partneři.

## Nejbližší odpojení od sousedů

V této tabulce se zobrazuje  $k$  nejbližší sousední země a vzdálenosti pouze pro cílové případy. Je k dispozici, je-li identifikátor ohniska zadán na kartě Proměnné a zobrazuje pouze cílové případy identifikované touto proměnnou.

Každý řádek:

- Sloupec **Focal Case** obsahuje hodnotu proměnné case opatřování pro ohnisko případu; nejsou-li štítky case definovány, tento sloupec obsahuje číslo případu cílového případu.
- Sloupec  $i$  pod položkou Nejbližší sousední skupina obsahuje hodnotu proměnné case popisování pro  $i$ . nejbližší sousedce cílového případu; pokud nejsou definovány návěští case, tento sloupec obsahuje číslo případu  $i$ . nejbližší sousední případ.
- Sloupec  $i$  pod skupinou Nejbližší vzdálenost obsahuje vzdálenost od  $i$ . nejbližšího souseda k ohniskovému případu.

## Kvadrantové mapy

Tento graf zobrazuje cílové případy a jejich  $k$  nejbližší sousedé na bodovém grafu (nebo dotpart, v závislosti na úrovni měření cíle) s cílem na ose  $y$  a funkci měřítka na ose  $x$  jsou součástí funkcí. Je k dispozici v případě, že je cíl a je-li v prostoru funkce vybrán cílový případ.

- Referenční čáry se kreslí pro souvislé proměnné, v proměnné prostředky v oblasti školení.

## Protokol chyb výběru funkcí

Body na grafu zobrazí chybu (buď chyba v chybovém poměru, nebo chyba součtu čtverců, v závislosti na úrovni měření cíle) na ose  $y$  pro model s funkcí uvedenou na ose  $x$  (plus všechny funkce vlevo na ose  $x$ ). Tento graf je k dispozici, pokud existuje cíl a výběr funkcí je v platnosti.

## k výběru protokolu chyb

Body na grafu zobrazují chybu (buď chyba v chybovém poměru, nebo chyba součtu čtverců, v závislosti na úrovni měření cíle) na ose  $y$  pro model s počtem nejbližších sousedů ( $k$ ) na ose  $x$ . Tento graf je dostupný, pokud existuje cíl a  $k$  výběr je v platnosti.

## Protokol k chybě výběru funkcí k a funkci

Jedná se o grafy výběru funkcí (viz "Protokol chyb výběru funkcí" na stránce 168), které jsou zobrazeny podle  $k$ . Tento graf je k dispozici, je-li v platnosti cíl a  $k$  a výběr funkcí.

## Tabulka klasifikací

Tato tabulka zobrazuje křížovou klasifikaci pozorovaných versus předpokládaných hodnot cíle, podle oblastí. Je k dispozici v případě, že existuje cíl a je kategorický.

- Řádek (**Chybí**) v oblasti Holdout obsahuje případy zadržení s chybějícími hodnotami v cíli. Tyto případy se podílejí na ukázce Zadržení: Celkové procentní hodnoty, nikoli však hodnoty Procento správných hodnot.

## Souhrn chyb

Tato tabulka je k dispozici v případě, že existuje cílová proměnná. Zobrazí chybu přidruženou k modelu; součet-čtverců pro souvislý cíl a četnost chyb (100% – celkové správné procento) pro kategorický cíl.



## Analýza diskriminátoru

Analýza diskriminantů vytváří prediktivní model pro členství ve skupinách. Model je složen z diskriminačních funkcí (nebo, pro více než dvě skupiny, sadu diskriminačních funkcí) na základě lineárních kombinací proměnných prediktorů, které zajišťují nejlepší diskriminaci mezi skupinami. Funkce jsou generovány z ukázky případů, pro které je známo členství ve skupině; tyto funkce pak mohou být použity na nové případy, které mají měření pro proměnné prediktoru, ale mají neznámé členství ve skupině.

*Poznámka:* Proměnná seskupení může mít více než dvě hodnoty. Kódy pro proměnnou seskupení musí být celá čísla, ale musíte uvést jejich minimální a maximální hodnoty. Případy s hodnotami mimo tyto meze jsou vyloučeny z analýzy.

**Příklad.** V průměru lidé v zemích mírného pásma spotřebovávají více kalorií denně než lidé v tropech a větší část obyvatel v mírných zónách jsou obyvatelé měst. Výzkumník chce spojit tyto informace do funkce, aby určil, jak dobře může jednotlivec diskriminovat mezi oběma skupinami zemí. Výzkumník se domnívá, že velikost populace a ekonomické informace mohou být také důležité. Discriminantní analýza umožňuje odhadnout koeficienty lineární diskriminačních funkcí, které vypadají jako pravá strana násobné lineární regresní rovnice. To znamená, že pomocí koeficientů  $a$ ,  $b$ ,  $c$  a  $d$ , funkce je:

$$D = a * climate + b * urban + c * population + d * gross domestic product per capita$$

Pokud jsou tyto proměnné užitečné pro rozlišování mezi dvěma klimatickými zónami, hodnoty  $D$  se budou lišit pro temperovanou a tropickou země. Pokud použijete metodu výběru krokové proměnné, můžete zjistit, že není třeba zahrnout do funkce všechny čtyři proměnné.

**Statistika.** Pro každou proměnnou: střední odchylky, jednorozměrné odchylky ANOVA. Pro každou analýzu:  $M$ , uvnitř skupin korelace matice, uvnitř-skupiny kovarianční matice, skupinová kovarianční matice, celková kovarianční matice. Pro každou kanonickou funkci discriminant: eigenvalue, percentage of variance, canonical correlation, Wilks'  $\lambda$ , chi-square. Pro každý krok: předchozí pravděpodobnosti, Fisher funkce koeficienty, nestandardizované funkční koeficienty, Wilks'  $\lambda$  pro každou kanonickou funkci.

Aspekty dat analýzy diskriminujících dat

**Data.** Proměnná seskupení musí mít omezený počet odlišených kategorií, kódovaných jako celá čísla. Nezávislé proměnné, které jsou nominální, musí být opačně nastaveny na fiktivní nebo kontrastní proměnné.

**Předpoklady.** Případy by měly být nezávislé. Proměnné prediktoru by měly mít normální rozdělení a v rámci skupiny by se matice kovariance měly rovnat napříč skupinami. Předpokládá se, že členství ve skupinách se vzájemně vylučuje (to znamená, že žádný případ nepatří do více než jedné skupiny) a hromadně vyčerpávající (to znamená, že všechny případy jsou členy skupiny). Postup je nejefektivnější, když členství ve skupině je skutečně kategorickou proměnnou; je-li členství ve skupině založeno na hodnotách souvislé proměnné (například vysoké IQ versus nízké IQ), zvažte použití lineární regrese k využití předností bohatších informací, které nabízí samotná spojitá proměnná.

Jak získat analýzu diskriminátoru

1. Z nabídky vyberte:

**Analyzovat > Klasifikovat > Diskriminant ...**

2. Vyberte proměnnou seskupení celočíselné hodnoty a klepněte na volbu **Definovat rozsah** a určete kategorie, které vás zajímají.

3. Vyberte nezávislý nebo prediktor, proměnné. (Pokud vaše proměnná seskupení nemá celočíselné hodnoty, automatická Recode v nabídce Transformace vytvoří proměnnou, která ano.)

4. Vyberte metodu pro zadání nezávislých proměnných.

- **Zadejte nezávislé složky.** Současně vstupuje do všech nezávislých proměnných, které splňují kritéria tolerance.

- **Použijte metodu krokové metody.** Používá nepostupnou analýzu k řízení záznamu proměnné a odebrání.

5. Volitelně můžete vybrat případy s proměnnou výběru.

## Definovat rozsah analýzy diskriminátoru

Určete minimální a maximální hodnotu proměnné seskupení pro analýzu. Případy s hodnotami mimo tento rozsah nejsou použity v analýze diskriminačních prostředků, ale jsou klasifikovány do jedné z existujících skupin na základě výsledků analýzy. Minimální a maximální hodnota musí být celá čísla.

## Výběr případů analýzy diskriminátoru

Chcete-li vybrat případy pro analýzu:

1. V dialogovém okně Analýza diskriminantů vyberte proměnnou výběru.
2. Chcete-li zadat celočíselnou hodnotu jako hodnotu výběru, klepněte na volbu **Hodnota**.

Pro odvození funkcí diskriminantů se používají pouze případy s uvedenou hodnotou pro proměnnou výběru. Výsledky statistiky a klasifikace jsou generovány pro vybrané i nevybrané případy. Tento proces poskytuje mechanismus pro klasifikaci nových případů na základě dříve existujících dat nebo pro rozdělení dat na oblasti na školení a testování podmnožin, aby bylo možné provést ověření platnosti na generovaném modelu.

## Statistika analýzy diskriminátoru

**Popisovače.** Dostupné možnosti jsou prostředky (včetně směrodatných odchylek), univariate ANOVAs a test  $M$  produktu Box.

- *Prostředky.* Zobrazí celkovou a skupinu středních hodnot a směrodatné odchylky pro nezávislé proměnné.
- *Jedinečná ANOVA.* Provádí jednosměrný test analýzy rozptylu pro rovnost skupin znamená pro každou nezávislou proměnnou.
- *M.* Test pro rovnost matic kovariance skupiny. U dostatečně velkých vzorků nevýznamná hodnota  $p$  znamená nedostatečné důkazy o tom, že se matice liší. Test je citlivý na odchylky od vícehodnotové normality.

**Koeficienty funkcí.** Dostupné možnosti jsou Fisher klasifikační koeficienty a nestandardizované koeficienty.

- *Fisher's.* Zobrazí Fisherovy funkce pro klasifikaci funkcí, které lze použít přímo pro klasifikaci. Pro každou skupinu se získá samostatná sada koeficientů funkcí klasifikace a případ je přiřazen ke skupině, pro kterou má největší diskriminační skóre (hodnota funkce klasifikace).
- *Nestandardizováno.* Zobrazí nestandardizované koeficienty funkce discriminant.

**Matrice.** Dostupné matice koeficientů pro nezávislé proměnné jsou korelační matice uvnitř skupin, uvnitř skupin kovarianční matice, kovarianční matice oddělených skupin a celkové kovarianční matice.

- *Korelace v rámci skupin.* Zobrazí směrovanou matici korelace uvnitř skupin, která je získána zprůměrováním oddělených matic kovariance pro všechny skupiny před výpočtem korelací.
- *covariance v rámci skupiny.* Zobrazí společnou matici kovariance v rámci skupiny, která se může lišit od celkové matice kovariance. Matice se získá zprůměrováním samostatného kovarianční matice pro všechny skupiny.
- *Rozdíl jednotlivých skupin.* Zobrazí samostatné kovarianční matice pro každou skupinu.
- *Celkem kovariance.* Zobrazí kovarianční matici ze všech případů, jako kdyby byly z jednoho vzorku.

## Metoda selektivní analýzy diskriminátoru

. Vyberte statistiku, která má být použita pro zadání nebo odebrání nových proměnných. Dostupné alternativy jsou Wilks'  $\lambda$ , nevysvětlitelné odchylky, Mahalanobis vzdálenost, nejmenší poměr  $F$  a Rao's  $V$ . Pomocí Rao's  $V$  můžete zadat minimální nárůst hodnoty proměnné  $V$  pro zadání proměnné.

- *Wilks'  $\lambda$* . Metoda výběru proměnných pro metodu discriminant discriminant, která vybírá proměnné pro vstup do rovnice na základě toho, jak výrazně nižší Wilks'  $\lambda$  je. V každém kroku je zadána proměnná, která minimalizuje celkový počet Wilks'  $\lambda$ .
- *Nevysvětlitelná odchylka*. V každém kroku je zadána proměnná minimalizující součet nevysvětlitelných odchylek mezi skupinami.
- *Mahalanobis distance*. Míra, v jaké míře se hodnoty jednotlivých případů na nezávislých proměnných liší od průměru všech případů. Velká Mahalanobis vzdálenost identifikuje případ, kdy mají extrémní hodnoty na jedné nebo více nezávislých proměnných.
- *Nejmenší poměr  $F$* . Metoda výběru proměnných v krocích po krocích založených na maximalizaci poměru  $F$  vypočteného z vzdálenosti Mahalanobis mezi skupinami.
- *Rao's  $V$* . Měřítka rozdílů mezi skupinami znamená. Nazývá se to i Lawley-Hodgingsovo trasování. V každém kroku je zadána proměnná, která maximalizuje zvýšení hodnoty Rao's  $V$ . Po výběru této volby zadejte minimální hodnotu, kterou musí proměnná muset zadat, aby mohla být zadána analýza.

**Kritéria.** Dostupné alternativy jsou **Použít hodnotu  $F$**  a **Použít pravděpodobnost  $F$** . Zadejte hodnoty pro zadání a odebrání proměnných.

- *Použít hodnotu  $F$* . Proměnná je zadána do modelu, je-li její hodnota  $F$  větší než hodnota Entry a je odebrána, pokud je hodnota  $F$  menší než hodnota odebrání. Položka musí být větší než odebrání, a obě hodnoty musí být kladné. Chcete-li do modelu zadat více proměnných, snižte hodnotu položky Entry. Chcete-li odebrat více proměnných z modelu, zvyšte hodnotu Odebrání.
- *Použít pravděpodobnost  $F$* . Proměnná je zadána do modelu, pokud je úroveň významnosti její hodnoty  $F$  menší než hodnota Entry a je odebrána, je-li úroveň významnosti větší než hodnota Odebrání. Položka musí být menší než odebrání, přičemž obě hodnoty musí být kladné. Chcete-li do modelu zadat více proměnných, zvyšte hodnotu Zadání. Chcete-li odebrat více proměnných z modelu, snižte hodnotu Odebrání.

**Obrazovka. Souhrn kroků** zobrazuje statistiky pro všechny proměnné po každém kroku;  **$F$  pro jednotlivé vzdálenosti po dvojicích** zobrazuje matici párů  $F$  po dvojicích pro každou dvojici skupin.

## Klasifikace analýzy diskriminátoru

**Předchozí pravděpodobnosti.** Tato volba určuje, zda jsou koeficienty klasifikace upraveny pro a priori znalosti o členství ve skupině.

- **Všechny skupiny jsou shodné.** Pro všechny skupiny se předpokládá stejné pravděpodobnosti, že to nemá žádný vliv na koeficienty.
- **Vypočítat ze velikosti skupiny.** Pozorované velikosti skupin ve vašem vzorku určují předchozí pravděpodobnosti členství ve skupině. Pokud například 50% z pozorování zahrnutých do analýzy spadá do první skupiny, 25% ve druhém a 25% ve třetím, jsou koeficienty klasifikace upraveny tak, aby se zvýšila pravděpodobnost členství v první skupině ve vztahu k ostatním dvěma.

**Obrazovka.** Dostupné volby zobrazení jsou výsledky s casewise, souhrnná tabulka a klasifikace opuštění-jedna.

- *Výsledky Casewise.* Pro každý případ jsou zobrazeny kódy skutečné skupiny, předpokládané skupiny, zadní pravděpodobnosti a skóre diskriminačních skóre.
- *Souhrnná tabulka.* Počet případů správně a nesprávně přiřazených každé ze skupin na základě analýzy diskriminantů. Někdy se nazývá "Zmatená matice."
- *Klasifikace typu Opustit-one-out.* Každý případ v analýze je klasifikován funkcemi odvozenými od všech případů jiných než tohoto případu. Je také známý jako "U-method."

**Nahradte chybějící hodnoty střední hodnotou.** Vyberte tuto volbu, chcete-li nahradit střední hodnotu nezávislé proměnné pouze během fáze klasifikace.

**Použijte matici Cvariance.** Můžete zvolit klasifikaci případů pomocí kovariance kovariance v rámci skupin nebo kovarianční matice samostatných skupin.

- *Uvnitř skupin.* Seskupení kovarianční matice v rámci skupin se používá ke klasifikaci případů.
- *Oddělené skupiny.* Matice kovariance oddělené skupiny se používají pro klasifikaci. Protože klasifikace je založena na diskriminačních funkcích (nezaložených na původních proměnných), tato volba není vždy ekvivalentní kvadratické diskriminaci.

**Ploty.** Dostupné volby výkresu jsou kombinované-skupiny, oddělené skupiny a teritoriální mapa.

- *Kombinované skupiny.* Vytvoří bodový graf all-groups prvních dvou hodnot funkce discriminant. Existuje-li pouze jedna funkce, bude zobrazen histogram.
- *Oddělené skupiny.* Vytvoří rozptylové grafy oddělené skupiny prvních dvou hodnot funkce discriminant. Existuje-li pouze jedna funkce, zobrazí se místo toho histogramy.
- *Teritoriální mapa.* Vykreslení hranic použitých ke klasifikaci případů do skupin založených na hodnotách funkcí. Čísla odpovídají skupinám, do kterých jsou případy klasifikovány. Střední hodnota pro každou skupinu je indikována hvězdičkou v rámci jeho hranic. Mapa se nezobrazí, pokud existuje pouze jedna funkce discriminant.

## Uložení analýzy diskriminátoru

Do svého aktivního datového souboru můžete přidat nové proměnné. Dostupné volby jsou predikované členství ve skupinách (jedna proměnná), skóre diskriminantů (jedna proměnná pro každou funkci discriminant v řešení) a pravděpodobnosti členství ve skupině poskytnuté skóre diskriminantů (jedna proměnná pro každou skupinu).

Informace o modelu můžete také exportovat do určeného souboru ve formátu XML. Tento modelový soubor můžete použít k použití informací modelu na jiné datové soubory pro účely hodnocení.

## Dodatečné funkce příkazu DISCRIMINANT

Jazyk syntaxe příkazu vám také umožňuje:

- Proveďte více discriminantních analýz (s jedním příkazem) a řídí pořadí, ve kterém se proměnné zadávají (s dílčím příkazem ANALYSIS).
- Uvedte předchozí pravděpodobnosti pro klasifikaci (s dílčím příkazem PRIORS).
- Zobrazit otočený vzor a matrice struktury (s dílčím příkazem ROTATE).
- Omezte počet extrahovaných diskriminačních funkcí (s dílčím příkazem FUNCTIONS).
- Omezte klasifikaci na případy, které jsou vybrány (nebo nevybrané) pro analýzu (pomocí dílčího příkazu SELECT).
- Číst a analyzovat korelační matici (s dílčím příkazem MATRIX).
- Zapište korelační matici pro pozdější analýzu (pomocí dílčího příkazu MATRIX).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Analýza faktoru

Analýza faktoru se pokouší identifikovat základní proměnné, neboli **faktory**, které vysvětlují strukturu korelací v sadě sledovaných proměnných. Analýza faktoru se často používá při snižování objemu dat k identifikaci malého počtu faktorů, které vysvětlují většinu rozptylu, který je pozorován v mnohem větším počtu proměnných manifestu. Analýza faktoru může být také použita ke generování hypotéz s ohledem na příčinné mechanismy nebo na proměnné obrazovky pro následnou analýzu (například k identifikaci kolineárnosti před provedením lineární regrese analýzy).

Postup analýzy faktoru nabízí vysoký stupeň flexibility:

- Je k dispozici sedm metod těžby faktorů.
- K dispozici je pět metod rotace, včetně přímé oblminace a promax pro neortogonální rotace.
- K dispozici jsou tři metody výpočtu skóre faktoru a skóre skóre lze uložit jako proměnné pro další analýzu.

**Příklad.** Jaké základní postoje vedou lidi k tomu, aby reagovali na otázky z politického průzkumu tak, jak to dělají? Zkoumání korelací mezi položkami průzkumu odhalí významné překrytí mezi různými podskupinami položek -- otázky o daních mají tendenci korelovat navzájem, otázky o vojenských otázkách vzájemně korelují, a tak dále. Pomocí analýzy faktoru můžete vyšetřit počet základních faktorů a v mnoha případech identifikovat, jaké faktory představují koncepčně. Kromě toho můžete vypočítat skóre faktoru pro každého respondenta, který lze následně použít v následných analýzách. Můžete například sestavit model pro logistickou regresi, abyste předpovídali chování při hlasování na základě skóre faktoru.

**Statistika.** Pro každou proměnnou: počet platných případů, střední hodnoty a směrodatné odchylky. Pro každou analýzu faktoru: korelační matice proměnných, včetně úrovní významnosti, determinantu a inverzního; reprodukované korelační matice, včetně antiimage; počátečního řešení (communities, eigenvalues, and percentage of variance explained); Kaiser-Meyer-Olkin variance of sampling fatility and Bartlett's test of sphericity; unrotated solution, including factor loadings, communities, and eigenvalues; and rotated solution, including rotated pattern matrix and transformation matrix. Pro šikmé rotaci: otočený vzor a matrice struktury; koeficient faktoru a matice kovariance faktoru. Ploty: křížící nákres eligních hodnot a zatěžovací graf prvních dvou nebo tří faktorů.

Aspekty dat analýzy faktoru

**Data.** Proměnné by měly být kvantitativní na úrovni *interval* nebo *poměr*. Kategorická data (např. náboženství nebo země původu) nejsou vhodná pro analýzu faktoru. Údaje, pro které mohou být Pearsonové korelační koeficienty citlivě vypočteny, by měly být vhodné pro analýzu faktoru.

**Předpoklady.** Data by měla mít bivariate normální distribuci pro každou dvojici proměnných a pozorování by měla být nezávislá. Model analýzy faktoru uvádí, že proměnné jsou určovány společnými faktory (faktory odhadovanými modelem) a jedinečnými faktory (které se mezi pozorovanými proměnnými nepřekrývají); vypočtené odhady vycházejí z předpokladu, že všechny jedinečné faktory jsou vzájemně nekorelované a se společnými faktory.

Jak získat analýzu faktoru

1. Z nabídky vyberte:

**Analyzovat > Zmenšení dimenze > Faktor ...**

2. Vyberte proměnné pro analýzu faktoru.

## Výběr případů analýzy faktoru

Chcete-li vybrat případy pro analýzu:

1. Vyberte proměnnou výběru.
2. Chcete-li zadat celočíselnou hodnotu jako hodnotu výběru, klepněte na volbu **Hodnota**.

V analýze faktoru se používají pouze případy s touto hodnotou pro danou proměnnou výběru.

## Deskriptory analýzy faktoru

**Statistika. Univariate descriptives** zahrnuje střední, směrodatnou odchylku a počet platných případů pro každou proměnnou. **Počáteční řešení** zobrazuje počáteční komunity, hodnoty eigenvalues a procento vysvětlených odchylek.

**Korelační matice.** Dostupné možnosti jsou koeficienty, úrovně významnosti, determinant, KMO a Bartlettův test kulovitosti, inverzní, reprodukované a anti-image.

- *KMO and Bartlett's Test of Sphericity.* Kaiser-Meyer-Olkin měřítko odběru vzorků adekvátnosti, zda jsou částečné korelace mezi proměnnými jsou malé. Bartlett test sférické zkoušky, zda srovnávací matrice je matrice identity, která by naznačovala, že model faktoru je nevhodný.

- *Znovu vytvořeno.* Odhadnutá korelační matice z řešení faktoru. Zobrazeny jsou také rezidua (rozdíl mezi odhadovanými a pozorovanými korelacemi).
- *Anti-image.* Antiimage korelační matice obsahuje negativy dílčích korelačních koeficientů a matice kovariance anti-image obsahuje negativy dílčích kovariance. V dobrém modelu, většina z off-diagonálních prvků budou malé. Měřitko odběru vzorků odpovídající proměnné se zobrazí na diagonále antiimage matice korelace.

## Extrakce analýzy faktoru

. Umožňuje vám uvést metodu extrakce faktoru. Dostupné metody jsou hlavní komponenty, nevážené nejmenších čtverců, generalizované nejmenších čtverců, maximální pravděpodobnosti, faktoring hlavní osy, faktoring alfa a faktoring obrazu.

- *Analýza hlavních komponent.* Metoda extrakce faktoru použitá k vytvoření nekorelovaných lineárních kombinací pozorovaných proměnných. První komponenta má maximální rozptyl. Následné komponenty vysvětlují postupně menší části rozptylu a vzájemně se vzájemně nekorelují. Analýza hlavních komponent se používá k získání počátečního řešení faktoru. Lze ji použít, je-li korelační matice singulární.
- *Metoda neváženého nejmenších čtverců.* Metoda extrakce faktoru, která minimalizuje součet druhých mocných rozdílů mezi pozorovanými a reprodukovánými maticemi korelace (ignorování diagonálách).
- *Generalizovaná metoda nejmenších čtverců.* Metoda extrakce faktoru, která minimalizuje součet druhých mocnin rozdílů mezi pozorovanými a reprodukovánými korelační matice. Korelace jsou váženy inverzní k jejich jedinečnosti, takže proměnné s vysokou jedinečností jsou dány menší váhu než ty s nízkou jedinečností.
- *Maximum-metoda pravděpodobnosti.* Metoda extrakce faktoru, která produkuje odhady parametrů, které s největší pravděpodobností vyrobí pozorovanou srovnávací matici, pokud je vzorek od vícerozměrného normálního rozdělení. Korelace jsou váženy inverzní k jedinečnosti proměnných, a iterativní algoritmus je použit.
- *Factoring hlavní osy.* Metoda extrahování faktorů z původní korelační matice s násobnými korelačními koeficienty umístěnými do diagonály jako počáteční odhady komunitních prvků. Tyto naložení faktoru se používají k odhadu nových komunitami, které nahradí původní odhady Společenství ve diagonálu. Iterace pokračují, dokud se změny v komunitách z jedné iterace na další nesplňují kritérium konvergence pro extrakci.
- *Alfa Factoring.* Metoda extrakce faktoru, která považuje proměnné v analýze za vzorek z vesmíru potenciálních proměnných. Tato metoda maximalizuje spolehlivost alfa u faktorů.
- *Factoring obrázků.* Metoda extrakčního faktoru vyvinutá společností Guttman a založená na obrazové teorii. Společná část proměnné, která se nazývá dílčí obraz, je definována jako její lineární regrese na zbývající proměnné, spíše než na funkci hypotetických faktorů.

**Analýza.** Umožňuje vám uvést buď korelační matici, nebo kovarianční matici.

- **Korelační matice.** Užitečné, pokud jsou proměnné ve vaší analýze měřeny na různých měřítcích.
- **Matice Covariance.** Užitečné, když chcete použít analýzu faktoru na více skupin s různými odchylkami pro každou proměnnou.

**Extrahovat.** Můžete buď zachovat všechny faktory, jejichž vlastní hodnoty překračují zadanou hodnotu, nebo si můžete zachovat specifický počet faktorů.

**Obrazovka.** Umožňuje vám vyžádat řešení neotočených faktorů a výkřik na eligenhodnotách.

- *Řešení neotočeného faktoru.* Zobrazí zatížení neotočeného faktoru (matrice vzoru faktoru), komuniti a hodnoty eigenvalues pro řešení součinitele.
- *Screen plot.* Vykreslení rozdílu, který je přidružen ke každému faktoru. Tento graf se používá k určení, kolik faktorů by mělo být uchováno. Typicky je zápletka znázorněna zřetelnou přestávkou mezi strmou svahem velkých faktorů a postupnými konci zbytku (suti).

**Maximum iterací pro cíl konvergence.** Umožňuje vám uvést maximální počet kroků, které může algoritmus provést k odhadu řešení.

## Rotace analýzy faktoru

. Umožňuje vybrat metodu rotace faktoru a to, zda použít normalizaci Kaiseru. Dostupné metody jsou varimax, direct oblmin, quartimax, equamax, nebo promax.

- *Metoda Varimax.* Metoda rotace ortogonální rotace, která minimalizuje počet proměnných, které mají v každém faktoru vysoké zatížení. Tato metoda zjednodušuje interpretaci faktorů.
- *Metoda Direct Oblimin Method.* Střídání metody šikmé (neortogonální). Je-li rozdílová hodnota 0 (výchozí hodnota), řešení jsou nejvíce šikmo. Vzhledem k tomu, že delta se stává více negativní, se faktory stávají méně šikmé. Chcete-li přepsat výchozí hodnotu delta 0, zadejte číslo menší nebo rovné hodnotě 0.8.
- *Metoda kvartimax.* Metoda rotace, která minimalizuje počet faktorů potřebných k vysvětlení každé proměnné. Tato metoda zjednodušuje interpretaci sledovaných proměnných.
- *Metoda Equamax.* Metoda rotace, která je kombinací metody varimax, která zjednodušuje faktory, a metoda quartimax, která zjednoduší proměnné. Počet proměnných, které se načítají velmi na základě faktoru a počtu faktorů potřebných k vysvětlení proměnné jsou minimalizovány.
- *Rotace promax.* šikmé otočení, které umožňuje korelaci faktorů. Tato rotace může být vypočítána rychleji než přímá oblité rotace, takže je užitečné pro velké datové sady.
- *Použití normalizaci Kaiser.* Ve výchozím nastavení vám umožňuje aplikovat normalizaci Kaiser, je-li zadána rotace.

**Obrazovka.** Umožňuje vám zahrnout výstup na otočené řešení, stejně jako zatěžovací zkušné plochy pro první dva nebo tři faktory.

- *Otočené řešení.* Je třeba vybrat metodu rotace, abyste získali otočené řešení. U ortogonálních rotací se zobrazí matice a matice transformace faktoru. Pro šikmé rotace se zobrazí vzor, struktura a korelační matice faktorů.
- *Zákresový graf faktoru.* Zakreslení faktoru rozměrového faktoru hodnocení prvních tří faktorů. U dvoufaktorového řešení se zobrazí dvourozměrný graf. Vykreslení se nezobrazí, pokud je extrahován pouze jeden faktor. Je-li požadována rotace, zobrazí se rotované řešení rotovaných aplikací.

**Maximum iterací pro cíl konvergence.** Umožňuje vám uvést maximální počet kroků, které může algoritmus provést k provedení rotace.

## Skóre analýzy faktoru

**Uložit jako proměnné.** Vytvoří jednu novou proměnnou pro každý faktor v konečném řešení.

. Alternativní metody pro výpočet skóre faktoru jsou regrese, Bartlett a Anderson-Rubin.

- *Regresní metoda.* Metoda pro odhad koeficientů faktoru faktoru. Vyprodukovaná skóre mají střední hodnotu 0 a odchylku rovnající se druhé mocnině korelace mezi odhadovaným skóre faktoru a skutečnými hodnotami faktoru. Skóre lze korelovat i tehdy, jsou-li faktory ortogonální.
- *Bartlett Scores.* Metoda odhadu koeficientů faktoru faktoru. Vytvořená skóre mají střední hodnotu 0. Součet čtverců jedinečných faktorů v rozsahu proměnných je minimalizován.
- *Metoda Anderson-Rubin.* Metoda odhadu koeficientů faktoru faktoru; úprava metody Bartlett, která zajišťuje ortogonátu odhadovaných faktorů. Vyprodukovaná skóre mají střední hodnotu 0, mají směrodatnou odchylku 1 a jsou nekorelovaná.

**Zobrazení matice koeficientu skóre faktoru.** Zobrazí koeficienty, podle kterých se proměnné násobí k získání skóre faktoru. Zobrazuje také korelace mezi skóre faktoru.

## Volby analýzy faktoru

**Chybějící hodnoty.** Umožňuje vám uvést, jak se zachází s chybějícími hodnotami. Dostupné volby jsou vyloučení případů *litwise*, vyloučení případů *pairwise*, nebo nahrazení se střední hodnotou.

**Coefficient Display Format.** Umožňuje vám řídit aspekty výstupní matice. Třídí koeficienty podle velikosti a potlačit koeficienty s absolutními hodnotami, které jsou menší než uvedená hodnota.

## Další funkce příkazu FACTOR

Jazyk syntaxe příkazu vám také umožňuje:

- Uvedte kritéria konvergence pro iteraci během extrakce a rotace.
- Určete jednotlivé zakreslené grafy.
- Určete, kolik skóre faktoru se má uložit.
- Uvedte diagonální hodnoty pro metodu faktoringu hlavní osy.
- Psát korelační matice nebo faktory-načtení matic na disk pro pozdější analýzu.
- Čtete a analyzujete matice korelace nebo matrice faktoru-načtení.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Výběr procedury pro klastrování

---

Analýzy klastru lze provádět pomocí procedury TwoStep, Hierarchical nebo K-Means Cluster Analysis. Každá procedura používá odlišný algoritmus pro vytváření klastrů a každý z nich má volby, které nejsou dostupné v ostatních.

**Nástroj TwoStep Cluster Analysis.** Pro mnoho aplikací bude procedura TwoStep Cluster Analysis zvolenou metodou volby. Poskytuje následující jedinečné funkce:

- Automatický výběr nejlepšího počtu klastrů, kromě opatření pro výběr mezi modely klastru.
- Schopnost vytvářet modely klastru zároveň založené na kategoriálních a spojitých proměnných.
- Schopnost uložit model klastru do externího souboru XML a poté tento soubor přečíst a aktualizovat model klastru s použitím novějších dat.

Procedura TwoStep Cluster Analysis kromě toho může analyzovat velké datové soubory.

**Hierarchická analýza klastru.** Procedura Hierarchical Cluster Analysis je omezena na menší datové soubory (stovky objektů, které mají být klastrované), ale má následující jedinečné funkce:

- Schopnost klastrovaných případů nebo proměnných.
- Schopnost vypočítat rozsah možných řešení a uložit členství v klastru pro každé z těchto řešení.
- Několik metod vytváření klastrů, transformace proměnných a měření nepodobnosti mezi klastry.

Pokud jsou všechny proměnné stejného typu, může procedura Hierarchická analýza klastru analyzovat interval (souvislý), počet nebo binární proměnné.

**Analýza K-Means Cluster.** Procedura K-Means Cluster Analysis je omezena na souvislá data a vyžaduje, abyste uvedli počet klastrů předem, ale má následující jedinečné funkce:

- Schopnost ukládat vzdálenosti od center klastrů pro každý objekt.
- Schopnost číst počáteční centra klastru a uložit finální klastrová centra do externího souboru IBM SPSS Statistics.

Procedura K-Means Cluster Analysis kromě toho může analyzovat velké datové soubory.

## Analýza klastru TwoStep

---

Procedura TwoStep Cluster Analysis je průzkumný nástroj určený k odhalení přirozených seskupení (nebo klastrů) v rámci datové sady, která by jinak nebyla zřejmá. Algoritmus použitý touto procedurou má několik žádoucích funkcí, které jej odlišují od tradičních technik klastrování:

- **Práce s kategoriálními a spojitými proměnnými.** Předpokládá se, že proměnné jsou nezávislé, společné mnohonomické-normální rozdělení může být umístěn na kategoriální a spojitě proměnné.
- **Automatický výběr počtu klastrů.** Porovnáváním hodnot kritéria modelového výběru mezi různými řešeními klastrů může procedura automaticky určovat optimální počet klastrů.



- **Škálovatelnost.** Při konstruování stromu s funkcemi klastru (CF), který shrnuje záznamy, umožňuje algoritmus TwoStep analyzovat velké datové soubory.

**Příklad.** Maloobchodní a spotřební společnosti běžně používají metody klastrování k údajům, které popisují kupní návyky svých zákazníků, pohlaví, věk, úroveň příjmů atd. Tyto společnosti přizpůsobují své strategie marketingu a vývoje produktů pro každou skupinu zákazníků za účelem zvýšení prodeje a budování loajality zákazníků.

**Měřítko vzdálenosti.** Tento výběr určuje, jak se vypočítá podobnost mezi dvěma klastry.

- **Log-Pravděpodobnost.** Ukazatel pravděpodobnosti umístí distribuční rozdělení na proměnné. Předpokládá se, že jsou normálně distribuovány souvislé proměnné, zatímco kategorické proměnné jsou považovány za mnohokomické. Předpokládá se, že všechny proměnné jsou nezávislé.
- **euklidovský.** Euklidovský ukazatel je "přímá čára" vzdálenost mezi dvěma klastry. Lze ji použít pouze tehdy, jsou-li všechny proměnné souvislé.

**Počet klastrů.** Tento výběr umožňuje určit, jak má být počet klastrů určen.

- **Určete automaticky.** Procedura automaticky určí "nejlepší" počet klastrů s použitím kritéria uvedeného ve skupině Kritérium klastrování. Volitelně můžete zadat kladné celé číslo určující maximální počet klastrů, které má procedura považovat za vhodné.
- **Určete pevnou hodnotu.** Umožňuje vám opravit počet klastrů v řešení. Zadejte kladné celé číslo.

**Počet spojitých proměnných.** Tato skupina poskytuje souhrn specifikací souvislé proměnné standardizace provedených v dialogovém okně Volby. Další informace naleznete v tématu "[Volby analýzy klastru TwoStep](#)" na stránce 178 .

**Kritérium klastrování.** Tato volba určuje, jak bude určovat počet klastrů algoritmus automatického klastrování. Může být uvedeno buď Bayesovo informační kritérium (BIC), nebo informační kritérium Akaike (AIC).

Aspekty dat analýzy klastru TwoStep

**Data.** Tato procedura pracuje se spojitými a kategoriálními proměnnými. Případy představují objekty, které se mají dělit do klastrů, a proměnné představují atributy, na kterých je klastrování založeno.

**Pořadí případů.** Všimněte si, že strom funkcí klastru a konečné řešení může záviset na pořadí případů. Chcete-li minimalizovat efekty objednávky, náhodně objednejte případy. Možná budete chtít získat několik různých řešení s případy seřazenými v různých náhodných příkazech k ověření stability daného řešení. V situacích, kdy je to obtížné kvůli extrémně velkým velikostem souborů, může být nahrazen více běhů se vzorkem případů seřazených v různých náhodných příkazech.

**Předpoklady.** Ukazatel pravděpodobného vzdálenosti předpokládá, že proměnné v modelu klastru jsou nezávislé. Dále, každá spojitá proměnná se předpokládá, že má normální (Gaussovu) distribuci, a každá kategorická proměnná se předpokládá, že má polynomiální distribuci. Empirické vnitřní testování ukazuje, že postup je poměrně spolehlivý na porušování předpokladů nezávislosti a distribuční předpoklady, ale měli byste si být vědomi toho, jak dobře jsou tyto předpoklady splněny.

Použijte proceduru [Bivariate Correlations](#) k testování nezávislosti dvou spojitých proměnných. Proceduru [Kontingenční tabulky](#) můžete použít k testování nezávislosti dvou kategoriálních proměnných. Proceduru [Způsoby](#) použijte k testování nezávislosti mezi souvislou proměnnou a kategorickou proměnnou. Použijte proceduru [Prozkoumat](#) k testování normality spojitě proměnné. Pomocí procedury [Chi-Square Test](#) otestujte, zda má kategorická proměnná zadanou mnoho-mickou distribuci.

Chcete-li získat analýzu klastru TwoStep

1. Z nabídky vyberte:

**Analyzovat > Klasifikovat > Klastr TwoStep ...**

2. Vyberte jednu nebo více kategoriálních nebo spojitých proměnných.

Volitelně můžete:

- Upravte kritéria, podle kterých jsou vytvořeny klastry.

- Vyberte nastavení pro zpracování hluku, přidělení paměti, standardizaci proměnných a vstup modelu klastru.
- Vyžádat výstup prohlížeče modelu požadavku.
- Uložte výsledky modelu do pracovního souboru nebo do externího souboru XML.

## Volby analýzy klastru TwoStep

**Odlehlá léčba.** Tato skupina vám umožňuje zacházet s odlišy zvláště během klastrování, pokud se zaplní strom funkcí klastru (CF). Strom CF je plný, pokud nemůže přijmout žádné další případy v listových uzlech a žádný koncový uzel nelze rozdělit.

- Pokud vyberete zpracování hluku a dojde k naplnění stromu CF, bude po uložení případů v řídkých listech do listu "noise" automaticky znovu vytvořen. List je považován za řídký, pokud obsahuje méně než určený procentní podíl případů maximální velikosti nejnižších prvků. Po opětovném zobrazení stromu budou odlehlé hodnoty umístěny ve stromu prostředku CF, je-li to možné. Pokud tomu tak není, odlehlé hodnoty budou vyřazeny.
- Pokud nevyberete zpracování hluku a strom CF se zaplní, bude znovu použit s použitím větší prahové hodnoty pro změnu vzdálenosti. Po konečném klastrování jsou hodnoty, které nemohou být přiřazeny k klastru, označeny jako odlehlé hodnoty. Odlehlější klastr má identifikační číslo -1 a není zahrnutý v počtu klastrů.

**Přidělení paměti.** Tato skupina vám umožňuje uvést maximální množství paměti v megabajtech (MB), které by měl algoritmus klastru použít. Pokud procedura překročí tuto maximální hodnotu, použije disk k uložení informací, které se nevejdou do paměti. Zadejte číslo větší nebo rovné 4.

- Obraťte se na administrátora systému se žádostí o největší hodnotu, kterou lze zadat ve vašem systému.
- Je-li tato hodnota příliš nízká, může algoritmus selhat při hledání správného nebo zadaného počtu klastrů.

**Standardizace proměnných.** Algoritmus klastrování pracuje se standardizovanými spojitými proměnnými. Jakékoliv souvislé proměnné, které nejsou standardizovány, by měly být ponechány jako proměnné v seznamu To be Standardized. Chcete-li ušetřit určitou dobu a výpočetní výkon, můžete vybrat všechny souvislé proměnné, které jste již normalizovali jako proměnné v seznamu Předpokládány standardizovaný seznam.

Rozšířené volby

**Kritéria ladění stromu CF.** Následující nastavení klastrového algoritmu platí specificky pro strom vlastností klastru (CF) a měla by být změněna s opatrností:

- **Prahová hodnota počáteční změny vzdálenosti.** Jedná se o počáteční prahovou hodnotu použitou k růstu stromu CF. Pokud vložení daného případu do listu ve stromu prostředku CF bude mít netěsnost menší než prahová hodnota, list se nerozdělí. Pokud napětí překročí prahovou hodnotu, je list rozdělen.
- **Maximální počet větví (na koncový uzel).** Maximální počet podřízených uzlů, které může mít koncový uzel.
- **Maximální hloubka stromu.** Maximální počet úrovní, které může mít strom prostředku CF.
- **Maximální počet možných uzlů.** To označuje maximální počet uzlů ve stromu prostředku CF, které mohou být potenciálně generovány procedurou, založené na funkci  $(b^{d+1} - 1) / (b - 1)$ , kde  $b$  je maximální počet větví a 44% je maximální hloubka stromu. Buďte si vědomi toho, že příliš velký strom CF může být odlišem na systémové prostředky a může nepříznivě ovlivnit výkon procedury. Minimálně každý uzel vyžaduje 16 bajtů.

**Aktualizace modelu klastru.** Tato skupina vám umožňuje importovat a aktualizovat model klastru generovaný v předchozí analýze. Vstupní soubor obsahuje strom CF ve formátu XML. Model se poté aktualizuje s daty v aktivním souboru. Názvy proměnných v hlavním dialogovém okně musíte vybrat ve stejném pořadí, v jakém byly uvedeny v předchozí analýze. Soubor XML zůstane nezměněný, pokud výslovně nezapisujete nové informace o modelu do stejného názvu souboru. Další informace naleznete v tématu [“Výstup analýzy klastru TwoStep”](#) na stránce 179 .

Je-li zadána aktualizace modelu klastru, použijí se volby týkající se generování stromu CF, které byly zadány pro původní model. Přesněji řečeno jsou použita kritéria vzdálenosti, zpracování šumu, přidělení paměti nebo nastavení kritérií vyladění ladění pro daný model, a všechna nastavení těchto voleb v dialogových oknech jsou ignorována.

*Poznámka:* Při provádění aktualizace modelu klastru procedura předpokládá, že žádný z vybraných případů v aktivní datové sadě nebyl použit k vytvoření původního modelu klastru. Procedura také předpokládá, že se případy použité v aktualizaci modelu pocházejí ze stejné populace jako případy použité k vytvoření původního modelu; to znamená, že se předpokládá, že se v obou souborech případů shodují proměnné a odchylky spojitých proměnných a úrovně kategoriálních proměnných. Pokud vaše "nové" a "staré" soubory případů pocházejí z heterogenních populací, měli byste spustit proceduru analýzy klastru TwoStep na kombinovaných sadách případů pro nejlepší výsledky.

## Výstup analýzy klastru TwoStep

**Výstup.** Tato skupina poskytuje volby pro zobrazení výsledků klastrování.

- **Kontingenční tabulky.** Výsledky se zobrazují v kontingenčních tabulkách.
- **Grafy a tabulky v prohlížeči modelů.** Výsledky se zobrazí v prohlížeči modelů.
- **Pole vyhodnocení.** Vypočítává data klastru pro proměnné, které nebyly použity při vytváření klastru. Hodnotící pole lze zobrazit spolu se vstupními funkcemi v prohlížeči modelů tím, že je vyberete v dílčím dialogovém okně Zobrazit. Pole s chybějícími hodnotami jsou ignorována.

**Pracovní datový soubor.** Tato skupina vám umožňuje uložit proměnné do aktivní datové sady.

- **Vytvořte proměnnou členství klastru.** Tato proměnná obsahuje identifikační číslo klastru pro každý případ. Název této proměnné je *tsc\_n*, kde *n* je kladné celé číslo udávající pořadové číslo aktivní operace uložení datové sady dokončené touto procedurou v dané relaci.

**Soubory XML.** Konečný model klastru a strom CF jsou dva typy výstupních souborů, které lze exportovat ve formátu XML.

- **Exportovat finální model.** Konečný model klastru se vyexportuje do uvedeného souboru ve formátu XML (PMML). Tento modelový soubor můžete použít k použití informací modelu na jiné datové soubory pro účely hodnocení.
- **Exportovat strom CF.** Tato volba umožňuje uložit aktuální stav stromu klastru a aktualizovat jej později pomocí novějších dat.

## Prohlížeč klastrů

Modely klastrů se obvykle používají k vyhledání skupin (nebo klastrů) podobných záznamů na základě zkoumaných proměnných, kde podobnost mezi členy stejné skupiny je vysoká a podobnost mezi členy různých skupin je nízká. Výsledky lze použít k identifikaci přidružení, která by jinak nebyla zřejmá. Např. prostřednictvím analýzy preferencí zákazníků, úrovně příjmů a nákupních návyků může být možné identifikovat typy zákazníků, kteří budou s vyšší pravděpodobností odpovídat na konkrétní marketingovou kampaň.

Při interpretaci výsledků na obrazovce klastru se používají dva přístupy:

- Zkontrolujte klastry, abyste určili charakteristiky, které jsou jedinečné pro daný klastr. *Jeden klastr obsahuje všechny dlužníky s vysokým příjmem? Obsahuje tento klastr více záznamů než ostatní?*
- Prověřte pole napříč klastry, abyste určili, jak jsou hodnoty rozdělovány mezi klastry. *Je v klastru jedna úroveň vzdělání určující členství v klastru? Rozlišuje vysoké skóre kreditu mezi členstvím v jednom klastru nebo jiným?*

Díky hlavním pohledům a různým propojeným pohledům v prohlížeči klastru můžete získat poznatky, které vám pomohou odpovědět na tyto otázky.

Chcete-li zobrazit informace o modelu klastru, aktivujte (dvakrát klepněte) objekt Prohlížeč modelů v prohlížeči.

## Prohlížeč klastrů

Prohlížeč klastru se skládá ze dvou panelů, z hlavního pohledu vlevo a z propojeného a pomocného pohledu vpravo. Existují dvě hlavní zobrazení:

- Souhrn modelu (předvolba). Další informace naleznete v tématu [“Souhrnné zobrazení modelu”](#) na stránce 180 .
- Klastry. Další informace naleznete v tématu [“Zobrazení klastrů”](#) na stránce 180 .

Existují čtyři propojené/pomocné pohledy:

- Důležitost prediktoru. Další informace naleznete v tématu [“Zobrazení důležitosti predikátu klastru”](#) na stránce 182 .
- Velikosti klastru (výchozí). Další informace naleznete v tématu [“Zobrazení velikostí klastru”](#) na stránce 182 .
- Distribuce buněk. Další informace naleznete v tématu [“Zobrazení distribuce buněk”](#) na stránce 182 .
- Porovnání klastru. Další informace naleznete v tématu [“Zobrazení porovnání klastru”](#) na stránce 182 .

### Souhrnné zobrazení modelu

Pohled Souhrn modelu zobrazuje snímek nebo souhrn modelu klastru, včetně míry Silhouette klastru soudržnosti a oddělení, které je stínované, aby označovalo špatné, spravedlivé nebo dobré výsledky. Tento snímek vám umožňuje rychle zkontrolovat, zda je kvalita špatná, a v tom případě se můžete rozhodnout vrátit se do modelovacího uzlu, abyste změnili nastavení modelu klastru, abyste vytvořili lepší výsledek.

Výsledky chudých, spravedlivých a dobrých výsledků jsou založeny na práci Kaufman a Rousseeuw (1990), pokud jde o výklad struktury clusteru. V pohledu Souhrn modelů se dobrým výsledkem rovná údaje, které odráží rating Kaufman a Rousseeuw jako přiměřené nebo silné důkazy o struktuře seskupení, spravedlivý odráží jejich rating slabých důkazů a slabé odráží jejich hodnocení bez významných důkazů.

Hodnota siluetu měří průměr za všechny záznamy  $(B - A)/\max(A, B)$ , kde A je vzdálenost záznamu ke svému středu klastru a B je vzdálenost záznamu k nejbližšímu středisku klastrů, ke kterému nepatří. Silueta s koeficientem 1 by znamenala, že všechny případy se nacházejí přímo na svých klastrových centrech. Hodnota - 1 znamená, že všechny případy se nacházejí na serverovém centru nějakého jiného klastru. Hodnota 0 znamená, že v průměru jsou případy ve stejné vzdálenosti mezi svým vlastním klastrem a nejbližším jiným klastrem.

Souhrn obsahuje tabulku, která obsahuje následující informace:

- **Algoritmus.** Použitý kompresní algoritmus, například "TwoStep".
- **Vstupní funkce.** Počet polí, také známý jako **vstupy** nebo **prediktory**.
- **Klastry.** Počet klastrů v řešení.

### Zobrazení klastrů

Pohled Klastry obsahuje mřížku s klastrem, která obsahuje názvy klastrů, velikosti a profily pro každý klaster.

Sloupce v mřížce obsahují následující informace:

- **Klaster.** Číslo klastrů vytvořená algoritmem.
- **Popisek.** Libovolné štítky použité na každém klastru (toto je při výchozím nastavení prázdné). Poklepejte na buňku, chcete-li zadat popisek, který popisuje obsah klastru, například "Luxusní kupující auta".
- **Popis.** Libovolný popis obsahu klastru (toto je standardně prázdné). Poklepejte na buňku, chcete-li zadat popis klastru, například "55 + let, profesionálové, vydělávat více než 100 000 dolarů".
- **Velikost.** Velikost každého klastru jako procentní část z celkového vzorku klastru. Každá velikost buňky v mřížce zobrazuje svislý pruh, který zobrazuje procentní část velikosti v rámci klastru, procentní část velikosti v číselném formátu a počty případů klastru.

- **Vlastnosti.** Jednotlivé vstupy nebo prediktory, seřazené podle celkového významu při výchozím nastavení. Pokud mají jakékoli sloupce stejné velikosti, jsou zobrazeny ve vzestupném pořadí řazením čísel klastrů.

Celkový význam funkce je označen barvou stínování pozadí buňky; nejdůležitější funkce je nejtmaší; nejméně důležitá funkce je nestínovaná. Nad tabulkou je uvedena důležitost, která je připojena ke každé barvě buňky funkce.

Když podržíte ukazatel myši nad buňkou, zobrazí se celé jméno/popisek funkce a hodnota důležitosti pro danou buňku. Další informace mohou být zobrazeny v závislosti na typu pohledu a typu funkce. V pohledu *Klastry Center* to zahrnuje statistiku buňky a hodnotu buňky; například: "Střed: 4,32". V případě kategoriálních funkcí buňka zobrazuje název nejčastější (modální) kategorie a její procento.

V pohledu *Klastry* můžete vybrat různé způsoby, jak zobrazit informace o klastru:

- **Transpozice klastrů a funkcí.** Další informace naleznete v tématu "[Transpozice klastrů a funkcí](#)" na stránce 181 .
- **Funkce řazení.** Další informace naleznete v tématu "[Řadit funkce](#)" na stránce 181 .
- **Třídění klastrů.** Další informace naleznete v tématu "[Seřadit klastry](#)" na stránce 181 .
- **Vyberte obsah buňky.** Další informace naleznete v tématu "[Obsah buňky](#)" na stránce 181 .

### *Transpozice klastrů a funkcí*

Ve výchozím nastavení se klastry zobrazují jako sloupce a funkce se zobrazují jako řádky. Chcete-li tuto obrazovku vrátit, klepněte na tlačítko **Transponovat klastry a funkce** na levé straně tlačítek **Řadit funkce podle** . Můžete to například chtít udělat, když se zobrazí mnoho klastrů, chcete-li snížit množství horizontálního rolování potřebného pro zobrazení dat.

### *Řadit funkce*

Tlačítka **Řadit funkce podle** vám umožňují vybrat způsob zobrazení buněk funkcí:

- **Celková důležitost.** Jedná se o výchozí pořadí řazení. Funkce jsou seřazeny v sestupném pořadí podle celkového významu a pořadí řazení je stejné napříč klastry. Mají-li některé funkce hodnoty svázané důležitosti, jsou vázané funkce uvedeny ve vzestupném pořadí podle názvů funkcí.
- **Důležitost v rámci klastru.** Funkce jsou seřazeny s ohledem na jejich důležitost pro každý klastr. Mají-li některé funkce hodnoty svázané důležitosti, jsou vázané funkce uvedeny ve vzestupném pořadí podle názvů funkcí. Je-li tato volba vybrána, pořadí řazení se obvykle pohybuje mezi klastry.
- **Název.** Funkce jsou seřazeny podle názvu v abecedním pořadí.
- **Pořadí dat.** Funkce jsou seřazeny podle jejich pořadí v datové sadě.

### *Seřadit klastry*

Standardně jsou klastry seřazeny v sestupném pořadí podle velikosti. Tlačítka **Seřadit klastry podle** vám umožňují je řadit podle názvu v abecedním pořadí, nebo, pokud jste vytvořili jedinečné popisky, místo toho v alfanumerickém pořadí.

Vlastnosti, které mají stejný popisek, jsou řazeny podle názvu klastru. Pokud jsou klastry seřazeny podle popisku a upravíte-li popisek klastru, pořadí řazení se automaticky aktualizuje.

### *Obsah buňky*

Tlačítka **Cells** umožňují změnit zobrazení obsahu buňky pro funkce a pole vyhodnocení.

- **Klastrová centra.** Při výchozím nastavení buňky zobrazují názvy a popisky funkcí a centrální tendenci pro každou kombinaci klastro/funkce. Střední hodnota je zobrazena pro souvislá pole a režim (nejčastěji se vyskytující kategorie) s procentní částí kategorií pro kategoriální pole.
- **Absolutní distribuce.** Zobrazí názvy funkcí/štítky a absolutní distribuce funkcí v rámci každého klastru. Pro účely kategoriálních funkcí se v zobrazení zobrazují pruhové grafy překryné kategoriemi seřazenými vzestupně podle datových hodnot. Pro spojitě funkce zobrazení ukazuje hladký graf hustoty, který používá stejné koncové body a intervaly pro každý klastr.

Plná červená barevná obrazovka ukazuje rozdělení klastru, zatímco zobrazení paroru představuje celkové údaje.

- **Relativní distribuce.** Zobrazí názvy funkcí/štítky a relativní distribuce v buňkách. Obecně jsou obrazovky podobné těm, které jsou zobrazeny pro absolutní distribuce, kromě toho, že se místo toho zobrazí relativní distribuce.

Plná červená barevná obrazovka ukazuje rozdělení klastru, zatímco zobrazení paroru představuje celkové údaje.

- **Základní zobrazení.** Tam, kde je mnoho klastrů, může být obtížné vidět všechny detaily bez posouvání. Chcete-li snížit množství posouvání, vyberte tento pohled, abyste změnili zobrazení na kompaktnější verzi tabulky.

### **Zobrazení důležitosti predikátu klastru**

Zobrazení důležitosti predikátu zobrazuje relativní důležitost každého pole při odhadu modelu.

### **Zobrazení velikostí klastru**

V zobrazení Velikosti klastru se zobrazí výšečový graf, který obsahuje každý klastr. Velikost procentní části každého klastru se zobrazí na každém řezu. Přesunete-li ukazatel myši nad každou výseč, zobrazí se počet v této výseči.

Pod grafem se v tabulce zobrazí následující informace o velikosti:

- Velikost nejmenšího klastru (počet i procentní část celku).
- Velikost největšího klastru (jak počet, tak procento celku).
- Poměr velikosti největšího klastru k nejmenšímu klastru.

### **Zobrazení distribuce buněk**

Pohled Distribuce buňky zobrazuje rozbalený, podrobnější graf o rozložení dat pro libovolnou buňku funkce, kterou vyberete v tabulce v hlavním panelu Klastry.

### **Zobrazení porovnání klastru**

Pohled Porovnání klastru se skládá z rozvržení ve stylu mřížky, které má funkce v řádcích a vybraných klastrů ve sloupcích. Toto zobrazení vám pomůže lépe pochopit faktory, které tvoří klastry; také vám umožní vidět rozdíly mezi klastry nejen ve srovnání s celkovými daty, ale s ostatními.

Chcete-li vybrat klastry pro zobrazení, klepněte na horní část sloupce klastru v hlavním panelu Klastry. Chcete-li vybrat nebo zrušit výběr více než jednoho klastru pro porovnání, použijte klepnutí s přidržením klávesy Ctrl nebo Shift.

*Poznámka:* Pro zobrazení můžete vybrat až pět klastrů.

Klastry se zobrazují v pořadí, ve kterém byly vybrány, zatímco pořadí polí je určeno volbou **Řadit funkce podle**. Vyberete-li volbu **Důležitost v rámci klastru**, budou pole vždy seřazena podle celkového významu.

Na pozadí se zobrazují celkové distribuce jednotlivých funkcí:

- Kategorické funkce se zobrazují jako bodové grafy, kde velikost tečky označuje nejčastější/modální kategorii pro každý klastr (podle funkce).
- Nepřetržité funkce se zobrazují jako krabičkové grafy, které zobrazují celkové mediány a interkvartilové rozsahy.

Překrývající se pro tyto pohledy na pozadí jsou krabičkové grafy pro vybrané klastry:

- V případě spojitých funkcí se čtvercovými a vodorovnými čarami vyznačují střední a interkvartilní rozsah pro každý klastr.
- Každý klastr je představován jinou barvou, která je zobrazena v horní části pohledu.

## Navigace v prohlížeči klastru


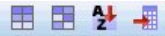
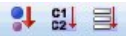

Prohlížeč klastru je interaktivní obrazovka. Můžete provést následující akce:

- Chcete-li zobrazit více podrobností, vyberte pole nebo klastr.
- Porovnejte klastry pro výběr položek, které vás zajímají.
- Změňte zobrazení.
- Transpozice os.

Použití panelů nástrojů

Informace zobrazené v levém i pravém panelu můžete řídit pomocí voleb na panelu nástrojů. Orientaci zobrazení (shora dolů, zleva doprava nebo zprava doleva) můžete změnit pomocí ovládacích prvků panelu nástrojů. Kromě toho můžete také resetovat prohlížeč na výchozí nastavení a otevřít dialogové okno pro uvedení obsahu pohledu Klastru v hlavním panelu.

Volby **Řadit funkce podle**, **Seřadit klastry podle**, **Buňky Zobrazit** jsou dostupné pouze tehdy, když vyberete zobrazení **Klastry** v hlavním panelu. Další informace naleznete v tématu [“Zobrazení klastrů”](#) na stránce 180 .

Ikona	Téma
	Viz <a href="#">Transponování klastrů a funkcí</a>
	Viz <a href="#">Řadit funkce podle</a>
	Viz <a href="#">Seřadit klastry podle</a>
	Viz <a href="#">Buňky</a>

Zobrazení řídicího zobrazení klastru

Chcete-li řídit, co se zobrazí v pohledu Klastru na hlavním panelu, klepněte na tlačítko **Zobrazit** ; otevře se dialogové okno Zobrazit.

**Vlastnosti.** Standardně vybráno. Chcete-li skrýt všechny vstupní funkce, zrušte zaškrtnutí tohoto políčka.

**Pole vyhodnocení.** Vyberte pole pro vyhodnocení (pole nepoužívaná k vytvoření modelu klastru, ale odeslaná do prohlížeče modelů za účelem vyhodnocení klastrů), které se mají zobrazit; při výchozím nastavení nejsou žádné zobrazeny. *Poznámka* Pole vyhodnocení musí být řetězec s více než jednou hodnotou. Toto zaškrtačkové políčko je nedostupné, nejsou-li k dispozici žádná pole vyhodnocení.

**Popisy klastrů.** Standardně vybráno. Chcete-li skrýt všechny buňky s popisem klastru, zrušte zaškrtnutí tohoto políčka.

**Velikosti klastrů.** Standardně vybráno. Chcete-li skrýt všechny buňky velikosti klastru, zrušte zaškrtnutí tohoto políčka.

**Maximální počet kategorií.** Zadejte maximální počet kategorií, které se mají zobrazit v grafech kategoričkových funkcí; výchozí hodnota je 20.

## Filtrování záznamů

Pokud chcete vědět více o případech v konkrétním klastru nebo skupině klastrů, můžete vybrat podmnožinu záznamů pro další analýzu založenou na vybraných klastrech.

1. Vyberte klastry v pohledu Klastr v prohlížeči klastru. Chcete-li vybrat více klastrů, použijte kombinaci Ctrl-click.
2. Z nabídky vyberte:

### Generovat > Filtrovat záznamy ...

3. Zadejte název proměnné filtru. Záznamy z vybraných klastrů obdrží hodnotu 1 pro toto pole. Všechny ostatní záznamy budou mít hodnotu 0 a budou vyloučeny z následných analýz, dokud nezměníte stav filtru.
4. Klepněte na tlačítko **OK**.

## Analýza hierarchického klastru

Tato procedura se pokouší identifikovat relativně homogenní skupiny případů (nebo proměnné) na základě vybraných charakteristik s použitím algoritmu, který začíná s každým případem (nebo proměnnou) v samostatném klastru a kombinuje klastry, dokud nezůstane pouze jedna. Můžete analyzovat prvotní proměnné, nebo můžete vybrat z různých standardizací transformací. Opatření týkající se vzdálenosti nebo podobnosti se generují pomocí procedury Proximities. Statistika se zobrazí v každé fázi, která vám pomůže vybrat nejlepší řešení.

**Příklad.** Jsou zde identifikovatelné skupiny televizních pořadů, které přitahují podobné publikum v rámci každé skupiny? Při analýze hierarchického klastru můžete v klastrové show (případy) do homogenních skupin vycházet z charakteristických vlastností prohlížeče. To lze použít k identifikaci segmentů pro marketing. Nebo můžete klastrovat města (případy) do homogenních skupin tak, aby bylo možné vybrat srovnatelná města k testování různých marketingových strategií.

**Statistika.** Plán agregace, vzdálenost (nebo podobnost) a členství klastru pro jediné řešení nebo rozsah řešení. Plošina: dendrogramy a číšní plochy.

Aspekty hierarchických dat analýzy klastru

**Data.** Proměnné mohou být kvantitativní, binární nebo spočítat data. Scaling of variables is an important issue -- differences in scaling may affect your cluster solution (s). Pokud mají vaše proměnné velké rozdíly ve škálování (například jedna proměnná se měří v dolarech a druhá se měří v letech), měli byste zvážit jejich standardizaci (to lze provést automaticky pomocí procedury Hierarchical Cluster Analysis).

**Pořadí případů.** Pokud ve vstupních datech existují spojované vzdálenosti nebo podobnosti nebo se při spojování vyskytují mezi aktualizovanými klastry, výsledné řešení klastru může záviset na pořadí případů v souboru. Možná budete chtít získat několik různých řešení s případy setříděnými v různých náhodných příkazech k ověření stability daného řešení.

**Předpoklady.** Použitá vzdálenost nebo podobnost opatření by měla být vhodná pro analyzovaná data (viz postup Proximities, kde získáte další informace o rozhodnutích o vzdálenosti a podobnosti). Do analýzy byste také měli zahrnout všechny relevantní proměnné. Omise vlivných proměnných může vést ke klamavému řešení. Protože hierarchická klastrová analýza je průzkumná metoda, mělo by se s výsledky zacházet jako s nezávislými výsledky, dokud se nepotvrdí nezávislým vzorkem.

Jak získat hierarchickou analýzu klastru

1. Z nabídky vyberte:

### Analyzovat > Klasifikovat > Hierarchický klastr ...

2. Pokud používáte klastrovací případy, vyberte alespoň jednu číselnou proměnnou. Pokud jste proměnné klastrování, vyberte alespoň tři číselné proměnné.

Volitelně můžete vybrat identifikační proměnnou pro případy popisku.

## Hierarchická metoda analýzy klastrů

**Metoda klastru.** Dostupné alternativy jsou sestavení mezi skupinami, propojení uvnitř skupin, nejbližší soused, nejvzdálenější soused, centroid klastrování, median dělení do klastrů a Ward.

**Ukazatel.** Umožňuje vám uvést vzdálenost nebo míru podobnosti, které se mají použít v klastrování. Vyberte typ dat a příslušnou vzdálenost nebo míru podobnosti:

- **Interval.** Dostupné alternativy jsou euklidovský vzdálenost, čtvercový euklidovský vzdálenost, kosinus, Pearson korelace, Chebychev, blok, Minkowského, a přizpůsobené.



- **Počty.** Dostupné alternativy jsou chí-kvadrát měřítko a Fí-square ukazatel.
- **Binární.** Dostupné alternativy jsou euklidovský vzdálenost, čtvercový euklidovský vzdálenost, velikost rozdíl, rozdíl ve vzorku, odchylka, rozptyl, tvar, jednoduché porovnávání, Fí 4-bodová korelace, lambda, Anderberg *D*, kostky, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance a Williams, Ochiai, Rogers a Tanimoto, Sokal a Snest 1, Sokal a Sedo3, Sokal a Sneth3, Sokal a Sneska 5, Yule's Ya Yule's *Q*.

**Transformace hodnot.** Umožňuje vám standardizovat datové hodnoty buď pro případy, nebo hodnoty před výpočtem proximity (nejsou k dispozici pro binární data). Dostupné standardizační metody jsou z skóre, rozsah – 1 až 1, rozsah 0 až 1, maximální velikost 1, střední hodnota 1 a směrodatná odchylka 1.

**Transformace měřítek.** Umožňuje vám transformovat hodnoty generované ukazatelem vzdálenosti. Jsou použity po vypočtených vzdálenoměrných opatřeních. Dostupné alternativy jsou absolutní hodnoty, změna znaménka a změna ve měřítku na 0-1 rozsah.

## Hierarchická statistika analýzy klastru

**Plán agregace.** Zobrazuje případy nebo klastry kombinované v jednotlivých fázích, vzdálené vzdálenosti mezi jednotlivými případy nebo klastry a poslední úroveň klastru, na které se případ (nebo proměnná) připojil ke klastru.

**Maticе Blízkoѕt.** Poskytuje vzdálenosti nebo podobnosti mezi položkami.

**Členství v klastru.** Zobrazuje klaster, ke kterému je každý případ přiřazen v jedné nebo více fázích v kombinaci klastrů. Dostupné volby jsou jednotlivé řešení a rozsah řešení.

## Plot pro analýzu hierarchického klastru

**Dendrogram.** Zobrazí *dendrogram*. Dendrogramy lze použít k posouzení kvality vytvořených klastrů a mohou poskytovat informace o odpovídajícím počtu klastrů, které mají být uchovány.

**Icicle.** Zobrazí *zákresový graf*, včetně všech klastrů, nebo určený rozsah klastrů. Icicle plots display information about how cases are combined into clusters at each iteration of the analysis. Orientace vám umožňuje vybrat vertikální nebo horizontální výkres.

## Hierarchická analýza klastru-Uložit nové proměnné

**Členství v klastru.** Umožňuje uložit členství klastru pro jediné řešení nebo pro rozsah řešení. Uložené proměnné lze poté použít v následných analýzách k prozkoumání dalších rozdílů mezi skupinami.

## Další funkce syntaxe příkazu CLUSTER

Procedura hierarchického klastru používá syntaxi příkazu CLUSTER . Jazyk syntaxe příkazu vám také umožňuje:

- V jedné analýze lze použít několik metod klastrování.
- Číst a analyzovat bezdotykovou matici.
- Zapsat matici blízkosti na disk pro pozdější analýzu.
- Uvedte libovolné hodnoty pro napájení a uživatele root v přizpůsobeném (Power) distanze vzdálenosti.
- Zadejte názvy uložených proměnných.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Analýza K-Means Cluster

Tato procedura se pokouší identifikovat relativně homogenní skupiny případů na základě vybraných charakteristik s použitím algoritmu, který dokáže zpracovat velký počet případů. Algoritmus však vyžaduje, abyste zadali počet klastrů. Pokud tyto informace znáte, můžete určit počáteční centra klastrů. Můžete vybrat jednu ze dvou metod klasifikace případů, buď aktualizovat centry klastru iterativně, nebo pouze klasifikaci. Můžete uložit členství klastru, informace o vzdálenosti a konečná centra klastrů. Volitelně můžete určit proměnnou, jejíž hodnoty se používají k označení výstupu na základě případu.

Můžete také vyžádat analýzu statistiky odchylky  $F$ . Zatímco tyto statistiky jsou oportunní (procedura se pokouší formovat skupiny, které se liší), relativní velikost statistiky poskytuje informace o příspěvku každé proměnné k oddělení skupin.

**Příklad.** Jaké jsou některé identifikovatelné skupiny televizních pořadů, které přitahují podobné publikum v rámci každé skupiny? S  $k$ -znamená klastrové analýzy, můžete v klastru (případech) do  $k$  homogenních skupin založených na charakteristikách prohlížeče. Tento proces může být použit k identifikaci segmentů pro marketing. Nebo můžete klastrovaná města (případy) do homogenních skupin tak, aby bylo možné vybrat srovnatelná města k testování různých marketingových strategií.

**Statistika.** Kompletní řešení: počáteční středová centra, tabulka ANOVA. Každý případ: informace o klastru, vzdálenost od centra klastru.

Pokyny k datům analýzy klastru K-Means

**Data.** Proměnné by měly být kvantitativní na úrovni intervalu nebo poměru. Pokud jsou vaše proměnné binární nebo počítají, použijte proceduru Hierarchical Cluster Analysis.

**Pořadí v případě malých a velkých písmen v klastru.** Výchozí algoritmus pro výběr počátečních center klastru není invariantní na řazení případů. Volba **Použít spuštěné prostředky** v dialogovém okně Iterovat učiní výsledné řešení potenciálně závislé na pořadí případů, bez ohledu na to, jak jsou zvolena počáteční centra klastru. Pokud používáte některou z těchto metod, můžete chtít získat několik různých řešení s případy setříděnými v různých náhodných objednávkách a ověřit tak stabilitu daného řešení. Určením výchozích center klastru a nepoužití volby **Použít spuštěné prostředky** se vyhnete problémům souvisejícím s objednávkou případů. Uspořádání počátečních center klastru však může toto řešení ovlivnit, je-li mezi případy na centra klastru svázány velké vzdálenosti. Chcete-li posoudit stabilitu daného řešení, můžete porovnat výsledky z analýz s různými permutací z výchozích hodnot středu.

**Předpoklady.** Vzdálenosti se počítají pomocí jednoduché euklidovské vzdálenosti. Chcete-li použít jinou vzdálenost nebo míru podobnosti, použijte proceduru Hierarchical Cluster Analysis. Rozšiřitelnost proměnných je důležitá úvaha. Pokud se vaše proměnné měří na různých měřítcích (například jedna proměnná je vyjádřena v dolarech a jiná proměnná je vyjádřena v rocích), vaše výsledky mohou být zavádějící. V takových případech byste měli zvážit standardizaci proměnných před provedením analýzy klastru  $k$  (tuto úlohu lze provést v rámci procedury s deskriptivními ukazateli). Procedura předpokládá, že jste vybrali odpovídající počet klastrů a že jste zahrnuli všechny relevantní proměnné. Pokud jste zvolili nevhodný počet klastrů nebo vynechali důležité proměnné, vaše výsledky mohou být zavádějící.

Jak získat klastrovou analýzu K-Means

1. Z nabídky vyberte:

**Analyzovat > Klasifikovat > K-Střední klastr ...**

2. Vyberte proměnné, které mají být použity v analýze klastru.

3. Zadejte počet klastrů. (Počet klastrů musí být nejméně 2 a nesmí být větší než počet případů v datovém souboru.)

4. Vyberte volbu **Iterovat a klasifikovat** nebo **Pouze klasifikovat**.

5. Volitelně můžete vybrat identifikační proměnnou pro případy popisku.

## Efektivita analýzy klastru KK

Příkaz  $k$ -znamená, že příkaz analýzy klastrů je efektivní primárně proto, že nevypočítávají vzdálenosti mezi všemi páry případů, stejně jako mnoho klastrovaných algoritmů, včetně algoritmu, který používá hierarchický příkaz klastrování.

Chcete-li dosáhnout maximální efektivity, použijte ukázkou případů a vyberte metodu **Iterovat a klasifikovat** pro určení center klastru. Vyberte **Write final as**. Poté obnovte celý datový soubor a vyberte **Klasifikovat pouze** jako metodu a vyberte **Přečíst počáteční z**, abyste klasifikovali celý soubor s použitím středisek, která jsou odhadována na základě ukázky. Můžete zapisovat a číst ze souboru nebo datové sady. Datové sady jsou k dispozici pro další použití ve stejné relaci, ale nejsou uloženy jako soubory, dokud nejsou explicitně uloženy před koncem relace. Názvy datových sad musí být v souladu s pravidly pro pojmenování proměnných. Další informace naleznete v tématu .

## Iterace analýzy K-Means Iterate

*Poznámka:* Tyto volby jsou dostupné pouze tehdy, když vyberete metodu **Iterovat a klasifikovat** z dialogového okna Analýza klanu K-Means.

**Maximum iterací.** Omezuje počet iterací v algoritmu *k*-means. Iterace se zastaví po tolika iteracích i v případě, že kritérium konvergence není splněno. Toto číslo musí být mezi 1 a 999.

Chcete-li reprodukovat algoritmus použitý příkazem Quick Cluster před verzí 5.0, nastavte parametr **Maximum iterací** na hodnotu 1.

**Kritérium konvergence.** Určuje, kdy iterace skončí. Reprezentuje proporcí minimální vzdálenosti mezi počátečními centry klastru, takže musí být větší než 0, ale ne větší než 1. Je-li kritérium například rovno 0.02, iterace se zastaví, když se celá iterace nepřesune do žádné z center klastru o vzdálenost větší než 2% nejmenší vzdálenosti mezi libovolnými počátečními centry klastru.

**Použijte spuštěné prostředky.** Umožňuje vám požádat, aby byla centra klastru aktualizována po každém přiřazení každého případu. Pokud tuto volbu nevyberete, budou nová centra klastru počítána až po přiřazení všech případů.

## K-Means Uložení analýzy klastru

Informace o řešení můžete uložit jako nové proměnné, které mají být použity v následných analýzách:

**Členství v klastru.** Vytvoří novou proměnnou označující konečné členství klastru každého případu. Hodnoty nové proměnné jsou v rozsahu 1 až po počet klastrů.

**Vzdálenost od centra klastrů.** Vytvoří novou proměnnou označující euklidovskou vzdálenost mezi jednotlivými případy a její klasifikační centrum.

## Volby analýzy klastrů K-Means

**Statistika.** Můžete vybrat následující statistiky: počáteční centra klastrů, tabulka ANOVA a informace o klastru pro každý případ.

- **Počáteční centra klastrů.** První odhad proměnné znamená pro každý z klastrů. Ve výchozím nastavení je z dat vybrán počet případů s dobře rozloženými případy, které jsou shodné s počtem klastrů. Počáteční centra klastrů se používají pro první kolo klasifikace a jsou poté aktualizována.
- **Tabulka ANOVA.** Zobrazí tabulku analýzy rozptylu, která zahrnuje testy Univariate F pro každou proměnnou dělení do klastrů. Testy F jsou popisné a výsledné pravděpodobnosti by neměly být vykládány. Tabulka ANOVA se nezobrazí, jsou-li všechny případy přiřazeny k jednomu klastru.
- **Informace o klastru pro každý případ.** Zobrazí pro každý případ konečné přiřazení klastru a euklidovskou vzdálenost mezi případem a centrem klastru, které se používá ke klasifikaci případu. Také zobrazí euklidovskou vzdálenost mezi konečnými centry klastru.

**Chybějící hodnoty.** Dostupné volby jsou **Vyloučit případy podle lisu** nebo **Vyloučit případy po dvojicích**.

- **Vyloučit případy jako litwise.** Vyloučí případy s chybějícími hodnotami pro jakoukoli proměnnou klastrování z analýzy.
- **Vyloučit případy po dvojicích.** Přiřadí případy na klastry založené na vzdálenostech, které jsou vypočteny ze všech proměnných s nechybějícími hodnotami.

## Další funkce příkazu QUICK CLUSTER

Procedura K-Means Cluster používá syntaxi příkazu QUICK CLUSTER. Jazyk syntaxe příkazu vám také umožňuje:

- Přijměte první *k* případy jako počáteční centra klastru, čímž se vyhnete předání dat, které se normálně používá k jejich odhadu.
- Určete počáteční centra klastrů přímo jako část syntaxe příkazu.
- Zadejte názvy uložených proměnných.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Neparametrické testy

---

Neparametrické testy provádějí minimální předpoklady o základní distribuci dat. Testy, které jsou k dispozici v těchto dialogových oknech, lze seskupit do tří širokých kategorií podle toho, jak jsou data uspořádána:

- Jedna vzorová zkouška analyzuje jedno pole.
- Test souvisejících ukázek porovnává dvě nebo více polí se stejnou sadou případů.
- Test nezávislých ukázek analyzuje jedno pole, které je seskupeno podle kategorií jiného pole.

### Jedna-ukázka neparametrických testů

Jedna-ukázková neparametrická testy identifikují rozdíly v jednotlivých polích pomocí jednoho nebo více neparametrických testů. Neparametrické testy nepředpokládají, že vaše data sledují normální distribuci.

**Jaký je váš cíl?** Cíle vám umožňují rychle uvést různé, ale běžně používané nastavení testu.

- **Automaticky porovnat zjištěná data s hypotézovanou.** Tento cíl používá Binomický test pro kategoriická pole pouze se dvěma kategoriemi, test chí-kvadrát na všechna ostatní kategoriální pole a test Kolmogorov-Smirnov do souvislých polí.
- **Testovací posloupnost pro náhodný výběr.** Tento záměr používá test Spuštění k testování pozorované posloupnosti hodnot dat pro náhodnost.
- **Vlastní analýza.** Chcete-li ručně změnit nastavení testu na kartě Nastavení, vyberte tuto volbu. Všimněte si, že toto nastavení je automaticky vybráno, pokud následně změníte volby na kartě Nastavení, které nejsou kompatibilní s aktuálně vybraným záměrem.

### Získání jednovzorných neparametrických testů

Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Jedna ukázka ...**

1. Klepněte na volbu **Spustit**.

Volitelně můžete:

- Zadejte cíl na kartě Cíl.
- Na kartě Pole zadejte přiřazení polí.
- Zadejte odborné nastavení na kartě Nastavení.

### Karta Pole

Karta Pole uvádí, která pole by měla být testována.

**Použit předdefinované role.** Tato volba používá existující informace o poli. Všechna pole s předdefinovanou rolí jako Vstup, Cíl nebo Obě budou použity jako testovací pole. Je požadováno alespoň jedno testovací pole.

**Použit vlastní přiřazení polí.** Tato volba umožňuje přepsat role polí. Po výběru této volby zadejte níže uvedená pole:

- **Testovací pole.** Vyberte jedno nebo více polí.

### Karta Nastavení

Karta Nastavení se skládá z několika různých skupin nastavení, které můžete upravit k finalizaci způsobu, jakým algoritmus zpracovává vaše data. Provedete-li jakékoli změny výchozích nastavení, které nejsou kompatibilní s aktuálně vybraným záměrem, bude karta Cíl automaticky aktualizována, abyste vybrali volbu **Přízpůsobit analýzu** .

## Vybrat testy

Tato nastavení určují testy, které mají být provedeny v polích uvedených na kartě Pole.

**Automaticky zvolit testy založené na datech.** Toto nastavení použije Binomický test na kategorická pole s pouze dvěma platnými (nechybějícími) kategoriemi, testem chí-kvadrát na všechna ostatní kategorická pole a testem Kolmogorov-Smirnov do souvislých polí.

**Upravte testy.** Toto nastavení umožňuje výběr specifických testů, které mají být provedeny.

- **Porovnejte pozorovanou binární pravděpodobnost k hypotéze (Binomická zkouška).** Test Binomického testu lze použít na všechna pole. Tím se vytvoří jednovzorkový test, který testuje, zda pozorovaná distribuce pole příznaku (kategorické pole s pouze dvěma kategoriemi) je stejné jako to, co se očekává od zadané binomického rozdělení. Kromě toho můžete požadovat intervaly spolehlivosti. Podrobnosti o nastavení testu viz [“Volby testu Binomial”](#) na stránce 189 .
- **Porovnejte sledované pravděpodobnosti s hypotézám (Chi-Square test).** Test chí-kvadrát se používá na nominální a ordinální pole. Tím se vytvoří jedn-vzorový test, který vypočítá statistiku chí-kvadrát na základě rozdílů mezi pozorovanými a očekávanými frekvencemi kategorií pole. Podrobnosti o nastavení testu viz [“Volby testu chí-kvadrát”](#) na stránce 190 .
- **Testujte pozorovanou distribuci proti hypotéze (zkouška Kolmogorov-Smirnov).** Test Kolmogorov-Smirnov se používá u spojitých polí a ordinal . Tím se vytvoří jeden ukázkový test, zda je ukázková funkce rozdělení pro pole homogenní s uniformní, normální, Poissonovo nebo exponenciální distribucí. Podrobnosti o nastavení testu viz [“Kolmogorov-Smirnov Options”](#) na stránce 190 .
- **Porovnání mediánu s hypotézám (Wilcoxonův test-ohodnocený test).** The Wilcoxon signed-rank test is applied to continuous and ordinal fields. Tím se vytvoří jedn-vzorový test mediánu hodnoty pole. Uvedte číslo jako medián hypotézy.
- **Testovací posloupnost pro výběr náhodných hodnot (Spuštění testu).** Test spuštění se použije na všechna pole. Tím se vytvoří jedn-vzorkovací test, zda je posloupnost hodnot difoktomizovaného pole náhodná. Podrobnosti o nastavení testu viz [“Spustí testovací volby”](#) na stránce 190 .

### Volby testu Binomial

Zkouška binomického testu je určena pro pole příznaku (kategorická pole s pouze dvěma kategoriemi), ale je použita pro všechna pole pomocí pravidel pro definování "úspěchu".

**Hypotézám podíl.** Tato hodnota určuje očekávaný podíl záznamů definovaných jako "úspěchy" nebo  $p$ . Zadejte hodnotu větší než 0 a menší než 1. Výchozí hodnota je 0,5.

**Interval spolehlivosti.** K dispozici jsou následující metody výpočtu intervalů spolehlivosti pro binární data:

- **Cloupper-Pearson (exact).** Přesný interval na základě kumulativní binomického rozdělení.
- **Jeffreys.** Bayesovský interval založený na zadní distribuci  $p$  pomocí Jeffreys před.
- **Poměr pravděpodobnosti.** Interval založený na funkci pravděpodobnosti pro  $p$ .

**Definujte úspěch pro kategorická pole.** To uvádí, jak "úspěch", testovaná datová hodnota ( $y$ ) proti hypotetizované proporci, jsou definována pro kategoriální pole.

- Volba **Použit první kategorii nalezenou v datech** provede binomický test s použitím první hodnoty nalezené ve vzorku za účelem definování "úspěchu". Tato volba je použitelná pouze pro nominálně nebo ordinální pole pouze s dvěma hodnotami; všechna ostatní kategorická pole zadaná na kartě Pole, kde se tato volba použije, nebude testována. Jedná se o výchozí nastavení.
- Volba **Zadat hodnoty úspěchu** provede test binomického testu s použitím určeného seznamu hodnot k definování "úspěchu". Zadejte seznam řetězcových nebo číselných hodnot. Hodnoty v seznamu nemusí být přítomny ve vzorku.

**Definujte úspěch pro souvislá pole.** To uvádí, jak "úspěch" je datová hodnota ( $y$ ) testována proti hodnotě testu, definovaná pro souvislá pole. Úspěch je definován jako hodnoty rovné nebo menší než bod řezu.

- **Ukázkový střed** nastavuje bod řezu v průměru na minimální a maximální hodnotu.
- **Vlastní řezaný bod** vám umožňuje uvést hodnotu pro bod řezu.

**Všechny kategorie mají stejnou pravděpodobnost.** Tím se vytvoří stejné frekvence mezi všemi kategoriemi ve vzorku. Jedná se o výchozí nastavení.

**Upravte očekávanou pravděpodobnost.** To vám umožňuje uvést nestejně frekvence pro uvedený seznam kategorií. Zadejte seznam řetězcových nebo číselných hodnot. Hodnoty v seznamu nemusí být přítomny ve vzorku. Ve sloupci **Kategorie** zadejte hodnoty kategorie. Ve sloupci **Relativní frekvence** zadejte hodnotu větší než 0 pro každou kategorii. S vlastními frekvencemi se zachází jako s poměry, takže například při uvedení frekvencí 1, 2 a 3 je ekvivalentní uvedení frekvencí 10, 20 a 30, a obojí uvádí, že 1/6 záznamů se očekává, že spadají do první kategorie, 1/3 do druhé, a 1/2 do třetí. Jsou-li určeny vlastní očekávané pravděpodobnosti, musí hodnoty vlastních kategorií zahrnovat všechny hodnoty polí v datech; jinak se test pro dané pole neprovede.

#### Kolmogorov-Smirnov Options

Toto dialogové okno uvádí, které distribuce by měly být testovány, a parametry hypotéz distribuce.

Mají-li být určité parametry distribuce odhadovány na základě vzorku, zkouška Kolmogorov-Smirnov se již nepoužije. V těchto případech lze statistiku testu Lilliefors použít k odhadu hodnoty *p* za použití vzorkování Monte Carlo pro testování normality a neznámého rozptylu. Test Lilliefors se vztahuje na tři souvislé distribuce (**Normální, Exponenciální a Uniformní**). Všimněte si, že test se nepoužije, je-li základní distribuce diskrétní (**Poisson**). Test je definován pouze pro jeden vzorek inference, když nejsou zadány odpovídající parametry distribuce.

#### Normální

Volba **Použití vzorová data** používá pozorovanou střední a směrodatnou odchylku a poskytuje volby pro výběr existujících výsledků **Asymptotic** nebo použití **testu Lilliefors na základě vzorkování Monte Carlo**. Volba **Vlastní** umožňuje zadat hodnoty.

#### Uniformní

**Použití vzorová data** používá pozorované minimum a maximum a používá test Lilliefors na základě vzorkování Monte Carlo. Volba **Vlastní** umožňuje určit minimální a maximální hodnoty.

#### Exponenciální

**Ukázka střední hodnoty** používá sledovaný průměr a používá test Lilliefors na základě vzorkování Monte Carlo. **Vlastní** vám umožňuje uvést pozorovanou střední hodnotu.

#### POISSON

**Střed** vám umožňuje uvést pozorovanou střední hodnotu.

#### Spustí testovací volby

Test spuštění je určen pro pole parametrů (kategorická pole s pouze dvěma kategoriemi), ale lze jej použít na všechna pole pomocí pravidel pro definování skupin.

**Definujte skupiny pro kategorická pole.** K dispozici jsou následující volby:

- **Ve vzorku jsou pouze 2 kategorie** provádí test spuštěním s použitím hodnot nalezených ve vzorku pro definování skupin. Tato volba je použitelná pouze pro nominálně nebo ordinální pole pouze s dvěma hodnotami; všechna ostatní kategorická pole zadaná na kartě Pole, kde se tato volba použije, nebude testována.
- **Recode data into 2 categories** provede test runs pomocí zadaného seznamu hodnot pro definování jedné ze skupin. Všechny ostatní hodnoty ve vzorku definují druhou skupinu. Hodnoty v seznamu nemusí být všechny přítomné ve vzorku, ale alespoň jeden záznam musí být v každé skupině.

**Definování řezu pro souvislá pole.** Uvádí, jak jsou skupiny definovány pro souvislá pole. První skupina je definována jako hodnoty rovné nebo menší než bod řezu.

- **Ukázka mediánu** nastavuje bod řezu na mediánu vzorku.
- **Ukázka střední hodnoty** nastavuje bod řezu na střední hodnotu vzorku.
- Volba **Vlastní** umožňuje určit hodnotu pro bod řezu.

## Volby testu

### Úroveň významnosti

Tato hodnota určuje úroveň významnosti (alfa) pro všechny testy. Uved'te číselnou hodnotu mezi 0 a 1. 0.05 je výchozí hodnota.

### Interval spolehlivosti (%)

Tato hodnota určuje úroveň důvěry pro všechny produkované intervaly spolehlivosti. Uved'te číselnou hodnotu mezi 0 a 100. Výchozí hodnota je 95.

### Vyloučené případy

Tato hodnota určuje, jak určit základ případu pro testy.

#### Vyloučit případy testy podle testu

Záznamy s chybějícími hodnotami pro pole, které se používají pro specifický test, jsou z tohoto testu vynechány. Je-li v analýze uvedeno několik testů, každý test se vyhodnotí samostatně.

#### Vyloučit případy jako litwise

Záznamy s chybějícími hodnotami pro jakékoli pole, které jsou pojmenované na kartě Pole, jsou vyloučeny ze všech analýz.

### Ukázka Monte Carlo

Mají-li být určité parametry distribuce odhadovány na základě vzorku, zkouška Kolmogorov-Smirnov se již nepoužije. V těchto případech lze statistiku testu Lilliefors použít k odhadu hodnoty  $p$  za použití vzorkování Monte Carlo pro testování normality a neznámého rozptylu. Test Lilliefors se vztahuje na tři souvislé distribuce (**Normální**, **Exponenciální** a **Uniformní**). Všimněte si, že test se nepoužije, je-li základní distribuce diskrétní (**Poisson**). Test je definován pouze pro jeden vzorek inference, když nejsou zadány odpovídající parametry distribuce.

#### Nastavit vlastní původ

Je-li tato volba povolena, poskytuje tato volba možnost resetování náhodného výběru hodnoty **Seed**, který se používá pro vzorkování Monte Carlo. Hodnota musí být jediné celé číslo mezi 1 a 2.147.483.647 (výchozí hodnota je 2 000 000). Při výchozím nastavení není nastavení povoleno, což generuje náhodné počáteční hodnoty.

#### Počet vzorků

Resetuje počet vzorků odběrů Monte Carla, které používá test Lilliefors. Hodnota musí být jediné celé číslo mezi 100 a největším celým číslem. Výchozí hodnota je 10 000.

#### Úroveň důvěryhodnosti simulace (%)

Obnoví odhadovanou úroveň intervalu spolehlivosti testu Kolmogorov-Smirnov. Hodnota musí být jednotlivá hodnota mezi 0 a 100. Výchozí hodnota je 99.

## Uživatel-chybějící hodnoty

**Uživatel-Chybějící hodnoty pro kategorická pole.** Kategorická pole musí mít platné hodnoty pro záznam, který má být zahrnut do analýzy. Tyto ovládací prvky umožňují rozhodnout, zda s uživatelem chybějícími hodnotami je zacházeno jako s kategoriálními poli. Systémem-chybějící hodnoty a chybějící hodnoty pro souvislá pole jsou vždy považovány za neplatné.

## Další funkce příkazu NPTESTS

Jazyk syntaxe příkazu vám také umožňuje:

- Určete jednovzorkové, nezávislé-ukázky a související testy vzorků v rámci jednoho spuštění procedury. Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Nezávislí-Ukázkové neparametrické testy

Nezávislé-vzorky neparametrické testy identifikují rozdíly mezi dvěma nebo více skupinami pomocí jednoho nebo více neparametrických testů. Neparametrické testy nepředpokládají, že vaše data sledují normální distribuci.

**Jaký je váš cíl?** Cíle vám umožňují rychle uvést různé, ale běžně používané nastavení testu.

- **Automaticky porovnat distribuce mezi skupinami.** Tento cíl aplikuje Mann-Whitney U test na data se 2 skupinami, nebo Kruskal-Wallis 1-way ANOVA k datům s  $k$  skupin.
- **Porovnejte střední a více skupin.** Tento cíl používá Medián test k porovnání pozorovaných mediánů mezi skupinami.
- **Vlastní analýza.** Chcete-li ručně změnit nastavení testu na kartě Nastavení, vyberte tuto volbu. Všimněte si, že toto nastavení je automaticky vybráno, pokud následně změníte volby na kartě Nastavení, které nejsou kompatibilní s aktuálně vybraným záměrem.

## Chcete-li získat neparametrické testy nezávislé na ukázkách

Z nabídky vyberte:

### Analyzovat > Neparametrické testy > Nezávislé ukázky ...

1. Klepněte na volbu **Spustit**.

Volitelně můžete:

- Zadejte cíl na kartě Cíl.
- Na kartě Pole zadejte přiřazení polí.
- Zadejte odborné nastavení na kartě Nastavení.

## Karta Pole

Karta Pole uvádí, která pole by měla být testována, a pole použité pro definování skupin.

**Použit předdefinované role.** Tato volba používá existující informace o poli. Všechna souvislá a ordinální pole s předdefinovanou rolí jako Cíl, nebo Obě budou použity jako testovací pole. Existuje-li jedno kategoričné pole s předdefinovanou rolí jako Vstup, bude použito jako pole seskupení. Jinak se ve výchozím nastavení nepoužije žádné pole seskupení a vy musíte použít přiřazení vlastních polí. Je požadováno alespoň jedno testovací pole a seskupovací pole.

**Použit vlastní přiřazení polí.** Tato volba umožňuje přepsat role polí. Po výběru této volby zadejte níže uvedená pole:

- **Testovací pole.** Vyberte jedno nebo více spojitých polí nebo ordinal .
- **Skupiny.** Vyberte kategoričné pole.

## Karta Nastavení

Karta Nastavení se skládá z několika různých skupin nastavení, které můžete upravit, a ladit tak způsob, jakým algoritmus zpracovává vaše data. Provedete-li jakékoli změny výchozích nastavení, které nejsou kompatibilní s aktuálně vybraným záměrem, bude karta Cíl automaticky aktualizována, abyste vybrali volbu **Přízpůsobit analýzu** .

### Zvolte testy

Tato nastavení určují testy, které mají být provedeny v polích uvedených na kartě Pole.

**Automaticky zvolit testy založené na datech.** Toto nastavení aplikuje Mann-Whitney U test na data se 2 skupinami nebo Kruskal-Wallis 1-way ANOVA k datům s  $k$  skupinami.

**Upravte testy.** Toto nastavení umožňuje výběr specifických testů, které mají být provedeny.

- **Porovnání distribucí napříč skupinami.** Tyto testy vytvářejí nezávislé-vzorky zkoušek, zda vzorky pocházejí ze stejné populace.

**Mann-Whitney U (2 ukázky)** používá pořadí jednotlivých případů k testování, zda jsou skupiny vykresleny ze stejné populace. První hodnota ve vzestupném pořadí pole seskupení definuje první skupinu a druhou definuje druhou skupinu. Má-li pole seskupení více než dvě hodnoty, tento test se nevytvorí.



**Kolmogorov-Smirnov (2 vzorky)** je citlivý na jakýkoli rozdíl v mediánu, disperzi, šikmost, a tak dále mezi dvěma distribucemi. Má-li pole seskupení více než dvě hodnoty, tento test se nevytvorí.

Volba **Testovací posloupnost pro náhodný výběr (Wald-Wolfowitz pro 2 ukázky)** vygeneruje test spuštění se členem skupiny jako kritériem. Má-li pole seskupení více než dvě hodnoty, tento test se nevytvorí.

**Kruskal-Wallis 1-way ANOVA (k samples)** is an extension of the Mann-Whitney U test and the nonparametrické analog of one-way analysis of variance. Volitelně můžete požadovat více porovnání ukázek  $k$ , buď **po dvojicích**, více porovnání, nebo porovnání **kroková úroveň dolů**.

**Test pro objednané alternativy (Jonckheere-Terpstra for k samples)** je mocnější alternativou k Kruskal-Wallis, když  $k$  ukázky mají přirozené pořadí. Například populace  $k$  může představovat  $k$  zvýšení teploty. Hypotéza, že různé teploty produkují stejnou distribuci odezvy, je testována proti alternativě, která se zvyšuje při zvyšování teploty, a zvyšuje se tak rozsah odezvy. Zde je třeba objednat alternativní hypotézu. Jonckheere-Terpstrana je tedy nejvhodnější zkouškou pro použití. **Nejmenší největší** určuje alternativní hypotézu, že parametr umístění první skupiny je menší nebo roven druhému, který je menší než nebo roven třetímu, atd. **Největší a nejmenší** uvádí alternativní hypotézu, že parametr umístění první skupiny je větší nebo roven druhému, který je větší než nebo roven třetímu, atd. U obou možností také alternativní hypotéza také předpokládá, že umístění nejsou všechny stejná. Volitelně můžete požadovat více porovnání ukázek  $k$  buď **Všechny po dvojicích**, více porovnání, nebo porovnání **Postupná postupná dolů**.

- **Porovnat rozsahy mezi skupinami.** Tím se vytvoří testy nezávislé na vzorcích, zda mají vzorky stejný rozsah. **Moses extreme reaction (2 samples)** tests a control group versus a comparison group. První hodnota ve vzestupném pořadí pole seskupení definuje řídicí skupinu a druhé definuje skupinu porovnání. Má-li pole seskupení více než dvě hodnoty, tento test se nevytvorí.
- **Porovnání Medianů napříč skupinami.** Tím se vytvoří testy nezávislých vzorků, zda vzorky mají stejný medián. **Medián testu (k vzorky)** může použít buď střední medián vzorku (vypočteno na základě všech záznamů v datové sadě), nebo vlastní hodnotu jako hypothesized median. Volitelně můžete požadovat více porovnání ukázek  $k$  buď **Všechny po dvojicích**, více porovnání, nebo porovnání **Postupná postupná dolů**.
- **Odhadnout intervaly spolehlivosti napříč skupinami. Hodges-Lehman odhad (2 vzorky)** vytvoří nezávislý odhad vzorků a interval spolehlivosti pro rozdíl mezi prostředníky dvou skupin. Má-li pole seskupení více než dvě hodnoty, tento test se nevytvorí.

## Volby testu

**Hladina významnosti.** Tato hodnota určuje úroveň významnosti (alfa) pro všechny testy. Uvedte číselnou hodnotu mezi 0 a 1. 0.05 je výchozí hodnota.

**Interval spolehlivosti (%).** Toto určuje úroveň důvěry pro všechny produkované intervaly spolehlivosti. Uvedte číselnou hodnotu mezi 0 a 100. Výchozí hodnota je 95.

**Vyloučené případy.** Tato hodnota určuje, jak určit základ případu pro testy. Volba **Vyloučit případy lišise** znamená, že záznamy s chybějícími hodnotami pro jakékoli pole uvedené na libovolném podpříkazu jsou vyloučeny ze všech analýz. Volba **Vyloučit případy podle testu** znamená, že záznamy s chybějícími hodnotami pole, které se používají pro specifický test, jsou z tohoto testu vynechány. Je-li v analýze uvedeno několik testů, každý test se vyhodnotí samostatně.

## Uživatel-chybějící hodnoty

**Uživatel-Chybějící hodnoty pro kategorická pole.** Kategorická pole musí mít platné hodnoty pro záznam, který má být zahrnut do analýzy. Tyto ovládací prvky umožňují rozhodnout, zda s uživatelem chybějícími hodnotami je zacházeno jako s kategoriálními poli. Systémem-chybějící hodnoty a chybějící hodnoty pro souvislá pole jsou vždy považovány za neplatné.

## Další funkce příkazu NPTESTS

Jazyk syntaxe příkazu vám také umožňuje:

- Určete jednovzorkové, nezávislé-ukázky a související testy vzorků v rámci jednoho spuštění procedury.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Související-Neparametrické testy ukázek

Identifikuje rozdíly mezi dvěma nebo více souvisejícími poli pomocí jednoho nebo více neparametrických testů. Neparametrické testy nepředpokládají, že vaše data sledují normální distribuci.

**Aspekty dat.** Každý záznam odpovídá danému předmětu, pro který jsou dvě nebo více souvisejících měření uloženy v oddělených polích v datové sadě. Například studie týkající se účinnosti plánu diety lze analyzovat pomocí neparametrických testů souvisejících s příbuznými vzorky, je-li hmotnost jednotlivých předmětů měřena v pravidelných intervalech a skladována v polích jako *Váha před stravou*, *Dočasná váha* a *Váha po stravě*. Tato pole jsou "související".

**Jaký je váš cíl?** Cíle vám umožňují rychle uvést různé, ale běžně používané nastavení testu.

- **Automaticky porovnat zjištěná data s údaji o hypotéze.** Tento cíl používá McNemarova testu na kategorická data, když jsou uvedena 2 pole, Cochran's Q pro kategorická data, když je uvedeno více než 2 pole, test Wilcoxon Matched-Pair Pair-Pair na základě souvislých dat, jsou-li uvedena 2 pole, a Friedman's 2-Way ANOVA k souvislým datům, když jsou uvedena více než dvě pole.
- **Vlastní analýza.** Chcete-li ručně změnit nastavení testu na kartě Nastavení, vyberte tuto volbu. Všimněte si, že toto nastavení je automaticky vybráno, pokud následně změníte volby na kartě Nastavení, které nejsou kompatibilní s aktuálně vybraným záměrem.

Jsou-li uvedena pole s odlišnou úrovní měření, jsou nejprve odděleny úrovní měření a pak se pro každou skupinu použije příslušná zkouška. Vyberete-li například volbu **Automaticky porovnat zjištěná data s údaji o hypotéze** jako cíl a zadejte 3 spojitá pole a 2 nominální pole, pak se test Friedman použije na pole se spojitými poli a McNemarova zkouška se použije na nominální pole.

## Chcete-li získat související-neparametrické testy-ukázky

Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Související ukázky ...**

1. Klepněte na volbu **Spustit**.

Volitelně můžete:

- Zadejte cíl na kartě Cíl.
- Na kartě Pole zadejte přiřazení polí.
- Zadejte odborné nastavení na kartě Nastavení.

### Karta Pole

Karta Pole uvádí, která pole by měla být testována.

**Použit předdefinované role.** Tato volba používá existující informace o poli. Všechna pole s předdefinovanou rolí jako Cíl nebo Obě budou použita jako testovací pole. Jsou vyžadována alespoň dvě testovací pole.

**Použit vlastní přiřazení polí.** Tato volba umožňuje přepsat role polí. Po výběru této volby zadejte níže uvedená pole:

- **Testovací pole.** Vyberte dvě nebo více polí. Každé pole odpovídá samostatné související ukázce.

### Karta Nastavení

Karta Nastavení se skládá z několika různých skupin nastavení, které můžete upravit tak, aby doladujete, jak procedura zpracovává vaše data. Provedete-li jakékoli změny výchozích nastavení, která nejsou kompatibilní s ostatními cíli, karta Cíl se automaticky aktualizuje, aby byla vybrána volba **Upravit analýzu** .

## Vybrat testy

Tato nastavení určují testy, které mají být provedeny v polích uvedených na kartě Pole.

**Automaticky zvolit testy založené na datech.** Toto nastavení použije McNemarova testu na kategorická data, když jsou uvedena 2 pole, Cochran's Q pro kategorická data, když je uvedeno více než 2 pole, test Wilcoxon-Pair-Rank-rank při určení 2 polí a Friedman's 2-Way ANOVA k souvislým datům pro souvislá data, je-li uvedeno více než 2 pole.

**Upravte testy.** Toto nastavení umožňuje výběr specifických testů, které mají být provedeny.

- **Test pro změnu v binárních datech. McNemarský test (2 ukázky)** lze použít na kategorická pole. Tím se vytvoří test souvisejících ukázek, zda kombinace hodnot mezi dvěma poli příznaku (kategorická pole s pouze dvěma hodnotami) jsou stejně pravděpodobné. Je-li na kartě Pole určeno více než dvě pole, tento test se neprovede. Podrobnosti o nastavení testu viz "McNemarský test: Define Success ." na stránce 195 . **Cochran's Q (k samples)** lze použít na kategoriální pole. Tím se vytvoří test souvisejících ukázek, zda kombinace hodnot mezi poli parametru  $k$  (kategorická pole s pouze dvěma hodnotami) jsou stejně pravděpodobné. Volitelně můžete požadovat více porovnání ukázek  $k$  , buď **po dvojicích** , více porovnání, nebo porovnání **kroková úroveň dolů** . Podrobnosti o nastavení testu viz "Cochran's Q: Define Success " na stránce 195 .
- **Test pro změny v objektu Multinomial Data. Marginální test homogeneity (2 vzorky)** vytvoří související testovací test, zda kombinace hodnot mezi dvěma párovanými pořadovými poli je stejně pravděpodobné. Zkouška mezní homogeneity se obvykle používá v situacích opakovaných měření. Tento test je rozšířením testu McNemar z binární odpovědi na mnohoomickou odpověď. Je-li na kartě Pole určeno více než dvě pole, tento test se neprovede.
- **Compare Median Difference to Hypothesized.** Tyto testy každý z nich vytvoří související-vzorkový test, zda střední rozdíl mezi dvěma poli se liší od 0. Test se používá pro souvislá a ordinální pole. Pokud je na kartě Pole určeno více než dvě pole, tyto testy se neprovedou.
- **Odhadnout interval spolehlivosti.** Tím se vytvoří související odhad vzorků a interval spolehlivosti pro medián rozdílu mezi dvěma párovanými poli. Test se používá pro souvislá a ordinální pole. Pokud je na kartě Pole určeno více než dvě pole, tento test se neprovede.
- **Množstevní přidružení. Kendall's coefficient of concordance (k samples)** produkuje měřítko shody mezi soudci nebo hodnotiteli, kde každý záznam obsahuje hodnocení jednoho soudce z několika položek (polí). Volitelně můžete požadovat více porovnání ukázek  $k$  buď **Všechny po dvojicích** , více porovnání, nebo porovnání **Postupná postupná dolů** .
- **Porovnání distribucí. Friedman's 2-way ANOVA by ranks (k samples)** produkuje související vzorky, z toho, zda  $k$  související vzorky byly odebírány ze stejné populace. Volitelně můžete požadovat více porovnání ukázek  $k$  buď **Všechny po dvojicích** , více porovnání, nebo porovnání **Postupná postupná dolů** .

*McNemarský test: Define Success .*

McNemarský test je určen pro pole parametrů (kategorická pole s pouze dvěma kategoriemi), ale používá se pro všechna kategorická pole pomocí pravidel pro definování "úspěchu".

**Definujte úspěch pro kategorická pole.** Určuje, jak je definován "úspěch" pro kategoriální pole.

- Volba **Použít první kategorii nalezenou v datech** provede test s použitím první hodnoty nalezené ve vzorku za účelem definování "úspěchu". Tato volba je použitelná pouze pro nominálně nebo ordinální pole pouze s dvěma hodnotami; všechna ostatní kategorická pole zadaná na kartě Pole, kde se tato volba použije, nebude testována. Jedná se o výchozí nastavení.
- Volba **Zadat hodnoty úspěchu** provede test s použitím určeného seznamu hodnot k definování "úspěchu". Zadejte seznam řetězcových nebo číselných hodnot. Hodnoty v seznamu nemusí být přítomny ve vzorku.

*Cochran's Q: Define Success*

Cochran je Q test je určen pro vlajka pole (kategorická pole s pouze dvěma kategoriemi), ale je aplikována na všechna kategorická pole pomocí pravidel pro definování "úspěchu".

**Definujte úspěch pro kategorická pole.** Určuje, jak je definován "úspěch" pro kategoriální pole.

- Volba **Použít první kategorii nalezenou v datech** provede test s použitím první hodnoty nalezené ve vzorku za účelem definování "úspěchu". Tato volba je použitelná pouze pro nominálně nebo ordinální pole pouze s dvěma hodnotami; všechna ostatní kategorická pole zadaná na kartě Pole, kde se tato volba použije, nebude testována. Jedná se o výchozí nastavení.
- Volba **Zadat hodnoty úspěchu** provede test s použitím určeného seznamu hodnot k definování "úspěchu". Zadejte seznam řetězcových nebo číselných hodnot. Hodnoty v seznamu nemusí být přítomny ve vzorku.

### **Volby testu**

**Hladina významnosti.** Tato hodnota určuje úroveň významnosti (alfa) pro všechny testy. Uveďte číselnou hodnotu mezi 0 a 1. 0.05 je výchozí hodnota.

**Interval spolehlivosti (%).** Toto určuje úroveň důvěry pro všechny produkované intervaly spolehlivosti. Uveďte číselnou hodnotu mezi 0 a 100. Výchozí hodnota je 95.

**Vyloučené případy.** Tato hodnota určuje, jak určit základ případu pro testy.

- Volba **Vyloučit případy lišise** znamená, že záznamy s chybějícími hodnotami pro jakékoli pole uvedené na libovolném podpříkazu jsou vyloučeny ze všech analýz.
- Volba **Vyloučit případy podle testu** znamená, že záznamy s chybějícími hodnotami pole, které se používají pro specifický test, jsou z tohoto testu vynechány. Je-li v analýze uvedeno několik testů, každý test se vyhodnotí samostatně.

### **Uživatel-chybějící hodnoty**

**Uživatel-Chybějící hodnoty pro kategorická pole.** Kategorická pole musí mít platné hodnoty pro záznam, který má být zahrnut do analýzy. Tyto ovládací prvky umožňují rozhodnout, zda s uživatelem chybějícími hodnotami je zacházeno jako s kategoriálními poli. Systémem-chybějící hodnoty a chybějící hodnoty pro souvislá pole jsou vždy považovány za neplatné.

## **Další funkce příkazu NPTESTS**

Jazyk syntaxe příkazu vám také umožňuje:

- Určete jednovzorkové, nezávislé-ukázky a související testy vzorků v rámci jednoho spuštění procedury. Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## **Další funkce příkazu NPTESTS**

Jazyk syntaxe příkazu vám také umožňuje:

- Určete jednovzorkové, nezávislé-ukázky a související testy vzorků v rámci jednoho spuštění procedury. Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## **Starší dialogová okna**

Existuje celá řada dialogových oken "legacy", které také provádějí neparametrické testy. Tato dialogová okna podporují funkce poskytované volbou Exact Tests.

**Chi-Square Test.** Tabuluje proměnnou do kategorií a vypočítá statistiku chí kvadrát na základě rozdílů mezi pozorovanými a očekávanými frekvencemi.

**Binomické testování.** Porovná pozorovanou frekvenci v každé kategorii dichotomní proměnné s očekávanými frekvencemi od binomického rozdělení.

**Spustí test.** Ověří, zda pořadí výskytu dvou hodnot proměnné je náhodné.

**One-Sample Kolmogorov-Smirnov Test.** Porovná zjištěné kumulativní distribuční funkce pro proměnnou s určenou teoretickou distribucí, což může být normální, rovnoměrné, exponenciální, nebo Poissonovo.

**Testy Two-Independent-Samples.** Porovná dvě skupiny případů na jedné proměnné. Mann-Whitney  $U$  test, dva-vzorek Kolmogorov-Smirnov test, Moses test extrémní reakce, a Wald-Wolfowitz runs test jsou k dispozici.

**Testy dvou souvisejících ukázek.** Porovná rozdělení dvou proměnných. K dispozici je test Wilcoxonův test, test podpisu a test McNemar.

**Testy pro několik nezávislých ukázek.** Porovná dvě nebo více skupin případů na jedné proměnné. Test Kruskal-Wallis, test Medián, a Jonckheere-Terpstra test jsou k dispozici.

**Testy pro několik souvisejících ukázek.** Porovná rozdělení dvou nebo více proměnných. Friedman test, Kendall je  $W$ , a Cochran  $Q$  jsou k dispozici.

Pro všechny výše uvedené zkoušky jsou k dispozici kvartily a průměrná směrodatná odchylka, minimum, maximum a počet nechybějících případů.

## test chí-kvadrát

Procedura testu chí-kvadrát roztavuje proměnnou do kategorií a vypočítá statistiku chí kvadrát. Tento test dobročinnosti porovnává pozorované a očekávané frekvence v každé kategorii za účelem testování, že všechny kategorie obsahují stejný podíl hodnot nebo testu, že každá kategorie obsahuje uživatelem specifikovaný podíl hodnot.

**Příklady.** Test chí kvadrát může být použit k určení, zda vak z fazolí typu bean obsahuje stejné proporce modře, hnědé, zelené, oranžové, červené a žluté nástrahy. Můžete také otestovat, zda vak s fazolemi obsahuje 5% modře, 30% hnědé, 10% zelené, 20% oranžové, 15% červeně a 15% žlutých kandiů.

**Statistika.** Střední, směrodatná odchylka, minimum, maximum a kvartily. Počet a procentní podíl nechybějících a chybějících případů; počet případů pozorovaných a očekávaných pro každou kategorii; zbytkové chyby a statistiky chí kvadrát.

Aspekty testovacích dat chí-kvadrát

**Data.** Použijte objednané nebo neobjednané číselné kategoriální proměnné (ordinální nebo nominální úrovně měření). Chcete-li převést řetězcové proměnné na číselné proměnné, použijte proceduru Automatic Recode, která je k dispozici v nabídce Transformace.

**Předpoklady.** Neparametrické testy nevyžadují předpoklady týkající se tvaru základní distribuce. Předpokládá se, že data jsou náhodná ukázkou. Očekávané frekvence pro každou kategorii by měly být alespoň 1. Ne více než 20% z kategorií by mělo mít očekávané frekvence menší než 5.

Získání testu chí-kvadrát

1. Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Starší dialogová okna > Chi-Square ...**

2. Vyberte jednu nebo více proměnných testu. Každá proměnná vytvoří samostatný test.

3. Volitelně klepněte na volbu **Volby** pro deskriptivní statistiky, kvartily a kontrolu zacházení s chybějícími daty.

## Očekávaný rozsah a očekávané hodnoty testu chí-kvadrát

**Očekávaný rozsah.** Při výchozím nastavení je každá odlišná hodnota proměnné definována jako kategorie. Chcete-li zřídit kategorie ve specifickém rozsahu, vyberte volbu **Použít určený rozsah** a zadejte celočíselné hodnoty pro dolní a horní meze. Kategorie se stanoví pro každou celočíselnou hodnotu v rámci rozsahu zahrnutí a jsou vyloučeny případy s hodnotami mimo hranice. Například, pokud uvedete hodnotu 1 pro Dolní a hodnotu 4 pro Upper, použijí se pro test chí-kvadrát pouze hodnoty typu integer od 1 do 4.

**Očekávané hodnoty.** Ve výchozím nastavení mají všechny kategorie stejné očekávané hodnoty. Kategorie mohou mít uživatelsky definované očekávané proporce. Vyberte **Hodnoty**, zadejte hodnotu, která je větší než 0 pro každou kategorii testovací proměnné, a poté klepněte na tlačítko **Přidat**. Pokaždé, když přidáte

nějakou hodnotu, zobrazí se na konci seznamu hodnot. Pořadí hodnot je důležité; odpovídá vzestupnému pořadí hodnot kategorie testovací proměnné. První hodnota v seznamu odpovídá nejnižší hodnotě skupiny testovací proměnné a poslední hodnota odpovídá nejvyšší hodnotě. Jsou sečteny prvky hodnotového seznamu a každá hodnota je vydělena tímto součtem pro výpočet podílu případů očekávaných v příslušné kategorii. Např. seznam hodnot 3, 4, 5, 4 určuje očekávané proporce 3/16, 4/16, 5/16 a 4/16.

### **Volby testu *chi-kvadrát***

**Statistika.** Můžete vybrat jednu nebo obě souhrnné statistiky.

- **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet případů, kdy nejsou chybějící případy.
- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy-podle-testu.** Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.
- **Vyloučit případy jako listwise.** Případy s chybějícími hodnotami pro každou proměnnou jsou vyloučeny ze všech analýz.

### **Další funkce příkazu *NPAAR (Chi-Square Test)***

Jazyk syntaxe příkazu vám také umožňuje:

- Uvedte odlišné minimální a maximální hodnoty nebo očekávané frekvence pro různé proměnné (s dílčím příkazem `CHISQUARE`).
- Testujte stejnou proměnnou s různými očekávanými frekvencemi nebo použijte různé rozsahy (s dílčím příkazem `EXPECTED`).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

### **Binomické testování**

Procedura Binomial Test porovnává pozorované frekvence dvou kategorií dichotomické proměnné s frekvencemi, které se očekávají pod binomickou distribucí s uvedeným parametrem pravděpodobnosti. Při výchozím nastavení je parametr pravděpodobnosti pro obě skupiny 0,5. Chcete-li změnit pravděpodobnosti, můžete zadat testovací část pro první skupinu. Pravděpodobnost pro druhou skupinu bude 1 minus uvedená pravděpodobnost pro první skupinu.

**Příklad.** Když budete házet desetník, pravděpodobnost, že se hlava rovná 1/2. Na základě této hypotézy se desetinasobek desetinasobně vyhodí a výsledky se zaznamenávají (hlavy nebo ocase). Z binomického testu, můžete zjistit, že 3/4 z tosses byli hlavy a že pozorovaný význam úroveň je malá (0.0027). Tyto výsledky naznačují, že není pravděpodobné, že pravděpodobnost hlavy se rovná 1/2; mince je pravděpodobně zkreslený.

**Statistika.** Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů a kvartily.

Podmínky pro Binomické testovací data

**Data.** Proměnné, které se testují, by měly být číselné a dichotomové. Chcete-li převést řetězcové proměnné na číselné proměnné, použijte proceduru Automatic Recode, která je k dispozici v nabídce Transformace. **Dichotomous variable** je proměnná, která může mít pouze dvě možné hodnoty: *yes* nebo *no*, *true* nebo *false*, 0 nebo 1, atd. První hodnota zjištěná v datové sadě definuje první skupinu a druhá hodnota definuje druhou skupinu. Nejsou-li proměnné dichotomové, musíte zadat bod řezu. Bod řezu přiřazuje případy s hodnotami, které jsou menší nebo rovny bodu řezu, do první skupiny a přiřadí zbytek případů do druhé skupiny.

**Předpoklady.** Neparametrické testy nevyžadují předpoklady týkající se tvaru základní distribuce. Předpokládá se, že data jsou náhodná ukázka.

Získání Binomického testu

1. Z nabídky vyberte:

### **Analyzovat > Neparametrické testy > Starší dialogová okna > Binomické ...**

2. Vyberte jednu nebo více číselných proměnných testů.
3. Volitelně klepněte na volbu **Volby** pro deskriptivní statistiky, kvartily a kontrolu zacházení s chybějícími daty.

#### **Volby testu Binomického testu**

**Statistika.** Můžete vybrat jednu nebo obě souhrnné statistiky.

- **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet případů, kdy nejsou chybějící případy.
- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy-podle-testu.** Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.
- **Vyloučit případy jako litwise.** Případy s chybějícími hodnotami pro každou proměnnou, která je testována, jsou vyloučeny ze všech analýz.

#### **Další funkce příkazu NPAR TESTS (Binomický test)**

Jazyk syntaxe příkazu vám také umožňuje:

- Vyberte konkrétní skupiny (a vyloučit jiné skupiny), pokud má proměnná více než dvě kategorie (s dílčím příkazem BINOMIAL ).
- Určete různé výstupní body nebo pravděpodobnosti pro různé proměnné (s dílčím příkazem BINOMIAL ).
- Otestovat stejnou proměnnou proti různým bodům nebo pravděpodobnostem (s dílčím příkazem EXPECTED ).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

### **Spouští test**

Procedura spuštění testů testuje, zda je pořadí výskytu dvou hodnot proměnné náhodné. Spuštění je poslovnost podobných pozorování. Ukázka s příliš mnoha nebo příliš málo běhy naznačuje, že vzorek není náhodný.

**Příklady.** Předpokládejme, že 20 lidí se dotazuje na to zjistit, zda by si koupit produkt. Předpokládáný náhodný výběr vzorku by byl vážně zpochybněn, pokud by všech 20 lidí bylo stejného pohlaví. Běží test lze použít k určení, zda byl vzorek nakreslen náhodně.

**Statistika.** Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů a kvartily.

Spouští aspekty testovacích dat

**Data.** Proměnné musí být číselné. Chcete-li převést řetězcové proměnné na číselné proměnné, použijte proceduru Automatic Recode, která je k dispozici v nabídce Transformace.

**Předpoklady.** Neparametrické testy nevyžadují předpoklady týkající se tvaru základní distribuce. Použijte ukázky z průběžné pravděpodobnosti rozdělení.

Chcete-li získat test spuštění

1. Z nabídky vyberte:

### **Analyzovat > Neparametrické testy > Starší dialogová okna > Spustit ...**

2. Vyberte jednu nebo více číselných proměnných testů.
3. Volitelně klepněte na volbu **Volby** pro deskriptivní statistiky, kvartily a kontrolu zacházení s chybějícími daty.

## Spouští testovací bod testu

**Vyjmout bod.** Určuje bod, který se má použít jako dichotomizaci proměnných, které jste vybrali. Můžete použít sledovaný střední, medián nebo režim, nebo můžete použít zadanou hodnotu jako bod řezu. Případy s hodnotami, které jsou nižší než bod omezení, jsou přiřazeny k jedné skupině a případy s hodnotami, které jsou větší než nebo rovny bodu přerušení, jsou přiřazeny do jiné skupiny. U každého vybraného bodu řezu se provede jedna zkouška.

## Spouští volby testu

**Statistika.** Můžete vybrat jednu nebo obě souhrnné statistiky.

- **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet případů, kdy nejsou chybějící případy.
- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy-podle-testu.** Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.
- **Vyloučit případy jako listwise.** Případy s chybějícími hodnotami pro každou proměnnou jsou vyloučeny ze všech analýz.

## Další funkce příkazu pro testování NPAR (test spuštění)

Jazyk syntaxe příkazu vám také umožňuje:

- Určete různé mezní body pro různé proměnné (s dílčím příkazem RUNS ).
- Otestovat stejnou proměnnou proti různým zakázaným bodům (s dílčím příkazem RUNS ).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## One-Ukázka Kolmogorov-Smirnov test

The One-Sample Kolmogorov-Smirnov Test procedure porovnává pozorovanou kumulativní distribuční funkci pro proměnnou s určenou teoretickou distribucí, která může být normální, uniformní, Poissonovo nebo exponenciální. Kolmogorov-Smirnov Z je počítán z největšího rozdílu (v absolutní hodnotě) mezi pozorovanými a teoretickými kumulativními distribučními funkcemi. Tento dobročinný test testuje, zda připomínky mohly přiměřeně pocházet z uvedené distribuce.

Počínaje verzí 27.0 lze statistiku testu Lilliefors použít k odhadu hodnoty  $p$  pomocí vzorkování Monte Carlo pro testování proti normální distribuci s odhadovanými parametry (tato funkčnost byla dříve možná pouze prostřednictvím procedury Explore).

### Příklad

Mnoho parametrických testů vyžaduje normálně distribuované proměnné. The one-sample Kolmogorov-Smirnov test can be used to test that a variable (for example, *income*) is normally distributed.

### Statistika

Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů, kvartilové, Lillieforsův test a simulace Monte Carlo.

## Jedna-ukázka aspektů testovacích dat Kolmogorov-Smirnov

### Data

Použit kvantitativní proměnné (interval nebo poměr úrovně měření).

### Předpoklady

Test Kolmogorov-Smirnov předpokládá, že parametry distribuce testu jsou specifikovány předem. Tento postup odhadne parametry ze vzorku. Střední a vzorová směrodatná odchylka jsou parametry pro normální rozložení, minimální a maximální hodnoty vzorku určují rozsah rovnoměrného rozdělení, střední hodnota vzorku je parametr Poissonova rozdělení a vzorek střední hodnoty je parametr pro



exponenciální distribuci. Síla testu na zjištění odchylek od hypotézované distribuce může být vážně zmenšena.

Mají-li být určité parametry distribuce odhadovány na základě vzorku, zkouška Kolmogorov-Smirnov se již nepoužije. V těchto případech lze statistiku testu Lilliefors použít k odhadu hodnoty  $p$  za použití vzorkování Monte Carlo pro testování normality a neznámého rozptylu. Test Lilliefors se vztahuje na tři souvislé distribuce (**Normální, Exponenciální a Uniformní**). Všimněte si, že test se nepoužije, je-li základní distribuce diskrétní (**Poisson**). Test je definován pouze pro jeden vzorek inference, když nejsou zadány odpovídající parametry distribuce.

## Získání jednovzorkového testu Kolmogorovan-Smirnov

1. Z nabídky vyberte:

**Analyzovat > Neparаметrické testy > Starší dialogová okna > 1-Ukázka K-S. ..**

2. Vyberte jednu nebo více číselných proměnných testu. Každá proměnná vytvoří samostatný test.

3. Volitelně můžete vybrat metodu rozdělení testu:

### Normální

Je-li tato volba vybrána, můžete určit, zda mají být distribuční parametry odhadovány na základě ukázkových dat (výchozí nastavení) nebo z vlastních nastavení. Je-li vybrána volba **Použit ukázková data**, použijí se obě existující asymptotické výsledky a oprava významnosti Lilliefors založená na vzorkování Monte Carlo. Je-li vybrána volba **Vlastní**, zadejte hodnoty pro hodnoty **Střední** a **Std Dev**.

### Uniformní

Je-li tato volba vybrána, můžete určit, zda mají být distribuční parametry odhadovány na základě ukázkových dat (výchozí nastavení) nebo z vlastních nastavení. Je-li vybrána volba **Použit ukázková data**, použije se test Lilliefors. Je-li vybrána volba **Vlastní**, zadejte hodnoty pro hodnoty **Minimum** a **Maximum**.

### POISSON

Je-li vybrána tato volba, zadejte hodnotu parametru **Střední hodnota**.

### Exponenciální

Je-li tato volba vybrána, můžete určit, zda mají být parametry distribuce odhadovány z průměrné hodnoty vzorku (výchozí nastavení) nebo z vlastních nastavení. Je-li vybrána volba **Použit ukázková data**, použije se test Lilliefors. Je-li vybrána volba **Vlastní**, zadejte hodnotu parametru **Střední hodnota**.

4. Volitelně můžete klepnutím na volbu **Simulace** určit parametry simulace Monte Carlo, klepnutím na volbu **Přesná** určit přesné testovací parametry nebo klepnout na volbu **Volby** pro deskriptivní statistiky, kvantily a řízení zpracování chybějících dat.

## One-Ukázka Kolmogorovan-Smirnov test: Simulace

Mají-li být určité parametry distribuce odhadovány na základě vzorku, zkouška Kolmogorov-Smirnov se již nepoužije. V těchto případech lze statistiku testu Lilliefors použít k odhadu hodnoty  $p$  za použití vzorkování Monte Carlo pro testování normality a neznámého rozptylu. Test Lilliefors se vztahuje na tři souvislé distribuce (**Normální, Exponenciální a Uniformní**). Všimněte si, že test se nepoužije, je-li základní distribuce diskrétní (**Poisson**). Test je definován pouze pro jeden vzorek inference, když nejsou zadány odpovídající parametry distribuce.

### Parametry simulace Monte Carla

#### Úroveň důvěry

Toto volitelné nastavení vynuluje úroveň intervalu spolehlivosti, kterou odhaduje test Kolmogorov-Smirnov při použití simulací Monte Carlo. Hodnota musí být v rozmezí 0 až 100. Předvolené nastavení je 99.

### Počet vzorků

Toto volitelné nastavení resetuje počet replik, které test Lilliefors používá pro vzorkování Monte Carlo. Hodnota musí být celé číslo mezi 10000 a největším počtem vzorkní hodnoty. Výchozí hodnota je 10000.

### Potlačit výsledky Monte Carla pro normální distribuci

Toto volitelné nastavení potlačuje vzorkování Monte Carlo pro normální distribuční výsledky. Při výchozím nastavení není toto nastavení vybráno (což znamená, že jsou prezentovány existující asymptotické výsledky a výsledky testu Lilliefors, které jsou založeny na vzorkování Monte Carlo).

## One-Ukázka Kolmogorov-Smirnov Test: Volby

### Statistika

Můžete vybrat jednu nebo obě souhrnné statistiky.

### Popisné

Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet případů, kdy nejsou chybějící případy.

### kvartily

Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

### Chybějící hodnoty

Řídí zpracování chybějících hodnot.

### Vyloučit případy-podle-testu

Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.

### Vyloučit případy jako litwise

Případy s chybějícími hodnotami pro každou proměnnou jsou vyloučeny ze všech analýz.

## Další funkce příkazu NPAR a další funkce (One-Ukázka Kolmogorov-Smirnov Test)

Jazyk syntaxe příkazu vám také umožňuje uvést parametry distribuce testu (s dílčím příkazem K-S).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Testy Two-Independent-Samples

Postup testování dvou nezávislých ukázek porovnává dvě skupiny případů v jedné proměnné.

**Příklad.** Byly vyvinuty nové stomatologické složené závorky, které mají být pohodlnější, lépe vypadají a poskytují rychlejší pokrok při přeskupení zubů. Chcete-li zjistit, zda mají být nové složené závorky používány tak dlouho jako staré složené závorky, bude náhodně vybráno 10 podřízených prvků, aby bylo možné nosit staré složené závorky, a dalších 10 podřízených prvků je zvoleno tak, aby nosily nové složené závorky. Z Mann-Whitney *U* testu můžete zjistit, že v průměru děti s novými závorkami nebyly muset nosit rovnátka tak dlouho jako děti s dávnými rovnátky.

**Statistika.** Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů a kvartily. Testy: Mann-Whitney *U*, Mojžiš extrémní reakce, Kolmogorov-Smirnov *Z*, Wald-Wolfowitz běží.

Poznámky k datům testů dvou nezávislých vzorků

**Data.** Použijte číselné proměnné, které lze objednat.

**Předpoklady.** Používejte nezávislé, náhodné vzorky. Mann-Whitney *U* test testuje rovnost dvou distribucí. Aby bylo možné jej použít k testování rozdílů v umístění mezi dvěma distribucemi, je třeba předpokládat, že distribuce má stejný tvar.

Chcete-li získat testy dvou nezávislých vzorků, postupujte takto:

1. Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Starší dialogová okna > 2 nezávislé ukázky ...**

2. Vyberte jednu nebo více číselných proměnných.

3. Vyberte proměnnou seskupení a klepněte na tlačítko **Definovat skupiny** , abyste rozdělili soubor do dvou skupin nebo ukázek.

## Typy testů *Two-Independent-Samples*

**Typ testu.** K dispozici jsou čtyři testy, které testují, zda dva nezávislé vzorky (skupiny) pocházejí ze stejné populace.

**Mann-Whitney U test** je nejpobulárnější ze dvou testů nezávislých vzorků. Je to stejné jako test skupiny Wilcoxon Rank a test Kruskal-Wallis pro dvě skupiny. Mann-Whitney testy, že dvě vzorkované populace jsou rovnocenné v místě. Připomínky obou skupin jsou zkombinovány a ohodnoceny vzhledem k tomu, že průměrná řadová skupina se přiřadí v případě kravatních vazeb. Počet vazeb by měl být malý vzhledem k celkovému počtu pozorování. Pokud jsou populace stejné jako v místě, je třeba se mezi těmito dvěma vzorky náhodně smíchat. Test počítá počet případů, kdy skóre ze skupiny 1 předchází skóre ze skupiny 2 a počet, kolikrát skóre ze skupiny 2 předcházelo skóre ze skupiny 1. Statistika Mann-Whitney *U* je menší z těchto dvou čísel. Zobrazí se také statistika *W* statistiky Wilcoxon. *W* je součet řad pro skupinu s menší střední hodnotou, pokud skupiny nemají stejnou střední hodnotu, v tom případě se jedná o skupinu očíslování pořadí od skupiny, která je pojmenována jako poslední v dialogovém okně Definovat skupiny dvou nezávislých ukázek.

**Test Kolmogorov-Smirnov a Wald-Wolfowitz běží test** jsou obecnější testy, které zjišťují rozdíly jak v umístění, tak ve formě tvarů distribucí. Test Kolmogorov-Smirnov je založen na maximálním absolutním rozdílu mezi pozorovanými kumulativními funkcemi distribuce pro oba vzorky. Je-li tento rozdíl výrazně větší, jsou tyto dvě distribuce považovány za odlišné. Wald-Wolfowitz run-Wolfowitz test kombinuje a hodnotí pozorování z obou skupin. Pokud jsou dva vzorky ze stejné populace, měly by být tyto dvě skupiny náhodně roztroušeny v celém žebříčku.

Zkouška **Moses extreme reakcíh** předpokládá, že experimentální proměnná bude mít vliv na některé předměty v jednom směru a další předměty v opačném směru. Zkušební testy pro extrémní odpovědi se porovnávají s řídicí skupinou. Tato zkouška se zaměřuje na rozpětí kontrolní skupiny a je měřítkem toho, jak velmi extrémní hodnoty v experimentální skupině ovlivňují rozpětí při kombinaci s řídicí skupinou. Řídicí skupina je definována hodnotou skupiny 1 v dialogovém okně Definovat skupiny dvou nezávislých ukázek. Připomínky obou skupin jsou zkombinovány a ohodnoceny. Rozpětí řídicí skupiny se vypočítá jako rozdíl mezi ohodnotami největších a nejmenších hodnot v kontrolní skupině plus 1. Vzhledem k tomu, že šance odlehlé hodnoty mohou snadno narušit rozsah rozpětí, 5% z kontrolních případů se automaticky ořízne od každého konce.

## Dva nezávislé-testy testů definují skupiny

Chcete-li rozdělit soubor do dvou skupin nebo ukázek, zadejte celočíselnou hodnotu pro skupinu 1 a další hodnotu pro skupinu 2. Případy s ostatními hodnotami jsou vyloučeny z analýzy.

## Volby testů *Two-Independent-Samples*

**Statistika.** Můžete vybrat jednu nebo obě souhrnné statistiky.

- **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet nechybějících případů.
- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy-podle-testu.** Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.
- **Vyloučit případy jako litwise.** Případy s chybějícími hodnotami pro každou proměnnou jsou vyloučeny ze všech analýz.

## Další funkce příkazu *NP*AR (*Two-Independent-Samples Tests*)

Jazyk syntaxe příkazu vám také umožňuje uvést počet případů, které mají být oříznuty pro test Moses (s dílčím příkazem MOSES).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Testy dvou souvisejících ukázek

Procedura Two-Related-Samples Tests porovnává rozdělení dvou proměnných.

**Příklad.** Obecně platí, že rodiny dostávají žádající cenu, když prodávají své domovy? Použitím Wilcoxonova testu na data po 10 domovech se můžete dozvědět, že sedm rodin dostane méně než žádající cenu, jedna rodina obdrží více než se požaduje cena, a dvě rodiny obdrží kladené ceny.

**Statistika.** Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů a kvartily. Testy: Wilcoxon signed-rank, sign, McNemar. Je-li nainstalována volba Exact Tests (dostupná pouze na operačních systémech Windows), je k dispozici i mezní test homogenity.

Aspekty dat testů souvisejících s dvěma souvisejícími s ukázkami

**Data.** Použijte číselné proměnné, které lze objednat.

**Předpoklady.** Ačkoli nejsou pro tyto dvě proměnné předpokládány žádné konkrétní distribuce, předpokládá se, že rozdělení populace ve spárovaných rozdílech je symetrické.

Chcete-li získat testy dvou souvisejících ukázek, postupujte takto:

1. Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Starší dialogová okna > 2 Související ukázky ...**

2. Vyberte jednu nebo více dvojic proměnných.

### Typy testů ve dvou souvisejících ukázkách

Testy v této sekci porovnávají distribuce dvou souvisejících proměnných. Odpovídající test, který bude použit, závisí na typu dat.

Pokud jsou vaše data spojitá, použijte test podpisu nebo test podepsaný Wilcoxon. **Test podpisu** vypočítá rozdíly mezi dvěma proměnnými pro všechny případy a klasifikuje rozdíly jako kladné, záporné nebo svázané. Jsou-li tyto dvě proměnné rozděleny podobně, počet kladných a záporných rozdílů se výrazně neliší. **Test podepsaný Wilcoxem** bere v úvahu informace o obou znaménkách rozdílů a o rozsahu rozdílů mezi dvojicemi. Vzhledem k tomu, že test podepsaný-rank Wilcoxon obsahuje více informací o datech, je výkonnější než test podpisu.

Pokud jsou vaše data binární, použijte příkaz **McNemar test**. Tento test se obvykle používá v situaci opakovaných měřitek, kdy je odezva každé jednotlivé části vyvolána dvakrát, jednou a jednou poté, co nastane uvedená událost. Test McNemar zjišťuje, zda je počáteční rychlost odezvy (před událostí) rovna konečné rychlosti odezvy (po události). Tento test je užitečný pro zjišťování změn v odpovědích v důsledku experimentální intervence v nepřed-po-cích (průmyslových) vzorů.

Jsou-li vaše data kategorická, použijte **test mezní homogenity**. Tento test je rozšířením testu McNemar z binární odpovědi na mnohoomickou odpověď. Testuje změny v odezvě (pomocí rozdělení chí-square) a je užitečný pro zjišťování změn odezvy v důsledku experimentální intervence v před-ově po-po (průmyslových) vzorů. Zkouška mezní homogenity je k dispozici pouze tehdy, pokud jste nainstalovali přesné testy.

### Volby testů ve dvou souvisejících s ukázkami

**Statistika.** Můžete vybrat jednu nebo obě souhrnné statistiky.

- **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet nechybějících případů.
- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy-podle-testu.** Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.
- **Vyloučit případy jako litwise.** Případy s chybějícími hodnotami pro každou proměnnou jsou vyloučeny ze všech analýz.

## ***Další funkce příkazu NPAR a další funkce (dva související ukázky)***

Jazyk syntaxe příkazu vám také umožňuje testovat proměnnou s každou proměnnou na seznamu.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## **Testy pro několik nezávislých vzorků**

Procedura Testy pro několik nezávislých ukázek porovnává dvě nebo více skupin případů na jedné proměnné.

**Příklad.** Jsou tři značky 100 W žárovky se liší v průměrné době, kdy se žárovky spálí? Od Kruskal-Wallis-Wallis analýza rozptylu, můžete se dozvědět, že tři značky se liší v průměrné životnosti.

**Statistika.** Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů a kvartily. Testy: Kruskal-Wallis  $H$ , medián.

Testování několika posouzení dat nezávislých vzorků

**Data.** Použijte číselné proměnné, které lze objednat.

**Předpoklady.** Používejte nezávislé, náhodné vzorky. Test Kruskal-Wallis  $H$  vyžaduje, aby testované vzorky byly podobné ve tvaru.

Získání testů pro několik nezávislých ukázek

1. Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Starší dialogová okna > K nezávislým vzorům ...**

2. Vyberte jednu nebo více číselných proměnných.

3. Vyberte proměnnou seskupení a klepněte na volbu **Definovat rozsah** , chcete-li zadat minimální a maximální celočíselné hodnoty pro proměnnou seskupení.

## ***Testy pro několik typů testů nezávislých ukázek***

K určení, zda několik nezávislých vzorků pochází ze stejné populace, jsou k dispozici tři zkoušky. Test Kruskal-Wallis  $H$  , medián testu a Jonckheere-Terpstruna test vyzkoušel všechny zkoušky, zda několik nezávislých vzorků pochází ze stejné populace.

Test **Kruskal-Wallis H test**, rozšíření testu Mann-Whitney  $U$  , je neparametrický analog jednosměrné analýzy rozptylu a zjišťuje rozdíly v umístění distribuce. **Medián test**, který je obecnějším testem (ale ne tak silný), detekuje distribuční rozdíly v umístění a tvaru. Test Kruskal-Wallis  $H$  a medián testu předpokládají, že neexistuje žádný *a priori* počet skupin  $k$  , ze kterých se odebírají vzorky.

Když je přirozené *a priori* uspořádání (vzestupné nebo sestupné) populace  $k$  , je **Jonckheere-Terpstranda test** výkonnější. Například populace  $k$  může představovat  $k$  zvýšení teploty. Hypotéza, že různé teploty produkují stejnou distribuci odezvy, je testována proti alternativě, která se zvyšuje při zvyšování teploty, a zvyšuje se tak rozsah odezvy. Zde je třeba objednat alternativní hypotézu. Jonckheere-Terpstrana je tedy nejvhodnější zkouškou pro použití. Test Jonckheere-Terpstra je k dispozici pouze v případě, že jste nainstalovali modul add-on Tests Exact Tests.

## ***Testy pro definování rozsahu pro několik nezávislých ukázek***

Chcete-li definovat rozsah, zadejte celočíselné hodnoty pro **Minimum** a **Maximum** , které odpovídají nejnižšímu a nejvyššímu kategoriím proměnné seskupení. Případy s hodnotami mimo rozsah jsou vyloučeny. Zadáte-li například minimální hodnotu 1 a maximální hodnotu 3, budou použity pouze celočíselné hodnoty od 1 do 3. Minimální hodnota musí být menší než maximální hodnota a musí být zadány obě hodnoty.

## ***Testy pro několik voleb nezávislých vzorků***

**Statistika.** Můžete vybrat jednu nebo obě souhrnné statistiky.

• **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet nechybějících případů.

- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

**Chybějící hodnoty.** Řídí zpracování chybějících hodnot.

- **Vyloučit případy-podle-testu.** Je-li zadáno několik testů, každý test se vyhodnotí samostatně pro chybějící hodnoty.
- **Vyloučit případy jako listwise.** Případy s chybějícími hodnotami pro každou proměnnou jsou vyloučeny ze všech analýz.

### **Další funkce příkazu NPAR a další funkce (K nezávislým vzorkům)**

Jazyk syntaxe příkazu také umožňuje zadat jinou hodnotu než sledovaný medián pro medián testu (s dílčím příkazem MEDIAN).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

### **Testy pro několik souvisejících ukázek**

Procedura Testy pro několik souvisejících ukázek porovnává rozdělení dvou nebo více proměnných.

**Příklad.** Umí veřejnost spojovat různé částky prestiže s lékařem, právníkem, policejním důstojníkem a učitelem? Deset lidí je požádána, aby tato čtyři povolání hodnotil v pořadí prestiže. Friedmanův test naznačuje, že veřejnost spojuje s těmito čtyřmi povoláními různé částky prestiže.

**Statistika.** Průměr, směrodatná odchylka, minimum, maximum, počet nechybějících případů a kvartily. Testy: Friedman, Kendall's  $W$  a Cochran's  $Q$ .

Testy pro několik souvisejících dat souvisejících s daty ukázek

**Data.** Použijte číselné proměnné, které lze objednat.

**Předpoklady.** Neparametrické testy nevyžadují předpoklady týkající se tvaru základní distribuce. Používejte závislé a náhodné vzorky.

Chcete-li získat testy pro několik souvisejících ukázek,

1. Z nabídky vyberte:

**Analyzovat > Neparametrické testy > Starší dialogová okna > Související ukázky K ...**

2. Vyberte dvě nebo více číselných proměnných testů.

### **Testy pro několik typů testů souvisejících ukázek**

K porovnání distribucí několika souvisejících proměnných jsou k dispozici tři testy.

**Friedman test** je neparametrický ekvivalent jednovzorkového návrhu opakovaných ukazatelů nebo dvoufaktorové analýzy rozptylu s jedním pozorováním na buňku. Friedman testuje nulovou hypotézu, že související proměnné  $k$  pocházejí ze stejné populace. Pro každý případ jsou proměnné  $k$  seřazeny od 1 do  $k$ . Statistika testu je založena na těchto řadách.

**Kendall's  $W$**  je normalizace statistiky Friedman. Kendall's  $W$  je interpretovatelný jako součinitel shody, což je míra souhlasu mezi hodnotiteli. Každý případ je soudce nebo rater, a každá proměnná je položka nebo osoba, která je posuzována. Pro každou proměnnou se vypočítá součet skupin disků rank. Kendall's  $W$  se pohybuje mezi 0 (žádná dohoda) a 1 (úplná smlouva).

**Cochran's  $Q$**  je identický s testem Friedman, ale je použitelný, když jsou všechny odpovědi binární. Tento test je rozšířením testu McNemar na situaci  $k$ -sample. Cochran's  $Q$  testuje hypotézu, že několik souvisejících dichotomických proměnných má stejnou střední hodnotu. Proměnné se měří na stejném jedinci nebo na shodných osobách.

### **Testy pro několik souvisejících statistik ukázek**

Můžete vybrat statistiky.

- **Popisné.** Zobrazí střední, směrodatnou odchylku, minimum, maximum a počet nechybějících případů.
- **Kvartil.** Zobrazí hodnoty odpovídající hodnotě 25th, 50th a 75th percentilů.

## Další funkce příkazu NPAR ZKOUŠKY (K souvisejícím ukázkám)

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Analýza vícenásobné odezvy

### Analýza vícenásobné odezvy

K dispozici jsou dvě procedury pro analýzu více dichotomie a více sad kategorií. Procedura vícenásobné frekvence odezvy zobrazuje tabulky frekvence. Procedura Kontingenční tabulky s více odpovědí zobrazuje dvourozměrné a tři rozměrové kontingenční tabulky. Před použitím jedné z procedur je třeba definovat více sad odpovědí.

**Příklad.** Tento příklad ilustruje použití více položek odpovědí v průzkumu průzkumu trhu. Data jsou fiktivní a neměla by být interpretována jako skutečná. Letecká společnost může cestujícím, kteří létají určitou trasu, vyhodnotit, aby vyhodnotili konkurenční dopravce. V tomto příkladu chce společnost American Airlines vědět o využití jiných leteckých společností na trase Chicago-New York a relativním významu plánování a služby při výběru letecké společnosti. Letová obsluha předá každému cestujícímu krátký dotazník při nástupu na palubu. První otázka zní: Kruh všechny letecké společnosti, které jste letěli nejméně jednou za posledních šest měsíců na této trase -- American, United, TWA, USAir, Ostatní. To je více otázek odpovědi, protože cestující může kroužit více než jednu odpověď. Tato otázka však nemůže být kódována přímo, protože proměnná může mít pro každý případ pouze jednu hodnotu. Chcete-li mapovat odpovědi na jednotlivé otázky, musíte použít několik proměnných. Existují dva způsoby, jak to udělat. Jednou z nich je definovat proměnnou odpovídající každé z voleb (například American, United, TWA, USAir a Other). Pokud se cestující v kruzích United, je proměnná *sjednocená* přiřazena ke kódu 1, jinak 0. Jedná se o **vícenásobnou metodu dichotomie** pro mapování proměnných. Druhým způsobem mapování odpovědí je **metoda více kategorií**, ve které odhadujete maximální počet možných odpovědí na otázku a nastavíte stejný počet proměnných, s kódy použitými pro určení prolétávané letecké společnosti. Prověrováním vzorku dotazníků můžete zjistit, že žádný uživatel v posledních šesti měsících nelétal více než tři různé letecké společnosti na této trase. Dále jste zjistili, že v důsledku deregulace leteckých společností je 10 dalších leteckých společností pojmenováno v kategorii Ostatní. Pomocí metody vícenásobné odezvy byste definovali tři proměnné, každý kódovaný jako 1 = *american*, 2 = *union*, 3 = *twa*, 4 = *usair*, 5 = *delta* tak dále. Má-li určitý osobní kroužek American a TWA, má první proměnná kód 1, druhá má kód 3 a třetí má kód chybějící hodnoty. Další cestující mohl zakroužku na Američana a vstoupil do Delta. Takže první proměnná má kód 1, druhá má kód 5 a třetí kód chybějící hodnoty. Použijete-li více metod dichotomie, na druhé straně, ukončíte 14 samostatných proměnných. Ačkoli je pro tento průzkum možné použít buď metodu mapování, metoda, kterou zvolíte, závisí na distribuci odpovědí.

### Definovat více sad odpovědí

Procedura Definovat více sad odpovědí seskupuje základní proměnné do více dichotomie a více sad kategorií, pro které můžete získat tabulky frekvencí a křížové tabulky. Můžete definovat až 20 více sad odpovědí. Každá sada musí mít jedinečný název. Chcete-li sadu odebrat, zvýrazněte ji v seznamu několika sad odpovědí a klepněte na tlačítko **Odebrat**. Chcete-li sadu změnit, zvýrazněte ji na seznamu, upravte libovolné charakteristiky definice sady a klepněte na tlačítko **Změnit**.

Můžete kódovat své základní proměnné jako dichotomie nebo kategorie. Chcete-li použít dichotomické proměnné, vyberte volbu **Dichotomie**, abyste vytvořili více dichotomické sady. Zadejte celočíselnou hodnotu pro hodnotu Counted. Každá proměnná, která má alespoň jeden výskyt započítané hodnoty, se stane kategorií násobné sady dichotomie. Vyberte volbu **Kategorie**, chcete-li vytvořit více sad kategorií se stejným rozsahem hodnot jako proměnné komponenty. Zadejte celočíselné hodnoty pro minimální a maximální hodnotu rozsahu pro kategorie sady více kategorií. Procedura sčítá každou jednotlivou celočíselnou hodnotu v inkluzivním rozsahu napříč všemi proměnnými komponenty. Prázdné kategorie nejsou tabulky s tabulátorem.

Každé vícenásobné sadě odpovědí musí být přiřazen jedinečný název až sedm znaků. Proceduru použijte předponu dolaru (\$) k názvu, který jste přiřadili. Nelze použít následující vyhrazené názvy: *casenum*, *sysmis*, *jdate*, *date*, *time*, *lengtha* *šířka*. Název vícenásobné sady odpovědí existuje pouze pro použití ve více procedurách odezvy. V jiných procedurách se nemůžete odkazovat na více názvů sad odpovědí.

Volitelně můžete zadat popisnou proměnnou popisku pro sadu více odezev. Popisek může mít délku až 40 znaků.

Chcete-li definovat více sad odezev

1. Z nabídky vyberte:

**Analyzovat > Více odpovědí > Definovat sady proměnných ...**

2. Vyberte dvě nebo více proměnných.

3. Pokud jsou vaše proměnné kódovány jako dichotomie, označte hodnotu, kterou chcete spočítat. Pokud jsou vaše proměnné kódovány jako kategorie, definujte rozsah kategorií.

4. Zadejte jedinečný název pro každou sadu více odpovědí.

5. Klepnutím na tlačítko **Přidat** přidejte více sad odpovědí do seznamu definovaných sad.

## Frekvence více odpovědí

Procedura vícenásobné frekvence odezev vytváří tabulky frekvence pro více sad odpovědí. Nejprve je třeba definovat jednu nebo více sad odpovědí (viz téma "Více sad odpovědí").

Pro více sad dichotomie se názvy kategorií zobrazené ve výstupu zobrazují z proměnných popisků definovaných pro elementární proměnné ve skupině. Pokud nejsou popisky proměnných definovány, názvy proměnných se používají jako popisky. V případě více sad kategorií pochází popisky kategorií z popisků hodnot první proměnné ve skupině. Pokud kategorie chybí pro první proměnnou jsou přítomny pro ostatní proměnné ve skupině, definujte popisek hodnoty pro chybějící kategorie.

**Chybějící hodnoty.** Případy s chybějícími hodnotami jsou vyloučeny z tabulky podle jednotlivých tabulek. Případně můžete zvolit jednu nebo obě z následujících možností:

- **Vyloučit případy na seznamu dichotomie.** Vyloučí případy s chybějícími hodnotami pro každou proměnnou z tabulky vícenásobné dichotomické sady. Toto platí pouze pro více sad odpovědí definovaných jako sady dichotomie. Ve výchozím nastavení je případ považován za chybějící pro vícenásobnou sadu dichotomie, pokud žádná z jejích proměnných komponent neobsahuje spočítanou hodnotu. Případy s chybějícími hodnotami pro některé (ale ne všechny proměnné) jsou zahrnuty v tabulacích skupiny, pokud alespoň jedna proměnná obsahuje započítanou hodnotu.
- **Vyloučit případy na základě seznamu v kategoriích.** Vyloučí případy s chybějícími hodnotami pro jakoukoli proměnnou z tabulky vícenásobné sady kategorií. Toto platí pouze pro více sad odpovědí definovaných jako sady kategorií. Ve výchozím nastavení je případ považován za chybějící pro více kategorií, pouze pokud žádná z jeho komponent nemá platné hodnoty v definovaném rozsahu.

**Příklad.** Každá proměnná vytvořená z dotazníkové otázky je elementární proměnná. Chcete-li analyzovat více položek odezvy, musíte proměnné sloučit do jednoho ze dvou typů více sad odpovědí: násobného nastavení dichotomie nebo jedné sady kategorií. Například, pokud se průzkum aerolinií zeptal, které ze tří leteckých společností (American, United, TWA) jste za posledních šest měsíců přeletěli, a použili jste dichotomické proměnné a definovali jste **více dichotomie**, každá ze tří proměnných v sadě by se stala kategorií proměnné skupiny. Počty a procentní podíly pro tři letecké společnosti se zobrazují v jedné tabulce frekvence. Zjistíte-li, že žádný respondent nezminil více než dvě letecké společnosti, mohli byste vytvořit dvě proměnné, z nichž každý má tři kódy, jeden pro každou leteckou společnost. Definujete-li **více sad kategorií**, budou hodnoty tabulovány přidáním stejných kódů do základních proměnných společně. Výsledná sada hodnot je stejná jako výsledná sada hodnot pro každou z elementárních proměnných. Například 30 odpovědí na United je součtem pěti odpovědí Spojených států amerických pro leteckou společnost 1 a 25 Spojené odpovědi na leteckou společnost 2. Počty a procentní podíly pro tři letecké společnosti se zobrazují v jedné tabulce frekvence.

**Statistika.** Frekvenční tabulky zobrazující počty, procentní části odpovědí, procentní části případů, počet platných případů a počet chybějících případů.

Více informací o frekvencích odpovědí

**Data.** Použijte více sad odpovědí.

**Předpoklady.** Počty a procenta poskytují užitečný popis pro data z libovolného rozdělení.



**Související postupy.** Procedura nastavení více sad odpovědí vám umožňuje definovat více sad odpovědí.

Jak získat více frekvencí odezev

1. Z nabídky vyberte:

**Analyzovat > Více odezev > Frekvence ...**

2. Vyberte jednu nebo více sad odpovědí.

## Kontingenční tabulky s více odpovědí

Procedura křížových tabulek s více odpověďmi překračuje definovanou více sad odpovědí, elementárních proměnných nebo jejich kombinace. Také můžete získat procentní části buněk založené na případech nebo odpovědích, upravit obsluhu chybějících hodnot nebo získat párované křížové tabulky. Nejprve je třeba definovat jednu nebo více sad odpovědí (viz "Chcete-li definovat více sad odezev").

Pro více sad dichotomie se názvy kategorií zobrazené ve výstupu zobrazují z proměnných popisků definovaných pro elementární proměnné ve skupině. Pokud nejsou popisky proměnných definovány, názvy proměnných se používají jako popisky. V případě více sad kategorií pochází popisky kategorií z popisků hodnot první proměnné ve skupině. Pokud kategorie chybí pro první proměnnou jsou přítomny pro ostatní proměnné ve skupině, definujte popisek hodnoty pro chybějící kategorie. Procedura zobrazí popisky kategorií pro sloupce na třech řádcích s až osmi znaky na řádek. Chcete-li se vyhnout dělení slov, můžete změnit pořadí řádků a sloupců nebo předefinovat popisky.

**Příklad.** S dalšími proměnnými v této proceduře může být více dichotomie a více sad kategorií sbočným způsobem. Průzkum cestujících v letecké dopravě žádá cestující o následující informace: Circle all of the following airlines you have flown at least once in the last six months (American, United, TWA). Co je důležitější při výběru letu -- plánu nebo služby? Vyberte pouze jednu položku. Po zadání údajů jako dichotomie nebo více kategorií a jejich sloučení do sady můžete převést výběr aerolinek na jednu otázku týkající se služby nebo plánu.

**Statistika.** Převedená hodnota s buňkou, řádkem, sloupcem a celkovým počtem buněk a buňkou, řádkem, sloupcem a celkovým procentem. Procenta buněk mohou být založena na případech nebo odezvách.

Aspekty více kontingenčních dat odpovědi

**Data.** Použijte více sad odpovědí nebo číselných kategorických proměnných.

**Předpoklady.** Počty a procentní části poskytují užitečný popis dat z libovolného rozdělení.

**Související postupy.** Procedura nastavení více sad odpovědí vám umožňuje definovat více sad odpovědí.

Jak získat více kontingenčních tabulek

1. Z nabídky vyberte:

**Analyzovat > Více odpovědí > Kontingenční tabulky ...**

2. Vyberte jednu nebo více číselných proměnných nebo více sad odpovědí pro každou dimenzi s kontingenčními tabulkami.

3. Definujte rozsah každé základní proměnné.

Volitelně můžete získat dvoucestný přechod pro každou kategorii řídicí proměnné nebo více sad odpovědí. Vyberte jednu nebo více položek pro seznam Vrstva (ů).

## Definovat rozsahy více odezev na kontingenční tabulky

Rozsahy hodnot musí být definovány pro každou elementární proměnnou v rámci napříč kontingenčními hodnotami. Zadejte celočíselnou minimální a maximální hodnotu kategorie, kterou chcete vytvořit tabulátem. Kategorie mimo rozsah jsou vyloučeny z analýzy. Předpokládá se, že hodnoty v rámci rozsahu zahrnutí jsou celá čísla (necelá čísla jsou zkrácena).

## Volby kontingenčních tabulek s více odpovědí

Procentuální části buňky . Počty buněk jsou vždy zobrazeny. Můžete zvolit zobrazení procentní části řádků, procentní části sloupců a dvoucestných procentních hodnot tabulky (celkem).

**Procenta založené na.** Procentuální podíl buněk můžete založit na případech (nebo respondenty). Tato volba není k dispozici, pokud jste vybrali shodu proměnných v rámci více sad kategorií. Procentuální podíl buněk můžete také založit na odpovědích. Pro více sad dichotomie se počet odpovědí rovná počtu spočítaných hodnot mezi jednotlivými případy. Pro sady více kategorií je počtem odpovědí počet hodnot v definovaném rozsahu.

**Chybějící hodnoty.** Můžete zvolit jednu nebo obě z následujících možností:

- **Vyloučit případy na seznamu dichotomie.** Vyloučí případy s chybějícími hodnotami pro každou proměnnou z tabulky vícenásobné dichotomické sady. Toto platí pouze pro více sad odpovědí definovaných jako sady dichotomie. Ve výchozím nastavení je případ považován za chybějící pro vícenásobnou sadu dichotomie, pokud žádná z jejích proměnných komponent neobsahuje spočítanou hodnotu. Případy s chybějícími hodnotami pro některé, ale ne všechny, proměnné jsou zahrnuty v tabulacích skupiny, pokud alespoň jedna proměnná obsahuje počítanou hodnotu.
- **Vyloučit případy na základě seznamu v kategoriích.** Vyloučí případy s chybějícími hodnotami pro jakoukoli proměnnou z tabulky vícenásobné sady kategorií. Toto platí pouze pro více sad odpovědí definovaných jako sady kategorií. Ve výchozím nastavení je případ považován za chybějící pro více kategorií, pouze pokud žádná z jeho komponent nemá platné hodnoty v definovaném rozsahu.

Při přechodu mezi dvěma množinami více kategorií je standardně každá proměnná v první skupině s každou proměnnou v druhé skupině a součet počtů pro každou buňku. Některé odpovědi se tedy mohou v tabulce objevit více než jednou. Můžete vybrat následující volbu:

**Odpovídat na proměnné v rámci sad odpovědí.** Dvojice první proměnné v první skupině s první proměnnou ve druhé skupině, atd. Vyberete-li tuto volbu, bude procedura základem procentní části buněk pro odpovědi spíše než respondenty. Párování není k dispozici pro více sad dichotomie nebo elementárních proměnných.

## Dodatečné funkce příkazu MULT RESPONSE

Jazyk syntaxe příkazu vám také umožňuje:

- Získejte křížové tabulky s až pěti dimenzemi (s dílčím příkazem BY ).
- Změnit volby formátování výstupu, včetně potlačení popisků hodnot (s dílčím příkazem FORMAT ).

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Výsledky vykazování

---

### Výsledky vykazování

Výpis případů a popisná statistika jsou základní nástroje pro studium a prezentaci dat. Můžete získat výpisy případů pomocí editoru dat nebo procedury Shrnout, frekvenci četnosti a popisnou statistiku s procedurou Frekvence a statistiku subpopulace s postupem Způsob. Každý z nich používá formát, který je navržen pro vyčištění informací. Chcete-li zobrazit informace v jiném formátu, souhrny sestav v řádcích a souhrny sestav ve sloupcích vám dají ovládací prvek, který potřebujete nad prezentací dat.

### Souhrny sestav v řádcích

Souhrny sestav v řádcích vytvářejí sestavy, v nichž jsou v řádcích uvedeny různé souhrnné statistiky. Seznam případů je také k dispozici, se souhrnnými statistikami nebo bez ní.

**Příklad.** Společnost s řetězem maloobchodních prodejen uchovává záznamy o zaměstnaneckých informacích, včetně platu, funkčního období a skladu a rozdělení, v nichž každý zaměstnanec pracuje. Můžete generovat sestavu, která poskytuje informace o jednotlivých zaměstnancích (výpisy) rozdělené

podle prodejny a dělení (proměnné přerušení), se souhrnnou statistikou (například průměrný plat) pro každé úložiště, oddělení a oddělení v rámci každého úložiště.

**Datové sloupce.** Vypisuje proměnné sestavy, pro které chcete zobrazit výpisy případů nebo souhrnné statistiky a řídí formát zobrazení datových sloupců.

**Přerušující sloupec.** Zobrazí seznam volitelných proměnných přerušení, které rozdělují sestavu do skupin a řídí souhrnné statistiky a formáty zobrazení sloupců přerušení. Pro více proměnných přerušení existuje samostatná skupina pro každou kategorii každé proměnné skupiny v rámci kategorií předchozí proměnné skupinové úrovně v seznamu. Přerušující proměnné by měly být samostatné kategorické proměnné, které rozdělují případy na omezený počet smysluplných kategorií. Jednotlivé hodnoty každé proměnné přerušení jsou seřazeny, seřazeny, v samostatném sloupci vlevo od všech datových sloupců.

**Sestava.** Řídí charakteristiky celkové sestavy včetně celkové souhrnné statistiky, zobrazení chybějících hodnot, číslování stránek a titulků.

**Zobrazit případy.** Zobrazí skutečné hodnoty (nebo popisky hodnot) proměnných datového sloupce pro každý případ. Tím se vytvoří sestava výpisu, která může být mnohem delší než souhrnná sestava.

**Náhled.** Zobrazí pouze první stránku sestavy. Tato volba je užitečná pro zobrazení náhledu formátu sestavy bez zpracování celé sestavy.

**Data jsou již seřazena.** Pro sestavy s proměnnými přerušení musí být datový soubor setříděn podle hodnot proměnných před generováním sestavy. Je-li váš datový soubor již seřazen podle hodnot proměnných skupiny, můžete ušetřit čas zpracování výběrem této volby. Tato volba je užitečná zejména po spuštění sestavy náhledu.

## Chcete-li získat souhrnnou sestavu: souhrny v řádcích

1. Z nabídky vyberte:

**Analyzovat > Sestavy > Souhrny sestav v řádcích ...**

2. Vyberte jednu nebo více proměnných pro datové sloupce. Pro každou vybranou proměnnou je generován jeden sloupec v sestavě.

3. Pro sestavy seřazené a zobrazené podskupinami vyberte jednu nebo více proměnných pro Přerušující sloupec.

4. Pro sestavy se souhrnnými statistikami pro podskupiny definované zalomením proměnných vyberte proměnnou přerušení v seznamu Proměnné sloupce přerušení a klepněte na **Souhrn** ve skupině sloupců skupiny pro uvedení souhrnného ukazatele (ukazatelů).

5. Pro sestavy s celkovými souhrnnými statistikami klepněte na **Souhrn**, abyste uvedli souhrnné ukazatele.

## Sloupec dat sestavy/Formát skupiny

Dialogová okna Formát řídí titulky sloupců, šířku sloupce, zarovnání textu a zobrazení datových hodnot nebo popisků hodnot. Formát datových sloupců řídí formát datových sloupců na pravé straně stránky sestavy. Formát Break Format řídí formát sloupců se zalomením na levé straně.

**Titulek sloupce.** Pro vybranou proměnnou řídí nadpis sloupce. Dlouhé nadpisy se automaticky zabalí do sloupce. Použijte klávesu Enter k ručnímu vložení zalomení řádku tam, kde chcete zalamovat nadpisy.

**Pozice hodnoty v rámci sloupce.** Pro vybranou proměnnou řídí zarovnání datových hodnot nebo popisků hodnot v rámci sloupce. Zarovnání hodnot nebo popisků neovlivňuje zarovnání záhlaví sloupců. Můžete buď odsadit obsah sloupce o uvedený počet znaků nebo zarovnat obsah na střed.

**Obsah sloupce.** Pro vybranou proměnnou řídí zobrazení buď datových hodnot, nebo definovaných popisků hodnot. Datové hodnoty jsou vždy zobrazeny pro všechny hodnoty, které nemají nadefinované popisky hodnot. (Není k dispozici pro datové sloupce v souhrnných sestavách sloupců.)

## Řádky souhrnu sestavy pro/Finální souhrnné řádky

Dialogová okna se souhrnnými řádky řídí zobrazení souhrnné statistiky pro skupiny přerušení a pro celou sestavu. Souhrnné řádky řídí statistiky podskupin pro každou kategorii definovanou proměnnou (ami) skupiny. Konečné souhrnné řádky řídí celkové statistiky, které se zobrazí na konci sestavy.

Dostupné souhrnné statistiky jsou součet, průměr, minimum, maximum, počet případů, procentní část případů nad nebo pod určenou hodnotou, procentní část případů v určeném rozsahu hodnot, směrodatná odchylka, špičatost, odchylka a šikmosti.

## Volby zalomení sestavy

Volby přerušení řídí rozteč a stránkování informací o kategorii přerušení.

**Řízení stránky.** Řídí odsazení a stránkování pro kategorie vybrané proměnné přerušení. Můžete uvést počet prázdných řádků mezi kategoriemi zářezů nebo začátek každé kategorie přerušení na nové stránce.

**Počet prázdných řádků před souhrny.** Řídí počet prázdných řádků mezi popisky skupin zářezů nebo statistikou dat a souhrnu. To je užitečné zejména pro kombinované sestavy, které obsahují jak jednotlivé výpisy případů, tak souhrnné statistiky pro kategorie přerušení; v těchto sestavách můžete vložit mezeru mezi výpisy případů a souhrnnou statistikou.

## Volby sestavy

Volby sestavy řídí zpracování a zobrazení chybějících hodnot a číslování stránek sestavy.

**Vyloučit případy s chybějícími hodnotami listwise.** Eliminuje (ze sestavy) každý případ s chybějícími hodnotami pro jakoukoli z proměnných sestavy.

**Chybějící hodnoty se zobrazí jako.** Umožňuje vám uvést symbol, který znázorňuje chybějící hodnoty v datovém souboru. Symbol může být pouze jeden znak a používá se k reprezentaci hodnot *system-missing* a *user-missing*.

**Počet stránek od.** Umožňuje vám uvést číslo stránky pro první stránku sestavy.

## Rozvržení sestavy

Rozvržení sestavy řídí šířku a délku každé stránky sestavy, umístění sestavy na stránce a vložení prázdných řádků a popisků.

**Rozvržení stránky.** Řídí okraje stránky vyjádřené v řádcích (nahore a dole) a znacích (vlevo a vpravo) a zarovnání sestav v rámci okrajů.

**Názvy stránek a zápatí.** Řídí počet řádků, které oddělují titulky stránek a zápatí od těla sestavy.

**Přerušující sloupec.** Řídí zobrazení sloupců přerušení. Je-li uvedeno více proměnných přerušení, mohou být v samostatných sloupcích nebo v prvním sloupci. Umístění všech proměnných přerušení do prvního sloupce vytvoří užší sestavu.

**Názvy sloupců.** Řídí zobrazení nadpisů sloupců, včetně podtržení titulku, mezer mezi titulky a tělem sestavy a vertikální zarovnání titulků sloupců.

**Řádky sloupce dat a zalomení zářezů.** Řídí umístění informací ve sloupci dat (hodnoty dat a/nebo souhrnné statistiky) ve vztahu k štítkům zářezů na začátku každé kategorie přerušení. První řádek dat datového sloupce může být spuštěn buď na stejném řádku jako popisek kategorie přerušení, nebo na uvedeném počtu řádků za popiskem kategorie přerušení. (Není k dispozici pro souhrnné sestavy sloupce.)

## Názvy sestav

Názvy sestav řídí obsah a umístění titulků a zápatí sestavy. Můžete uvést až 10 řádků titulků stránek a až 10 řádků zápatí stránek, s zarovnáním vlevo, zarovnáním vlevo a vpravo, na každém řádku.

Vložíte-li proměnné do nadpisů nebo do zápatí, zobrazí se v nadpisu nebo zápatí aktuální popisek hodnoty nebo hodnota proměnné. V titulcích se zobrazí popisek hodnoty odpovídající hodnotě proměnné na

začátku stránky. V zápatích se v zápatí zobrazí popisek hodnoty odpovídající hodnotě proměnné na konci stránky. Pokud zde není popisek hodnoty, zobrazí se skutečná hodnota.

**Speciální proměnné.** Speciální proměnné *DATE* a *PAGE* umožňují vložit aktuální datum nebo číslo stránky do libovolné řádky záhlaví nebo zápatí sestavy. Pokud váš datový soubor obsahuje proměnné s názvem *DATE* nebo *PAGE*, nemůžete tyto proměnné použít ve titulcích sestav nebo v zápatích.

## Souhrny sestav ve sloupcích

Souhrny sestav ve sloupcích vytváří souhrnné sestavy, ve kterých se různé souhrnné statistiky zobrazují v samostatných sloupcích.

**Příklad.** Společnost s řetězem maloobchodních prodejen uchovává záznamy o zaměstnancích, včetně platu, funkčního období a oddělení, ve kterém každý zaměstnanec pracuje. Můžete generovat sestavu, která poskytuje souhrnné statistiky platů (například střední, minimální a maximální) pro každou divizi.

**Datové sloupce.** Vypisuje proměnné sestavy, pro které požadujete souhrnné statistiky a řídí formát zobrazení a souhrnné statistiky zobrazené pro každou proměnnou.

**Zalomení sloupců.** Zobrazí seznam volitelných proměnných přerušení, které rozdělují sestavu do skupin a řídí formáty zobrazení sloupců přerušení. Pro více proměnných přerušení existuje samostatná skupina pro každou kategorii každé proměnné skupiny v rámci kategorií předchozí proměnné skupinové úrovně v seznamu. Přerušující proměnné by měly být samostatné kategorické proměnné, které rozdělují případy na omezený počet smysluplných kategorií.

**Sestava.** Řídí charakteristiky celkové sestavy, včetně zobrazení chybějících hodnot, číslování stránek a titulků.

**Náhled.** Zobrazí pouze první stránku sestavy. Tato volba je užitečná pro zobrazení náhledu formátu sestavy bez zpracování celé sestavy.

**Data jsou již seřazena.** Pro sestavy s proměnnými přerušení musí být datový soubor setříděn podle hodnot proměnných před generováním sestavy. Je-li váš datový soubor již seřazen podle hodnot proměnných skupiny, můžete ušetřit čas zpracování výběrem této volby. Tato volba je užitečná zejména po spuštění sestavy náhledu.

## Chcete-li získat souhrnnou sestavu: souhrny ve sloupcích

1. Z nabídky vyberte:

**Analyzovat > Sestavy > Souhrny sestav ve sloupcích ...**

2. Vyberte jednu nebo více proměnných pro datové sloupce. Pro každou vybranou proměnnou je generován jeden sloupec v sestavě.
3. Chcete-li změnit souhrnné měřítko pro proměnnou, vyberte proměnnou v seznamu Proměnné sloupce dat a klepněte na volbu **Souhrn**.
4. Chcete-li získat více než jeden souhrnný ukazatel pro určitou proměnnou, vyberte proměnnou ze seznamu zdrojů a přesuňte ji do seznamu Proměnné datového sloupce vícekrát, jeden pro každé souhrnné měřítko, které chcete.
5. Chcete-li zobrazit sloupec obsahující součet, průměr, poměr nebo jinou funkci existujících sloupců, klepněte na volbu **Vložte celkem**. To umístí proměnnou s názvem *total* do seznamu datových sloupců.
6. Pro sestavy seřazené a zobrazené podskupinami vyberte jednu nebo více proměnných pro Přerušující sloupce.

## Souhrnná funkce datových sloupců

Shrnující řádky řídí souhrnnou statistiku zobrazenou pro vybranou proměnnou datového sloupce.

Dostupné souhrnné statistiky jsou součet, průměr, minimum, maximum, počet případů, procentní část případů nad nebo pod určenou hodnotou, procentní část případů v určeném rozsahu hodnot, směrodatná odchylka, odchylka, špičatost a šikmosti.

## Souhrn datových sloupců pro celkový sloupec

Souhrnný sloupec řídí celkové souhrnné statistiky, které shrnují dva nebo více datových sloupců.

Dostupné celkové souhrnné statistiky jsou součet sloupců, průměr sloupců, minimum, maximum, rozdíl mezi hodnotami ve dvou sloupcích, podíl hodnot v jednom sloupci děleno hodnotami v jiném sloupci a součin hodnot sloupců vynásobených společně.

**Součet sloupců.** Sloupec *total* je součtem sloupců v seznamu Sloupec souhrnu.

**Střední hodnota sloupců.** Sloupec *total* je průměr sloupců ze seznamu Souhrn sloupců.

**Minimum sloupců.** Sloupec *total* představuje minimální počet sloupců v seznamu Sloupec souhrnu.

**Maximum sloupců.** Sloupec *total* je maximum sloupců ze seznamu Souhrn sloupců.

**1st sloupec- 2nd sloupec.** Sloupec *total* je rozdíl sloupců v seznamu Sloupec souhrnu. Seznam Sloupec souhrnu musí obsahovat přesně dva sloupce.

**1st sloupec/ 2nd sloupec.** Sloupec *total* je kvocient sloupců v seznamu Souhrn sloupců. Seznam Sloupec souhrnu musí obsahovat přesně dva sloupce.

**% 1st sloupec/ 2nd sloupec.** Sloupec *total* představuje procentní část prvního sloupce v druhém sloupci v seznamu Sloupec souhrnu. Seznam Sloupec souhrnu musí obsahovat přesně dva sloupce.

**Produkt sloupců.** Sloupec *total* je součinem sloupců ze seznamu Souhrn sloupců.

## Formát sloupce sestavy

Volby formátování sloupců dat a zarážek pro Souhrny sestav ve sloupcích jsou stejné, jako ty, které jsou popsány pro souhrny sestav v řádcích.

## Souhrny sestav ve volbách zalomení sloupců

Volby přerušení řídí mezisoučet zobrazení, řádkování a stránkování pro kategorie přerušení.

**Mezisoučet.** Řídí zobrazení mezisoučtů pro kategorie přerušení.

**Řízení stránky.** Řídí odsazení a stránkování pro kategorie vybrané proměnné přerušení. Můžete uvést počet prázdných řádků mezi kategoriemi zarážek nebo začátek každé kategorie přerušení na nové stránce.

**Počet prázdných řádků před mezisoučet.** Řídí počet prázdných řádků mezi daty kategorie přerušení a mezisoučty.

## Souhrny sestav ve volbách sloupců

Volby řídí zobrazení celkových součtů, zobrazení chybějících hodnot a stránkování v souhrnných sestavách sloupce.

**Celkový součet.** Zobrazí a označí celkový součet pro každý sloupec; zobrazí se v dolní části sloupce.

**Chybějící hodnoty.** Chybějící hodnoty můžete vyloučit ze sestavy nebo můžete vybrat jeden znak, který bude označovat chybějící hodnoty v sestavě.

## Rozvržení sestavy pro souhrny ve sloupcích

Volby rozvržení sestavy pro souhrny sestav ve sloupcích jsou stejné, jako ty, které jsou popsány pro souhrny sestav v řádcích.

## Další funkce příkazu REPORT

Jazyk syntaxe příkazu vám také umožňuje:

- Zobrazení různých součtových funkcí ve sloupcích jedné souhrnné řádky.
- Vložit řádky souhrnu do datových sloupců pro proměnné jiné než proměnnou sloupce dat nebo pro různé kombinace (složené funkce) součtových funkcí.

- Použijte funkci Medián, Mode, Frequency a Procenta jako součtovací funkce.
- Ovládejte přesněji zobrazovaný formát souhrnné statistiky.
- Vložte prázdné řádky do různých bodů v sestavách.
- Vkládat prázdné řádky za každých  $n$ -tého případu při vypisování sestav.

Vzhledem ke složitosti syntaxe produktu REPORT může být užitečné při sestavování nové sestavy se syntaxí přiblížit sestavu generovanou z dialogových oken, kopírovat a vložit příslušnou syntaxi a zpřesnit syntaxi tak, aby získávala přesnou sestavu, kterou chcete.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Analýza spolehlivosti

Analýza spolehlivosti vám umožňuje studovat vlastnosti měřítek měření a položky, které tvoří váhy. Procedura analýzy spolehlivosti vypočítá počet běžně používaných ukazatelů spolehlivosti měřítka a také poskytuje informace o vztazích mezi jednotlivými položkami v měřítku. Korelační koeficienty v rámci třídy lze použít k vypočítání odhadů spolehlivosti mezi rasobem.

Spolehlivost analýzy také poskytuje Fleiss ' Multiple Rater Kappa statistiky, které hodnotí interrater smlouvu na určení spolehlivosti mezi různými hodnotiteli. Vyšší dohoda poskytuje větší důvěru v ratingy odrážející skutečnou okolnost. The Fleiss ' Multiple Rater Kappa options are available in the [“Analýza spolehlivosti: Statistika”](#) na stránce 216 dialog.

### Příklad

Postiňuje můj dotazník spokojenost zákazníků s užitečným způsobem? Pomocí analýzy spolehlivosti můžete určit, do jaké míry se jednotlivé položky ve vašem dotazníku vzájemně souvisejí, můžete získat celkový index opakovatelnosti nebo vnitřní konzistence stupnice jako celku a můžete identifikovat problémové položky, které by měly být vyloučeny z rozsahu.

### Statistika

Popisovače pro každou proměnnou a pro stupnici, souhrnnou statistiku mezi položkami, mezipoložkovou korelatory a kovariance, odhady spolehlivosti, tabulku ANOVA, korelační koeficienty mezi jednotlivými body, Hotelův  $T^2$ , Tučův test pro aditivitu a Fleiss ' Multiple Rater Kappa.

### Modely

K dispozici jsou následující modely spolehlivosti:

#### Alfa (Cronbach)

Tento model je mírou vnitřní konzistence založené na průměrné korelaci mezi položkami.

#### Omega (McDonald's)

Tento model předpokládá, že model je jednorozměrný včetně jediného faktoru bez závislosti na lokální položce ve formě chybových kovariances. Model znamená, že kovariance dvou různých položek je součin jejich naložení.

#### Rozdělit-polovinu

Tento model rozděluje měřítko na dvě části a zkoumá korelaci mezi částmi.

#### Guttman

Tento model vypočítá dolní meze parametru Guttman pro skutečnou spolehlivost.

#### Paralelní

Tento model předpokládá, že všechny položky mají stejné rozptyly a stejné odchylky chyb napříč replikací.

#### Striktní paralelní

Tento model provádí předpoklady paralelního modelu a také předpokládá, že se mezi položkami rovnají rovny prostředky.

## Aspekty dat analýzy spolehlivosti

### Data

Data mohou být dichotomové, ordinální, nebo interval, ale data by měla být kódována numericky.

## Předpoklady

Pozorování by měla být nezávislá a chyby by měly být nekorelované mezi položkami. Každá dvojice položek by měla mít normální distribuci bivariate. Váhy by měly být aditivní, takže každá položka je lineárně spojená s celkovým skóre. Pro statistiky Fleiss "Multiple Rater Kappa" platí tyto předpoklady:

- Pro spuštění statistiky spolehlivosti musí být vybrány alespoň dvě proměnné položky.
- Jsou-li vybrány alespoň dvě proměnné hodnocení, bude vložena syntaxe Fliiss ' Multiple Rater Kappa.
- Mezi hodnotiteli není žádné spojení.
- Počet hodnotitelů je konstantní.
- Každý předmět je hodnocen stejnou skupinou obsahující pouze jeden nástroj.
- K různým neshodám nelze přiřadit žádné váhy.

## Související postupy

Chcete-li prozkoumat dimenzionalitu položek stupnice (chcete-li zjistit, zda je třeba vytvořit více než jeden konstrukt k zohlednění vzorů skóre položek), použijte analýzu faktorů nebo vícerozměrné škálování. Chcete-li identifikovat homogenní skupiny proměnných, použijte hierarchickou analýzu klastrů k proměnným klastru.

## Chcete-li získat analýzu spolehlivosti

1. Z nabídky vyberte:

**Analyzovat > Měřítko > Analýza spolehlivosti ...**

2. Vyberte dvě nebo více proměnných jako potenciální komponenty aditivního měřítka.

3. Vyberte model z rozevíracího seznamu **Model** .

4. Volitelně klepněte na volbu **Statistika** a vyberte různé statistické údaje popisující vaši položku měřítka nebo smlouvu o interrateru.

## Analýza spolehlivosti: Statistika

Můžete vybrat různé statistiky, které popisují měřítko, položky a mezihodnotící smlouvu k určení spolehlivosti mezi různými hodnotiteli. Statistika, které jsou ve výchozím nastavení vykazovány, zahrnují počet případů, počet položek a odhady spolehlivosti takto:

### Modely alfa

Coefficient alfa; pro dichotomická data se jedná o ekvivalent koeficientu Kuder-Richardson 20 (KR20).

### Modely omega

Odhad McDonaldu omega k vyhodnocení spolehlivosti.

### Rozdělené modely

Korelace mezi formuláři, Guttman split-polovina spolehlivosti, Spearman-Brown spolehlivost (stejná a nerovná délka), a koeficient alfa pro každou polovinu.

### Modely Guttman

Spolehlivost koeficientů lambda 1 až lambda 6.

### Paralelní a Striktní paralelní modely

Zkouška na proslunost modelu; odhady rozptylu chyb, běžné odchylky a skutečné odchylky; odhadovaná běžná korelace mezi položkami; odhad spolehlivosti; a nezaujatý odhad spolehlivosti.

### deskriptivní ukazatele pro

Vytváří deskriptivní statistiky pro měřítko nebo položky v různých případech.

#### Položka

Vytváří deskriptivní statistiky pro položky v různých případech.

#### Měřítko

Vytváří deskriptivní statistiky pro měřítko.



### **Změnit měřítko, pokud položka odstraněna**

Zobrazí souhrnnou statistiku porovnávající každou položku s rozsahem, který se skládá z ostatních položek. Statistika zahrnuje střední hodnotu měřítka a odchylku, pokud měla být položka odstraněna z měřítka, korelace mezi položkou a měřítkem, která se skládá z jiných položek, a alfa objektu Cronbach, pokud má být položka odstraněna z rozsahu.

### **Souhrny**

Poskytuje popisnou statistiku rozdělení položek v rámci všech položek v měřítku.

#### ***Prostředky***

Souhrnná statistika pro položku znamená. Nejmenší, největší a průměrná položka znamená rozsah a rozptyl položek znamená, že se zobrazí poměr největšího k nejmenšímu zobrazenému prostředku.

#### ***Varianty***

Souhrnná statistika pro odchylky položek. Jsou zobrazeny nejmenší, největší a průměrné odchylky položek, rozsah a odchylka rozptylů položek a poměr největšího k nejmenším odchylkám položek.

#### ***Korelace***

Souhrnná statistika pro korelace mezi položkami. Jsou zobrazeny nejmenší, největší a průměrné korelace mezi položkami, rozsah a rozptyl vzájemných korelací mezi položkami a poměr největšího k nejmenšímu vzájemnému korelacím mezi položkami.

#### ***Covariances***

Souhrnná statistika pro mezipoložkové kovariance. Jsou zobrazeny nejmenší, největší a průměrné mezipoložkové kovariance, rozsah a rozptyl mezipoložkové kovariance a poměr největšího k nejmenšímu mezipoložkového rozdílu mezi položkami.

### **Vnitřní položka**

Vytvoří matice korelací nebo kovariance mezi položkami.

### **Tabulka ANOVA**

Vytvoří testy shodných středních hodnot.

#### ***F test***

Zobrazí tabulku analýzy opakujících se ukazatelů v tabulce odchylek.

#### ***Friedman chí kvadrát***

Zobrazí koeficient Friedmanova chí-kvadrát a Kendall na shodě. Tato volba je vhodná pro data, která jsou ve formě skupin disků rank. Test chí kvadrát nahradí obvyklou zkoušku F v tabulce ANOVA.

#### ***Cochran chi-square***

Zobrazí program Cochran's Q. Tato volba je vhodná pro data, která jsou dichotomová. Statistika Q nahrazuje obvyklou statistiku F v tabulce ANOVA.

### **Interrater Agreement: Fleiss ' Kappa**

Vyhodnocuje smlouvu mezi různými hodnotiteli a určuje spolehlivost jednotlivých hodnotitelů. Vyšší dohoda poskytuje větší důvěru v ratingy odrážející skutečnou okolnost. Generalizovaná nevážená kvapla statistika měří smlouvu mezi libovolným konstantním počtem hodnotitelů za předpokladu, že:

- Pro spuštění statistiky spolehlivosti musí být určeny alespoň dvě proměnné položky.
- Musí být určeny alespoň dvě proměnné hodnocení.
- Proměnné vybrané jako položky lze také vybrat jako hodnocení.
- Mezi hodnotiteli není žádné spojení.
- Počet hodnotitelů je konstantní.
- Každý předmět je hodnocen stejnou skupinou obsahující pouze jeden nástroj.
- K různým neshodám nelze přiřadit žádné váhy.

### **Zobrazit smlouvu na jednotlivých kategoriích**

Uvádí, zda se má nebo nemá výstup na základě výstupu na jednotlivých kategoriích. Standardně výstup potlačuje odhad na všech jednotlivých kategoriích. Je-li povoleno, ve výstupu se zobrazí více tabulek.

### **Ignorovat případy řetězce**

Řídí, zda řetězcové proměnné rozlišují velikost písmen. Při výchozím nastavení hodnoty hodnocení řetězců rozlišují velká a malá písmena.

### **Popisky kategorie řetězce se zobrazují velkými písmeny**

Řídí, zda se popisky kategorií ve výstupních tabulkách zobrazují velkými nebo malými písmeny. Při výchozím nastavení je toto nastavení povoleno, což zobrazí popisky kategorie řetězce velkými písmeny.

### **Hladina významnosti asymptotické významnosti (%)**

Určuje úroveň významnosti pro asymptotické intervaly spolehlivosti. Výchozí nastavení je 95 .

## **Chybějící**

### **Vyloučit jak uživatelské, tak systémové chybějící hodnoty**

Řídí vyloučení uživatelských hodnot a systémových chybějících hodnot. Ve výchozím nastavení jsou hodnoty chybějících a systémových hodnot vyloučeny.

### **Uživatel-chybějící hodnoty jsou považovány za platné**

Je-li tato volba povolena, považuje uživatele-chybějící a systémem-chybějící hodnoty za platná data. Nastavení je ve výchozím nastavení vypnuto.

## **Vyprášení na náměstí**

Vytvoří vícehodnotový test nulové hypotézy, že všechny položky v měřítku mají stejnou střední hodnotu.

## **Test na aditivitu Tukeyho**

Vytvoří test hypotézy, že mezi položkami neexistuje žádná multiplikativní interakce.

## **Korelační koeficient Intrachess**

Produkují měřítka konzistence nebo shody hodnot v případech.

### **Model**

Vyberte model pro výpočet korelačního koeficientu uvnitř třídy. Dostupné modely jsou Two-Way Mixed, Two-Way Random a One-Way Random. Vyberte **Two-Way Mixed** , když jsou efekty lidí náhodné a efekty položky jsou pevné, vyberte **Two-Way Random** , když se efekty osob a efekty položky jsou náhodné, nebo vyberte **One-Way Random** , když jsou efekty lidí náhodné.

### **Typ**

Vyberte typ indexu. Dostupné typy jsou Soudržnost a Absolutní dohoda.

### **Interval spolehlivosti (%)**

Uveďte úroveň pro interval spolehlivosti. Předvolba je 95%.

### **Testová hodnota**

Uveďte hypotetizovanou hodnotu koeficientu pro test hypotézy. Tato hodnota je hodnota, na kterou se porovnává pozorovaná hodnota. Výchozí hodnota je 0.

## **Dodatečné funkce příkazu RELIABILITY**

Jazyk syntaxe příkazu vám také umožňuje:

- Číst a analyzovat korelační matici.
- Zapište korelační matici pro pozdější analýzu.
- Uveďte rozdělení jiné než rovnoměrné půlky pro rozdělenou polovinu metody.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## **Vážené Kappy**

Cohen je vážený kappa se široce používá v křížové klasifikaci jako měřítko dohody mezi pozorovanými hodnotiteli. Je to vhodný index dohody, jsou-li ratingy nominálními stupnicemi a bez struktury objednávky. Vývoj Cohena váženého kappa byl motivován faktorem, že některá přiřazení v rezervních tabulkách mohou být větší gravitace než ostatní. Statistika se opírá o předem definované váhy buněk odrážející buď shodu, nebo nesouhlasné neshody.

Postup Vážené Kappy poskytuje volby pro odhadování vážené kappy Cohen, což je důležitá generalizace statistiky kappla, která měří souhlas dvou řadových předmětů se stejnými kategoriemi.

**Poznámka:** Procedura Weighted Kappa nahrazuje funkci, která již byla dříve poskytována rozšířením `STATS WEIGHTED KAPPA.spe`.

#### Příklad

Existují situace, kdy by rozdíly mezi hodnotiteli neměly být považovány za stejně důležité. Příkladem může být ve zdravotnickém průmyslu, kde více lidí sbírá výzkum nebo klinické údaje. V takových případech může spolehlivost dat přijít na otázku vzhledem k variabilitě mezi těmito shromažďováním údajů.

#### Statistika

Cohen je vážený kapa, lineární stupnice, kvadratické měřítko, asymptotický interval spolehlivosti.

## Aspekty týkající se vážených aplikací Kappa

#### Data

Dvousměrná tabulka, která je založena na aktivní datové sadě, je nezbytná k odhadu statistiky Cohen ve váženém stavu.

Proměnné hodnocení musí být stejného typu (všechny řetězce nebo všechny číslice).

Odhad váhy Cohen má smysl pouze tehdy, když kategorie dvou ratingových proměnných reprezentovaných řádkem a sloupcem tabulky jsou správně uspořádané (pro dvojici číselných proměnných, je použita numerická objednávka; pro dvojici řetězcových proměnných se použije abecední pořadí).

#### Předpoklady

Když jsou vybrány smíšené páry proměnných, Cohen je vážený kapa se neodhaduje.

Předpokládá se, že ratingové proměnné sdílejí stejnou sadu kategorií.

## Chcete-li získat analýzu váženého Kappa

1. Z nabídky vyberte:

**Analyzovat > Měřítko > Vážené Kappy ...**

2. Vyberte dva nebo více řetězcových nebo číselných proměnných, které mají být zadány ve tvaru **Hodnotící nástroje po dvojicích**.

**Poznámka:** Je třeba vybrat buď všechny řetězcové proměnné, nebo všechny číselné proměnné.

3. Volitelně povolte nastavení **Určit hodnotitele pro řádky a sloupce**, které řídí zobrazení rámců po dvojicích nebo rachteček řádků/sloupců.

- Je-li tato volba povolena, jsou potlačovány rastery po dvojicích a zobrazují se hodnotitelé řádků a sloupců. Uživatelské rozhraní se aktualizuje a bude poskytovat pole **Rozpětí řádků a Sloupec (y) sloupce** (efektivně nahradí pole **Hodnotníky po dvojicích**).
- Když je zakázáno, rasterní řádků a sloupců jsou potlačovány a zobrazovače po dvojicích se zobrazují (výchozí nastavení)

Je-li aktivována volba **Určit hodnotitele pro řádky a sloupce**, zadejte alespoň jednu proměnnou pro **Roter rater (s)** a **Column racter (s)**.

**Poznámka:** Pokud obsahují obě proměnné **Ročovací rater (s)** a **Sloupec (y) sloupce** pouze jednu proměnnou, nemohou být vybrané proměnné stejné pro obě proměnné.

4. Volitelně můžete klepnutím na volbu **Kritéria** určit měřítko váhy a nastavení chybějících hodnot, nebo můžete klepnutím na tlačítko **Tisknout** určit nastavení formátu zobrazení a kontingenční tabulky.

## Vážené Kappa: Kritéria

Dialogové okno Kritéria poskytuje volby pro uvedení odhadu statistiky Cohen na základě vážených kappů.

### **Měřitko váhy**

Poskytuje volby pro určení lineární nebo kvadratické váhy pro smlouvu. Použití váhy lineárního (také známého jako Cicchetti-Allison) je výchozí nastavení.

### **Chybějící hodnoty**

Poskytuje volby pro odebrání případů s chybějícími hodnotami na dvojicích a zacházení s uživateli-chybějící hodnoty jako platné.

### **Proměnné hodnocení řetězce jsou citlivé na velikost písmen**

Je-li tato volba vybrána, budou řetězcové proměnné považovány za citlivé na velikost písmen

### **Interval spolehlivosti Asymptotické spolehlivosti (%)**

Toto volitelné nastavení uvádí úroveň důvěry pro odhad asymptotických intervalů spolehlivosti. Musí se jednat o jednoduchou hodnotu typu double mezi 0 a 100 (95 je výchozí nastavení).

## **Vážené Kappa: Tisk**

Dialogové okno Tisk poskytuje volby pro řízení tabulek crosstaulace.

### **Zobrazit a formátovat**

Poskytuje volby pro řízení zobrazení a formátování tabulky crosstablace.

#### **Kategorie hodnocení se zobrazují ve vzestupném pořadí**

Je-li vybráno, kategorie hodnocení v tabulkách crosstaulation se zobrazí ve vzestupném pořadí. Není-li toto nastavení vybráno, kategorie hodnocení v tabulkách crosstablation se zobrazí v sestupném pořadí. Nastavení je ve výchozím nastavení povoleno.

#### **Popisky kategorie řetězce se zobrazují velkými písmeny**

Je-li tato volba vybrána, tabulka s kontingenční tabulkou se zobrazí jako velká písmena. Není-li toto nastavení vybráno, tabulky crosstablation se zobrazí jako malá písmena. Nastavení je ve výchozím nastavení povoleno.

### **Skrocení**

Poskytuje volby pro určení ratingových proměnných, které se používají v kontingenčních cubaku. Ve výchozím nastavení nejsou povolena nastavení crosstabulace, která potlačuje křížové zpracování všech ratingových proměnných.

#### **Zobrazit crosstakulaci ratingových proměnných**

Je-li tato volba vybrána, toto nastavení umožňuje křížové zpracování všech nebo uživatelem zadaných ratingových proměnných.

#### **Zahrnout všechny proměnné hodnocení**

Je-li vybrána tato volba, jsou tabulky s kontingenčními tabulkami tištěny pro všechny definované páry ratingových proměnných.

#### **Zahrnout uživatelem zadané hodnoty hodnocení**

Jsou-li vybrány, použijte pole **Dostupné proměnné**, **Rozpětí řádků** a **Sloupec (s) sloupce** k výběru, které proměnné hodnocení se zahrnou do tabulek crosstablation.

## **Multidimenzionální škálování**

Vícedimenzionální změna měřítka se pokouší najít strukturu v sadě vzdáleností ukazatelů mezi objekty nebo případy. Tato úloha se provádí přiřazením pozorování do specifických lokalit v koncepčním prostoru (obvykle dvourozměrném nebo trojrozměrném) tak, že vzdálenosti mezi body v prostoru odpovídají daným nepodobnostem co nejlépe. V mnoha případech mohou být dimenze tohoto konceptuálního prostoru interpretovány a použity k dalšímu pochopení vašich dat.

Pokud máte objektivně naměřené proměnné, můžete použít vícerozměrné škálování jako techniku redukce dat (Procedura Multidimensional Scaling na vás v případě potřeby vypočítá vzdálenosti od více dat od více dodavatelů dat). Vícerozměrné škálování může být také použito na subjektivní hodnocení nepodobnosti mezi objekty nebo koncepcemi. Kromě toho může procedura Multidimensional Scaling zpracovávat nepodobnost dat z více zdrojů, protože můžete mít několik hodnotitelů nebo dotazovaných respondentů.

**Příklad.** Jak lidé vnímají vztahy mezi různými auty? Pokud máte data od respondentů, kteří označují podobnost mezi různými způsoby a modely automobilů, lze použít vícerozměrné škálování k identifikaci rozměrů, které popisují vnímání spotřebitelů. Můžete například zjistit, že cena a velikost vozidla definuje dvourozměrný prostor, který odpovídá podobnostem, které vaše respondenti vykazují.

**Statistika.** Pro každý model: datová matice, optimálně škálované datové matice, S-stress (Young's), stres (Kruskal's), RSQ, stimulační souřadnice, průměrný stres a RSQ pro každý podnět (modelyRMDS). U individuálních rozdílů (INDSCAL) models: subject weights and weirdness index for each subject. Pro každou matici v replikovaných modelech multidimenzionálního škálování: stres a RSQ pro každý stimul. Ploty: stimulační souřadnice (dvojměrný nebo trojměrný), rozptyl rozdílů oproti vzdálenosti.

Aspekty použití více rozměrových dat

**Data.** Pokud jsou data nepodobná datům, měly by být všechny znepodobnosti kvantitativní a měly by být měřeny ve stejné metrice. Pokud jsou data vícerozměrná, proměnné mohou být kvantitativní, binární nebo spočítat data. Rozšiřitelnost proměnných je důležitým problémem-rozdíly ve škálování mohou ovlivnit vaše řešení. Pokud mají proměnné velké rozdíly ve škálování (například jedna proměnná se měří v dolarech a druhá proměnná se měří v rocích), zvažte jejich standardizaci (tento proces může být proveden automaticky pomocí procedury Multidimensional Scaling).

**Předpoklady.** Procedura Multidimensional Scaling je relativně prostá distribuční assumptions. Ujistěte se, že jste vybrali příslušnou úroveň měření (ordinal, interval, nebo ratio) v dialogovém okně Multidimensional Scaling Options tak, aby byly výsledky vypočteny správně.

**Související postupy.** Je-li vaším cílem snížení objemu dat, je třeba zvážit alternativní metodu analýzy faktoru, a to zejména tehdy, jsou-li vaše proměnné kvantitativní. Chcete-li identifikovat skupiny podobných případů, zvažte doplnění vícerozměrové analýzy škálování pomocí hierarchického nebo *k*-znamená klastrovou analýzu.

Získání vícedimenzionální analýzy škálování

1. Z nabídky vyberte:

**Analyzovat > Měřítka > Vícerozměrné škálování ...**

2. Vyberte alespoň čtyři číselné proměnné pro analýzu.

3. Ve skupině Vzdálenosti vyberte buď volbu **Data jsou vzdálenostmi**, nebo volbu **Vytvořit vzdálenosti z dat**.

4. Vyberete-li volbu **Vytvořit vzdálenosti z dat**, můžete také vybrat proměnnou seskupení pro jednotlivé matice. Proměnná seskupení může být číselná nebo řetězcová.

Volitelně můžete také:

- Uved'te tvar matice vzdálenosti, když jsou data vzdálena.
- Uved'te vzdálenost, která se má použít při vytváření vzdáleností od dat.

## Vícerozměrný tvar měřítka dat

Pokud vaše aktivní datová sada představuje vzdálenost mezi sadou objektů nebo představuje vzdálenosti mezi dvěma sadami objektů, určete tvar matice dat, abyste získali správné výsledky.

*Poznámka:* Nemůžete-li v dialogovém okně Model určit podmíněnost řádků, nemůžete vybrat **Čtvercový symetrický**.

## Ukazatel vytvoření vícerozměrného škálování

Vícedimenzionální škálování používá data nesoupodobnosti k vytvoření řešení škálování. Pokud jsou vaše data vícerozměrná data (hodnoty naměřených proměnných), musíte pro výpočet vícerozměrného řešení škálování vytvořit data dispodobnosti. Můžete uvést podrobnosti o vytváření dispodobnosti ukazatelů z vašich dat.

**Ukazatel.** Umožňuje určit míru nepodobnosti pro analýzu. Vyberte jednu alternativu ze skupiny ukazatelů, odpovídající vašemu typu dat, a pak zvolte jedno z ukazatelů z rozevíracího seznamu odpovídajícího typu ukazatele. Dostupné alternativy jsou:

- **Interval.** euklidovský odstup, Squared euklidovský odstup, Chebychev, Block, Minkowského, nebo Upravené.
- **Počty.** Chi-kvadrát ukazatel nebo Phi-square opatření.
- **Binární.** euklidovský odstup, Squared euklidovský vzdálenost, Velikost rozdíl, Vzor rozdíl, Odchylka, nebo Lance a Williams.

**Vytvořit matici vzdálenosti.** Umožňuje zvolit jednotku analýzy. Alternativy jsou Mezi proměnnými nebo Mezi případy.

**Transformace hodnot.** V některých případech, například při měření proměnných na velmi odlišných měřících, můžete chtít standardizovat hodnoty před výpočtem proximities (nevztahuje se na binární data). Vyberte standardizační metodu z rozevíracího seznamu Standardizace. Není-li nutná žádná standardizace, vyberte volbu **Žádná**.

## Vícerozměrný model škálování

Správný odhad vícerozměrového modelu škálování závisí na aspektech dat a na samotném modelu.

**Úroveň měření.** Umožňuje vám uvést úroveň vašich dat. Alternativy jsou Ordinální, Interval nebo Poměr. Pokud jsou vaše proměnné ordinální, vyberte volbu **Odvázet vázané pozorování**, aby proměnné byly považovány za souvislé proměnné, takže vazby (stejně hodnoty pro různé případy) jsou řešeny optimálně.

**Podmíněnost.** Umožňuje vám uvést, která porovnání jsou smysluplná. Alternativy jsou Matice, Řádek nebo Nepodmíněné.

**Dimenze.** Umožňuje určit dimenzionalitu řešení škálování. Jedno řešení se vypočítá pro každé číslo v rozsahu. Uveďte celá čísla mezi 1 a 6; minimální hodnota 1 je povolena pouze tehdy, když vyberete **euklidovský vzdálenost** jako model škálování. Pro jedno řešení uveďte stejné číslo jako minimum a maximum.

**Model škálování.** Umožňuje vám uvést předpoklady, podle kterých se provádí změna měřítka. Dostupné alternativy jsou euklidovský vzdálenost nebo Individuální rozdíly euklidovský vzdálenost (také známý jako INDSCAL). Pro jednotlivé rozdíly Euklidovský model vzdálenosti, můžete vybrat **Povolit záporné váhy předmětu**, je-li to vhodné pro vaše data.

## Vícerozměrné volby škálování

Můžete určit volby pro vícerozměrnou analýzu škálování.

**Obrazovka.** Umožňuje vám vybrat různé typy výstupů. Dostupné volby jsou Sklady skupin, Jednotlivé předměty předmětů, matice dat a souhrn údajů o modelu a volbách.

**Kritéria.** Umožňuje určit, kdy má být iterace zastavena. Chcete-li výchozí hodnoty změnit, zadejte hodnoty pro konvergenci **S-stress**, **Minimum s-stress valuea** **Maximum iterations**.

**Zacházet se vzdáleností kratším než n jako chybějící.** Vzdálenosti, které jsou nižší než tato hodnota, jsou vyloučeny z analýzy.

## Další funkce příkazu ALSCAL

Jazyk syntaxe příkazu vám také umožňuje:

- Použijte tři další typy modelu, známé jako ASCAL, AINDS, a GEMSCAL v literatuře o vícedimenzionálním škálování.
- Provádět polynomiální transformace na interval a poměr dat.
- Analyzujte podobnosti (spíše než vzdálenosti) s ordinálními daty.
- Analýza nominálních dat.

- Ukládat různé souřadnice souřadnic a váhy do souborů a číst je zpětně pro analýzu.
- Omezuje vícerozměrné skládání.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference* .

## Poměr statistik

Procedura Poměrové statistiky poskytuje souhrnný seznam souhrnných statistik pro popis poměru mezi dvěma proměnnými měřítky.

Výstup lze řadit podle hodnot proměnné seskupení ve vzestupném nebo sestupném pořadí. Sestava poměru statistiky může být potlačena ve výstupu a výsledky lze uložit do externího souboru.

### Příklad

Je v poměru mezi cenou ocenění a prodejní cenou nemovitostí v každém z pěti okresů dobré jednotnosti? Z výstupu se můžete dozvědět, že rozložení poměrů se značně liší od krajské a okresní.

### Statistika

Medián, střední, vážený průměr, intervaly spolehlivosti, koeficient rozptylu (COD), variační koeficient (COV), variační koeficient variace, střední hodnota rozptylu v průměru (PRB), rozdíl v cenovém rozdílu (PRD), směrodatná odchylka, průměrná absolutní odchylka (AAD), rozsah, minimální a maximální hodnoty a index koncentrace vypočtený pro uživatelem specifikovaný rozsah nebo procento v rámci střední hodnoty poměru.

## Aspekty dat

### Data

Použijte číselné kódy nebo řetězce k seskupení proměnných seskupení (měření nominální nebo ordinální úrovně).

### Předpoklady

Proměnné, které definují čitatele a jmenovatele tohoto poměru, by měly být proměnné stupnice, které berou kladné hodnoty.

## Získání poměru statistik

1. Z nabídky vyberte:

**Analyzovat > Deskriptivní statistika > Poměr ...**

2. Vyberte proměnnou čitatele.

3. Vyberte proměnnou jmenovatele.

Volitelně:

- Vyberte proměnnou seskupení a zadejte řazení skupin ve výsledcích.
- Zvolte, zda mají být zobrazeny výsledky v prohlížeči.
- Zvolte přidání klíčového slova **N** do dílčího příkazu **PRINT** . Toto nastavení přidá velikost vzorku do tabulky statistiky ve výstupu.
- Vyberte, zda chcete výsledky uložit do externího souboru pro pozdější použití, a zadejte název souboru, do kterého se uloží výsledky.

## Poměrové statistiky: Statistika

### Střední tendnost

Měřítka centrální tendence jsou statistiky, které popisují rozdělení poměrů.

#### Medián

Hodnota taková, že počet poměrů, které jsou menší než tato hodnota, a počet poměrů, které jsou větší než tato hodnota, jsou stejné.

### **Střední**

Výsledek součtování poměrů a vydělením výsledku celkovým počtem poměrů.

### **Vážený průměr**

Výsledek vydělením střední hodnoty čitatele střední hodnotou jmenovatele. Vážené průměrné hodnoty jsou také průměrem poměrů váženém jmenovatelem.

### **Intervaly spolehlivosti pro PRB a ukazatele centrální tendence**

Zobrazí intervaly spolehlivosti pro PRB, střední hodnotu, medián a vážený průměr (je-li požadován). Uveďte hodnotu, která je větší než nebo rovna 0 a menší než 100 jako úroveň důvěry. Předvolené nastavení je 95%.

### **Disperze**

Tyto statistické údaje měří množství variace nebo rozptylu v pozorovaných hodnotách.

### **AADINA**

Průměrná absolutní odchylka je výsledkem součtu absolutních odchylek poměrů o střední hodnotě a vydělení výsledku celkovým počtem poměrů.

### **TRESKA OBECNÁ**

Koeficient disperze je výsledkem vyjadřování průměrné absolutní odchylky vyjádřené jako procentní podíl střední hodnoty.

### **KV.**

Variační koeficient. Varianta variačního variačního faktoru je výsledkem vyjádření kořene střední kvadratické odchylka od mediánu střední hodnoty jako procento střední hodnoty. Variační koeficient střední hodnoty je výsledek vyjadřování směrodatné odchylky jako procentní podíl střední hodnoty.

### **PB**

Posudky vztahující se k ceně. Index toho, zda je posouzení cenových poměrů systematicky vyšší nebo nižší u vysoce ceněných nemovitostí. Obléká procentní rozdíly v posouzení poměru mezi mediálním poměrem a logaritmickou hodnotou v rozmezí 2-logaritmická hodnota vypočítaná jako průměr prodejních cen a poměr hodnocených hodnot k průměrnému poměru. Poskytuje procentuální změnu poměrů posouzení pro změnu hodnoty 100 procent.

### **PRD**

Cenový rozdíl, známý také jako index regresivity, je výsledkem vydělení střední hodnoty váženého průměru.

### **Směrodatná odchylka**

Směrodatná odchylka je výsledkem součtování čtvercové odchylky poměrů o průměru, vydělením výsledku celkovým počtem poměrů minus jedna a s kladnou druhou odmocninou.

### **Rozsah**

Rozsah je výsledkem odečtení minimálního poměru od maximálního poměru.

### **Minimální**

Minimální poměr je nejmenší poměr.

### **Maximální**

Maximální poměr je největší.

### **Index koncentrace**

Koeficient koncentrace měří procento z převodových poměrů, které spadají do intervalu. Lze ji vypočítat dvěma různými způsoby:

#### **Mezi proporcí**

Interval je definován explicitně uvedením nízkých a vysokých hodnot intervalu. Zadejte hodnoty pro nízký podíl a vysoký podíl a klepněte na tlačítko **Přidat**, abyste získali interval.

#### **V procentním podílu mediánu**

Zde je interval definován implicitně uvedením procentního podílu mediánu. Zadejte hodnotu v rozsahu 0 až 100 a klepněte na tlačítko **Přidat**. Spodní konec intervalu je roven  $(1 - 0.01 \times \text{hodnota}) \times \text{medián}$  a horní konec se rovná  $(1 + 0.01 \times \text{hodnota}) \times \text{medián}$ .



## Poměr statistik

Procedura Poměrové statistiky poskytuje souhrnný seznam souhrnných statistik pro popis poměru mezi dvěma proměnnými měřítka.

Výstup lze řadit podle hodnot proměnné seskupení ve vzestupném nebo sestupném pořadí. Sestava poměru statistiky může být potlačena ve výstupu a výsledky lze uložit do externího souboru.

### Příklad

Je v poměru mezi cenou ocenění a prodejní cenou nemovitostí v každém z pěti okresů dobré jednotnosti? Z výstupu se můžete dozvědět, že rozložení poměrů se značně liší od krajské a okresní.

### Statistika

Medián, střední, vážený průměr, intervaly spolehlivosti, koeficient rozptylu (COD), variační koeficient (COV), variační koeficient variace, střední hodnota rozptylu v průměru (PRB), rozdíl v cenovém rozdílu (PRD), směrodatná odchylka, průměrná absolutní odchylka (AAD), rozsah, minimální a maximální hodnoty a index koncentrace vypočtený pro uživatelem specifikovaný rozsah nebo procento v rámci střední hodnoty poměru.

## Aspekty dat

### Data

Použijte číselné kódy nebo řetězce k seskupení proměnných seskupení (měření nominální nebo ordinální úrovně).

### Předpoklady

Proměnné, které definují čitatele a jmenovatele tohoto poměru, by měly být proměnné stupnice, které berou kladné hodnoty.

## Získání poměru statistik

1. Z nabídky vyberte:

**Analyzovat > Deskriptivní statistika > Poměr ...**

2. Vyberte proměnnou čitatele.

3. Vyberte proměnnou jmenovatele.

Volitelně:

- Vyberte proměnnou seskupení a zadejte řazení skupin ve výsledcích.
- Zvolte, zda mají být zobrazeny výsledky v prohlížeči.
- Zvolte přidání klíčového slova **N** do dílčího příkazu **PRINT**. Toto nastavení přidá velikost vzorku do tabulky statistiky ve výstupu.
- Vyberte, zda chcete výsledky uložit do externího souboru pro pozdější použití, a zadejte název souboru, do kterého se uloží výsledky.

## Poměrové statistiky: Statistika

### Střední tendnost

Měřítka centrální tendence jsou statistiky, které popisují rozdělení poměrů.

#### Medián

Hodnota taková, že počet poměrů, které jsou menší než tato hodnota, a počet poměrů, které jsou větší než tato hodnota, jsou stejné.

#### Střední

Výsledek součtování poměrů a vydělením výsledku celkovým počtem poměrů.

#### Vážený průměr

Výsledek vydělením střední hodnoty čitatele střední hodnotou jmenovatele. Vážené průměrné hodnoty jsou také průměrem poměrů váženém jmenovatelem.

### Intervaly spolehlivosti pro PRB a ukazatele centrální tendence

Zobrazí intervaly spolehlivosti pro PRB, střední hodnotu, medián a vážený průměr (je-li požadován). Uveďte hodnotu, která je větší než nebo rovna 0 a menší než 100 jako úroveň důvěry. Předvolené nastavení je 95%.

### Disperze

Tyto statistické údaje měří množství variace nebo rozptylu v pozorovaných hodnotách.

#### AADINA

Průměrná absolutní odchylka je výsledkem součtu absolutních odchylek poměrů o střední hodnotě a vydělení výsledku celkovým počtem poměrů.

#### TRESKA OBECNÁ

Koeficient disperze je výsledkem vyjadřování průměrné absolutní odchylky vyjádřené jako procentní podíl střední hodnoty.

#### KV.

Variační koeficient. Varianta variačního variačního faktoru je výsledkem vyjádření kořene střední kvadratické odchylka od mediánu střední hodnoty jako procento střední hodnoty. Variační koeficient střední hodnoty je výsledek vyjadřování směrodatné odchylky jako procentní podíl střední hodnoty.

#### PB

Posudky vztahující se k ceně. Index toho, zda je posouzení cenových poměrů systematicky vyšší nebo nižší u vysoce ceněných nemovitostí. Obléká procentní rozdíly v posouzení poměru mezi mediálním poměrem a logaritmickou hodnotou v rozmezí 2-logaritmická hodnota vypočítaná jako průměr prodejních cen a poměr hodnocených hodnot k průměrnému poměru. Poskytuje procentuální změnu poměrů posouzení pro změnu hodnoty 100 procent.

#### PRD

Cenový rozdíl, známý také jako index regresivity, je výsledkem vydělení střední hodnoty váženého průměru.

#### Směrodatná odchylka

Směrodatná odchylka je výsledkem součtování čtvercové odchylky poměrů o průměru, vydělením výsledku celkovým počtem poměrů minus jedna a s kladnou druhou odmocninou.

#### Rozsah

Rozsah je výsledkem odečtení minimálního poměru od maximálního poměru.

#### Minimální

Minimální poměr je nejmenší poměr.

#### Maximální

Maximální poměr je největší.

### Index koncentrace

Koeficient koncentrace měří procento z převodových poměrů, které spadají do intervalu. Lze ji vypočítat dvěma různými způsoby:

#### Mezi proporcí

Interval je definován explicitně uvedením nízkých a vysokých hodnot intervalu. Zadejte hodnoty pro nízký podíl a vysoký podíl a klepněte na tlačítko **Přidat**, abyste získali interval.

#### V procentním podílu mediánu

Zde je interval definován implicitně uvedením procentního podílu mediánu. Zadejte hodnotu v rozsahu 0 až 100 a klepněte na tlačítko **Přidat**. Spodní konec intervalu je roven  $(1 - 0.01 \times \text{hodnota}) \times \text{median}$  a horní konec se rovná  $(1 + 0.01 \times \text{hodnota}) \times \text{median}$ .

## Výkres P-P

---

Procedura P-P zakresluje proceduru pravděpodobnosti jedné nebo více proměnných posloupnosti nebo časové řady. Proměnné mohou být standardizovány, odlišovat a transformovány před vykreslením.

Dostupné testovací distribuce zahrnují beta, chí kvadrát, exponenciální, gama, polonormální, Laplaceova, Logistický, logormální, normální, pareto, Studentovo t, Weibullovo a uniformní. V závislosti na vybrané distribuci můžete určit stupně volnosti a další parametry.

- Získáte pravděpodobnostní grafy pro transformované hodnoty. Volby transformace zahrnují přirozený protokol, standardizace hodnot, rozdílu a sezónně odlišnost.
- Můžete uvést metodu pro výpočet očekávaných distribucí a pro řešení "vazeb" (více pozorování se stejnou hodnotou).

### Rozdělení testu

Uveďte typ rozdělení pro vaše data. Rozevírací seznam poskytuje následující volby:

#### Beta

*Distribuce beta.* Parametry shape1 a shape2  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty střední hodnoty odběru vzorku a vychýlení vzorku. Všechna pozorování musí být v rozsahu 0 až 1 včetně.

#### Chí-kvadrát

*Distribuce chí-kvadrát.* Je třeba určit stupně volnosti ( $df$ ). Záporná pozorování nejsou povolena.

#### Exponenciální

*Exponenciální rozdělení.* Parametr měřítka  $a$  musí být kladný. Není-li parametr zadán, produkt DISTRIBUTION jej odhadne ze střední hodnoty. Záporná pozorování nejsou povolena.

#### Gama

*Gamma rozdělení.* Parametry tvaru a měřítka  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty střední hodnoty odběru vzorku a vychýlení vzorku. Záporná pozorování nejsou povolena.

#### Polonormální

*Half-normální rozdělení.* Předpokládá se, že data jsou ve volném umístění nebo jsou centralizovaná. (Lokalita parameter=0.) Můžete uvést parametr měřítka  $a$  nebo nechat DISTRIBUTION odhadnout to pomocí metody maximální pravděpodobnosti.

#### Laplaceovo

*Laplaceovo nebo dvojitě exponenciální rozdělení.* LAPLACE vezme místo a parametr scale ( $a$  a  $b$ ). Parametr měřítka ( $b$ ) musí být kladný. Nejsou-li parametry zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty vzorku a směrodatné odchylky vzorků.

#### Logistický

*Logistická distribuce.* LOGISTIC vezme místo a parametr scale ( $a$  a  $b$ ). Parametr měřítka ( $b$ ) musí být kladný. Nejsou-li parametry zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty vzorku a směrodatné odchylky vzorků.

#### Logaritmicko-normální

*Logaritmicko-normální rozdělení.* Parametry měřítka a tvaru  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední a směrodatné odchylky přirozeného logaritmu dat vzorku. Záporná pozorování nejsou povolena.

#### Normální

*Normální rozdělení.* Parametr umístění  $a$  může být libovolná číselná hodnota, zatímco parametr měřítka  $b$  musí být kladný. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty střední hodnoty odběru vzorku a vychýlení vzorku.

#### Paretův

*Paretova distribuce.* Parametry prahové hodnoty a tvaru  $a$  a  $b$  musí být kladné. Pokud nejsou zadány, předpokládá DISTRIBUTION, že  $a$  se rovná minimální hodnotě pozorování a  $b$  podle maximální pravděpodobnosti. Záporná pozorování nejsou povolena.

#### Student t

*Studentova t rozdělení.* Je třeba určit stupně volnosti ( $df$ ).

#### Uniformní

*Rovnoměrné rozdělení.* UNIFORM vezme minimální a maximální parametr ( $a$  a  $b$ ). Parametr  $a$  musí být roven nebo větší než  $b$ . Nejsou-li parametry uvedeny, DISTRIBUTION je předpokládá z ukázkových dat.

## WEIBULL

*Weibullovo rozdělení.* Parametry měřítka a tvaru  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne s použitím metody nejmenších čtverců. Záporná pozorování nejsou povolena.

### Parametry distribuce

Poskytuje strategii distribuce a volby parametrů.

#### Odhadnout z dat

Je-li tato volba vybrána, toto nastavení odhaduje distribuční parametry na základě dat a vybraného typu rozdělení.

#### Zadat

Je-li tato volba vybrána, můžete zadat parametry distribuce pro vybraný typ distribuce.

**Poznámka:** Dostupné parametry se liší na základě vybraného typu rozdělení.

### Transformovat

Poskytnuté volby nastavují nastavení transformace a periodicity.

#### Transformace přirozeného protokolu

Transformuje data pomocí přirozeného logaritmu (základu  $e$ ) k odstranění různé amplitudy.

#### Standardizovat hodnoty

Transformuje proměnné posloupnosti nebo proměnné časové řady do vzorku se střední hodnotou 0 a směrodatnou odchylkou 1.

#### Rozdíl

Uvádí stupeň diferenciaci, který se používá před vykreslením k převodu nestacionární proměnné do stacionární proměnné s konstantní střední hodnotou a rozptylem. Zadejte do pole příslušnou hodnotu.

#### Sezonní rozdíl

Pokud má proměnná sezónní nebo periodickou strukturu, můžete použít toto nastavení pro sezonní rozdíl dané proměnné před zakresetím.

**Poznámka:** Toto nastavení je povoleno pouze tehdy, je-li jako jedna z kvantitativních proměnných vybrána proměnná posloupnosti nebo proměnná časové řady s definovanou periodicitou.

### Vzorec odhadu dílčí dávky

Poskytnutá volby nastavují vzorec, který se používá k odhadu podílů.

#### Blomovy

Vytvoří novou hodnotou proměnnou založenou na odhadech poměru, které používají vzorec  $(r-3/8) / (w+1/4)$ , kde  $w$  je součet vah případu a  $r$  je očíslování pořadí.

#### Sada Rankit

Používá vzorec  $(r-1/2) / w$ , kde  $w$  je počet pozorování a  $r$  je očíslování pořadí, řazeno od 1 do  $w$ .

#### Tukeyho

Používá vzorec  $(r-1/3) / (w+1/3)$ , kde  $r$  je očíslování pořadí a  $w$  je součet vah velkých a malých písmen.

#### Van der Waerden

Van der Waerden transformace, definované vzorcem  $r / (w+1)$ , kde  $w$  je součet vah případu a  $r$  je očíslování pořadí, řazeno od 1 do  $w$ .

### Pořadí přiřazené k vazbám

Poskytnutá volba řídí způsob, jak určit, jak zpracovat hodnoty tie-tie. Následující tabulka ukazuje, jak různé metody přiřazují rank k vázaným hodnotám.

*Tabulka 3. Metody a výsledky hodnocení*

Hodnota	Střední	Nízký	Vysoké	Přerušit vazby libovolně
10	1	1	1	1
15	3	2	4	2

Tabulka 3. Metody a výsledky hodnocení (pokračování)

Hodnota	Střední	Nízký	Vysoké	Přerušit vazby libovolně
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

## Získání pravděpodobnostních grafů P-P

1. Z nabídky vyberte:

**Analýza > Deskriptivní statistika > P-P Plots ...**

2. Vyberte jednu nebo více číselných proměnných a přesuňte je do pole Proměnné.

3. Vyberte distribuci testu.

Volitelně můžete vybrat volby transformace pro získání pravděpodobnostních grafů pro transformované hodnoty a určit metodu pro výpočet očekávaných distribucí.

## Zakreslení

Kvantilový graf kvantilové grafy vygeneruje pravděpodobnostní grafy pro transformované hodnoty. Dostupné testovací distribuce zahrnují beta, chí kvadrát, exponenciální, gama, polonormální, Laplaceova, Logistický, logormální, normální, pareto, Studentovo t, Weibullovo a uniformní. V závislosti na vybrané distribuci můžete určit stupně volnosti a další parametry.

- Získáte pravděpodobnostní grafy pro transformované hodnoty. Volby transformace zahrnují přirozený protokol, standardizace hodnot, rozdílu a sezónně odlišnost.
- Můžete uvést metodu pro výpočet očekávaných distribucí a pro řešení "vazeb" (více pozorování se stejnou hodnotou).

### Rozdělení testu

Uveďte typ rozdělení pro vaše data. Rozevírací seznam poskytuje následující volby:

#### Beta

*Distribuce beta.* Parametry shape1 a shape2  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty střední hodnoty odběru vzorku a vychýlení vzorku. Všechna pozorování musí být v rozsahu 0 až 1 včetně.

#### Chí-kvadrát

*Distribuce chí-kvadrát.* Je třeba určit stupně volnosti ( $df$ ). Záporná pozorování nejsou povolena.

#### Exponenciální

*Exponenciální rozdělení.* Parametr měřítka  $a$  musí být kladný. Není-li parametr zadán, produkt DISTRIBUTION jej odhadne ze střední hodnoty. Záporná pozorování nejsou povolena.

#### Gama

*Gamma rozdělení.* Parametry tvaru a měřítka  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty střední hodnoty odběru vzorku a vychýlení vzorku. Záporná pozorování nejsou povolena.

#### Polonormální

*Half-normální rozdělení.* Předpokládá se, že data jsou ve volném umístění nebo jsou centralizovaná. (Lokalita parameter=0.) Můžete uvést parametr měřítka  $a$  nebo nechat DISTRIBUTION odhadnout to pomocí metody maximální pravděpodobnosti.

#### Laplaceovo

*Laplaceovo nebo dvojitě exponenciální rozdělení.* LAPLACE vezme místo a parametr scale ( $a$  a  $b$ ). Parametr měřítka ( $b$ ) musí být kladný. Nejsou-li parametry zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty vzorku a směrodatné odchylky vzorků.

### **Logistický**

*Logistická distribuce.* LOGISTIC vezme místo a parametr scale ( $a$  a  $b$ ). Parametr měřítka ( $b$ ) musí být kladný. Nejsou-li parametry zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty vzorku a směrodatné odchylky vzorků.

### **Logaritmicko-normální**

*Logaritmicko-normální rozdělení.* Parametry měřítka a tvaru  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední a směrodatné odchylky přirozeného logaritmu dat vzorku. Záporná pozorování nejsou povolena.

### **Normální**

*Normální rozdělení.* Parametr umístění  $a$  může být libovolná číselná hodnota, zatímco parametr měřítka  $b$  musí být kladný. Nejsou-li zadány, produkt DISTRIBUTION je odhadne ze střední hodnoty střední hodnoty odběru vzorku a vychýlení vzorku.

### **Paretův**

*Paretova distribuce.* Parametry prahové hodnoty a tvaru  $a$  a  $b$  musí být kladné. Pokud nejsou zadány, předpokládá DISTRIBUTION, že  $a$  se rovná minimální hodnotě pozorování a  $b$  podle maximální pravděpodobnosti. Záporná pozorování nejsou povolena.

### **Student t**

*Studentova t rozdělení.* Je třeba určit stupně volnosti ( $df$ ).

### **Uniformní**

*Rovnoměrné rozdělení.* UNIFORM vezme minimální a maximální parametr ( $a$  a  $b$ ). Parametr  $a$  musí být roven nebo větší než  $b$ . Nejsou-li parametry uvedeny, DISTRIBUTION je předpokládá z ukázkových dat.

### **WEIBULL**

*Weibullovo rozdělení.* Parametry měřítka a tvaru  $a$  a  $b$  musí být kladné. Nejsou-li zadány, produkt DISTRIBUTION je odhadne s použitím metody nejmenších čtverců. Záporná pozorování nejsou povolena.

## **Parametry distribuce**

Poskytuje strategii distribuce a volby parametrů.

### **Odhadnout z dat**

Je-li tato volba vybrána, toto nastavení odhaduje distribuční parametry na základě dat a vybraného typu rozdělení.

### **Zadat**

Je-li tato volba vybrána, můžete zadat parametry distribuce pro vybraný typ distribuce.

**Poznámka:** Dostupné parametry se liší na základě vybraného typu rozdělení.

## **Transformovat**

Poskytnuté volby nastavují nastavení transformace a periodicity.

### **Transformace přirozeného protokolu**

Transformuje data pomocí přirozeného logaritmu (základu  $e$ ) k odstranění různé amplitudy.

### **Standardizovat hodnoty**

Transformuje proměnné posloupnosti nebo proměnné časové řady do vzorku se střední hodnotou 0 a směrodatnou odchylkou 1.

### **Rozdíl**

Uvádí stupeň diferenciacce, který se používá před vykreslením k převodu nestacionární proměnné do stacionární proměnné s konstantní střední hodnotou a rozptylem. Zadejte do pole příslušnou hodnotu.

### **Sezonní rozdíl**

Pokud má proměnná sezónní nebo periodickou strukturu, můžete použít toto nastavení pro sezonní rozdíl dané proměnné před zakresetím.

**Poznámka:** Toto nastavení je povoleno pouze tehdy, je-li jako jedna z kvantitativních proměnných vybrána proměnná posloupnosti nebo proměnná časové řady s definovanou periodicitou.

### Vzorec odhadu dílčí dávky

Poskytnutá volba nastavují vzorec, který se používá k odhadu podílů.

#### Blomov

Vytvoří novou hodnotou proměnnou založenou na odhadech poměru, které používají vzorec  $(r-3/8) / (w+1/4)$ , kde  $w$  je součet vah případu a  $r$  je očíslování pořadí.

#### Sada Rankit

Používá vzorec  $(r-1/2) / w$ , kde  $w$  je počet pozorování a  $r$  je očíslování pořadí, řazeno od 1 do  $w$ .

#### Tukeyho

Používá vzorec  $(r-1/3) / (w+1/3)$ , kde  $r$  je očíslování pořadí a  $w$  je součet vah velkých a malých písmen.

#### Van der Waerden

Van der Waerden transformace, definované vzorcem  $r/(w+1)$ , kde  $w$  je součet vah případu a  $r$  je očíslování pořadí, řazeno od 1 do  $w$ .

### Pořadí přiřazené k vazbám

Poskytnutá volba řídí způsob, jak určit, jak zpracovat hodnoty tie-tie. Následující tabulka ukazuje, jak různé metody přiřazují rank k vázaným hodnotám.

Tabulka 4. Metody a výsledky hodnocení

Hodnota	Střední	Nízký	Vysoké	Přerušit vazby libovolně
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

### Získání grafů pravděpodobnosti Q-Q

1. Z nabídky vyberte:

**Analyzovat > Deskriptivní statistika > kvantilový graf ...**

2. Vyberte jednu nebo více číselných proměnných a přesuňte je do pole Proměnné.

3. Vyberte distribuci testu.

Volitelně můžete vybrat volby transformace pro získání pravděpodobnostních grafů pro transformované hodnoty a určit metodu pro výpočet očekávaných distribucí.

## Analýza ROC

Analýza funkce ROC (Receiver operating characteristic) je užitečným způsobem pro posouzení přesnosti modelových předpovědí zakreslením senzitivity versus (1-specifičnosti) klasifikačního testu (jelikož prahová hodnota se liší v celém rozsahu výsledků diagnostických testů). Plná plocha pod danou křivkou ROC nebo AUC formuluje důležitou statistiku, která představuje pravděpodobnost, že předpověď bude ve správném pořadí, když je pozorována testovací proměnná (u jednoho předmětu, který je náhodně vybrán ze skupiny případů a druhý náhodně vybraný z kontrolní skupiny). Analýza ROC podporuje inferenci týkající se jednotlivých hodnot AUC, precision-recall (PR) a poskytuje volby pro porovnání dvou křivek ROC, které jsou generovány buď z nezávislých skupin, nebo z párových předmětů.

Starý postup ROC Curve podporuje statistickou inferenci o jediné křivce ROC. Tato operace může být také nahrazena novým postupem analýzy ROC. Nová procedura analýzy ROC dále může porovnávat dva křivky ROC generované buď z nezávislých skupin, nebo z párových subjektů.

Výkřiky v křivce nákupních křivek versus opětovné vyvolání mají tendenci být více informativní, jsou-li pozorované vzorky dat velmi posunuty a poskytují alternativu k křivku ROC pro data s velkým posunem v distribuci třídy.

### Příklad

Je v zájmu banky správně klasifikovat zákazníky do těch zákazníků, kteří budou a nebudou přednastaveny na své půjčky, takže jsou vyvinuty speciální modely pro provedení těchto rozhodnutí. Analýza ROC může být použita k vyhodnocení a vyhodnocení přesnosti modelových předpovědí.

### Statistika

AUC, záporná skupina, chybějící hodnoty, pozitivní klasifikace, hodnota uzavření, pevnost přesvědčení, Dvoustranný asymptotický interval spolehlivosti, distribuce, standardní chyba, nezávislý-skupinový návrh, párování-ukázkový návrh, neparametrický předpoklad, bi-negativní exponenciální rozložení šíření, střed, bod řezu, PR křivka, kroková interpolace, asymptotické významnosti (2-tail), Citlivost a (1-Specita), Precision a Recall.

### Metody

Porovnávají se oblasti pod dvěma oblouky ROC, které se vytvářejí buď z nezávislých skupin, nebo z párových subjektů. Porovnání dvou křivek ROC může poskytnout více informací v přesnosti výsledků ze dvou srovnávacích diagnostických přístupů.

## Aspekty dat analýzy ROC

### Data

Křivky NŽ znázorňují přesnost versus opětovné vyvolání a mají tendenci být více informativní, když jsou pozorované vzorky dat vysoce zkreslené. Jednoduchá lineární interpolace může omylem přinést příliš optimistický odhad PR křivky.

### Předpoklady

Předpověď bude ve správném pořadí, když je pozorována testovací proměnná pro jeden předmět, který je náhodně vybrán ze skupiny případu a druhý je náhodně vybrán z řídicí skupiny. Každá definovaná skupina bude obsahovat alespoň jedno platné pozorování. Pro jednu proceduru se používá pouze jedna proměnná seskupení.

## Získání analýzy ROC

1. Z nabídky vyberte:

**Analyzovat > Klasifikovat > Analýza ROC**

2. Vyberte jednu nebo více proměnných pravděpodobnosti testu.

3. Vyberte jednu proměnnou stavu.

4. Identifikujte *kladnou* hodnotu pro proměnnou stavu.

5. Volitelně vyberte volbu **Paired-ukázkový návrh**, nebo vyberte jednu proměnnou seskupení (nelze vybrat obě volby).

- Použijte nastavení **Paired-sample design** k vyžádání párování-ukázkového návrhu pro testovací proměnnou (proměnné). V párování-ukázkový návrh porovnává dva křivky ROC ve scénáři párovaných ukázek, když se více testovacích hodnot měří na stejných předmětech, které jsou přidruženy ke stavové proměnné.

**Poznámka:** Je-li vybrána volba **Paired-sample design**, volby **Grouping Variable** a **Distribution Assumption** (v dialogovém okně Volby) jsou zablokovány.

- Je-li vybrána numerická proměnná seskupení, můžete klepnout na volbu **Definovat skupiny ...** požadovat nezávislý návrh skupiny pro zkušební proměnnou (proměnné) a uvést dvě hodnoty, střed nebo bod řezu.
6. Volitelně můžete klepnutím na tlačítko Volby definovat klasifikaci, směr testu, standardní chybové parametry a chybějící nastavení hodnot.
7. Volitelně klepněte na Zobrazit, abyste definovali nastavení a nastavení tisku (které zahrnují ROC Curve, Precision-Recall Curve and model quality settings).



8. Klepněte na tlačítko **OK**.

## Analýza ROC: Volby

Pro analýzu ROC můžete zadat následující volby:

### Klasifikace

Umožňuje určit, zda by měla být hodnota uzavření zahrnuta nebo vyloučena při vytváření *kladné* klasifikace. Toto nastavení momentálně nemá žádný vliv na výstup.

### Směr testu

Poskytuje volby pro určení, který směr proměnné výsledku testu označuje zvyšující se sílu přesvědčení, že předmět je pozitivní test.

### Parametry pro standardní chybu oblasti

Umožňuje vám uvést metodu odhadu standardní chyby oblasti pod křivkou. Dostupné metody jsou neparametrické a bi-negativní exponenciální hodnoty. Výchozí nastavení **Neparametrické** poskytuje odhady pod neparametrickým předpokladem. Nastavení **Bi-negative exponencial** poskytuje odhady v rámci binegativního exponenciálního rozložení distribuce.

Tato sekce vám také umožňuje určit úroveň důvěry pro dvoustranný asymptotický interval spolehlivosti AUC. Dostupný rozsah je 0,0% až 100,0% (výchozí hodnota je 95%).

**Poznámka:** Nastavení se vztahuje pouze na návrh nezávislé skupiny a nemá žádný vliv na návrh párovaných vzorů.

### Chybějící hodnoty

Umožňuje vám uvést, jak se zachází s chybějícími hodnotami. Není-li toto nastavení vybráno, jsou vyloučeni jak uživatelé, tak i systémově chybějící hodnoty. Je-li vybráno nastavení, uživateli chybí hodnoty, které jsou považovány za platné, systémově chybějící hodnoty jsou vyloučeny. Případy se systémem chybějícími hodnotami, buď v testovací proměnné, nebo ve stavové proměnné, jsou vždy vyloučeny z analýzy.

## Analýza ROC: Zobrazit

Pro analýzu ROC můžete určit následující nastavení zobrazení:

### Graf

Poskytuje volby pro vykreslování křivek ROC a Precision-Recall.

### Křivka ROC

Je-li tato volba vybrána, ve výstupu se zobrazí graf ROC Curve. Vyberte **S diagonálou vztažné čáry** pro vykreslení diagonální vztažné čáry na grafu ROC Curve.

### Přesnost-Křivka opětovného vyvolání

Je-li tato volba vybrána, bude ve výstupu zobrazen graf Curs-Recall Curve. Precision-Opětovně vyvolat křivky mají tendenci být více informativní, když jsou pozorované vzorky dat jsou vysoce zkršené a poskytují alternativu ROC Křivky pro data s velkým posunem v distribuci třídy. Výchozí nastavení **Interpolate along the true positives** nastavuje postupnou interpolaci společně s pozitivními pozitivními výsledky. Nastavení **Interpolace podél falešných pozitiv** provádí postupku interpolace podél falešných pozitiv.

### Celková kvalita modelu

Nastavení určuje, zda se má či nemá vytvořit sloupcový graf, aby se zobrazila hodnota dolní hranice intervalu spolehlivosti odhadovaného AUC. Ve výchozím nastavení není toto nastavení vybráno, což potlačuje sloupcový graf.

### Tisk

Poskytuje volby pro definování výstupu pro odpovídající statistiku.

### Standardní chyba a interval spolehlivosti

Nastavení řídí, které statistiky se zobrazí v tabulce "Oblast pod křivkou". Není-li toto nastavení vybráno, zobrazí se pouze odhadovaná hodnota AUC. Když je vybráno nastavení, zobrazí se další statistické údaje včetně standardní chyby AUC, asymptotického významu (2-tail) a intervalu Asymptotic Confidence pod nulou hypotézou.

### Souřadnice křivky ROC

Nastavení řídí souřadnice křivky ROC Curve společně s hodnotami uzavření objektu. Není-li toto nastavení vybráno, bude výstup souřadnic potlačen. Je-li vybráno nastavení, jsou dvojice citlivosti a (1-Specificity) zadány s hodnotami uzavření pro každou křivku ROC.

### Index nástroje Youden

Zobrazí hodnotu Youden Index pro každou hodnotu uzavření křivky ROC.

### Souřadnice souřadnic Expres-Recall Curve

Nastavení řídí souřadnice křivky Precision-Recall Curve, spolu s hodnotami uzavření objektu. Není-li toto nastavení vybráno, bude výstup souřadnic potlačen. Když je vybráno nastavení, dvojice hodnot Přesnost a Vyvolání jsou uvedeny s hodnotami uzavření pro každou křivku Recall-Recall Curve.

### Metriky vyhodnocení klasifikátoru

Nastavení řídí zobrazení tabulky metrik vyhodnocení klasifikátoru ve výstupu. Tabulka ukazuje, jak dobře klasifikační model odpovídá datům ve srovnání s náhodným přiřazením, a poskytuje následující informace:

- Uživatelem zadané testovací proměnné
- Informace o skupině
- Gini index (Gini index je  $2 \cdot \text{AUC} - 1$ , kde AUC je plocha pod křivkou ROC)
- Maximální počet hodnot K-S a Cutoff

## Analýza ROC: Definovat skupiny (řetězec)

U řetězcových proměnných seskupení zadejte řetězec pro skupinu 1 a další hodnotu pro skupinu 2, jako je *ano* a *ne*. Případy s ostatními řetězci jsou z analýzy vyloučeny.

**Poznámka:** Uvedené hodnoty musí existovat v proměnné, jinak se zobrazí chybová zpráva, která označuje, že alespoň jedna ze skupin je prázdná.

## Analýza ROC: Definovat skupiny (číselné)

Pro číselné proměnné seskupení definujte dvě skupiny pro test *t* určením dvou hodnot, středového bodu nebo bodu řezu.

**Poznámka:** Uvedené hodnoty musí existovat v proměnné, jinak se zobrazí chybová zpráva, která označuje, že alespoň jedna ze skupin je prázdná.

- **Použijte zadané hodnoty.** Zadejte hodnotu pro skupinu 1 a další hodnotu pro skupinu 2. Případy s jakýmkoli jinými hodnotami jsou vyloučeny z analýzy. Čísla nemusí být celá čísla (např. 6.25 a 12.5 jsou platné).
- **Použití hodnotu středu.** Je-li tato volba vybrána, skupiny jsou odděleny do hodnot  $< a$  a  $\geq$  střed.
- **Použijte mezní bod.**
  - **Místo uzavření objektu.** Zadejte číslo, které rozdělí hodnoty proměnné seskupení do dvou sad. Všechny případy s hodnotami, které jsou nižší než jedna skupina, a případy s hodnotami, které jsou větší než nebo rovny ve výstřiku, tvoří druhou skupinu.

## ROC Křivky

Tento postup je užitečným způsobem pro vyhodnocení výkonnosti klasifikačních systémů, v nichž existuje jedna proměnná se dvěma kategoriemi, podle nichž jsou předměty klasifikovány.

**Příklad.** Je v zájmu banky správně klasifikovat zákazníky do těch zákazníků, kteří budou a nebudou přednastaveny na své půjčky, takže jsou vyvinuty speciální metody pro provedení těchto rozhodnutí. Křivky ROC lze použít k vyhodnocení, jak dobře tyto metody fungují.

**Statistika.** Plocha pod křivkou ROC s intervalem spolehlivosti a souřadnicovou křivkou křivky ROC. Plot: Křivka ROC.

**Metody.** Odhad plochy pod křivkou ROC lze vypočítat buď neparametricky, nebo parametricky pomocí binegativního exponenciálního modelu.

Pokyny k datům křivky ROC

Starý postup ROC Curve podporuje statistickou inferenci o jediné křivce ROC. Tato operace může být také nahrazena novým postupem analýzy ROC. Nová procedura analýzy ROC dále může porovnávat dva křivky ROC generované buď z nezávislých skupin, nebo z párových subjektů.

**Data.** Testovací proměnné jsou kvantitativní. Testovací proměnné se často skládají z pravděpodobnosti diskriminace na základě diskriminačních nebo logistických regresí nebo složené ze skóre na libovolné stupnici naznačující raterové "sílu přesvědčení", že předmět spadá do jedné kategorie nebo do jiné kategorie. Proměnná stavu může být libovolného typu a označuje skutečnou kategorii, do které předmět patří. Hodnota proměnné stavu označuje, která kategorie by měla být považována za *kladnou*.

**Předpoklady.** Předpokládá se, že rostoucí počet na stupnici rváče představuje rostoucí přesvědčení, že předmět patří do jedné kategorie, zatímco klesající počty na stupnici představují zvyšující se přesvědčení, že toto téma patří do jiné kategorie. Uživatel musí zvolit, který směr je *kladný*. Předpokládá se také, že je známá kategorie *true*, ke které je každý předmět přiřazen.

Získání křivky ROC

1. Z nabídky vyberte:

**Analyzovat > Klasifikovat > Křivka ROC ...**

2. Vyberte jednu nebo více proměnných pravděpodobnosti testu.

3. Vyberte jednu proměnnou stavu.

4. Identifikujte *kladnou* hodnotu pro proměnnou stavu.

## Volby křivky ROC

Pro analýzu ROC můžete zadat následující volby:

### Klasifikace

Umožňuje určit, zda by měla být hodnota uzavření zahrnuta nebo vyloučena při vytváření *kladné* klasifikace. Toto nastavení momentálně nemá žádný vliv na výstup.

### Směr testu

Umožňuje určit směr měřítka ve vztahu k kategorii *pozitivní*.

### Parametry pro standardní chybu oblasti

Umožňuje vám uvést metodu odhadu standardní chyby oblasti pod křivkou. Dostupné metody jsou neparametrické a binegativní exponenciální hodnoty. Také vám umožňuje nastavit úroveň intervalu spolehlivosti. Dostupný rozsah je 50, 1% až 99, 9%.

### Chybějící hodnoty

Umožňuje vám uvést, jak se zachází s chybějícími hodnotami.

## Simulace

Prediktivní modely, jako např. lineární regrese, vyžadují sadu známých vstupů pro předpověď výsledku nebo cílové hodnoty. V mnoha reálných aplikacích však hodnoty vstupů jsou nejisté. Simulace umožňuje počítat s nejistotou ve vstupech do prediktivních modelů a hodnotit pravděpodobnost různých výsledků modelu v přítomnosti této nejistoty. Předpokládejme například, že máte model zisku, který zahrnuje náklady na materiály jako vstup, ale v důsledku nestálosti trhu existuje nejistota v těchto nákladech. Simulaci můžete použít k modelování této nejistoty a k určení efektu, který má na zisku.

Simulace v produktu IBM SPSS Statistics používá metodu Monte Carlo. Neurčité vstupy jsou modelovány s pravděpodobností distribucí (jako je trojúhelníková distribuce) a simulované hodnoty těchto vstupů jsou generovány vykreslením z těchto distribucí. Vstupy, jejichž hodnoty jsou známy, jsou zadrženy na základě známých hodnot. Prediktivní model je vyhodnocován pomocí simulované hodnoty pro každou nejistou vstupní a pevnou hodnotu pro známé vstupy k výpočtu cíle (nebo cílů) modelu. Proces se opakuje mnohokrát (obvykle desetitisíce nebo stovky tisícrát), což vede k rozdělení cílových hodnot,

keré lze použít k zodpovězení otázek pravděpodobnostní povahy. V kontextu produktu IBM SPSS Statistics generuje každé opakování procesu samostatný případ (záznam) dat, který se skládá ze sady simulovaných hodnot pro nejisté vstupy, hodnot pevných vstupů a předpovídaného cíle (nebo cílů) modelu.

Data v případě absence prediktivního modelu můžete také simulovat zadáním rozdělení pravděpodobnosti pro proměnné, které mají být simulovány. Každý generovaný případ dat se skládá ze sady simulovaných hodnot pro zadané proměnné.

Chcete-li spustit simulaci, musíte uvést podrobnosti, jako např. prediktivní model, rozdělení pravděpodobnosti pro nejisté vstupy, korelace mezi těmito vstupy a hodnotami pro jakékoli pevné vstupy. Jakmile uvedete všechny podrobnosti pro simulaci, můžete ji spustit a volitelně uložit specifikace do souboru **simulačního plánu**. Simulační plán můžete sdílet s ostatními uživateli, kteří pak mohou simulaci spouštět, aniž by bylo nutné porozumět jejím informacím.

Pro práci se simulacemi jsou k dispozici dvě rozhraní. Tvůrce simulací je rozšířené rozhraní pro uživatele, kteří navrhují a provádějí simulace. Poskytuje úplnou sadu funkcí pro návrh simulace, uložení specifikací do souboru s plánem simulace, určení výstupu a spuštění simulace. Simulaci můžete sestavit na základě souboru modelu IBM SPSS nebo na sadu vlastních rovnic, které definujete v tvůrci simulace. Můžete také načíst existující plán simulace do Tvůrce simulací, upravit veškerá nastavení a spustit simulaci a volitelně uložit aktualizovaný plán. Pro uživatele, kteří mají simulační plán a primárně chtějí spustit simulaci, je k dispozici jednodušší rozhraní. Umožňuje vám upravit nastavení, která vám umožní spustit simulaci za různých podmínek, ale neposkytuje úplné schopnosti tvůrce simulací pro navrhování simulací.

## Chcete-li navrhnout simulaci na základě souboru modelu

1. Z nabídky vyberte:

**Analyzovat > Simulace ...**

2. Klepněte na volbu **Vybrat soubor modelu SPSS** a klepněte na tlačítko **Pokračovat**.

3. Otevřete soubor modelu.

Modelový soubor je soubor XML, který obsahuje model PMML vytvořený z produktu IBM SPSS Statistics nebo IBM SPSS Modeler. Další informace naleznete v tématu [“Karta Model” na stránce 238](#).

4. Na kartě Simulace (v tvůrci simulací) určete rozdělení pravděpodobnosti pro simulované vstupy a hodnoty pro pevné vstupy. Pokud aktivní datová sada obsahuje historická data pro simulované vstupy, klepnutím na volbu **Přízpůsobit vše** automaticky určete rozdělení, které nejvíce odpovídá datům pro každý takový vstup, a také ke stanovení korelací mezi nimi. Pro každý simulovaný vstup, který se nevejde do historických dat, musíte výslovně uvést distribuci výběrem typu distribuce a zadáním požadovaných parametrů.

5. Klepnutím na tlačítko **Spustit** spustíte simulaci. Standardně se plán simulace, který určuje podrobnosti simulace, uloží do umístění uvedeného v nastavení Uložit.

K dispozici jsou následující volby:

- Upravte umístění pro uložený plán simulace.
- Určete známé korelace mezi simulovanými vstupy.
- Automaticky vypočítat pohotovostní tabulku přidružení mezi kategorickými vstupy a použít tato přidružení, když jsou data generována pro tyto vstupy.
- Uvedte analýzu citlivosti, abyste prozkoumali vliv proměnné hodnoty pevného vstupu nebo proměnné distribučního parametru pro simulovaný vstup.
- Uvedte rozšířené volby, jako je nastavení maximálního počtu případů pro generování nebo požadování vzorkování tail.
- Přízpůsobení výstupu.
- Uložit simulovaná data do datového souboru.

## Chcete-li navrhnout simulaci na základě vlastních rovnic,

1. Z nabídky vyberte:

**Analyzovat > Simulace ...**

2. Klepněte na **Typ v rovnicích** a klepněte na **Pokračovat**.
3. Klepnutím na volbu **Nová rovnice** na kartě Model (v tvůrci simulací) definujte každou rovnici ve vašem předpovědním modelu.
4. Klepněte na kartu Simulace a určete rozdělení pravděpodobnosti pro simulované vstupy a hodnoty pro pevné vstupy. Pokud aktivní datová sada obsahuje historická data pro simulované vstupy, klepnutím na volbu **Přizpůsobit vše** automaticky určete rozdělení, které nejvíce odpovídá datům pro každý takový vstup, a také ke stanovení korelací mezi nimi. Pro každý simulovaný vstup, který se nevejde do historických dat, musíte výslovně uvést distribuci výběrem typu distribuce a zadáním požadovaných parametrů.
5. Klepnutím na tlačítko **Spustit** spustíte simulaci. Standardně se plán simulace, který určuje podrobnosti simulace, uloží do umístění uvedeného v nastavení Uložit.

K dispozici jsou následující volby:

- Upravte umístění pro uložený plán simulace.
- Určete známé korelace mezi simulovanými vstupy.
- Automaticky vypočítat pohotovostní tabulku přidružení mezi kategoriickými vstupy a použít tato přidružení, když jsou data generována pro tyto vstupy.
- Uvedte analýzu citlivosti, abyste prozkoumali vliv proměnné hodnoty pevného vstupu nebo proměnné distribučního parametru pro simulovaný vstup.
- Uvedte rozšířené volby, jako je nastavení maximálního počtu případů pro generování nebo požadování vzorkování tail.
- Přizpůsobení výstupu.
- Uložit simulovaná data do datového souboru.

## Chcete-li navrhnout simulaci bez prediktivního modelu

1. V nabídce vyberte:

**Analyzovat > Simulace ...**

2. Klepněte na volbu **Vytvořit simulovaná data** a klepněte na tlačítko **Pokračovat**.
3. Na kartě Model (v tvůrci simulace) vyberte pole, která chcete simulovat. Můžete vybrat pole z aktivní datové sady nebo můžete definovat nová pole klepnutím na **Nová**.
4. Klepněte na kartu Simulace a uveďte rozdělení pravděpodobnosti pro pole, která se mají simulovat. Pokud aktivní datová sada obsahuje historická data pro jakékoli z těchto polí, klepněte na **Přizpůsobit vše**, aby se automaticky určilo rozdělení, které nejlépe odpovídá datům a určilo korelace mezi poli. U polí, která se nevejdu do historických dat, musíte výslovně uvést distribuci výběrem typu distribuce a zadáním požadovaných parametrů.
5. Klepnutím na tlačítko **Spustit** spustíte simulaci. Standardně se simulovaná data ukládají do nové datové sady uvedené v nastavení Uložení. Kromě toho se plán simulace, který určuje podrobnosti simulace, uloží do umístění uvedeného v nastavení Uložit.

K dispozici jsou následující volby:

- Upravte umístění pro simulovaná data nebo uložený simulační plán.
- Určete známé korelace mezi simulovanými poli.
- Automaticky vypočítat pohotovostní tabulku přidružení mezi kategoriálními poli a použít tato přidružení, když jsou data vygenerována pro tato pole.
- Zadejte analýzu citlivosti, abyste prošetřili vliv logického distribučního parametru na simulované pole.
- Uvedte rozšířené volby, jako je nastavení počtu případů, které se mají generovat.

## Chcete-li spustit simulaci ze simulačního plánu

Pro spuštění simulace z plánu simulace jsou k dispozici dvě volby. Můžete použít dialogové okno Spustit simulaci, které je primárně určeno pro spouštění ze simulačního plánu, nebo můžete použít tvůrce simulace.

Chcete-li použít dialogové okno Spustit simulaci:

1. Z nabídky vyberte:

**Analyzovat > Simulace ...**

2. Klepněte na volbu **Otevřít existující plán simulace**.
3. Ujistěte se, že zaškrťovací políčko **Otevřít v tvůrci simulací** není zaškrtnuto, a klepněte na **Pokračovat**.
4. Otevřete plán simulace.
5. Klepněte na tlačítko **Spustit** v dialogovém okně Spustit simulaci.

Chcete-li spustit simulaci z tvůrce simulací, postupujte takto:

1. Z nabídky vyberte:

**Analyzovat > Simulace ...**

2. Klepněte na volbu **Otevřít existující plán simulace**.
3. Označte zaškrťovací políčko **Otevřít v tvůrci simulací** a klepněte na **Pokračovat**.
4. Otevřete plán simulace.
5. Upravte všechna nastavení, která chcete upravit, na kartě Simulace.
6. Klepnutím na tlačítko **Spustit** spustíte simulaci.

Volitelně můžete provést následující akce:

- Nastavení nebo úprava analýzy citlivosti za účelem zkoumání vlivu změny hodnoty pevného vstupu nebo proměnné distribučního parametru pro simulovaný vstup.
- Založení distribucí a korelací pro simulované vstupy na nová data.
- Změňte distribuci simulovaného vstupu.
- Přizpůsobení výstupu.
- Uložit simulovaná data do datového souboru.

## Tvůrce simulace

Tvůrce simulací poskytuje úplnou sadu schopností pro navrhování a spouštění simulací. Umožňuje provádět následující obecné úlohy:

- Navrhujte a spusťte simulaci pro model IBM SPSS definovaný v souboru modelu PMML.
- Navrhujte a spusťte simulaci pro prediktivní model definovaný sadou vlastních rovnic, které jste zadali.
- Navrhněte a spusťte simulaci, která generuje data v nepřítomnosti prediktivního modelu.
- Spusťte simulaci na základě existujícího plánu simulace, případně upravte libovolné nastavení plánu.

## Karta Model

U simulací založených na modelu prediktivního modelu je na kartě Model uveden zdroj modelu. V případě simulací, které nezahrnují prediktivní model, určuje karta Model pole, která mají být simulována.

**Vyberte soubor modelu SPSS** . Tato volba uvádí, že prediktivní model je definován v souboru modelu IBM SPSS . Modelový soubor IBM SPSS je soubor XML nebo komprimovaný archiv souborů (soubor .zip), který obsahuje model PMML vytvořený z produktu IBM SPSS Statistics nebo IBM SPSS Modeler. Prediktivní modely jsou vytvářeny procedurami, jako je například lineární regrese a rozhodnutí o rozhodnutí v rámci produktu IBM SPSS Statistics lze exportovat do souboru modelu. Můžete použít jiný soubor modelu klepnutím na tlačítko **Procházet** a navigací k souboru, který chcete.

## Modely PMML podporované simulací

- Lineární regrese
- Automatický lineární model
- Generalizovaný lineární model
- Generalizovaný lineární smíšený model
- Obecný lineární model
- Binární logistické regrese
- Multinomická Logistická regrese
- Ordinal Multnomial Regression
- Regrese Cox
- Strom
- Boosted Tree (C5)
- Diskriminant
- Dvoukrokový klastr
- Klastr středních hodnot
- Neuronová síť
- Ruleset (Seznam rozhodnutí)

### Poznámka:

- Modely PMML, které mají více cílových polí (proměnných) nebo rozdělení, nejsou pro použití v simulaci podporovány.
- Hodnoty řetězcových vstupů do binárních logistických regresních modelů jsou v modelu omezeny na 8 bajtů. Pokud se do aktivní datové sady hodí takové řetězcové vstupy, ujistěte se, že hodnoty v datech nepřekračují 8 bajtů. Hodnoty dat, které překračují 8 bajtů, jsou vyloučeny z přidružené kategorické distribuce pro vstup a jsou zobrazeny jako nepoužité ve výstupní tabulce Neshodující se kategorie.

**Zadejte rovnice pro model.** Tato volba uvádí, že prediktivní model se skládá z jedné nebo více vlastních rovnic, které mají být vytvořeny vámi. Vytvořte rovnice klepnutím na volbu **Nové rovnice**. Tím se otevře Editor rovnic. Existující rovnice můžete upravit, zkopírovat je, aby bylo možné je použít jako šablony pro nové rovnice, změnit jejich pořadí a odstranit je.

- Tvůrce simulací nepodporuje systémy simultánních rovnic nebo rovnic, které nejsou v cílové proměnné nelineární.
- Vlastní rovnice jsou vyhodnocovány v pořadí, ve kterém jsou určeny. Pokud rovnice pro daný cíl závisí na jiném cíli, pak musí být druhý cíl definován předchozí rovnicí.

Například, je-li dána sada tří rovnic níže, rovnice pro *profit* závisí na hodnotách *revenue* a *expenses*, takže rovnice pro *revenue* a *expenses* musí být předcházet rovnicí pro *profit*.

```
revenue = price*volume
```

```
expenses = fixed + volume*(unit_cost_materials + unit_cost_labor)
```

```
profit = revenue - expenses
```

**Vytvářejte simulovaná data bez modelu.** Vyberte tuto volbu, chcete-li simulovat data bez prediktivního modelu. Uveďte pole, která se mají simulovat výběrem polí z aktivní datové sady, nebo klepnutím na tlačítko **Nový**, abyste definovali nová pole.

### Editor rovnic

Editor rovnic vám umožňuje vytvořit nebo upravit vlastní rovnici pro prediktivní model.

- Výraz pro rovnici může obsahovat pole z aktivní datové sady nebo z nových vstupních polí, které definujete v editoru rovnic.

- Můžete uvést vlastnosti cíle, jako je jeho úroveň měření, jmenovka hodnot a zda je výstup generován pro cíl.
  - Můžete použít cíle z dříve definovaných rovnic jako vstupy pro aktuální rovnici, což vám umožní vytvořit spřažené rovnice.
  - K rovnici můžete připojit popisný komentář. Komentáře se zobrazí spolu s rovnicí na kartě Model.
1. Zadejte název cíle. Volitelně můžete klepnutím na tlačítko **Upravit** pod textovým polem Cíl otevřít dialogové okno Definované vstupy a změnit výchozí vlastnosti cíle.
  2. Chcete-li sestavit výraz, buď vložte komponenty do pole Numerického výrazu, nebo zadejte přímo do pole Numerický výraz.
- Výraz můžete sestavit pomocí polí z aktivní datové sady nebo můžete definovat nové vstupy klepnutím na tlačítko **Nový**. Otevře se dialogové okno Definovat vstupy.
  - Funkce můžete vložit tak, že vyberete skupinu ze seznamu Skupina funkcí a klepnete na funkci v seznamu Funkce (nebo vyberete funkci a klepnete na šipku vedle seznamu skupin funkcí). Zadejte libovolné parametry označené otazníky. Skupina funkcí označená **Vše** poskytuje výpis všech dostupných funkcí. Krátký popis aktuálně vybrané funkce se zobrazí v rezervované oblasti v dialogovém okně.
  - Řetězcové konstanty musí být uzavřeny v uvozovkách.
  - Pokud hodnoty obsahují desetinná čísla, musí být jako desetinný indikátor použita tečka (.)

*Poznámka:* Simulace nepodporuje vlastní rovnice s řetězcovými cíli.

### *Definované vstupy*

Dialogové okno Definované vstupy vám umožňuje definovat nové vstupy a nastavovat vlastnosti pro cíle.

- Pokud vstup, který má být použit v rovnici, v aktivní datové sadě neexistuje, musíte jej definovat dříve, než bude možné jej použít v rovnici.
- Pokud simulujete data bez prediktivního modelu, musíte definovat všechny simulované vstupy, které v aktivní datové sadě neexistují.

**Název.** Zadejte název cíle nebo vstupu.

**Cíl.** Můžete určit úroveň měření cíle. Výchozí úroveň měření je souvislá. Můžete také určit, zda bude pro tento cíl vytvořen výstup. Například, pro sadu spřažené rovnice můžete mít zájem pouze o výstup z cíle pro konečnou rovnici, takže byste měli potlačit výstup z jiných cílů.

**Vstup, který má být simulován.** To znamená, že hodnoty vstupu budou simulovány podle zadané rozdělení pravděpodobnosti (rozdělení pravděpodobnosti je uvedeno na kartě Simulace). Úroveň měření určuje výchozí sadu distribucí, které se berou v úvahu při hledání distribuce, která nejvíce odpovídá datům pro vstup (klepnutím na **Přizpůsobit** nebo **Přizpůsobit vše** na kartě Simulace). Například, je-li úroveň měření spojitá, pak by se mělo uvažovat o normální distribuci (vhodné pro průběžné údaje), ale binomická distribuce by nebyla.

**Poznámka:** Vyberte úroveň měření řetězce pro vstupní řetězce. Řetězcové vstupy, které mají být simulovány, jsou omezeny na kategorickou distribuci.

**Vstup pevné hodnoty.** Tato hodnota určuje, že je známa hodnota vstupu a že bude opravena na známou hodnotu. Pevné vstupy mohou být číselné nebo řetězcové. Zadejte hodnotu pro pevný vstup. Řetězcové hodnoty by neměly být uzavřeny v uvozovkách.

**Popisky hodnot.** Můžete určit popisky hodnot pro cíle, simulované vstupy a pevné vstupy. Popisky hodnot se používají ve výstupních grafech a tabulkách.

## **Karta Simulace**

Karta Simulace určuje všechny vlastnosti simulace jiné než prediktivní model. Na kartě Simulace můžete provádět následující obecné úlohy:

- Určete rozdělení pravděpodobnosti pro simulované vstupy a hodnoty pro pevné vstupy.



- Určete korelace mezi simulovanými vstupy. Pro kategoriální vstupy můžete uvést, že přidružení, která existují mezi těmito vstupy v aktivní datové sadě, se použijí, když se vygenerují data pro tyto vstupy.
- Uvedte rozšířené volby, jako je odběr vzorků a kritéria pro instalaci distribucí do historických dat.
- Přizpůsobení výstupu.
- Určete, kam má být uložen plán simulace, a volitelně uložte simulovaná data.

## Simulovaná pole

Chcete-li spustit simulaci, musí být každé vstupní pole určeno jako pevné nebo simulované. Simulované vstupy jsou ty, jejichž hodnoty jsou nejisté a budou generovány vykreslením z uvedené pravděpodobnosti rozdělení. Jsou-li k dispozici historická data pro simulované vstupy, je možné automaticky určit distribuce, které nejlépe odpovídají datům, spolu se všemi korelacemi mezi těmito vstupy. Můžete také ručně uvést distribuce nebo korelace, pokud historická data nejsou k dispozici nebo požadujete specifické distribuce nebo korelace.

Pevné vstupy jsou ty, jejichž hodnoty jsou známy a zůstávají konstantní pro každý případ generovaný v rámci simulace. Máte například lineární regresní model pro prodej jako funkce řady vstupů, včetně ceny, a chcete držet cenu stanovenou za aktuální tržní cenu. Pak byste zadali cenu jako pevný vstup.

Pro simulace založené na předpovědním modelu je každý prediktor v modelu vstupním polem pro simulaci. U simulací, které nezahrnují prediktivní model, jsou pole, která jsou uvedena na kartě Model, vstupy simulace.

**Automatická instalace distribucí a výpočet korelací pro simulované vstupy.** Pokud aktivní datová sada obsahuje historická data pro vstupy, které chcete simulovat, pak můžete automaticky vyhledat distribuce, které nejvíce odpovídají datům pro tyto vstupy, a také určit případné korelace mezi nimi. Kroky jsou následující:

1. Ověřte, že každý ze vstupů, které chcete simulovat, je porovnán se správným polem v aktivní datové sadě. Vstupy jsou vypsány ve sloupci Vstup a Přizpůsobit sloupci zobrazuje odpovídající pole v aktivní datové sadě. Můžete porovnat vstup s jiným polem v aktivní datové sadě tak, že vyberete jinou položku ze seznamu Přizpůsobit na rozevírací seznam.

Hodnota *-None-* ve sloupci Přizpůsobit sloupci označuje, že vstup nemohl být automaticky porovnán s polem v aktivní datové sadě. Standardně se vstupy shodovaly s poli datové sady na názvu, úrovni měření a typu (číselné nebo řetězcové). Pokud aktivní datová sada neobsahuje historická data pro vstup, pak ručně zadejte distribuci pro vstup, nebo uveďte vstup jako pevný vstup, jak je popsáno níže.

2. Klepněte na volbu **Přizpůsobit vše**.

Nejbližší odpovídající distribuce a její přidružené parametry se zobrazí ve sloupci Distribuce spolu s zákresovým grafem na histogramu (nebo sloupcovém grafu) historických dat. Korelace mezi simulovanými vstupy se zobrazují v nastavení Korelace. Výběrem řádku pro vstup a klepnutím na **Přizpůsobit podrobnosti** můžete zkontrolovat výsledky přizpůsobení a přizpůsobit automatické rozložení distribuce pro určitý vstup. Další informace naleznete v tématu [“Přizpůsobit podrobnosti”](#) na stránce 243.

Pro konkrétní vstup můžete spustit automatické rozložení distribuce tak, že vyberete řádek pro vstup a klepnete na **Přizpůsobit**. Korelace pro všechny simulované vstupy, které odpovídají polím v aktivní datové sadě, jsou také automaticky vypočítány.

### Poznámka:

- Případy s chybějícími hodnotami pro všechny simulované vstupy jsou vyloučeny z distribuce, výpočtu korelací a výpočtem volitelné rezervní tabulky (pro vstupy s kategoričnou distribucí). Volitelně můžete určit, zda uživatelské hodnoty vstupů s kategoričnou distribucí jsou považovány za platné. Při výchozím nastavení se s nimi zachází jako s chybějícími. Další informace naleznete v tématu [“Rozšířené volby”](#) na stránce 244.
- Pro souvislé a ordinální vstupy, pokud nelze najít přijatelné vhodné vybavení pro některou z testovaných distribucí, je jako nejbližší umístění doporučováno Empirical distribution. Pro souvislé vstupy je Empirical distribuce kumulativní distribuční funkcí historických dat. Pro ordinální vstupy je Empirická distribuce kategoričnou distribucí historických dat.

**Manuální uvedení distribucí.** Můžete ručně zadat rozdělení pravděpodobnosti pro libovolný simulovaný vstup tak, že vyberete rozdělení z rozevíracího seznamu **Typ** a zadáte parametry distribuce v mřížce parametrů. Jakmile zadáte parametry pro distribuci, zobrazí se ukázkový výkres distribuce, založený na uvedených parametrech, sousedící s mřížkou parametrů. Níže jsou uvedeny některé poznámky týkající se jednotlivých distribucí:

- **Kategorické.** Kategorické rozdělení popisuje vstupní pole, které má pevný počet hodnot, které se označují jako kategorie. Každá kategorie má přidruženou pravděpodobnost tak, že součet pravděpodobností se u všech kategorií rovná jednomu. Chcete-li zadat kategorii, klepněte na levý sloupec v mřížce parametrů a určete hodnotu kategorie. Zadejte pravděpodobnost přidruženou ke kategorii ve sloupci na pravé straně.

**Poznámka:** Kategorické vstupy z modelu PMML mají kategorie, které jsou určeny z modelu a nelze je upravit.

- **Negativní Binomické-Selhání.** Popisuje rozdělení počtu selhání v posloupnosti pokusů před určeným počtem úspěšných pokusů. Parametr *thresh* je určený počet úspěchů a parametr *prob* je pravděpodobnost úspěchu v jakémkoli pokusu.
- **Negativní binomické-pokusy.** Popisuje rozdělení počtu požadovaných pokusů, které se požadují před uvedeným počtem úspěšných pokusů. Parametr *thresh* je určený počet úspěchů a parametr *prob* je pravděpodobnost úspěchu v jakémkoli pokusu.
- **Rozsah.** Tato distribuce se skládá ze sady intervalů s pravděpodobností přiřazenou každému intervalu tak, aby součet pravděpodobnosti za všechny intervaly se rovná 1. Hodnoty v daném intervalu jsou vykresleny z rovnoměrného rozdělení definovaného v daném intervalu. Intervaly jsou určovány zadáním minimální hodnoty, maximální hodnotou a přidruženou pravděpodobností.

Například se domníváte, že náklady na suroviny mají 40% šanci na snížení v rozmezí 10 až 15 dolarů za jednotku a 60% šanci na pokles v rozmezí \$15-\$20 za jednotku. Modeloval byste náklady s distribucí Range sestávající ze dvou intervalů [ 10-15] a [ 15-20], přičemž nastavíte pravděpodobnost spojenou s prvním intervalem na 0,4 a pravděpodobnost přidruženou k druhému intervalu na 0.6. Intervaly nemusí být souvislé a mohou se dokonce překrývat. Mohli jste například zadat intervaly \$10-\$15 a \$20-\$25 nebo \$10-\$15 a \$13-\$16.

- **Weibullovo.** Parametr *c* je volitelný parametr umístění, který uvádí, kde se nachází původ distribuce.

Parametry pro následující distribuce mají stejný význam jako v přidružených náhodných proměnných funkcích dostupných v dialogovém okně Compute Variable: Bernoulli, beta, binomická, exponenciální, gama, logormální, negativní binomické (pokusy a selhání), normální, Poisson a uniforma.

**Určení pevných vstupů.** Uvedte pevný vstup tak, že vyberete volbu Pevné z rozevíracího seznamu **Typ** ve sloupci Distribuce a zadejte pevnou hodnotu. Hodnota může být číselná nebo řetězcová v závislosti na tom, zda je vstup číselný nebo řetězcový. Řetězcové hodnoty by neměly být uzavřeny v uvozovkách.

**Určení omezení pro simulované hodnoty.** Většina distribucí podporuje určení horních a dolních hranic na simulovaných hodnotách. Dolní mez můžete zadat zadáním hodnoty do textového pole **Min** a můžete zadat horní mez zadáním hodnoty do textového pole **Maximum**.






**Zamykání vstupů.** Zamykání vstupu označením zaškrtačacího políčka ve sloupci s ikonou zámku vylučuje vstup z automatické instalace distribuce. To je nejužitečnější, když ručně uvedete distribuci nebo pevnou hodnotu a chcete se ujistit, že to nebude ovlivněno automatickou distribucí. Zamykání je také užitečné, pokud zamýšlíte sdílet svůj plán simulace s uživateli, kteří jej budou spouštět v dialogovém okně Spustit simulaci a chcete zabránit změnám v některých vstupech. V tomto ohledu nelze v dialogovém okně Spustit simulaci upravit specifikace pro uzamčené vstupy.

**Analýza citlivosti.** vám umožňuje vyšetřit vliv systematických změn na pevný vstup nebo v distribučním parametru pro simulovaný vstup generováním nezávislé sady simulovaných případů-efektivně samostatnou simulací-pro každou zadanou hodnotu. Chcete-li určit analýzu citlivosti, vyberte pevný nebo simulovaný vstup a klepněte na volbu **Analýza citlivosti**. Analýza citlivosti je omezena na jednoduchý pevný vstup nebo jeden distribuční parametr pro simulovaný vstup. Další informace naleznete v tématu "[Analýza citlivosti](#)" na stránce 244.

Přízpůsobit ikony stavu

Ikony ve sloupci Přizpůsobit sloupci označují stav přizpůsobení pro každé vstupní pole.

Tabulka 5. Ikony stavu

Ikona	Popis
	Nebyla zadána žádná distribuce pro vstup a vstup nebyl zadán jako pevný. Chcete-li spustit simulaci, musíte buď uvést distribuci pro tento vstup, nebo definovat její pevnou hodnotu a uvést pevnou hodnotu.
	Vstup byl dříve vhodný pro pole, které neexistuje v aktivní datové sadě. Není nutná žádná akce, pokud nechcete, aby byla distribuce znovu přizpůsobena pro vstup do aktivní datové sady.
	Nejbližší distribuce byla nahrazena alternativní distribucí z dialogu Přizpůsobit podrobnosti.
	Vstup je nastaven na nejblížejší distribuci.
	Distribuce byla zadána ručně nebo pro tento vstup byly zadány iterace analýzy citlivosti.

#### Přizpůsobit podrobnosti

Dialogové okno Přizpůsobit podrobnosti zobrazí výsledky automatického rozložení distribuce pro konkrétní vstup. Distribuce jsou seřazeny podle dobroty pro přizpůsobení, s nejbližší distribucí distribuce uvedená jako první. Můžete přepsat nejbližší distribuci tím, že vyberete přepínač pro distribuci, kterou chcete ve sloupci Použít. Výběrem přepínače ve sloupci Použít se také zobrazí zakres rozložení na histogramu (nebo pruhovém grafu) historických dat pro daný vstup.

**Přizpůsobit statistiky.** Ve výchozím nastavení a pro souvislá pole se používá test Anderson-Darling pro určení dobroty pro uložení. Alternativně, a pouze pro souvislá pole, můžete zadat test Kolmogorov-Smirnov pro dobročinnou volbu tím, že vyberete tuto volbu v nastavení rozšířených voleb. U spojitých vstupů se výsledky obou testů zobrazují ve sloupci Přizpůsobit statistiku (A pro Anderson-Darling a K pro Kolmogorov-Smirnov), s vybranou zkouškou použitou k objednání distribucí. Pro pořadové a nominální vstupy se použije zkouška chí kvadrát. Zobrazeny jsou také p-hodnoty přidružené k testům.

**Parametry.** Parametry distribuce přidružené k každé instalované distribuci jsou zobrazeny ve sloupci Parametry. Parametry pro následující distribuce mají stejný význam jako v přidružených náhodných proměnných funkcích dostupných v dialogovém okně Compute Variable: Bernoulli, beta, binomická, exponenciální, gama, logormální, negativní binomické (pokusy a selhání), normální, Poisson a uniforma. Pro kategorickou distribuci jsou názvy parametrů kategorií a hodnotami parametrů jsou přidružené pravděpodobnosti.

**Pronařizování s vlastní sadou distribucí.** Standardně se úroveň měření pro vstup používá k určení sady distribucí uvažovaných pro automatické rozdělení distribuce. Například souvislé distribuce, jako jsou logormální a gama, se berou v úvahu při přizpůsobení souvislého vstupu, ale diskrétní distribuce jako Poisson a binomial nejsou. Výběrem distribucí ve sloupci Refit můžete vybrat podmnožinu výchozích distribucí. Výchozí sadu distribucí můžete také přepsat výběrem jiné úrovně měření z rozevíracího seznamu **Považovat jako (Ukazatel)** a výběrem rozdělení ve sloupci Přeložení. Chcete-li se znovu přizpůsobit vlastní sadě distribuce, klepněte na volbu **Spustit přeložení**.

#### Poznámka:

- Případy s chybějícími hodnotami pro všechny simulované vstupy jsou vyloučeny z distribuce, výpočtu korelací a výpočtem volitelné rezervní tabulky (pro vstupy s kategorickou distribucí). Volitelně můžete určit, zda uživatelské hodnoty vstupů s kategorickou distribucí jsou považovány za platné. Při výchozím nastavení se s nimi zachází jako s chybějícími. Další informace naleznete v tématu [“Rozšířené volby” na stránce 244.](#)

- Pro souvislé a ordinální vstupy, pokud nelze najít přijatelné vhodné vybavení pro některou z testovaných distribucí, je jako nejbližší umístění doporučováno Empirical distribution. Pro souvislé vstupy je Empirical distribuce kumulativní distribuční funkcí historických dat. Pro ordinální vstupy je Empirická distribuce kategoričnou distribucí historických dat.

### *Analýza citlivosti*

Analýza citlivosti vám umožňuje vyšetřit účinek simulovaného vstupu nebo distribučního parametru pro simulovaný vstup přes zadanou sadu hodnot. Samostatná sada simulovaných případů-efektivně, samostatná simulace-je generována pro každou uvedenou hodnotu, což vám umožňuje vyšetřit efekt, který se mění na vstup. Každá sada simulovaných případů se odkazuje jako na **iteraci**.

**Iterujte.** Tato volba vám umožňuje uvést sadu hodnot, přes které bude vstup logicky aktivován.

- Pokud se liší hodnota parametru distribuce, pak vyberte parametr z rozevíracího seznamu. Zadejte sadu hodnot v hodnotě parametru podle iterace mřížky. Klepnutím na tlačítko **Pokračovat** přidáte uvedené hodnoty do mřížky Parameters pro přidružený vstup, s indexem uvedením číslo iterace hodnoty.
- U distribucí Categorical a Range mohou být pravděpodobnosti jednotlivých kategorií nebo intervalů různé, ale hodnoty kategorií a koncových bodů intervalů nemohou být různé. Vyberte kategorii nebo interval z rozevíracího seznamu a určete sadu pravděpodobností v hodnotě parametru podle iterace mřížky. Součet pravděpodobností pro ostatní kategorie nebo intervaly bude odpovídajícím způsobem automaticky upraven.

**Žádné iterace.** Tuto volbu použijte ke zrušení iterací pro vstup. Klepnutím na tlačítko **Pokračovat** odeberete iterace.

### **Korelace**

Vstupní pole, která se mají simulovat, jsou často známy jako korelovaná pole -- například výška a váha. Korelace mezi vstupními výstupy, které budou simulovány, musí být začítovány, aby se zajistilo, že simulované hodnoty zachovají tyto korelace.

**Přepočítat korelace při instalaci.** Tato volba určuje, že korelace mezi simulovanými vstupy se automaticky počítají při instalaci distribucí do aktivní datové sady pomocí akcí **Přízpusobit vše** nebo **Přízpusobit** na nastavení simulovaného pole.

**Při montáži nepřepočítávat korelace.** Vyberte tuto volbu, chcete-li ručně zadat korelace a zabránit jejich přepsání při automatické instalaci distribucí do aktivní datové sady. Hodnoty zadané v mřížce Correlations musí být mezi -1 a 1. Hodnota 0 určuje, že mezi přidruženou dvojicí vstupů neexistuje žádná korelace.

**Resetovat.** Tím se resetují všechny korelace na 0.

**Pro vstupy s kategoričnou distribucí použijte zabudovaný vícesměrný rezervní tabulka pro vstupy.**

Pro vstupy s kategoričnou distribucí můžete automaticky vypočítat vícecestnou tabulku pro nepředvídané události z aktivní datové sady, která popisuje přidružení mezi těmito vstupy. Nouzová tabulka se pak použije, když jsou data generována pro tyto vstupy. Pokud zvolíte uložení plánu simulace, rezervní tabulka se uloží do souboru s plánem a použije se při spuštění plánu.

- **Výpočet rezervní tabulky z aktivní datové sady.** Pokud pracujete s existujícím simulačním plánem, který obsahuje tabulku pro nepředvídané události, můžete znovu vypočítat tabulku pro nepředvídané události z aktivní datové sady. Tato akce přepíše náhradní tabulku ze souboru načteného plánu.
- **Použit pohotovostní tabulku z načteného plánu simulace.** Při načítání simulačního plánu, který obsahuje tabulku pro nepředvídané události, se standardně použije tabulka z plánu. Pohotovostní tabulku můžete znovu vypočítat z aktivní datové sady tak, že vyberete volbu **Vypočítat rezervní tabulku z aktivní datové sady**.

### **Rozšířené volby**

**Maximální počet případů.** Tato hodnota určuje maximální počet případů simulovaných dat a přidružených cílových hodnot, které mají být generovány. Je-li uvedena analýza citlivosti, jedná se o maximální počet případů pro každou iteraci.

**Cíl pro zastavení kritérií.** Pokud váš prediktivní model obsahuje více než jeden cíl, pak můžete vybrat cíl, na který se použijí kritéria zastavení.

**Zastavení kritérií.** Tyto volby určují kritéria pro zastavení simulace, případně dříve, než bude vygenerován maximální počet přípustných případů.

- **Pokračujte až do dosažení maximálního počtu.** Tato hodnota určuje, že se budou generovat simulované případy, dokud nebude dosaženo maximálního počtu případů.
- **Zastavte, kdy byl odebrán vzorek drobků.** Tuto volbu použijte, chcete-li zajistit, aby jedna z ocasů určité cílové distribuce byla adekvátně vzorkována. Simulované případy budou generovány, dokud nebude dokončen odběr vzorků z ocasu nebo dokud nebude dosaženo maximálního počtu případů. Pokud váš prediktivní model obsahuje více cílů, pak vyberte cíl, na který se tato kritéria použijí, z rozevíracího seznamu **Cíl pro zastavení kritérií** .

**Typ.** Hranice oblasti koncového bodu můžete definovat určením hodnoty cíle, jako je 10.000.000 nebo percentilu, jako je například 99th percentil. Vyberete-li hodnotu v rozevíracím seznamu **Typ** , zadejte hodnotu okraje do textového pole Hodnota a pomocí rozevíracího seznamu **Boční** určete, zda se jedná o hranici oblasti Ole vlevo nebo oblasti pravého konce. Vyberete-li Percentil v rozevíracím seznamu **Typ** , zadejte hodnotu do textového pole Percentil.

**Frekvence.** Uvedte počet hodnot cíle, které musí ležet v oblasti zbytkové části, aby se zajistilo, že ocas byl odpovídajícím způsobem zařazen do vzorku. Generování případů se zastaví, když bude dosaženo tohoto čísla.

- **Zastavit, je-li interval spolehlivosti střední hodnoty střední hodnoty v rámci uvedené prahové hodnoty.** Tuto volbu použijte, chcete-li zajistit, aby průměr daného cíle byl znám s určenou mírou přesnosti. Simulované případy budou generovány, dokud nebude dosaženo uvedeného stupně přesnosti nebo dokud není dosaženo maximálního počtu případů. Chcete-li použít tuto volbu, uvedete úroveň důvěry a prahovou hodnotu. Simulované případy budou generovány, dokud nebude interval spolehlivosti přidružený k uvedené úrovni v rámci prahové hodnoty. Například můžete tuto volbu použít k uvedení, že se případy vygenerují, dokud se interval spolehlivosti střední hodnoty 95% neuvede do 5% střední hodnoty. Pokud váš prediktivní model obsahuje více cílů, pak vyberte cíl, na který se tato kritéria použijí, z rozevíracího seznamu **Cíl pro zastavení kritérií** .

**Typ prahové hodnoty.** Prahovou hodnotu můžete zadat jako číselnou hodnotu nebo jako procentní část střední hodnoty. Pokud vyberete hodnotu v rozevíracím seznamu **Typ prahové hodnoty** , pak zadejte prahovou hodnotu do textového pole Prahová hodnota jako Hodnota. Pokud vyberete procento v rozevíracím seznamu **Typ prahové hodnoty** , pak zadejte hodnotu do textového pole Prahová hodnota jako procento.

**Počet případů do vzorku.** Uvádí počet případů, které se použijí při automatické instalaci distribucí pro simulované vstupy do aktivní datové sady. Je-li vaše datová sada velmi velká, možná byste měli zvážit omezení počtu případů použitých pro montáž k distribuci. Vyberete-li volbu **Omezit na případy N**, bude použito prvních N případů.

**Souběžná kritéria (Souvislá).** Pro souvislé vstupy můžete použít test Anderson-Darling nebo test Kolmogorovov-Smirnoff, který se hodí ke kategoriím rozdělení při montáži distribucí pro simulované vstupy do aktivní datové sady. Test Anderson-Darling je při výchozím nastavení vybrán a je doporučen, když chcete zajistit nejlepší možné uložení do koncových oblastí.

**Empirická distribuce.** Pro souvislé vstupy je Empirical distribuce kumulativní distribuční funkcí historických dat. Můžete uvést počet příhrádek použitých pro výpočet Empirické distribuce pro souvislé vstupy. Výchozí hodnota je 100 a maximum je 1000.

**Proveďte replikaci výsledků.** Nastavení náhodného objektu seed vám umožní replikovat simulaci. Zadejte celé číslo nebo klepněte na tlačítko **Generovat**, které vytvoří pseudo-náhodné celé číslo v rozsahu 1 až 2147483647 včetně. Předvolba je 629111597.

**Poznámka:** Pro konkrétní náhodný řetězec se replikují výsledky, pokud se nezmění počet podprocesů. Na konkrétním počítači je počet podprocesů konstantní, pokud jej nezměníte spuštěním syntaxe příkazu SET THREADS . Počet podprocesů se může změnit, pokud spustíte simulaci na jiném počítači, protože se používá interní algoritmus k určení počtu podprocesů pro každý počítač.

**Uživatel-chybějící hodnoty pro vstupy s kategoričkou distribucí.** Tyto ovládací prvky určují, zda uživatelské hodnoty vstupů s kategoričkou distribucí jsou považovány za platné. Systémem-chybějící hodnoty a uživatelské hodnoty pro všechny ostatní typy vstupů se vždy považují za neplatné. Všechny vstupy musí mít platné hodnoty pro případ, který se má zahrnout do distribuce k distribuci, výpočtu korelací a výpočet volitelné rezervní tabulky.

## Hustota funkcí

Tato nastavení umožňují přizpůsobit výstup pro funkce hustoty pravděpodobnosti a kumulativní distribuční funkce pro souvislé cíle, stejně jako sloupcové grafy předpovídaných hodnot pro kategoriální cíle.

**Funkce Hustota pravděpodobnosti (PDF).** Funkce hustoty pravděpodobnosti zobrazí distribuci cílových hodnot. Pro souvislé cíle vám umožňuje určit pravděpodobnost, že se cíl nachází v daném regionu. Pro kategoriální cíle (cíle s úrovní měření nominálního nebo pořadového) se vygeneruje pruhový graf, který zobrazuje procentní část případů, které spadají do každé kategorie cíle. Další volby pro kategoriální cíle modelů PMMML jsou k dispozici s hodnotami kategorií, které jsou popsány níže.

U modelů klastrů Two-Step a modelů K-Means je vytvořen pruhový graf členství klastru.

**Kumulativní distribuční funkce (CDF).** Funkce kumulativní distribuce zobrazuje pravděpodobnost, že hodnota cíle je menší nebo rovna zadané hodnotě. Je k dispozici pouze pro souvislé cíle.

**Pozice posuvného ovladače.** Můžete určit počáteční pozice přesouvatelných referenčních řádků v grafech PDF a CDF. Hodnoty, které jsou určeny pro dolní a horní čáry, se odkazují na pozice podél vodorovné osy, nikoli percentily. Chcete-li odstranit dolní řádek, vyberte volbu **-Infinity** nebo můžete odstranit řádek s horním řádkem výběrem volby **Infinity**. Standardně jsou řádky umístěny na 5-té a 95. percentilech. Je-li v jednom diagramu zobrazeno více funkcí rozdělení (z důvodu více cílů nebo výsledků z iterací analýzy citlivosti), výchozí odkazuje na distribuci pro první iteraci nebo první cíl.

**Referenční čáry (Souvislé).** Můžete požadovat různé vertikální referenční čáry, které mají být přidány do funkcí hustoty pravděpodobnosti, a kumulativní distribuční funkce pro souvislé cíle.

- **Sigmy.** Můžete přidat referenční řádky za plus a minus uvedený počet směrodatných odchylek od střední hodnoty cíle.
- **Percentily.** Na jednu nebo dvě hodnoty percentilu distribuce cíle můžete přidat referenční řádky zadáním hodnot do textových polí Dolní a Horní. Například hodnota 95 v horním textovém poli představuje 95th percentilu, což je hodnota, pod níž 95% pozorování spadne. Podobně hodnota 5 v dolním textovém poli představuje 5th percentil, což je hodnota pod níž 5% pozorování spadne.
- **Vlastní referenční čáry.** Na uvedených hodnotách cíle můžete přidávat referenční čáry.

**Poznámka:** Je-li v jednom diagramu zobrazeno více funkcí rozdělení (z důvodu více cílů nebo výsledků z iterací analýzy citlivosti), jsou referenční čáry použity pouze pro distribuci pro první iteraci nebo první cíl. Do ostatních rozdělení můžete přidat referenční čáry z dialogového okna Volby grafu, které je přístupné z grafu PDF nebo CDF.

**Překrývané výsledky se samostatnými spojitými cíli.** V případě více souvislých cílů uvádí, zda jsou distribuční funkce pro všechny takové cíle zobrazeny v jednom grafu, s jedním grafem pro funkce hustoty pravděpodobnosti a další pro kumulativní distribuční funkce. Není-li tato volba vybrána, výsledky pro každý cíl se zobrazí na samostatném grafu.

**Hodnoty kategorie pro sestavu.** Pro modely PMMML s kategoriálními cíli je výsledek modelu sadou předpokládaných pravděpodobností, jedna pro každou kategorii, že cílová hodnota spadá do každé kategorie. Kategorie s nejvyšší pravděpodobností je převzata jako předpovězená kategorie a používá se při generování pruhového grafu popsaného výše pro nastavení **Funkce hustoty hustoty**. Vyberete-li volbu **Předpovězená kategorie**, bude pruhový graf vygenerován. Vyberete-li volbu **Predikovaná pravděpodobnost**, vygenerují se histogramy rozdělení předpokládaných pravděpodobností pro každou z kategorií cíle.

**Seskupení pro analýzu citlivosti.** Simulace, které zahrnují analýzu citlivosti, generují nezávislou sadu předpokládaných cílových hodnot pro každou iteraci definovanou analýzou (jedna iterace pro každou hodnotu vstupu, která se mění). Jsou-li přítomny iterace, sloupcový graf předpovídané kategorie pro

kategorický cíl se zobrazí jako klastrovaný pruhový graf, který zahrnuje výsledky pro všechny iterace. Můžete se rozhodnout seskupit skupiny dohromady nebo můžete seskupit iterace dohromady.

## Výstup

**Tornádové grafy.** Tornádovo grafy jsou pruhové grafy, které zobrazují vztahy mezi cíli a simulovanými vstupy pomocí různých metrik.

- **Korelace cíle se vstupem.** Tato volba vytvoří tornádo grafu korelačních koeficientů mezi daným cílem a každým ze svých simulovaných vstupů. Tento typ diagramu tornado nepodporuje cíle se jmenovitou nebo ordinální úrovní měření nebo simulovanými vstupy s kategorickou distribucí.
- **Příspěvek k rozptylu.** Tato volba vytvoří graf tornádo, který zobrazuje příspěvek k rozptylu cíle z každého ze simulovaných vstupů, což vám umožňuje posoudit míru, do jaké každý vstup přispívá k celkové nejistotě v cíli. Tento typ diagramu tornado nepodporuje cíle s pořadovými nebo nominálními úrovněmi měření nebo simulované vstupy s některou z následujících distribucí: kategorická, Bernoullim, binomická, Poissonova nebo záporná binomická.
- **Citlivost cíle, který má být změněn.** Tato volba vytvoří graf tornádo, který zobrazí efekt na cíl modulace každého simulovaného vstupu plus nebo minus uvedený počet směrodatných odchylek distribuce přidružených ke vstupu. Tento typ diagramu tornado nepodporuje cíle s pořadovými nebo nominálními úrovněmi měření nebo simulované vstupy s některou z následujících distribucí: kategorická, Bernoullim, binomická, Poissonova nebo záporná binomická.

**Políčko rámečku cílových distribucí.** Políčkové grafy jsou k dispozici pro souvislé cíle. Vyberte **Překryv výsledků ze samostatných cílů**, pokud má váš prediktivní model více souvislých cílů a vy chcete zobrazit krabičkové grafy pro všechny cíle na jednom grafu.

**Korelační grafy cílů oproti vstupům.** Korelační grafy cílů a simulované vstupy jsou k dispozici jak pro souvislé, tak i kategoriální cíle, a obsahují také rozptylové cíle cíle se spojitými vstupy a kategoriálními vstupy. Jako teplotní mapa se zobrazují rozptyl zahrnující kategorický cíl nebo kategorický vstup.

**Vytvořte tabulku hodnot percentilů.** Pro souvislé cíle můžete získat tabulku uvedených percentilů cílových distribucí. Kvartily (25th, 50th, 75th percentily) rozdělují pozorování do čtyř skupin stejné velikosti. Chcete-li mít stejný počet jiných skupin než čtyři, vyberte volbu **Intervaly** a zadejte číslo. Výběrem volby **Vlastní percentily** určete jednotlivé percentily -- například percentil 99th .

**Deskriptivní statistiky cílových distribucí.** Tato volba vytváří tabulky popisných statistik pro souvislé a kategoriální cíle, stejně jako pro souvislé vstupy. Pro souvislé cíle tabulka zahrnuje střední, směrodatnou odchylku, medián, minimum a maximum, interval spolehlivosti střední hodnoty na zadané úrovni a 5th a 95th percentily cílové distribuce. Pro kategoriální cíle tabulka zahrnuje procentní část případů, které spadají do každé kategorie cíle. Pro kategoriální cíle modelů PMML obsahuje tabulka také střední pravděpodobnost každé kategorie cíle. U souvislých vstupů tabulka obsahuje střední, směrodatné odchylky, minimum a maximum.

**Korelační a rezervní tabulka pro vstupy.** Tato volba zobrazuje tabulku korelačních koeficientů mezi simulovanými vstupy. Jsou-li vstupy s kategorickými distribucemi generovány z tabulky pro nepředvídané události, zobrazí se také rezervní tabulka dat generovaná pro tyto vstupy.

**Simulované vstupy, které mají být zahrnuty do výstupu.** Ve výchozím nastavení jsou do výstupu zahrnuty všechny simulované vstupy. Vybrané simulované vstupy můžete vyloučit z výstupu. To je vyloučí z tornádových grafů, bodových grafů a tabulkových výstupů.

**Omezit rozsahy pro souvislé cíle.** Můžete uvést rozsah platných hodnot pro jeden nebo více souvislých cílů. Hodnoty mimo určený rozsah jsou vyloučeny ze všech výstupů a analýz přidružených k cílům. Chcete-li nastavit dolní limit, vyberte ve sloupci Omezení hodnotu **Dolní** a zadejte hodnotu do sloupce Minimum. Chcete-li nastavit horní limit, vyberte **Upper** ve sloupci Limit a zadejte hodnotu do sloupce Maximum. Chcete-li nastavit dolní i horní limit, vyberte volbu **Obojí** ve sloupci Omezení a zadejte hodnoty do sloupců Minimum a Maximum.

**Formáty zobrazení.** Můžete nastavit formát, který se používá při zobrazování hodnot cílů a vstupů (pevných vstupů i simulovaných vstupů).

## Uložit

**Uložte plán pro tuto simulaci.** Aktuální specifikace pro simulaci můžete uložit do souboru s plánem simulace. Soubory plánu simulace mají příponu *.splan*. Plán můžete znovu otevřít v Tvůrci simulací, volitelně provést úpravy a spustit simulaci. Simulační plán můžete sdílet s dalšími uživateli, kteří jej poté mohou spouštět v dialogovém okně Spustit simulaci. Plány simulací zahrnují všechny specifikace kromě následujících: nastavení pro funkce hustoty; Nastavení výstupu pro grafy a tabulky; Nastavení rozšířených voleb pro Fitinky, Empirical Distribution a Random Seed.

**Uložte simulovaná data jako nový datový soubor.** Můžete uložit simulované vstupy, pevné vstupy a předpovídané cílové hodnoty do datového souboru SPSS Statistics, nové datové sady v aktuální relaci nebo souboru aplikace Excel. Každý případ (nebo řádek) datového souboru se skládá z předpokládaných hodnot cílů společně se simulovanými vstupy a pevnými vstupy, které generují cílové hodnoty. Je-li uvedena analýza citlivosti, každá iterace vede k souvislé sadě případů, které jsou označeny číslem iterace.

## Spustit dialogové okno simulace

Dialogové okno Spustit simulaci je navrženo pro uživatele, kteří mají simulační plán a primárně chtějí spustit simulaci. Poskytuje také funkce, které potřebujete ke spuštění simulace za různých podmínek. Umožňuje provádět následující obecné úlohy:

- Nastavení nebo úprava analýzy citlivosti za účelem zkoumání vlivu změny hodnoty pevného vstupu nebo proměnné distribučního parametru pro simulovaný vstup.
- Přeložení pravděpodobnostních rozdělení pro nejisté vstupy (a korelace mezi těmito vstupy) na nová data.
- Upravte distribuci pro simulovaný vstup.
- Přizpůsobení výstupu.
- Spusťte simulaci.

## Karta Simulace

Ouško Simulace vám umožňuje uvést analýzu citlivosti, přizpůsobit rozdělení pravděpodobnosti pro simulované vstupy a korelace mezi simulovanými vstupy do nových dat a upravit rozdělení pravděpodobnosti přidružené se simulovaným vstupem.

Mřížka Simulovaná vstupní mřížka obsahuje položku pro každé vstupní pole, které je definováno v plánu simulace. Každá položka zobrazuje název vstupu a typ rozdělení pravděpodobnosti přidružený ke vstupu, spolu s ukázkou grafu přidružené distribuční křivky. Každý vstup má také přidruženou ikonu stavu (barevné kolečko se zaškrtnutím), které je užitečné, když znovu vytváříte distribuce na nová data. Kromě toho mohou vstupy zahrnovat ikonu zámku, která označuje, že vstup je uzamčen a nelze jej upravit nebo přizpůsobit novým datům v dialogovém okně Spustit simulaci. Chcete-li upravit uzamčený vstup, bude třeba otevřít plán simulace v tvůrci simulace.

Každý vstup je buď simulován, nebo pevný. Simulované vstupy jsou ty, jejichž hodnoty jsou nejisté a budou generovány vykreslením z uvedené pravděpodobnosti rozdělení. Pevné vstupy jsou ty, jejichž hodnoty jsou známy a zůstávají konstantní pro každý případ generovaný v rámci simulace. Chcete-li pracovat s určitým vstupem, vyberte vstup pro vstup v mřížce Simulovaných vstupů.

### Určení analýzy citlivosti

vám umožňuje vyšetřit vliv systematických změn na pevný vstup nebo v distribučním parametru pro simulovaný vstup generováním nezávislé sady simulovaných případů-efektivně samostatnou simulací-pro každou zadanou hodnotu. Chcete-li určit analýzu citlivosti, vyberte pevný nebo simulovaný vstup a klepněte na volbu **Analýza citlivosti**. Analýza citlivosti je omezena na jednoduchý pevný vstup nebo jeden distribuční parametr pro simulovaný vstup. Další informace naleznete v tématu [“Analýza citlivosti” na stránce 244](#).

### Přeložení distribucí na nová data

Automatická změna pravděpodobnostní distribuce pro simulované vstupy (a korelace mezi simulovanými vstupy) na data v aktivní datové sadě:



1. Ověřte, že každý z modelů vstupů je porovnán se správným polem v aktivní datové sadě. Každý simulovaný vstup se vejde do pole v aktivní datové sadě určené v rozevíracím seznamu **Pole** přidruženém k tomuto vstupu. Vstupy, které si neodpovídají, můžete snadno identifikovat vyhledáním vstupů se stavem, který obsahuje zaškrtnutí s otazníkem, jak je zobrazeno níže.



2. Upravte všechny potřebné shody polí výběrem volby **Přizpůsobit poli v datové sadě** a výběrem pole ze seznamu.
3. Klepněte na volbu **Přizpůsobit vše**.

Pro každý vstup, který byl vhodný, se zobrazí rozdělení, které nejvíce odpovídá datům, spolu s zakreslením distribuce, která je uložena na histogramu (nebo sloupcovém grafu) historických dat. Pokud nelze nalézt přijatelné uložení, použije se Empirická distribuce. Pro vstupy, které se vejdou do Empirické distribuce, uvidíte pouze histogram historických dat, protože Empirická distribuce je ve skutečnosti reprezentována daným histogramem.

*Poznámka:* Úplný seznam ikon stavu naleznete v tématu [“Simulovaná pole”](#) na stránce 241.

Úprava rozdělení pravděpodobnosti

Můžete upravit rozdělení pravděpodobnosti pro simulovaný vstup a volitelně změnit simulovaný vstup na pevný vstup, nebo naopak.

1. Vyberte vstup a vyberte volbu **Ruční nastavení distribuce**.
2. Vyberte typ rozdělení a určete parametry distribuce. Chcete-li změnit simulovaný vstup na pevný vstup, vyberte volbu Pevná v rozevíracím seznamu **Typ**.

Jakmile zadáte parametry pro distribuci, ukázkový výkres distribuce (zobrazený v položce pro vstup) bude aktualizován, aby odrážel vaše změny. Další informace o ručním určení rozdělení pravděpodobnosti naleznete v tématu [“Simulovaná pole”](#) na stránce 241.

**Zahrnout uživatele-chybějící hodnoty kategoriálních vstupů při instalaci.** To uvádí, zda uživatelské hodnoty vstupů s kategoriálním distribucí jsou považovány za platné, když jste znovu odmítli data v aktivní datové sadě. Systémem-chybějící hodnoty a uživatelské hodnoty pro všechny ostatní typy vstupů se vždy považují za neplatné. Všechny vstupy musí mít platné hodnoty pro případ, který se má zahrnout do instalace distribuce a výpočtu korelací.

## Karta Výstup

Ouško Výstup vám umožňuje upravit výstup generovaný simulací.

**Hustota funkcí.** Charakteristiky hustoty jsou primární prostředky pro zkoušení sady výsledků ze simulace.

- **Funkce hustoty pravděpodobnosti.** Funkce hustoty pravděpodobnosti zobrazuje distribuci cílových hodnot, což vám umožňuje určit pravděpodobnost, že cíl je v daném regionu. Pro cíle s výsledky fixní sady-například "špatná služba", "fair service", "dobrá služba" a "vynikající služba" -- se vygeneruje sloupcový graf, který zobrazuje procento případů, které spadají do každé kategorie cíle.
- **Kumulativní distribuční funkce.** Funkce kumulativní distribuce zobrazuje pravděpodobnost, že hodnota cíle je menší nebo rovna zadané hodnotě.

**Grafy Tornado.** Tornadovo grafy jsou pruhové grafy, které zobrazují vztahy mezi cíli a simulovanými vstupy pomocí různých metrik.

- **Korelace cíle se vstupem.** Tato volba vytvoří tornádo grafu korelačních koeficientů mezi daným cílem a každým ze svých simulovaných vstupů.
- **Příspěvek k rozptylu.** Tato volba vytvoří graf tornádo, který zobrazuje příspěvek k rozptylu cíle z každého ze simulovaných vstupů, což vám umožňuje posoudit míru, do jaké každý vstup přispívá k celkové nejistotě v cíli.

- **Citlivost cíle, který má být změněn.** Tato volba vytvoří graf tornádo, který zobrazí efekt na cíl modulace každého simulovaného vstupu plus nebo minus jednu směrodatnou odchylku rozdělení přidruženého ke vstupu.

**Korelační grafy cílů oproti vstupům.** Tato volba generuje bodový graf cílů oproti simulovaným vstupům.

**Políčko rámečku cílových distribucí.** Tato volba generuje krabicové grafy cílových distribucí.

**Tabulka kvartilů.** Tato volba generuje tabulku kvartilů cílových distribucí. Kvartily distribuce jsou 25th, 50th a 75th percentily distribuce a rozdělují pozorování do čtyř skupin o stejné velikosti.

**Korelační a rezervní tabulka pro vstupy.** Tato volba zobrazuje tabulku korelačních koeficientů mezi simulovanými vstupy. Pohotovostní tabulka přidružení mezi vstupy s kategoričkou distribucí se zobrazí, když plán simulace uvádí generování kategoričkových dat z tabulky pro nepředvídané události.

**Překrývané výsledky se samostatnými cíli.** Pokud prediktivní model, který simulujete, obsahuje více cílů, můžete určit, zda se výsledky ze samostatných cílů zobrazí v jediném grafu. Toto nastavení se použije u grafů pro funkce hustoty pravděpodobnosti, kumulativní distribuční funkce a krabičkové grafy. Vyberete-li například tuto volbu, budou se v jednom grafu zobrazovat funkce hustoty pravděpodobnosti pro všechny cíle.

**Uložte plán pro tuto simulaci.** Veškeré úpravy provedené v simulaci můžete uložit do souboru s plánem simulace. Soubory plánu simulace mají příponu *.splan*. Plán můžete znovu otevřít v dialogovém okně Spustit simulaci nebo v tvůrci simulace. Plány simulace zahrnují všechny specifikace kromě nastavení výstupu.

**Uložte simulovaná data jako nový datový soubor.** Můžete uložit simulované vstupy, pevné vstupy a předpovídané cílové hodnoty do datového souboru SPSS Statistics, nové datové sady v aktuální relaci nebo souboru aplikace Excel. Každý případ (nebo řádek) datového souboru se skládá z předpokládaných hodnot cílů společně se simulovanými vstupy a pevnými vstupy, které generují cílové hodnoty. Je-li uvedena analýza citlivosti, každá iterace vede k souvislé sadě případů, které jsou označeny číslem iterace.

Pokud vyžadujete více přizpůsobení výstupu, než je zde k dispozici, zvažte možnost spuštění simulace z tvůrce simulace. Další informace naleznete v tématu [“Chcete-li spustit simulaci ze simulačního plánu”](#) na stránce 238.

## Práce s výstupem grafu ze simulace

Počet grafů vygenerovaných ze simulace má interaktivní funkce, které umožňují přizpůsobit zobrazení. Interaktivní funkce jsou dostupné aktivací (poklepání) na objekt grafu ve výstupním prohlížeči. Všechny simulační grafy jsou vizualizace v grafu.

**Grafy funkce hustoty pravděpodobnosti pro souvislé cíle.** Tento graf má dva posuvné svislé vztažné čáry, které rozdělí graf do samostatných oblastí. Tabulka pod grafem zobrazuje pravděpodobnost, že cíl je v každém z regionů. Je-li ve stejném grafu zobrazeno více funkcí hustoty, má tabulka samostatný řádek pro pravděpodobnosti přidružené k jednotlivým funkcím hustoty. Každý z referenčních čar má posuvník (obrácený trojúhelník), který vám umožňuje snadno pohybovat po řádku. Další funkce jsou k dispozici po klepnutí na tlačítko **Volby grafu** v diagramu. Zejména můžete explicitně nastavit polohy posuvných ovladačů, přidávat pevné referenční čáry a měnit zobrazení grafu z souvislé křivky na histogram nebo naopak. Další informace naleznete v tématu [“Volby grafu”](#) na stránce 251.

**Kumulativní grafy funkce distribuce pro souvislé cíle.** Tento graf má stejné dva přesunovatelné vertikální referenční čáry a přidružené tabulky popsané pro výše uvedený graf funkce pravděpodobnosti pravděpodobnosti. Poskytuje také přístup k dialogu Volby grafu, který vám umožňuje explicitně nastavit polohy posuvných ovladačů, přidávat pevné referenční čáry a specifikovat, zda se kumulativní distribuční funkce zobrazí jako rostoucí funkce (výchozí) nebo snižující se funkce. Další informace naleznete v tématu [“Volby grafu”](#) na stránce 251.

**Sloupcové grafy pro kategoriální cíle s iteracemi analýzy citlivosti.** Pro kategoriální cíle s iteracemi analýzy citlivosti se výsledky pro předpokládanou cílovou kategorii zobrazí jako klastrovaný pruhový graf, který zahrnuje výsledky pro všechny iterace. Graf obsahuje rozevírací seznam, který vám umožňuje klastrovat v kategorii nebo na iteraci. U modelů klastru Twin-Step a modelů K-Means můžete zvolit klastr na čísle klastru nebo iteraci.

**Krabicové grafy pro více cílů s iteracemi analýzy citlivosti.** Pro prediktivní modely s více spojitými cíli a iteracemi analýzy citlivosti výběrem zobrazení krabicových grafů pro všechny cíle na jednom grafu se vytvoří krabicový graf s klastry. Graf obsahuje rozevírací seznam, který umožňuje klastru v cíli nebo na iteraci.

## Volby grafu

Dialogové okno Volby grafu vám umožňuje přizpůsobit zobrazení aktivovaných grafů funkcí hustoty pravděpodobnosti a kumulativních distribučních funkcí generovaných ze simulace.

**Zobrazení.** Rozevírací seznam **Pohled** se používá pouze pro graf funkce hustoty pravděpodobnosti. Umožňuje přepnout zobrazení grafu z souvislé křivky na histogram. Tato funkce není k dispozici, je-li ve stejném diagramu zobrazeno více funkcí hustoty. V takovém případě mohou být funkce hustoty zobrazeny pouze jako souvislé křivky.

**Pořadí.** Rozevírací seznam **Pořadí** se používá pouze pro graf kumulativní distribuce. Uvádí, zda se funkce kumulativní distribuce zobrazí jako vzestupná funkce (výchozí) nebo sestupná funkce. Když se zobrazí jako sestupná funkce, hodnota funkce v daném bodě na vodorovné ose je pravděpodobnost, že cíl leží napravo od tohoto bodu.

**Pozice posuvného ovladače.** Pozice posuvných referenčních čar lze explicitně nastavit zadáním hodnot do horních a dolních textových polí. Můžete odstranit levou čáru tak, že vyberete **-Infinity**, čímž efektivně nastavíte polohu na zápornou nekonečno a můžete odstranit pravou čáru výběrem **Infinity**, čímž se nastaví její pozice do nekonečna.

**Referenční čáry.** Můžete přidat různé pevné svislé referenční čáry do funkcí hustoty pravděpodobnosti a kumulativních distribučních funkcí. Pokud je v jednom diagramu zobrazeno více funkcí (z důvodu více cílů nebo výsledků z iterací analýzy citlivosti), můžete určit konkrétní funkce, na které se čáry použijí.

- **Sigmy.** Můžete přidat referenční řádky za plus a minus uvedený počet směrodatných odchylek od střední hodnoty cíle.
- **Percentily.** Na jednu nebo dvě hodnoty percentilu distribuce cíle můžete přidat referenční řádky zadáním hodnot do textových polí Dolní a Horní. Například hodnota 95 v horním textovém poli představuje 95th percentilu, což je hodnota, pod níž 95% pozorování spadne. Podobně hodnota 5 v dolním textovém poli představuje 5th percentil, což je hodnota pod níž 5% pozorování spadne.
- **Vlastní pozice.** Na zadané hodnoty podél vodorovné osy můžete přidat referenční čáry.

**Referenční čáry popisků.** Tato volba určuje, zda mají být na vybrané referenční čáry použity popisky.

Referenční čáry se odeberou vymazáním přidruženého výběru v dialogovém okně Volby grafu a klepnutím na **Pokračovat**.

## Geoprostorové modelování

Techniky geoprostorového modelování jsou navrženy tak, aby zjišťoval vzorky v datech, která obsahují geoprostorovou (mapu) komponentu. Průvodce Geoprostorovým modelováním poskytuje metody pro analýzu geografických dat s časovou komponentou a bez ní.

### Vyhledat přidružení na základě událostí a geoprostorových dat (Geospatial Associations Rules)

Pomocí geografických pravidel přidružení můžete vyhledávat vzory v datech založených na prostorových i neprostorových vlastnostech. Můžete například identifikovat vzory v datech o trestné činnosti podle umístění a demografických atributů. Z těchto vzorů můžete vytvářet pravidla, která předpovídají, kde se mohou vyskytnout určité typy trestných činů.

### Učinit predikce pomocí časových řad a geoprostorových dat (Spacio-Temporal Prediction)

Prostorová časová předpověď používá data, která obsahují data umístění, vstupní pole pro predikci (prediktory), jedno nebo více časových polí a cílové pole. Každé umístění má mnoho řádků v datech, které představují hodnoty každého prediktoru a cíle v každém časovém intervalu.

## Použití průvodce Geoprostorovým modelováním

1. Z nabídky vyberte:

**Analyzovat > Prostorové a dočasné modelování > Prostorové modelování**

2. Postupujte podle kroků v průvodci.

### Příklady

Podrobné příklady jsou k dispozici v systému nápovědy.

- Geografická přidružení pravidel: **Nápověda > Témata > Případové studie > Základ statistiky > Prostorové přidružení pravidla**
- Prostorová časová předpověď: **Nápověda > Témata > Případové studie > Základ statistiky > Prostorový časový interval predikce**

## Výběr map

Geoprostorové modelování může použít jeden nebo více zdrojů dat mapy. Zdroje dat mapy obsahují informace, které definují zeměpisné oblasti a jiné zeměpisné prvky, jako jsou silnice nebo řeky. Mnoho zdrojů mapy také obsahuje demografická nebo jiná popisná data a data událostí, jako jsou zprávy o trestné činnosti nebo míry nezaměstnanosti. Můžete použít dříve definovaný soubor specifikace mapy nebo definovat specifikace mapy zde a uložit tyto specifikace pro následné použití.

### Načíst specifikaci mapy

Načte dříve definovaný soubor specifikace mapy (.mplan). Mapovat zdroje dat, které zde definujete, lze uložit do souboru specifikace mapy. Pokud u prostorové časové předpovědi vyberete soubor specifikace mapy, který identifikuje více než jednu mapu, budete vyzváni k výběru jedné mapy ze souboru.

### Přidat soubor mapování

Přidejte soubor shape ESRI (.shp) nebo .zip, který obsahuje soubor tvaru ESRI.

- Musí existovat odpovídající soubor .dbf ve stejném umístění jako soubor .shp, a tento soubor musí mít stejný kořenový název jako soubor .shp.
- Je-li soubor archivem .zip, musí mít soubory .shp a .dbf stejný kořenový název jako archiv .zip.
- Pokud neexistuje odpovídající projekční soubor (.prj), jste vyzváni k výběru projekčního systému.

### Vztah

U geografických pravidel přidružení tento sloupec definuje, jak události souvisí s funkcemi v mapě. Toto nastavení není k dispozici pro prostorové časové předpovědi.

### Přesunout nahoru, Přesunout dolů

Pořadí vrstev prvků mapy je určeno pořadím, ve kterém se objevují v seznamu. První mapa v seznamu je spodní vrstvou.

## Výběr mapy

Pokud u prostorové časové předpovědi vyberete soubor specifikace mapy, který identifikuje více než jednu mapu, bude váš dotaz vyzván k výběru jedné mapy ze souboru. Prostorová časová předpověď nepodporuje více map.

## Geografický vztah

Pro geografická pravidla přidružení definuje dialogové okno Geografický vztah, jak události souvisí s funkcemi v mapě.

- Toto nastavení platí pouze pro pravidla geografických přidružení.
- Toto nastavení má vliv pouze na zdroje dat přidružené k mapám, které jsou uvedeny jako kontextová data v kroku pro výběr zdrojů dat.

## Vztah

### Zavřít

Událost se vyskytuje v blízkosti zadaného bodu nebo oblasti na mapě.

### V rámci

Událost se vyskytuje v rámci určité oblasti na mapě.

### Obsahuje

Oblast událostí obsahuje objekt kontextu mapy.

### Intersecty

Místa, kde linky nebo oblasti z různých map se křížují navzájem.

### Křížek

Pro více map, místa, kde linky (pro silnice, řeky, železnice) z různých čar mezi sebou navzájem.

### Sever of, Sténe of, East of, West of

Událost se vyskytuje v oblasti sever, jih, východ nebo západ od určitého bodu na mapě.

## Nastavit souřadnicový systém

Pokud neexistuje žádný soubor projekce (.prj) s mapou nebo nadefinujete dvě pole ze zdroje dat jako sadu souřadnic, musíte nastavit souřadnicový systém.

### Výchozí zeměpisná šířka (zeměpisná délka a šířka)

Souřadnicový systém má zeměpisnou délku a šířku.

### Jednoduché kartézské (X a Y)

Souřadnicový systém je jednoduché souřadnice X a Y.

### Použít známý ID (WKID)

"No known ID" pro společné projekce.

### Použít název souřadnicového systému

Souřadnicový systém je založen na uvedené projekci. Název je uzavřen do závorek.

## Nastavení projekce

Pokud systém projekce nemůže být určen z informací poskytnutých s mapou, musíte uvést projekční systém. Nejběžnější příčinou této podmínky je absence souboru projekce (.prj) přidruženého k mapě nebo souboru projekce, který nelze použít.

- **Město, region nebo země (Mercator)**
- **Velká země, několik zemí nebo kontinenty (Winkel Tripel)**
- **Oblast velmi blízko k rovníku (Mercator)**
- **Oblast blízka jednomu ze sloupů (Stereographic)**

Projekce Mercator je společná projekce používaná v mnoha mapách. Tato projekce považuje zeměkouli za tlakovou láhev, která se válí na plochý povrch. Projekce Mercator zkresluje velikost a tvar velkých objektů. Toto zkreslení se zvyšuje, jak se pohybuje dál od rovníku a blíže k sloupům. Winkel Tripel a Stereografické projekce činí úpravy pro fakt, že mapa představuje část třírozměrné koule, která se zobrazí ve dvou rozměrech.

## Projekce a koordinovaný systém

Pokud vyberete více než jednu mapu a mapy mají různé projekce a souřadnicové systémy, musíte vybrat mapu s projektovým systémem, který chcete použít. Tento projekční systém se bude používat pro všechny mapy, jsou-li společně ve výstupu dohromady.

## Zdroje dat

Zdrojem dat může být soubor dBase, který je poskytován spolu s formálním souborem, datovým souborem IBM SPSS Statistics nebo otevřeným datovým souborem v aktuální relaci.

**Kontextová data.** Kontextová data identifikují funkce na mapě. Kontextová data mohou také obsahovat pole, která lze použít jako vstupy pro model. Chcete-li použít kontextový soubor dBase (.dbf), který je přidružen k souboru tvaru mapy (.shp), musí být kontextový dBase soubor ve stejném umístění jako soubor tvaru a musí mít stejný kořenový název. Je-li například soubor tvaru geodata . shp, musí mít soubor dBase název geodata . dbf

**Data události.** Data událostí obsahují informace o událostech, které se vyskytnou, jako jsou např. trestné činy nebo nehody. Tato volba je k dispozici pouze pro geografická pravidla přidružení.

**Hustota bodů.** Časový interval a souřadnice dat pro odhad hustoty jádra. Tato volba je k dispozici pouze pro prostorové časové předpovědi.

**Přidat.** Otevře dialogové okno pro přidání zdrojů dat. Zdrojem dat může být soubor dBase, který je poskytován spolu s formálním souborem, datovým souborem IBM SPSS Statistics nebo otevřeným datovým souborem v aktuální relaci.

**Přidružit.** Otevře dialogové okno pro určení identifikátorů (souřadnice nebo klíče) použité k přidružení dat k mapám. Každý zdroj dat musí obsahovat jeden nebo více identifikátorů, které přidruží data k mapě. Soubory dBase, které přicházejí do souboru tvaru, zpravidla obsahují pole, které se automaticky použije jako výchozí identifikátor. Pro ostatní zdroje dat musíte uvést pole, která se použijí jako identifikátory.

**Ověřit klíč.** Otevře dialogové okno pro ověření shody klíče mezi mapou a zdrojem dat.

## Geoprostorová pravidla pro přidružení

- Alespoň jeden zdroj dat musí být zdrojem dat události.
- Všechny zdroje dat událostí musí používat stejnou formu identifikátorů přidružení mapy: souřadnice nebo klíčové hodnoty.
- Jsou-li zdroje dat událostí přidruženy k mapám s klíčovými hodnotami, všechny zdroje událostí musí používat stejný typ funkce mapy (například mnohoúhelníky, body, čáry).

## prostorová časová prognóza

- Musí existovat kontextový zdroj dat.
- Pokud existuje pouze jeden zdroj dat (datový soubor s žádnou přidruženou mapou), musí obsahovat souřadnice.
- Máte-li dva zdroje dat, jeden zdroj dat musí být kontextová data a druhý zdroj dat musí být data hustoty bodu.
- Nemůžete zahrnout více než dva zdroje dat.

## Přidat zdroj dat

Zdrojem dat může být soubor dBase, který je poskytnut spolu s souborem tvaru a kontextem souboru, datovým souborem IBM SPSS Statistics nebo otevřeným datovým souborem v aktuální relaci.

Stejný zdroj dat můžete přidat vícekrát, chcete-li pro každou z nich použít jiné geografické přidružení.

## Přidružení dat a map

Každý zdroj dat musí obsahovat jeden nebo více identifikátorů, které přidruží data k mapě.

### Souřadnice

Zdroj dat obsahuje pole, která představují kartézské souřadnice, vyberte pole, která představují souřadnice X a Y. U geografických asociačních pravidel může být také souřadnicová souřadnice Z.

### Klíčové hodnoty

Klíčové hodnoty v polích ve zdroji dat odpovídají vybraným klávesám mapy. Mapa oblastí může mít například název identifikátoru (klíč mapy) pro každou oblast. Tento identifikátor odpovídá poli v datech, která také obsahuje názvy regionů (datový klíč). Pole se přiřazují k mapovým klíčům na základě pořadí, ve kterém jsou zobrazeny v obou seznamech.

## Validovat klíče

Dialogové okno Ověřit klíče poskytuje souhrn záznamu shodující se mezi mapou a zdrojem dat na základě vybraných klíčů identifikátoru. Pokud zde existují neodpovídající hodnoty datového klíče, můžete je ručně přiřadit k mapování hodnot klíče.

## Geoprostorová pravidla přidružení

U geografických pravidel přidružení platí, že po definování map a zdrojů dat jsou zbývající kroky v průvodci:

- Existuje-li více zdrojů dat událostí, definujte, jak budou sloučeny zdroje dat událostí.
- Vyberte pole, která mají být použita jako podmínky a předpovědi v analýze.

Volitelně můžete také:

- Vyberte jiné výstupní volby.
- Uložte soubor modelu hodnocení.
- Vytvořte nová pole pro předpovězené hodnoty a pravidla ve zdrojích dat použitých v modelu.
- Upravit nastavení pro pravidla přidružení sestavení.
- Úprava nastavení binning a agregace.

## Definovat datová pole událostí

Pokud existuje více než jeden zdroj dat událostí, jsou zdroje dat událostí sloučeny pro geografická pravidla přidružení, pokud existuje více než jeden zdroj dat událostí.

- Ve výchozím nastavení jsou zahrnuta pouze pole, která jsou společná pro všechny zdroje dat událostí.
- Můžete zobrazit seznam společných polí, polí pro určitý zdroj dat nebo polí ze všech zdrojů dat a vybrat pole, která chcete zahrnout.
- U společných polí musí být **Typ** a **Měření** stejné pro všechny zdroje dat. Pokud existují konflikty, můžete uvést typ a úroveň měření, které se mají použít pro každé společné pole.

## Vyberte pole

Seznam dostupných polí obsahuje pole ze zdrojů dat událostí a polí z kontextových zdrojů dat.

- Seznam zobrazených polí můžete nastavit tak, že vyberete zdroj dat ze seznamu **Zdroje dat**.
- Musíte vybrat alespoň dvě pole. Alespoň jedna musí být podmínka a alespoň jedna musí být predikce. Existuje několik způsobů, jak tento požadavek splnit, včetně výběru dvou polí pro seznam **Obojí (Podmínka a předpověď)**.
- Pravidla přidružení předpovídají hodnoty polí předpovědí, které jsou založené na hodnotách polí podmínky. Například v pravidle "Je-li  $x=1$  a  $y=2$ , pak  $z=3$ ", jsou hodnoty  $x$  a  $y$  podmínky a hodnota  $z$  je predikce.

## Výstup

### Tabulky pravidel

Každá tabulka pravidel zobrazuje nejvyšší pravidla a hodnoty pro důvěru, podporu pravidel, výtah, podporu podmínek a implementovatelnost. Každá tabulka je seřazena podle hodnot vybraného kritéria. Na základě vybraného kritéria můžete zobrazit všechna pravidla nebo nejvyšší **počet** pravidel.

### Sortable Word Cloud

Seznam horních pravidel založený na hodnotách vybraného kritéria. Velikost textu označuje relativní důležitost pravidla. Interaktivní výstupní objekt obsahuje nejvyšší pravidla pro sebedůvěru, podporu pravidel, výtah, podporu podmínek a implementovatelnost. Zvolené kritérium určuje, který seznam pravidel se standardně zobrazí. Ve výstupu můžete vybrat jiné kritérium interaktivně. Volba **Maximální počet zobrazovaných pravidel** určuje počet pravidel, která se zobrazí ve výstupu.

## Mapa

Interaktivní pruhový graf a mapování horních pravidel na základě vybraného kritéria. Každý interaktivní výstupní objekt obsahuje nejvyšší pravidla pro sebedůvěru, podporu pravidel, výtah, podporu podmínek a implementovatelnost. Zvolené kritérium určuje, který seznam pravidel se standardně zobrazí. Ve výstupu můžete vybrat jiné kritérium interaktivně. Volba **Maximální počet zobrazovaných pravidel** určuje počet pravidel, která se zobrazí ve výstupu.

## Modelové informační tabulky

### Polní transformace.

Popisuje transformace, které se použijí na pole použítá v analýze.

### Souhrn záznamů.

Počet a procento zahrnutých a vyloučených záznamů.

### Statistika pravidel.

Souhrnné statistiky pro podporu podmínek, důvěru, podporu pravidel, výtah a implementovatelnost. Statistické údaje zahrnují střední, minimální, maximální a směrodatnou odchylku.

### Nejčastější položky.

Položky, které se vyskytují nejčastěji. Položka je zahrnuta do podmínky nebo předpovědi v pravidle. Například stáří < 18 nebo gender=samice.

### Nejčastější pole.

Pole, která se vyskytují nejčastěji v pravidlech.

### Vyloučené vstupy.

Pole, která jsou vyloučena z analýzy, a důvod, proč byly všechny pole vyloučeny.

## Kritérium pro tabulky pravidel, aplikace Word Cloud a Mapy

### Důvěra.

Procento správných předpovědí pravidel.

### Podpora pravidel.

Procentní část případů, pro které má pravidlo hodnotu true. Například, je-li pravidlo "Je-li  $x=1$  a  $y=2$ , pak  $z=3$ ," podpora pravidla je skutečným procentem případů v datech, pro které platí  $x=1$ ,  $y=2$  a  $z=3$ .

### Zvednout.

Vlek je měřítkem toho, jak daleko pravidlo zlepšuje predikci ve srovnání s náhodným náhodným náhodným výběrem. Jedná se o poměr správných předpovědí k celkovému výskytu předpovězené hodnoty. Hodnota musí být větší než 1. Například, pokud se předpovězená hodnota vyskytne 20% času a důvěryhodnost v předpovědi je 80%, pak je hodnota zvednutí 4.

### Provozní podpora.

Procentní část případů, pro které podmínka pravidla existuje. Například, je-li pravidlo "Je-li  $x=1$  a  $y=2$ , pak  $z=3$ ," podpora provozního stavu je proporce případů v datech, pro která  $x=1$  a  $y=2$ .

### Implementovatelnost.

Procento chybných předpovědí, jsou-li podmínky pravidel. Implementovatelnost se rovná (1-důvěryhodnost) násobená podporou podmínky nebo podporou podmínek minus podpora pravidla.

## Uložit

### Uložit mapu a data kontextu jako specifikaci mapy

Uložte specifikace mapy do externího souboru (.mplan). Tento soubor specifikace mapy můžete načíst do průvodce pro následné analýzy. Specifikační soubor mapy můžete také použít s příkazem SPATIAL ASSOCIATION RULES.

### Kopírovat všechny mapy a datové soubory do specifikace

Data z souborů tvaru mapy, externí datové soubory a datové sady použité ve specifikaci mapy jsou uloženy v souboru specifikace mapy.



## **přidělení skóre**

Uloží nejlepší hodnoty pravidel, hodnoty důvěryhodnosti pro pravidla a číselné hodnoty ID pro pravidla jako nová pole v uvedeném zdroji dat.

### **Zdroj dat k skóre**

Zdroj dat nebo zdroje, kde jsou nová pole vytvořena. Není-li zdroj dat otevřen v aktuální relaci, otevře se v aktuální relaci. Chcete-li uložit nová pole, musíte upravený soubor explicitně uložit.

### **Cílové hodnoty**

Vytvořte nová pole pro vybraná cílová pole (predikce).

- Pro každé cílové pole jsou vytvořena dvě nová pole: předpovězená hodnota a hodnota důvěry.
- Pro trvalá cílová pole (měřítko) je předpovězená hodnota řetězec, který popisuje rozsah hodnot. Hodnota ve tvaru "(value1, value2]" znamená "větší než value1 a menší nebo rovna hodnotě value2."

### **Počet nejlepších pravidel**

Vytvořte nová pole pro počet doporučených pravidel. Pro každé pravidlo se vytvoří tři nová pole: hodnota pravidla, hodnota důvěry a hodnota numerického ID pro dané pravidlo.

### **Předpona názvu**

Předpona, která má být použita pro nové názvy polí.

## **Budova pravidel**

Parametry sestavení pravidla nastavují kritéria pro generovaná pravidla přidružení.

### **Položek na pravidlo**

Počet hodnot polí, které lze zahrnout do podmínek pravidla a předpovědí. Celkový počet položek nesmí překročit 10. Například, v pravidle "Pokud  $x=1$  a  $y=2$ , pak  $z=3$ ", existují dvě položky podmínky a jedna položka předpovědí.

### **Maximální počet předpovědí.**

Maximální počet hodnot polí, které se mohou vyskytnout v předpovědích pro pravidlo.

### **Maximální podmínky.**

Maximální počet hodnot polí, které se mohou vyskytnout v podmínkách pro pravidlo.

### **Vyloučit pár**

Vyloučí uvedené dvojice polí z toho, že jsou zahrnuty do stejného pravidla.

### **Kritéria pravidel**

#### **Důvěra.**

Do výstupu musí být zahrnuta minimální jistota, že musí být pravidlo zahrnuto. Důvěra je procento správných předpovědí.

#### **Podpora pravidel.**

Do výstupu musí být zahrnuta minimální pravidla pro podporu pravidla. Hodnota představuje procento případů, pro které pravidlo platí v pozorovaných datech. Například, je-li pravidlo "Je-li  $x=1$  a  $y=2$ , pak  $z=3$ ," podpora pravidla je skutečným procentem případů v datech, pro které platí  $x=1$ ,  $y=2$  a  $z=3$ .

#### **Provozní podpora.**

Do výstupu musí být zahrnut minimální stav, který musí být zahrnut do výstupu. Hodnota představuje procentní část případů, pro které podmínka existuje. Například, je-li pravidlo "Pokud  $x=1$  a  $y=2$ , pak  $z=3$ ," podpora podmínky je procentní část případů v datech, pro která  $x=1$  a  $y=2$ .

#### **Zvednout.**

Do výstupu musí být zahrnuta minimální zdvihací zařízení. Vlek je měřítkem toho, jak daleko pravidlo zlepšuje predikci předpovídanou náhodným náhodným ukazatelem. Jedná se o poměr správných předpovědí k celkovému výskytu předpovězené hodnoty. Například, pokud se předpovězená hodnota vyskytne 20% času a důvěryhodnost v předpovědi je 80%, pak je hodnota zvednutí 4.

### **Považovat za stejné**

Identifikuje dvojice polí, se kterými se má zacházet jako se stejným polem.

### **Fixování a agregace**

- Agregace je nezbytná, je-li v datech více záznamů, než jsou funkce v mapě. Máte např. datové záznamy pro jednotlivé okresy, ale máte mapu stavů.
- Pro souvislá a ordinální pole můžete určit metodu agregovaného souhrnného ukazatele. Nominální pole jsou agregována na základě modální hodnoty.

### **Nepřetržitá**

Pro souvislá pole (měřítko) může být měřítkem střední hodnota, medián nebo součet.

### **Pořadové**

Pro ordinální pole může být součtový ukazatel medián, režim, nejvyšší nebo nejnižší.

### **Počet příhrádek**

Nastaví maximální počet příhrádek pro souvislá pole (měřítko). Souvislá pole jsou vždy seskupena nebo "binned" do rozsahů hodnot. Např.: menší než nebo rovno 5, větší než 5 a menší nebo rovno 10 nebo větší než 10.

### **Agregovat mapu**

Použijte agregaci na data i mapy.

### **Vlastní nastavení pro specifická pole**

Můžete přepsat výchozí souhrnné ukazatele a počet příhrádek pro specifická pole.

- Klepnutím na ikonu otevřete dialogové okno **Volič pole** a vyberte pole, které chcete přidat do seznamu.
- Ve sloupci **Agregace** vyberte souhrnné měřítko.
- Pro souvislá pole klepněte na tlačítko ve sloupci **Bins**, abyste uvedli vlastní počet příhrádek pro pole v dialogovém okně **Bins**.

## **Prostorová časová předpověď**

Pro prostorové časové předpovědi po definování map a zdrojů dat jsou zbývající kroky v průvodci:

- Uved'te cílové pole, časová pole a volitelné prediktory.
- Definujte časové intervaly nebo cyklická období pro časová pole.

Volitelně můžete také:

- Vyberte jiné výstupní volby.
- Upravit parametry sestavení modelu.
- Upravit nastavení agregace.
- Ukládat předpovězené hodnoty do datové sady v aktuální relaci nebo do datového souboru s formátem produktu IBM SPSS Statistics.

### **Vyberte pole**

Seznam dostupných polí obsahuje pole z vybraných zdrojů dat. Seznam zobrazených polí můžete nastavit tak, že vyberete zdroj dat ze seznamu **Zdroje dat**.

### **Cíl**

Je požadováno cílové pole. Cíl je pole, pro které jsou předpovězeny hodnoty.

- Cílové pole musí být souvislé (scale), numerické pole.
- Pokud existují dva zdroje dat, cílem je odhad hustoty jádra a hodnota "Hustota" se zobrazí jako cílový název. Tento výběr nelze změnit.

### **Prediktory**

Je možné zadat jedno nebo více polí prediktoru. Toto nastavení je nepovinné.

## Pole času

Musíte vybrat jedno nebo více polí, která představují časová období, nebo vyberte **Cyklická období**.

- Pokud existují dva zdroje dat, musíte vybrat časová pole z obou zdrojů dat. Obě tato pole musí představovat stejný interval.
- V případě cyklických období je třeba určit pole, která definují periodické cykly, v **panelu Časové intervaly** průvodce.

## Časové intervaly

Volby na tomto panelu jsou založeny na volbě pole **Časová pole** nebo **Cyklické období** v kroku pro výběr polí.

## Pole času

**Vybraná časová pole.** Vyberete-li jedno nebo více časových polí v kroku pro výběr polí, tato pole se zobrazí v tomto seznamu.

**Časový interval.** Vyberte odpovídající časový interval ze seznamu. V závislosti na časovém intervalu můžete také zadat jiná nastavení, jako např. interval mezi pozorováními (přírůstek) a počáteční hodnotou. Tento časový interval se používá pro všechna vybraná časová pole.

- Procedura předpokládá, že všechny případy (záznamy) představují rovnoměrně rozložené intervaly.
- Na základě vybraného časového intervalu může procedura zjistit chybějící pozorování nebo více pozorování ve stejném časovém intervalu, který je třeba agregovat společně. Například, je-li časový interval dny a datum 2014-10-27 je následován 2014-10-29, pak chybí pozorování pro 2014-10-28. Je-li časový interval měsíc, pak se více dat ve stejném měsíci seskupí dohromady.
- Pro některé časové intervaly může další nastavení definovat přerušení v normálních intervalech rovnoměrně rozložených. Například, je-li časový interval dny, ale pouze pracovní dny jsou platné, můžete uvést, že je v týdnu pět dní a týden začíná v pondělí.
- Pokud není vybrané časové pole formát data nebo pole formátu času, časový interval se automaticky nastaví na **Období** a nelze jej změnit.

## Cyklus pole

Vyberete-li v kroku pro výběr polí položku **Cyklické období**, je třeba určit pole, která definují cyklická období. Cyklické období identifikuje opakovanou cyklickou variantu, jako například počet měsíců v roce nebo počet dnů v týdnu.

- Můžete určit až tři pole, která definují cyklická období.
- Pole prvního cyklu představuje nejvyšší úroveň cyklu. Například, pokud existuje cyklická variace podle roku, čtvrtletí a měsíce, pole, které představuje rok, je první pole cyklu.
- Délka cyklu pro první a druhá pole cyklu je periodičita na následující úrovni. Například, jsou-li pole cyklu rok, čtvrtletí a měsíc, délka prvního cyklu je 4 a délka druhého cyklu je 3.
- Počáteční hodnota pro pole druhého a třetího cyklu je první hodnotou v každém z těchto cyklických období.
- Délka cyklu a počáteční hodnoty musí být kladná celá čísla.

## agregace

- Vyberete-li v kroku pro výběr polí některý **Prediktory**, můžete pro prediktory vybrat metodu souhrnu agregace.
- Agregace je nezbytná, pokud je v definovaném časovém intervalu více než jeden záznam. Například, je-li časový interval měsíc, pak se více dat ve stejném měsíci seskupí dohromady.
- Můžete určit metodu agregace souhrnu agregace pro souvislá a ordinální pole. Nominální pole jsou agregována na základě modální hodnoty.

## Nepřetržitá

Pro souvislá pole (měřítko) může být měřítkem střední hodnota, medián nebo součet.

## Pořadové

Pro ordinální pole může být součtový ukazatel medián, režim, nejvyšší nebo nejnižší.

## Vlastní nastavení pro specifická pole

Výchozí souhrnné měřítko agregace pro specifické prediktory můžete potlačit.

- Klepnutím na ikonu otevřete dialogové okno **Volič pole** a vyberte pole, které chcete přidat do seznamu.
- Ve sloupci **Agregace** vyberte souhrnné měřítko.

## Výstup

### Mapy

#### Cílové hodnoty.

Mapa hodnot pro vybrané cílové pole.

#### Korelace

Mapa korelací.

#### Klastry

Mapa, která zvýrazňuje klastry umístění, které jsou si podobné navzájem. Mapy klastrů jsou k dispozici pouze pro empirické modely.

#### Prahová hodnota podobnosti umístění.

Podobnost, která je nezbytná k vytvoření klastrů. Hodnota musí být číslo větší než nula a menší než 1.

#### Zadejte maximální počet klastrů.

Maximální počet klastrů, které se mají zobrazit.

### Tabulky vyhodnocení modelu

#### Specifikace modelu.

Souhrn specifikací použitých ke spuštění analýzy, včetně cílových, vstupních a umístění polí.

#### Souhrn časových informací.

Identifikuje časová pole a časové intervaly, které se používají v modelu.

#### Test účinků ve střední struktuře.

Výstup obsahuje hodnotu statistiky testu, stupně volnosti a úroveň významnosti pro model a každý efekt.

#### Průměrná struktura modelových koeficienty.

Výstupem je hodnota koeficientu, směrodatná chyba, hodnota statistiky testu, úroveň významnosti a intervaly spolehlivosti pro každý vzorný výraz modelu.

#### Autoregresivní koeficienty.

Výstup obsahuje hodnotu koeficientu, standardní chybový výstup, hodnotu statistiky testu, úroveň významnosti a intervaly spolehlivosti pro jednotlivé prodlevy.

#### Testuje se prostorová rozptýl.

Pro parametrické modely založené na variogramech zobrazuje dohoda vyhovění výsledků testu pro strukturu prostorových kovariance. Výsledky testů mohou určit, zda se má model prostorové kovariance modelovat parametricky nebo aby používal neparametrický model.

#### Parametrický prostorový rozdíl.

Pro parametrické modely založené na variogramech se zobrazí odhady parametrů pro parametrický prostorový kovarianci.

## Volby modelu

### Nastavení modelu

#### **Automaticky zahrnout zachycení**

Zahrnout zachycení do modelu.

#### **Maximální prodleva automatické regrese**

Maximální prodleva autoregression. Hodnota musí být celé číslo mezi 1 a 5.

### **Geografický rozptyl**

Uvádí metodu odhadu pro geografickou kovarianci.

#### **Parametrický**

Metoda odhadu je parametrická. Tato metoda může být **Gaussovo**, **Exponenciální** nebo **Exponenciální napájení**. Pro Exponenciální napájení můžete zadat hodnotu **Power**.

#### **Neparametrický**

Metoda odhadu je neparametrická.

## Uložit

### **Uložit mapu a data kontextu jako specifikaci mapy**

Uložte specifikace mapy do externího souboru (.mplan). Tento soubor specifikace mapy můžete načíst do průvodce pro následnou analýzu. Specifikační soubor mapy můžete také použít s příkazem SPATIAL TEMPORAL PREDICTION.

### **Kopírovat všechny mapy a datové soubory do specifikace**

Data z souborů tvaru mapy, externích datových souborů a datových sad, které jsou použity ve specifikaci mapování, se uloží do souboru specifikace mapy.

### **přidělení skóre**

Uloží předpovězené hodnoty, odchylky a horní a dolní meze spolehlivosti pro cílové pole ve vybraném datovém souboru.

- Předpovězené hodnoty můžete uložit do otevřené datové sady v aktuální relaci nebo ve formátu datového souboru IBM SPSS Statistics.
- Datový soubor nemůže být zdrojem dat, který se používá v modelu.
- Datový soubor musí obsahovat všechna časová pole a prediktory, které se používají v modelu.
- Hodnoty času musí být větší než časové hodnoty použité v modelu.

## Rozšířené

### **Maximální počet případů s chybějícími hodnotami (%)**

Maximální procentní část případů s chybějícími hodnotami.

### **Úroveň významnosti**

Úroveň významnosti pro určení, zda je parametrický model založený na variogramu vhodný. Hodnota musí být větší než 0 a menší než 1. Výchozí hodnota je 0,05. Úroveň významnosti se používá pro zkoušku vyhovění struktury prostorové kovariance. Pro určení toho, zda použít parametrický nebo neparametrický model, se použije dobrá vhodná statistika.

### **Faktor nejistoty (%)**

Faktor nejistoty je procentní hodnota, která představuje růst nejistoty pro budoucí prognózy. Horní a dolní mezní hodnoty prognózy nejistoty se zvyšují o určený procentní podíl pro každý krok do budoucnosti.

## Dokončit

V posledním kroku průvodce geoprostorovým modelováním můžete buď spustit model, nebo vložit vygenerovanou syntaxi příkazu do okna syntaxe. Můžete upravit a uložit generovanou syntaxi pro následné použití.



## Upozornění

---

Tyto informace byly vytvořeny pro produkty a služby poskytované v USA. Tento materiál může být dostupný od IBM v jiných jazycích. K povolení přístupu však může být vyžadováno vlastnictví kopie produktu nebo verze produktu v tomto jazyce.

Společnost IBM nemusí nabízet produkty, služby nebo funkce uvedené v tomto dokumentu v jiných zemích. Informace o produktech a službách, které jsou aktuálně k dispozici ve vaší oblasti, získáte od lokálního zástupce společnosti IBM. Odkazy na produkty, programy nebo služby společnosti IBM neuvádí ani neimplikují, že lze použít pouze daný produkt, program nebo službu společnosti IBM. Lze použít libovolný funkčně ekvivalentní produkt, program nebo službu neporušující práva duševního vlastnictví společnosti IBM. Vyhodnocení a ověření funkčnosti produktů, programů nebo služeb, které nepatří společnosti IBM, je však zodpovědností uživatele.

Společnost IBM může vlastnit patenty nebo nevyřízené žádosti o patenty zahrnující předměty popsané v tomto dokumentu. Vlastnictví tohoto dokumentu neposkytuje licenci k těmto patentům. Dotazy na licence můžete písemně odeslat na následující adresu:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
USA*

S dotazy na licence týkající se dvoubajtových informací (DBCS) se obraťte na oddělení intelektuálního vlastnictví společnosti IBM v dané zemi, nebo je odešlete písemně na následující adresu:

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan*

SPOLEČNOST INTERNATIONAL BUSINESS MACHINES CORPORATION POSKYTUJE TUTO PUBLIKACI "TAKOVOU, JAKÁ JE", BEZ JAKÝCHKOLIV ZÁRUK, VYJÁDŘENÝCH NEBO ODVOZENÝCH VČETNĚ, MIMO JINÉ, ODVOZENÝCH ZÁRUK NEPORUŠENÍ PRÁV TŘETÍCH STRAN, ZÁRUKY PRODEJNOSTI NEBO VHODNOSTI PRO URČITÝ ÚČEL. Některé právní řády u určitých transakcí nepřipouštějí vyloučení záruk výslovně vyjádřených nebo vyplývajících z okolností, a proto se na vás výše uvedené omezení nemusí vztahovat, a proto se vás toto prohlášení nemusí týkat.

Uvedené údaje mohou obsahovat technické nepřesnosti nebo typografické chyby. Údaje zde uvedené jsou pravidelně upravovány a tyto změny budou zahrnuty v nových vydáních této publikace. Společnost IBM může kdykoli bez upozornění provádět vylepšení nebo změny v produktech či programech popsaných v této publikaci.

Jakékoliv odkazy v této publikaci na webové stránky jiných společností než IBM jsou poskytovány pouze pro pohodlí uživatele a nemohou být žádným způsobem vykládány jako doporučení těchto webových stránek. Materiály uvedené na těchto webových stránkách nejsou součástí materiálů pro tento produkt IBM a použití uvedených stránek je pouze na vlastní nebezpečí.

IBM může použít nebo distribuovat jakékoli informace, které jí poskytnete, libovolným způsobem, který společnost považuje za odpovídající, bez vzniku jakýchkoliv závazků vůči vám.

Vlastníci licence k tomuto programu, kteří chtějí získat informace o možnostech (i) výměny informací s nezávisle vytvořenými programy a jinými programy (včetně tohoto) a (ii) oboustranného využití vyměňovaných informací, mohou kontaktovat informační středisko na adrese:

*IBM Director of Licensing  
IBM Corporation*

North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
USA

Poskytnutí takových informací může být podmíněno dodržením určitých podmínek a požadavků zahrnujících v některých případech uhrazení stanoveného poplatku.

Licencovaný program popsáný v tomto dokumentu a veškerý licencovaný materiál k němu dostupný jsou společností IBM poskytovány na základě podmínek uvedených ve smlouvách IBM Customer Agreement, IBM International Program License Agreement nebo v jiné ekvivalentní smlouvě.

Citovaná data o výkonu a příklady klienta jsou uvedeny pouze pro názornost. Skutečné výsledky výkonu se mohou lišit v závislosti na specifických konfiguracích a provozních podmínkách.

Informace týkající se produktů jiných společností než IBM byly získány od dodavatelů těchto produktů, z jejich publikovaných sdělení, nebo z jiných veřejně dostupných zdrojů. IBM tyto produkty netestovala a nemůže potvrdit přesnost údajů o výkonu, kompatibilitě nebo jiná tvrzení týkající se produktů jiných společností než IBM. Otázky týkající se možností produktů jiných společností než IBM by měly být adresovány dodavatelům těchto produktů.

Prohlášení týkající se budoucího směru vývoje nebo záměrů společnosti IBM se mohou změnit nebo mohou být zrušena bez předchozího upozornění a představují pouze cíle a záměry.

Tyto údaje obsahují příklady dat a sestav používaných v běžných obchodních operacích. Aby byla představa úplná, používají se v příkladech jména osob, společností, značek a produktů. Všechna tato jména jsou fiktivní a jakákoliv podobnost se skutečnými lidmi nebo obchodními podniky je čistě náhodná.

#### COPYRIGHT - LICENCE:

Tyto informace obsahují ukázkové aplikační programy ve zdrojovém jazyku a ilustrují různé programovací techniky na různých operačních platformách. Tyto ukázkové programy můžete bez závazků vůči společnosti IBM jakýmkoli způsobem kopírovat, měnit a distribuovat za účelem vývoje, používání, odbytu či distribuce aplikačních programů odpovídajících rozhraní API pro operační platformu, pro kterou byly ukázkové programy napsány. Tyto příklady nebyly důkladně testovány ve všech podmínkách. Společnost IBM proto nemůže zaručit spolehlivost, upotřebitelnost nebo funkčnost těchto programů. Ukázkové programy jsou poskytovány "JAK JSOU", bez záruky jakéhokoli druhu. IBM nenes odpovědnost za žádné škody vzniklé ve spojení s Vaším užíváním ukázkových programů.

Jakákoli kopie nebo část těchto ukázkových programů nebo jakékoli odvozené dílo musí obsahovat následující poznámku o autorských právech:

© Copyright IBM Corp. 2021. Části tohoto kódu jsou odvozeny ze vzorových programů společnosti IBM Corp. Vzorové programy.

© Copyright IBM Corp. 1989-2021. Všechna práva vyhrazena.

## Ochranné známky

---

IBM, logo IBM a [ibm.com](http://ibm.com) jsou ochranné známky nebo registrované ochranné známky společnosti International Business Machines Corp., registrované v mnoha jurisdikcích po celém světě. Ostatní názvy produktů a služeb mohou být ochrannými známkami společnosti IBM nebo jiných společností. Aktuální seznam ochranných známek společnosti IBM je k dispozici na webu na stránce "Copyright and trademark information" na adrese [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, logo Adobe, PostScript a logo PostScript jsou buď registrované ochranné známky, nebo ochranné známky společnosti Adobe Systems Incorporated ve Spojených státech anebo v dalších zemích.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium a Pentium jsou ochranné známky nebo registrované ochranné známky společnosti Intel Corporation nebo jejich dceřiných společností ve Spojených státech a případně v dalších jiných zemích.

Linux je registrovaná ochranná známka Linuse Torvaldse ve Spojených státech a případně v dalších jiných zemích.



Microsoft, Windows, Windows NT a logo Windows jsou ochranné známky společnosti Microsoft Corporation ve Spojených státech a případně v dalších jiných zemích.

UNIX je registrovaná ochranná známka společnosti The Open Group ve Spojených státech a případně v dalších jiných zemích.

Java a všechny ochranné známky a loga založené na jazyce Java jsou ochranné známky nebo registrované ochranné známky společnosti Oracle anebo příbuzných společností.



# Rejstřík

## A

- Agresti-Caffo
  - in Independent-Samples Proportions [108](#)
- Agresti-Coull
  - v jednovzorkových porcích [102](#)
- Agresti-Min
  - v paired-rozměry vzorků [105](#)
- analýza citlivosti
  - v simulaci [244](#)
- Analýza časové řady
  - prognóza [159](#)
  - předpovídání případů [159](#)
- Analýza diskriminátoru
  - definování rozsahů [170](#)
  - Deskriptivní statistika [170](#)
  - export informací o modelu [172](#)
  - chybějící hodnoty [171](#)
  - koeficienty funkce [170](#)
  - kovarianční matice [171](#)
  - krokové metody [169](#)
  - Mahalanobis vzdálenost [171](#)
  - matice [170](#)
  - metody discriminant [171](#)
  - Nezávislé proměnné [169](#)
  - Podmínky [171](#)
  - předchozích pravděpodobností [171](#)
  - příkaz additional features [172](#)
  - příklad [169](#)
  - Raovo V. [171](#)
  - seskupování proměnných [169](#)
  - statistika [169](#), [170](#)
  - ukládání klasifikačních proměnných [172](#)
  - volby zobrazení [171](#)
  - výběr případů [170](#)
  - výkresy [171](#)
  - Wilks " lambda [171](#)
- Analýza faktoru
  - deskriptivní ukazatele [173](#)
  - extrakční metody [174](#)
  - formát zobrazení koeficientu [175](#)
  - chybějící hodnoty [175](#)
  - konvergence [174](#), [175](#)
  - metody rotace [175](#)
  - načítání grafů [175](#)
  - přehled [172](#)
  - příkaz additional features [176](#)
  - příklad [172](#)
  - skóre faktoru [175](#)
  - statistika [172](#), [173](#)
  - výběr případů [173](#)
- Analýza hierarchického klastru
  - aglomerační plány [185](#)
  - členství klastru [185](#)
  - dendrogramy [185](#)
  - matice vzdálenosti [185](#)
  - Analýza hierarchického klastru (*pokračování*)
    - metody klastrování [184](#)
    - opatření na vzdálenost [184](#)
    - opatření podobnosti [184](#)
    - orientace výkresu [185](#)
    - proměnné klastrování [184](#)
    - příkaz additional features [185](#)
    - příklad [184](#)
    - případy klastrování [184](#)
    - statistika [184](#), [185](#)
    - transformace hodnot [184](#)
    - transformace ukazatelů [184](#)
    - ukládání nových proměnných [185](#)
    - zaliční grafy [185](#)
- analýza hlavních komponent [172](#), [174](#)
- Analýza K-Means Cluster
  - členství klastru [187](#)
  - chybějící hodnoty [187](#)
  - Iterace [187](#)
  - Kritéria konvergence [187](#)
  - metody [185](#)
  - přehled [185](#)
  - příkaz additional features [187](#)
  - příklady [185](#)
  - statistika [185](#), [187](#)
  - účinnost [186](#)
  - uložení informací o klastru [187](#)
  - vzdálenosti klastrů [187](#)
- analýza klastru
  - Analýza hierarchického klastru [184](#)
  - Analýza K-Means Cluster [185](#)
  - účinnost [186](#)
- Analýza klastru TwoStep
  - statistika [179](#)
  - uložit do externího souboru [179](#)
  - uložit do pracovního souboru [179](#)
  - volby [178](#)
- Analýza napájení
  - statistika [1](#)
- Analýza nejbližšího souseda
  - Oblasti [164](#)
  - sousedé [163](#)
  - ukládání proměnných [165](#)
  - volby [166](#)
  - Výběr funkcí [164](#)
  - výstup [165](#)
  - Zobrazení modelu [166](#)
- Analýza ROC
  - statistiky a grafy [233](#)
- analýza rozptylu
  - in Means [97](#)
  - v jednofaktorové ANOVA [113](#)
  - v lineární regresi [143](#)
  - v odhadu křivky [157](#)
- Analýza spolehlivosti
  - deskriptivní ukazatele [216](#)
  - Hotelovo T 2 [216](#)

Analýza spolehlivosti (*pokračování*)  
korelace mezi položkami a kovariance [216](#)  
korelační koeficient uvnitř vlákna [216](#)  
Kuder-Richardson 20 [216](#)  
příkaz additional features [218](#)  
příklad [215](#)  
statistika [215](#), [216](#)  
tabulka ANOVA [216](#)  
Test na aditivitu Tukeyho [216](#)  
analýza více odpovědí  
Frekvence více odpovědí [208](#)  
frekvenční tabulky [208](#)  
Kontingenční tabulky s více odpovědí [209](#)  
zkřížžení [209](#)  
Ancombe  
v jednovzorkových porcích [102](#)  
Anderson-skóre faktoru Rubin [175](#)  
Andrewsův odhad vlny  
ve Prozkoumat [88](#)  
ANOVA  
in Means [97](#)  
model [119](#)  
v GLM Univariate [118](#)  
v jednofaktorové ANOVA [113](#)  
v lineárních modelech [138](#)  
automatická příprava dat  
v lineárních modelech [137](#)  
automatické rozložení distribuce  
v simulaci [241](#)

## B

Bartlettův test sférickosti  
Analýza faktoru [173](#)  
beta koeficienty  
v lineární regresi [143](#)  
Binární meta-analýza  
Bublina-karta [75](#)  
Dialogové okno Analýza [54](#)  
Dialogové okno Bias [57](#)  
Dialogové okno Inference [55](#)  
Dialogové okno kontrastu [56](#)  
Dialogové okno Kritéria [53](#)  
Dialogové okno Oříznout a Vyplnit [58](#)  
Dialogové okno Tisk [59](#)  
Dialogové okno vykreslení [61](#)  
Karta " L' Abb'e Plot [78](#)  
Karta Výkres funklu [76](#)  
Karta Výkres lesa [73](#)  
Kumulativní karta Zákresový strom [74](#)  
Ouško Galbraith [77](#)  
Uložit dialogové okno [60](#)  
binární výsledky  
data o velikosti nezpracovaného efektu [36](#)  
data o velikosti předvypočteného účinku [36](#)  
Binomické testování  
dichotomie [198](#)  
chybějící hodnoty [199](#)  
příkaz additional features [199](#)  
statistika [199](#)  
volby [199](#)  
binomický test  
Jedna-ukázka neparametrických testů [189](#)  
Bivariantní koráže

Bivariantní koráže (*pokračování*)  
chybějící hodnoty [130](#)  
interval spolehlivosti [129](#)  
intervaly spolehlivosti [130](#)  
korelační koeficienty [129](#)  
příkaz additional features [131](#)  
statistika [130](#)  
úroveň významnosti [129](#)  
volby [130](#)  
Bodová skóre Bartlett [175](#)  
bodové grafy  
v lineární regresi [141](#)  
bodový graf  
v simulaci [247](#)  
Bonett-cena  
v paired-rozměry vzorků [105](#)  
Bonferroniho  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)  
Brown-Li-Jeffreys  
in Independent-Samples Proportions  
[108](#)  
Bublina-karta  
Binární meta-analýza [75](#)  
Souvislá meta-analýza [75](#)  
Velikost binárního efektu metaanalýzy [75](#)  
Velikost souvislé analýzy meta-analýzy [75](#)

## C

celkové procento  
v kontingenčních tabulkách [92](#)  
celkové součty  
ve sloupcových souhrnných sestavách [214](#)  
cenový rozdíl (PRD)  
v měrových statistikách [223](#), [225](#)  
city-block distance  
v Nearest Sousední analýza [163](#)  
classification  
v analýze ROC [231](#)  
Clopper-Pearsonovy intervaly  
Jedna-ukázka neparametrických testů [189](#)  
Cloupper-Pearson (Přesná)  
v jednovzorkových porcích [102](#)  
Cochran's Q test  
Související-Neparametrické testy ukázek [195](#)  
Cochran's statistic  
v kontingenčních tabulkách [91](#)  
Cochranův Q  
v testech pro několik souvisejících ukázek [206](#)  
Cox a Snell R<sup>2</sup>  
v ordinální regresi [146](#)  
Cramér's V  
v kontingenčních tabulkách [91](#)  
Cronbachův alfa  
v Analytické spolehlivosti [215](#), [216](#)

## Č

Částečné korelace  
chybějící hodnoty [132](#)  
korelace s nulovým pořadím [132](#)  
příkaz additional features [132](#)

Částečné korelace (*pokračování*)

- statistika [132](#)
- v lineární regresi [143](#)
- volby [132](#)

číslování stránek

- v souhrnných sestavách řádku [212](#)
- ve sloupcových souhrnných sestavách [214](#)

## D

d

- v kontingenčních tabulkách [91](#)

Definovat více sad odpovědí

- dichotomie [207](#)
- kategorie [207](#)
- nastavit názvy [207](#)
- nastavit popisky [207](#)

dělení

- dělení mezi sloupci sestavy [214](#)

dendrogramy

- v Hierarchické analýze klastru [185](#)

dependent t test

- v paired-test vzorků T [111](#)

Deskriptivní statistika

- v analýze TwoStep Cluster [179](#)
- v deskriptivních [84](#)
- v frekvencích [83](#)
- v měrových statistikách [223](#), [225](#)
- v souhrnu [95](#)
- ve Prozkoumat [88](#)

detrended normal plots

- ve Prozkoumat [88](#)

DfBeta

- v lineární regresi [142](#)

diagnostické informace kolineárnosti

- v lineární regresi [143](#)

Diagnostické informace o případech

- v lineární regresi [143](#)

Dialogové okno Analýza

- Binární meta-analýza [54](#)
- Souvislá meta-analýza [39](#)
- Velikost binárního efektu metaanalýzy [63](#)
- Velikost souvislé analýzy meta-analýzy [46](#)

Dialogové okno Bias

- Binární meta-analýza [57](#)
- Souvislá meta-analýza [41](#)
- Velikost binárního efektu metaanalýzy [65](#)
- Velikost souvislé analýzy meta-analýzy [49](#)

Dialogové okno Inference

- Binární meta-analýza [55](#)
- Regrese meta-analýzy [71](#)
- Souvislá meta-analýza [40](#)
- Velikost binárního efektu metaanalýzy [64](#)
- Velikost souvislé analýzy meta-analýzy [47](#)

Dialogové okno kontrastu

- Binární meta-analýza [56](#)
- Souvislá meta-analýza [41](#)
- Velikost binárního efektu metaanalýzy [65](#)
- Velikost souvislé analýzy meta-analýzy [48](#)

Dialogové okno Kritéria

- Binární meta-analýza [53](#)
- Regrese meta-analýzy [70](#)
- Souvislá meta-analýza [38](#), [46](#)
- Velikost binárního efektu metaanalýzy [62](#)

Dialogové okno Oříznout a Vyplnit

- Binární meta-analýza [58](#)
- Souvislá meta-analýza [42](#)
- Velikost binárního efektu metaanalýzy [66](#)
- Velikost souvislé analýzy meta-analýzy [49](#)

Dialogové okno Tisk

- Binární meta-analýza [59](#)
- Regrese meta-analýzy [71](#)
- Souvislá meta-analýza [43](#)
- Velikost binárního efektu metaanalýzy [67](#)
- Velikost souvislé analýzy meta-analýzy [51](#)

Dialogové okno vykreslení

- Binární meta-analýza [61](#)
- Regrese meta-analýzy [73](#)
- Souvislá meta-analýza [44](#)
- Velikost binárního efektu metaanalýzy [69](#)
- Velikost souvislé analýzy meta-analýzy [52](#)

dílčí grafy

- v lineární regresi [141](#)

dobrota vložení

- v ordinální regresi [146](#)

dopředu krokové

- v lineárních modelech [136](#)

důležitost prediktoru

- lineární modely [138](#)

Duncanův test s více rozsahy

- v GLM [123](#)
- v jednofaktorové ANOVA [115](#)

Dunnnettův C

- v GLM [123](#)
- v jednofaktorové ANOVA [115](#)

Dunnnettův T3

- v GLM [123](#)
- v jednofaktorové ANOVA [115](#)

Dunnnettův test t

- v GLM [123](#)
- v jednofaktorové ANOVA [115](#)

Durbin-statistikaWatson

- v lineární regresi [143](#)

dvoufázový test t

- in Independent-Samples T Test [109](#)

## E

euklidovský odstup

- v Nearest Sousední analýza [163](#)
- ve Vzdálenosti [133](#)

exponenciální model

- v odhadu křivky [158](#)

extrémní hodnoty

- ve Prozkoumat [88](#)

## F

F statistika

- v lineárních modelech [136](#)

faktoring alfa [174](#)

faktoring hlavní osy [174](#)

faktoring obrázků [174](#)

Fí-metr čtvereční vzdálenost

- ve Vzdálenosti [133](#)

Fisherův LSD

Fisherův LSD (*pokračování*)  
v GLM [123](#)  
Fisherův přesný test  
v kontingenčních tabulkách [91](#)  
Fleiss ' Multiple Rater Kappa [215](#), [216](#)  
formátování  
sloupce v sestavách [211](#)  
Frekvence  
formáty [84](#)  
grafy [84](#)  
pořadí zobrazení [84](#)  
potlačení tabulek [84](#)  
statistika [83](#)  
frekvence klastru  
v analýze TwoStep Cluster [179](#)  
Frekvence více odpovědí  
chybějící hodnoty [208](#)  
frekvenční tabulky  
v frekvencích [82](#)  
ve Prozkoumat [88](#)  
Friedman test  
Související-Neparametrické testy ukázek [195](#)  
v testech pro několik souvisejících ukázek [206](#)  
funkce hustoty pravděpodobnosti  
v simulaci [246](#)

## G

Galbraithův náskres [73](#)  
Gamma  
v kontingenčních tabulkách [91](#)  
Gamman Goodman a Kruskal  
v kontingenčních tabulkách [91](#)  
generalizované nejnižší čtverce  
Analýza faktoru [174](#)  
geometrický průměr  
in Means [97](#)  
v krychlicích OLAP [99](#)  
v souhrnu [95](#)  
geoprostorové modelování [251–261](#)  
GLM  
model [119](#)  
post hoc testy [123](#)  
profilové grafy [121](#)  
součet čtverců [119](#)  
ukládání matic [126](#)  
ukládání proměnných [126](#)  
GLM Univariate  
kontrast [120](#), [121](#)  
Goodman a Kruskal's tau  
v kontingenčních tabulkách [91](#)  
grafy  
grafy pravděpodobnosti [226](#), [229](#)  
popisky případů [157](#)  
v analýze ROC [231](#)  
v ROC křivce [234](#)  
grafy pravděpodobnosti  
P-P [226](#)  
Q-Q [229](#)  
grafy s rozpětím na úrovni  
ve Prozkoumat [88](#)

## H

harmonický průměr  
in Means [97](#)  
v krychlicích OLAP [99](#)  
v souhrnu [95](#)  
Hauck-Anderson  
in Independent-Samples Proportions  
[108](#)  
Helmert kontrastuje  
v GLM [120](#), [121](#)  
Hierarchický [119](#)  
Histogramy  
v frekvencích [84](#)  
v lineární regresi [141](#)  
ve Prozkoumat [88](#)  
historie iterace  
v ordinální regresi [146](#)  
Hladina významnosti Asymptotic [216](#)  
hloubka stromu  
v analýze TwoStep Cluster [178](#)  
Hnědý-Forsythe statistiky  
v jednofaktorové ANOVA [117](#)  
Hodges-Lehman Brothers  
Související-Neparametrické testy ukázek [195](#)  
hodnoty eigenvalues  
Analýza faktoru [173](#), [174](#)  
v lineární regresi [143](#)  
Hochbergův GT2  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)  
Hotelovo T 2  
v Analytické spolehlivosti [215](#), [216](#)  
Huberův měřič M  
ve Prozkoumat [88](#)

## CH

Chebychev vzdálenost  
ve Vzdálenosti [133](#)  
chí kvadrát  
Fisherův přesný test [91](#)  
chybějící hodnoty [198](#)  
jedn-vzorový test [197](#)  
lineární přidružení [91](#)  
očekávané hodnoty [197](#)  
očekávaný rozsah [197](#)  
PEARSON [91](#)  
Pravděpodobnost-poměr [91](#)  
pro nezávislost [91](#)  
statistika [198](#)  
v kontingenčních tabulkách [91](#)  
volby [198](#)  
Yates ' korekce kontinuity [91](#)  
chí-kvadrát vzdálenost  
ve Vzdálenosti [133](#)  
chybějící hodnoty  
Analýza faktoru [175](#)  
in Bivariate Correlations [130](#)  
in Independent-Samples Proportions [109](#)  
in Independent-Samples T Test [111](#)  
in One-Sample Kolmogorov-Smirnov Test [201](#),  
[202](#)  
v analýze ROC [233](#)

chybějící hodnoty (*pokračování*)  
v Binomickém testu [199](#)  
v části Correlations [132](#)  
v jednofaktorové ANOVA [117](#)  
v jednovzorkového testu T [113](#)  
v jednovzorkových porcích [103](#)  
v kontingenčních tabulkách s více odpovědí [210](#)  
v lineární regresi [144](#)  
v Nearest Sousední analýza [166](#)  
v paired-rozměry vzorků [106](#)  
v paired-test vzorků T [112](#)  
v ROC křivce [235](#)  
v souhrnech sestav v řádcích [212](#)  
v testech pro několik nezávislých vzorků [205](#)  
v testech Two-Independent-Samples [203](#)  
v testu chí-kvadrát [198](#)  
V testu spuštění [200](#)  
ve dvou souvisejících-testech ukázek [204](#)  
ve Prozkoumat [89](#)  
ve sloupcových souhrnných sestavách [214](#)  
Ve vícenásobných frekvencích odpovědí [208](#)  
Chybějící hodnoty  
v percentilech [87](#)

## I

ICC. Viz korelační koeficient v rámci vnitrobuněčné [216](#)  
index koncentrace  
v měrových statistikách [223](#), [225](#)  
inflační faktor rozptylu  
v lineární regresi [143](#)  
informační kritéria  
v lineárních modelech [136](#)  
Informační kritérium společnosti Akaike  
v lineárních modelech [136](#)  
interval spolehlivosti  
in Bivariate Correlations [129](#)  
Intervaly Jeffreys  
Jedna-ukázka neparametrických testů [189](#)  
interval poměru pravděpodobnosti  
Jedna-ukázka neparametrických testů [189](#)  
interval předpovědi  
uložení v lineární regresi [142](#)  
uložení v odhadu křivky [159](#)  
interval spolehlivosti  
in Bivariate Correlations [130](#)  
in Independent-Samples T Test [111](#)  
uložení v lineární regresi [142](#)  
v analýze ROC [233](#)  
v GLM [120](#)  
v jednofaktorové ANOVA [117](#)  
v jednovzorkového testu T [113](#)  
v lineární regresi [143](#)  
v paired-test vzorků T [112](#)  
v ROC křivce [235](#)  
ve Prozkoumat [88](#)  
inverzní model  
v odhadu křivky [158](#)  
Iterace  
Analýza faktoru [174](#), [175](#)  
v analýze K-Means Cluster [187](#)

## J

Jedna-ukázka neparametrických testů  
binomický test [189](#)  
pole [188](#)  
spustí test [190](#)  
test chí-kvadrát [190](#)  
Test Kolmogorov-Smirnov [190](#)  
jednoduché kontrasty  
v GLM [120](#), [121](#)  
Jednosměrný test ANOVA  
chybějící hodnoty [117](#)  
kontrast [114](#)  
polynomu [114](#)  
post hoc testy [115](#)  
proměnné faktoru [113](#)  
příkaz additional features [117](#)  
statistika [117](#)  
více porovnání [115](#)  
volby [117](#)  
Jednovzorkový proporce [101](#), [102](#)  
Jednovzorkový test T  
chybějící hodnoty [113](#)  
intervaly spolehlivosti [113](#)  
příkaz additional features [112](#), [113](#)  
volby [113](#)  
Jedny-vzorky pro porce [105](#)  
Jeffreydae  
v jednovzorkových porcích [102](#)

## K

k a výběru funkcí  
v Nearest Sousední analýza [168](#)  
k výběru  
v Nearest Sousední analýza [168](#)  
kappa  
v kontingenčních tabulkách [91](#)  
Karta " L' Abb'e Plot  
Binární meta-analýza [78](#)  
Souvislá meta-analýza [78](#)  
Velikost binárního efektu metaanalýzy [78](#)  
Velikost souvislé analýzy meta-analýzy [78](#)  
Karta Výkres funklu  
Binární meta-analýza [76](#)  
Souvislá meta-analýza [76](#)  
Velikost binárního efektu metaanalýzy [76](#)  
Velikost souvislé analýzy meta-analýzy [76](#)  
Karta Výkres lesa  
Binární meta-analýza [73](#)  
Souvislá meta-analýza [73](#)  
Velikost binárního efektu metaanalýzy [73](#)  
Velikost souvislé analýzy meta-analýzy [73](#)  
kategorie odkazů  
v GLM [120](#), [121](#)  
Kendall je koeficient concordance (W)  
Související-Neparametrické testy ukázek [195](#)  
Kendall's tau-b  
in Bivariate Correlations [129](#)  
v kontingenčních tabulkách [91](#)  
Kendall's tau-c  
v kontingenčních tabulkách [91](#)  
Kendallův W  
v testech pro několik souvisejících ukázek [206](#)

Klasifikace  
 v ROC křivce [234](#)

Klastrování  
 celková obrazovka [180](#)  
 výběr procedury [176](#)  
 zobrazení klastrů [180](#)

Kódová kniha  
 statistika [81](#)  
 výstup [80](#)

koefficient korelace Spearman  
 in Bivariate Correlations [129](#)  
 v kontingenčních tabulkách [91](#)

koefficient nejistoty  
 v kontingenčních tabulkách [91](#)

koefficient rozptylu (COD)  
 v měrových statistikách [223](#), [225](#)

Kolmogorov-Smirnov Z  
 in One-Sample Kolmogorov-Smirnov Test [200](#)  
 v testech Two-Independent-Samples [203](#)

kombinování pravidel  
 v lineárních modelech [136](#)

komplety  
 v lineárních modelech [136](#)

Kontingenční tabulky  
 formáty [94](#)  
 klastrované pruhové grafy [90](#)  
 potlačení tabulek [89](#)  
 řídicí proměnné [90](#)  
 statistika [91](#)  
 vrstvy [90](#)  
 zobrazení buňky [92](#)

Kontingenční tabulky s více odpovědí  
 definování rozsahů hodnot [209](#)  
 chybějící hodnoty [210](#)  
 odpovídající proměnné v rámci sad odpovědí [210](#)  
 procentní části buněk [210](#)  
 procentní části založené na odpovědích [210](#)  
 procentní části založené na případech [210](#)

kontrast  
 v GLM [120](#), [121](#)  
 v jednofaktorové ANOVA [114](#)

konvergence  
 Analýza faktoru [174](#), [175](#)  
 v analýze K-Means Cluster [187](#)

Korelace  
 in Bivariate Correlations [129](#)  
 nula-objednávka [132](#)  
 v části Correlations [131](#)  
 v kontingenčních tabulkách [91](#)  
 v simulaci [244](#)

korelace s nulovým pořadím  
 v části Correlations [132](#)

korelační koeficient (ICC) v rámci vnitrostrany  
 v Analytické spolehlivosti [216](#)

korelační matice  
 Analýza faktoru [172](#), [173](#)  
 v analýze diskriminant [170](#)  
 v ordinální regresí [146](#)

kovarianční matice  
 v analýze diskriminant [170](#), [171](#)  
 v GLM [126](#)  
 v lineární regresí [143](#)  
 v ordinální regresí [146](#)

KR20  
 v Analytické spolehlivosti [216](#)

krabičkové grafy  
 porovnání proměnných [88](#)  
 porovnání úrovní faktoru [88](#)  
 v simulaci [247](#)  
 ve Prozkoumat [88](#)

kritéria výběru  
 v lineární regresí [143](#)

kroková volba  
 v lineární regresí [141](#)

Kruskal-Wallis H  
 v testech Two-Independent-Samples [205](#)

Kruskovo tau  
 v kontingenčních tabulkách [91](#)

Krychle OLAP  
 nadpisy [101](#)  
 statistika [99](#)

Křivka ROC  
 statistiky a grafy [235](#)

Kuder-Richardson 20 (KR20)  
 v Analytické spolehlivosti [216](#)

kumulativní distribuční funkce  
 v simulaci [246](#)

Kumulativní frekvence  
 v ordinální regresí [146](#)

Kumulativní karta Zákresový strom  
 Binární meta-analýza [74](#)  
 Souvislá meta-analýza [74](#)  
 Velikost binárního efektu metaanalýzy [74](#)  
 Velikost souvislé analýzy meta-analýzy [74](#)

Kumulativní vykreslení lesa [73](#)

kvadrant mapa  
 v Nearest Sousední analýza [168](#)

kvadrát euklidovský vzdálenost  
 ve Vzdálenosti [133](#)

kvadratický model  
 v odhadu křivky [158](#)

Kvantilový graf pravděpodobnosti [229](#)

kvartily  
 v frekvencích [83](#)

kvartimax rotace  
 Analýza faktoru [175](#)

**L**

L' Abb'e Plot [73](#)

lachtama cohenská  
 v kontingenčních tabulkách [91](#)

lambda  
 v kontingenčních tabulkách [91](#)

Lambda Goodmanova a Kruskina  
 v kontingenčních tabulkách [91](#)

Lilieffors test  
 ve Prozkoumat [88](#)

Lineární greso Regression  
 partition (oblast) [151](#)  
 proměnné [151](#)  
 volby [152](#)

lineární model  
 v odhadu křivky [158](#)

lineární modely  
 automatická příprava dat [135](#), [137](#)



- lineární modely (*pokračování*)
    - Cíle [134](#)
    - důležitost prediktoru [138](#)
    - informační kritérium [137](#)
    - Koeficienty [139](#)
    - kombinování pravidel [136](#)
    - komplety [136](#)
    - odhadované prostředky [139](#)
    - Odlehlé hodnoty [138](#)
    - předpovězeno pozorovanými [138](#)
    - replikace výsledků [137](#)
    - souhrn modelu [137](#)
    - souhrn sestavení modelu [139](#)
    - Statistika R-čtvereček [137](#)
    - tabulka ANOVA [138](#)
    - Úroveň důvěry [135](#)
    - volby modelu [137](#)
    - Výběr modelu [136](#)
    - zbytkové chyby [138](#)
  - lineární přidružování
    - v kontingenčních tabulkách [91](#)
  - Lineární regrese
    - bloky [140](#)
    - export informací o modelu [142](#)
    - chybějící hodnoty [144](#)
    - metody výběru proměnných [141](#), [144](#)
    - proměnná výběru [141](#)
    - příkaz additional features [145](#)
    - statistiky [143](#)
    - ukládání nových proměnných [142](#)
    - váhy [140](#)
    - výkresy [141](#)
    - zbytkové chyby [142](#)
  - Lineární regrese elastické sítě
    - partition (oblast) [148](#)
    - proměnné [148](#)
    - volby [149](#)
  - logaritmický model
    - v odhadu křivky [158](#)
  - logistický model
    - v odhadu křivky [158](#)
  - Logit
    - v jednovzorkových porcích [102](#)
- M**
- M-odhad
    - ve Prozkoumat [88](#)
  - Mahalanobis vzdálenost
    - v analýze diskriminant [171](#)
    - v lineární regresi [142](#)
  - Manhattan
    - v Nearest Sousední analýza [163](#)
  - Mann-Whitney U
    - v testech Two-Independent-Samples [203](#)
  - matrix vzorků
    - Analýza faktoru [172](#)
  - maximální
    - in Means [97](#)
    - porovnání sloupců sestavy [214](#)
    - v deskriptivních [85](#)
    - v frekvencích [83](#)
    - v krychlích OLAP [99](#)
  - maximální (*pokračování*)
    - v měrových statistikách [223](#), [225](#)
    - v souhrnu [95](#)
    - ve Prozkoumat [88](#)
  - maximální počet větví
    - v analýze TwoStep Cluster [178](#)
  - maximální pravděpodobnost
    - Analýza faktoru [174](#)
  - McFadden R2
    - v ordinální regresi [146](#)
  - McNemar
    - v paired-rozměry vzorků [105](#)
  - McNemar (continuity corrected)
    - v paired-rozměry vzorků [105](#)
  - McNemarský test
    - Související-Neparametrické testy ukázek [195](#)
    - v kontingenčních tabulkách [91](#)
    - ve dvou souvisejících-testech ukázek [204](#)
  - medián
    - in Means [97](#)
    - v frekvencích [83](#)
    - v krychlích OLAP [99](#)
    - v měrových statistikách [223](#), [225](#)
    - v souhrnu [95](#)
    - ve Prozkoumat [88](#)
  - medián testu
    - v testech Two-Independent-Samples [205](#)
  - měřitko rozdílu velikosti
    - ve Vzdálenosti [133](#)
  - Metaanalýza
    - binární výsledky [36](#)
    - Ploty [73](#)
    - průběžné výsledky [36](#)
    - Regrese [36](#)
  - Meziletová dohoda [216](#)
  - mezisoučty
    - ve sloupcových souhrnných sestavách [214](#)
  - mezni zkouška homogenity
    - Související-Neparametrické testy ukázek [195](#)
    - ve dvou souvisejících-testech ukázek [204](#)
  - minimum
    - in Means [97](#)
    - porovnání sloupců sestavy [214](#)
    - v deskriptivních [85](#)
    - v frekvencích [83](#)
    - v krychlích OLAP [99](#)
    - v měrových statistikách [223](#), [225](#)
    - v souhrnu [95](#)
    - ve Prozkoumat [88](#)
  - Minkovská vzdálenost
    - ve Vzdálenosti [133](#)
  - Model Guttman
    - v Analytické spolehlivosti [215](#), [216](#)
  - model krychlový
    - v odhadu křivky [158](#)
  - model napájení
    - v odhadu křivky [158](#)
  - model scale
    - v ordinální regresi [147](#)
  - model umístění
    - v ordinální regresi [147](#)
  - Mojžíš Extrémní test reakce

Mojžiš Extrémní test reakce (*pokračování*)  
v testech Two-Independent-Samples [203](#)  
Multidimenzionální škálování  
definice tvaru dat [221](#)  
dimenze [222](#)  
modely škálování [222](#)  
opatření na vzdálenost [221](#)  
podmíněnost [222](#)  
Podmínky [222](#)  
příkaz additional features [222](#)  
příklad [220](#)  
statistika [220](#)  
transformace hodnot [221](#)  
úrovně měření [222](#)  
volby zobrazení [222](#)  
vytváření matic vzdáleností [221](#)

## N

na druhou  
in Means [97](#)  
načítání grafů  
Analýza faktoru [175](#)  
nadpisy  
v krychlích OLAP [101](#)  
Nagelkerke R2  
v ordinální regresí [146](#)  
násobení  
násobení ve sloupcích sestavy [214](#)  
návrhy pro testy heteroskedasticity  
v GLM [128](#)  
nejbližší vzdálenosti od sousedů  
v Nearest Sousední analýza [168](#)  
nejlepší podmnožiny  
v lineárních modelech [136](#)  
nejméně významný rozdíl  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)  
neparametrické testy  
chí kvadrát [197](#)  
One-Ukázka Kolmogorovan-Smirnov test [200](#)  
Spouští test [199](#)  
Testy dvou souvisejících ukázek [204](#)  
Testy pro několik nezávislých vzorků [205](#)  
Testy pro několik souvisejících ukázek [206](#)  
Testy Two-Independent-Samples [202](#)  
nestandardizované zbytkové chyby  
v GLM [126](#)  
nevážené nejmenších čtverců  
Analýza faktoru [174](#)  
Newcombe (*pokračování opravené*)  
in Independent-Samples Proportions [108](#)  
Newman-Keuls  
v GLM [123](#)  
Nezávisle-části vzorků [106](#), [108](#)  
Nezávislí-Ukázkové neparametrické testy  
Karta Pole [192](#)  
Nezávislý-T-test vzorků  
definice skupin [110](#)  
řetězcové proměnné [110](#)  
seskupování proměnných [110](#)

Nezávislý-Test č. T  
chybějící hodnoty [111](#)  
intervaly spolehlivosti [111](#)  
volby [111](#)  
Normalizace čísla  
Analýza faktoru [175](#)  
normální grafy pravděpodobnosti  
P-P [226](#)  
Q-Q [229](#)  
v lineární regresí [141](#)  
ve Prozkoumat [88](#)  
nouzové tabulky [89](#)  
nouzový koeficient  
v kontingenčních tabulkách [91](#)  
Novcombe  
in Independent-Samples Proportions [108](#)  
v paired-rozměry vzorků [105](#)

## O

očekávané frekvence  
v ordinální regresí [146](#)  
očekávaný počet  
v kontingenčních tabulkách [92](#)  
Odhad křivky  
analýza rozptylu [157](#)  
modely [158](#)  
prognóza [159](#)  
uložení intervalů předpovědi [159](#)  
uložení předpokládaných hodnot [159](#)  
uložení zbytkových chyb [159](#)  
včetně konstanty [157](#)  
Odhadce klesání (M-estimator)  
ve Prozkoumat [88](#)  
odhady parametrů  
v ordinální regresí [146](#)  
odchylka  
in Means [97](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v krychlích OLAP [99](#)  
v souhrnech sestav v řádcích [212](#)  
v souhrnu [95](#)  
ve Prozkoumat [88](#)  
ve Souhrnech sestav ve sloupcích [213](#)  
odchylky kontrastů  
v GLM [120](#), [121](#)  
odkaz  
v ordinální regresí [146](#)  
Odlehlé hodnoty  
v analýze TwoStep Cluster [178](#)  
v lineární regresí [141](#)  
ve Prozkoumat [88](#)  
odstraněné zbytkové chyby  
v GLM [126](#)  
v lineární regresí [142](#)  
Okovy vzdálenost  
v GLM [126](#)  
v lineární regresí [142](#)  
Omega McDonald  
v Analytické spolehlivosti [215](#), [216](#)  
One-Ukázka Kolmogorovan-Smirnov  
test

One-Ukázka Kolmogorovan-Smirnov test (*pokračování*)  
 distribuce testu [200](#)  
 chybějící hodnoty [201](#), [202](#)  
 příkaz additional features [202](#)  
 statistika [201](#), [202](#)  
 Test Liliefors [200](#), [201](#)  
 volby [201](#), [202](#)  
 opakované kontrasty  
 v GLM [120](#), [121](#)  
 opatření na vzdálenost  
 v Hierarchické analýze klastru [184](#)  
 v Nearest Sousední analýza [163](#)  
 ve Vzdálenosti [133](#)  
 opatření podobnosti  
 v Hierarchické analýze klastru [184](#)  
 ve Vzdálenosti [133](#)  
 opatření pro distribuci  
 v deskriptivních [85](#)  
 v frekvencích [83](#)  
 opatření ústředních tendencí  
 v frekvencích [83](#)  
 v měrových statistikách [223](#), [225](#)  
 ve Prozkoumat [88](#)  
 Ordinální regrese  
 model scale [147](#)  
 model umístění [147](#)  
 odkaz [146](#)  
 příkaz additional features [148](#)  
 statistika [145](#)  
 volby [146](#)  
 oříznuté  
 ve Prozkoumat [88](#)  
 Ouško Galbraith  
 Binární meta-analýza [77](#)  
 Souvislá meta-analýza [77](#)  
 Velikost binárního efektu metaanalýzy [77](#)  
 Velikost souvislé analýzy meta-analýzy [77](#)

## P

P-P pravděpodobnostní grafy [226](#)  
 Paired-rozměry-porce [104](#)  
 Paired-test s ukázkami T  
 chybějící hodnoty [112](#)  
 volby [112](#)  
 výběr párových proměnných [111](#)  
 Paired-Ukázky t-Test [111](#)  
 pákové hodnoty  
 v GLM [126](#)  
 v lineární regresi [142](#)  
 paralelní model  
 v Analytické spolehlivosti [215](#), [216](#)  
 Pearson chi-square  
 v kontingenčních tabulkách [91](#)  
 v ordinální regresi [146](#)  
 Pearsonová korelace  
 in Bivariate Correlations [129](#)  
 v kontingenčních tabulkách [91](#)  
 Pearsonovy chyby  
 v ordinální regresi [146](#)  
 percentily  
 v frekvencích [83](#)  
 v simulaci [247](#)  
 ve Prozkoumat [88](#)

Phi  
 v kontingenčních tabulkách [91](#)  
 Ploty  
 Metaanalýza [73](#)  
 počáteční prahová hodnota  
 v analýze TwoStep Cluster [178](#)  
 počet desetinných míst  
 v Analytické spolehlivosti [215](#)  
 Vážené Kappy [218](#)  
 ve vícerozměrném škálování [220](#)  
 počet případů  
 in Means [97](#)  
 v krychlích OLAP [99](#)  
 v souhrnu [95](#)  
 Podepsaný test Wilcoxon  
 Jedna-ukázka neparametrických testů [189](#)  
 Související-Neparametrické testy ukázek [195](#)  
 ve dvou souvisejících-testech ukázek [204](#)  
 podepsat test  
 Související-Neparametrické testy ukázek [195](#)  
 ve dvou souvisejících-testech ukázek [204](#)  
 Podmíněná analýza  
 v simulaci [244](#)  
 podmínka prevence proti přeplnění  
 v lineárních modelech [136](#)  
 podmínky sestavení [119](#), [147](#), [148](#)  
 podskupina prostředků [96](#), [98](#)  
 polynomu  
 v GLM [120](#), [121](#)  
 v jednofaktorové ANOVA [114](#)  
 poměr kovariance  
 v lineární regresi [142](#)  
 Poměr statistik  
 statistika [223](#), [225](#)  
 pomocný regresní model  
 v GLM [128](#)  
 Popisovače  
 pořadí zobrazení [85](#)  
 příkaz additional features [86](#)  
 statistika [85](#)  
 uložení skóre z [84](#)  
 porovnání proměnných  
 v krychlích OLAP [100](#)  
 porovnání skupin  
 v krychlích OLAP [100](#)  
 poslední  
 in Means [97](#)  
 v krychlích OLAP [99](#)  
 v souhrnu [95](#)  
 post hoc vícenásobné porovnání [115](#)  
 pozorované frekvence  
 v ordinální regresi [146](#)  
 pozorovaný počet  
 v kontingenčních tabulkách [92](#)  
 Pravděpodobnost-poměr chí kvadrát  
 v kontingenčních tabulkách [91](#)  
 v ordinální regresi [146](#)  
 predikované hodnoty  
 uložení v lineární regresi [142](#)  
 uložení v odhadu křivky [159](#)  
 procent  
 v kontingenčních tabulkách [92](#)  
 procentní části řádků  
 v kontingenčních tabulkách [92](#)

procentní části sloupců  
     v kontingenčních tabulkách [92](#)  
 profilové grafy  
     v GLM [121](#)  
 prognóza  
     v odhadu křivky [159](#)  
 prohlížeč klastru  
     distribuce buněk [182](#)  
     důležitost prediktoru [182](#)  
     filtrování záznamů [183](#)  
     funkce řazení [181](#)  
     o modelech klastrů [179](#)  
     porovnání klastrů [182](#)  
     použití [183](#)  
     přehled [180](#)  
     překlopení klastrů a funkcí [181](#)  
     řazení klastrů [181](#)  
     řazení obsahu buněk [181](#)  
     řazení zobrazení funkcí [181](#)  
     řazení zobrazení klastru [181](#)  
     souhrn modelu [180](#)  
     souhrnné zobrazení [180](#)  
     transpozice klastrů a funkcí [181](#)  
     velikost klastrů [182](#)  
     základní zobrazení [181](#)  
     zobrazení distribuce buněk [182](#)  
     zobrazení důležitosti prediktoru klastru [182](#)  
     zobrazení klastrů [180](#)  
     zobrazení obsahu buňky [181](#)  
     zobrazení porovnání klastru [182](#)  
     zobrazení středisek klastru [180](#)  
     zobrazení velikosti klastru [182](#)  
 proměnná výběru  
     v lineární regresi [141](#)  
 Proměnná závažnost  
     v Nearest Sousední analýza [167](#)  
 prostorové modelování [251](#)  
 prostorový graf funkcí  
     v Nearest Sousední analýza [166](#)  
 Prostředky  
     statistika [97](#)  
     volby [97](#)  
 Proximity  
     v Hierarchické analýze klastru [184](#)  
 Prozkoumat  
     chybějící hodnoty [89](#)  
     příkaz additional features [89](#)  
     statistiky [88](#)  
     transformace napájení [89](#)  
     volby [89](#)  
     výkresy [88](#)  
 průběžné výsledky  
     data o velikosti nezpracovaného efektu [36](#)  
     data o velikosti předvypočteného účinku [36](#)  
     Metaanalýza [36](#)  
 průměrná absolutní odchylka (AAD)  
     v měrových statistikách [223](#), [225](#)  
 první  
     in Means [97](#)  
     v krychlích OLAP [99](#)  
     v souhrnu [95](#)  
 přebalování  
     v lineárních modelech [134](#)  
 předat výběr  
     v lineární regresi [141](#)  
     v Nearest Sousední analýza [164](#)  
 Předpokládaný zbývající čas:  
     in Means [97](#)  
     v kontingenčních tabulkách [91](#)  
 Přesná Binomické  
     v paired-rozměry vzorků [105](#)  
 Přidělení paměti  
     v analýze TwoStep Cluster [178](#)  
 přímá oblití rotace  
     Analýza faktorů [175](#)  
 případové studie  
     Paired-test s ukázkami T [111](#)  
 přísně paralelní model  
     v Analytické spolehlivosti [215](#), [216](#)

## R

r korelační koeficient  
     in Bivariate Correlations [129](#)  
     v kontingenčních tabulkách [91](#)  
 R statistika  
     in Means [97](#)  
     v lineární regresi [143](#)  
 R-čtverec  
     v lineárních modelech [137](#)  
 R-E-G-W F  
     v GLM [123](#)  
     v jednofaktorové ANOVA [115](#)  
 R-E-G-W Q  
     v GLM [123](#)  
     v jednofaktorové ANOVA [115](#)  
 R2  
     in Means [97](#)  
     v lineární regresi [143](#)  
     Změna R 2 [143](#)  
 Raovo V.  
     v analýze diskriminant [171](#)  
 Regrese  
     Lineární regrese [140](#)  
     Metaanalýza [36](#)  
     vícenásobná regrese [140](#)  
     výkresy [141](#)  
 Regrese částečných nejmenších čtverců  
     export proměnných [161](#)  
     model [161](#)  
 Regrese lineárního Ridge  
     partition (oblast) [154](#)  
     proměnné [154](#)  
     volby [155](#)  
 Regrese meta-analýzy  
     Dialogové okno Inference [71](#)  
     Dialogové okno Kritéria [70](#)  
     Dialogové okno Tisk [71](#)  
     Dialogové okno vykreslení [73](#)  
     Uložit dialogové okno [72](#)  
     Zákresový [75](#)  
 regresní koeficienty  
     v lineární regresi [143](#)  
 relativní riziko  
     v kontingenčních tabulkách [91](#)  
 režim  
     v frekvencích [83](#)

riziko  
v kontingenčních tabulkách [91](#)

ró  
in Bivariate Correlations [129](#)  
v kontingenčních tabulkách [91](#)

rotace equamax  
Analýza faktoru [175](#)

rotace varimax  
Analýza faktoru [175](#)

Rovnocenný partneri  
v Nearest Sousední analýza [167](#)

rozdíly mezi proměnnými  
v krychlích OLAP [100](#)

rozdíly mezi skupinami  
v krychlích OLAP [100](#)

rozdíly v kontrastu  
v GLM [120](#), [121](#)

rozsah  
in Means [97](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v krychlích OLAP [99](#)  
v měrových statistikách [223](#), [225](#)  
v souhrnu [95](#)

rozvodné vybavení  
v simulaci [241](#)

růstový model  
v odhadu křivky [158](#)

Ryan-Einot-Gabriel-Welsch více F  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

Ryan-Einot-Gabriel-Welsch vícenásobný rozsah  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

## Ř

řídící proměnné  
v kontingenčních tabulkách [90](#)

řízení stránky  
v souhrnných sestavách řádku [212](#)  
ve sloupcových souhrnných sestavách [214](#)

## S

S model  
v odhadu křivky [158](#)

S-stres  
ve vícerozměrném škálování [220](#)

seskupený medián  
in Means [97](#)  
v krychlích OLAP [99](#)  
v souhrnu [95](#)

sestavy  
celkový počet sloupců [214](#)  
dělení hodnot sloupce [214](#)  
násobení hodnot sloupce [214](#)  
porovnání sloupců [214](#)  
sestavy souhrnu řádku [210](#)  
sloupce se souhrnem sloupců [213](#)  
složené součty [214](#)

Shapiro-Wilkův test  
ve Prozkoumat [88](#)

Shrnout  
statistika [95](#)  
volby [94](#)

Sidakův test  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

simulace  
analýza citlivosti [244](#)  
bodové grafy [247](#)  
editor rovnic [239](#)  
formáty zobrazení pro cíle a vstupy [247](#)  
funkce hustoty pravděpodobnosti [246](#)  
funkce kumulativní distribuce [246](#)  
interaktivní grafy [250](#)  
korelace mezi vstupy [244](#)  
krabičkové grafy [247](#)  
percentily cílových distribucí [247](#)  
Podmíněná analýza [244](#)  
podporované modely [238](#)  
přizpůsobení instalace pro distribuci [243](#)  
rozvodné vybavení [241](#)  
specifikace modelu [238](#)  
spuštění plánu simulace [238](#), [248](#)  
tornádo, grafy [247](#)  
Tvůrce simulace [238](#)  
uložit plán simulace [248](#)  
uložit simulovaná data [248](#)  
Volby grafu [251](#)  
výsledky distribuce instalace [243](#)  
výstup [246](#), [247](#)  
vytváření distribučních distribucí pro nová data [248](#)  
vytvoření nových vstupů [240](#)  
vytvoření plánu simulace [236](#), [237](#)  
vzorkování ocasů [244](#)  
zastavení kritérií [244](#)

simulace Monte Carla [235](#)

Skóre [103](#)

Skóre (Vodivosti opravené) [103](#)

skóre faktoru [175](#)

Skóre Wilson  
v jednovzorkových porcích [102](#)

skupina znamená [96](#), [98](#)

sloupec se souhrnem sloupců [213](#)

Sloupec součtů  
v sestavách [214](#)

Slovník  
Kódová kniha [79](#)

složený model  
v odhadu křivky [158](#)

směrodatná chyba střední hodnoty  
in Means [97](#)  
v krychlích OLAP [99](#)  
v souhrnu [95](#)

směrodatná odchylka  
in Means [97](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v krychlích OLAP [99](#)  
v měrových statistikách [223](#), [225](#)  
v souhrnech sestav v řádcích [212](#)  
v souhrnu [95](#)  
ve Prozkoumat [88](#)  
ve Souhrnech sestav ve sloupcích [213](#)

Somers ' d

Somers ' d (pokračování)  
v kontingenčních tabulkách [91](#)  
součet  
in Means [97](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v krychlích OLAP [99](#)  
v souhrnu [95](#)  
součet čtverců  
v GLM [119](#)  
součinitel alfa  
v Analytické spolehlivosti [215](#), [216](#)  
součinitel korelace  
in Bivariate Correlations [129](#)  
souhrn chyb  
v Nearest Sousední analýza [168](#)  
Souhrny sestav v řádcích  
číslování stránek [212](#)  
Datové sloupce [210](#)  
formát sloupce [211](#)  
chybějící hodnoty [212](#)  
nadpisy [212](#)  
proměnné v titulcích [212](#)  
příkaz additional features [214](#)  
rozteč mezer [212](#)  
rozvržení stránky [212](#)  
řazení posloupností [210](#)  
řízení stránky [212](#)  
sloupce přerušení [210](#)  
zápatí [212](#)  
Souhrny sestav ve sloupcích  
celkový počet sloupců [214](#)  
celkový součet [214](#)  
číslování stránek [214](#)  
formát sloupce [211](#)  
chybějící hodnoty [214](#)  
mezisoučty [214](#)  
příkaz additional features [214](#)  
rozvržení stránky [212](#)  
řízení stránky [214](#)  
související ukázky [204](#), [206](#)  
Související-Neparametrické testy ukázek  
Cochran's Q test [195](#)  
McNemarský test [195](#)  
pole [194](#)  
Souvislá meta-analýza  
Bublina-karta [75](#)  
Dialogové okno Analýza [39](#)  
Dialogové okno Bias [41](#)  
Dialogové okno Inference [40](#)  
Dialogové okno kontrastu [41](#)  
Dialogové okno Kritéria [38](#), [46](#)  
Dialogové okno Oříznout a Vyplnit [42](#)  
Dialogové okno Tisk [43](#)  
Dialogové okno vykreslení [44](#)  
Karta " L' Abb'e Plot [78](#)  
Karta Výkres funklu [76](#)  
Karta Výkres lesa [73](#)  
Kumulativní karta Zákresový strom [74](#)  
Ouško Galbraithní [77](#)  
Uložit dialogové okno [44](#)  
Spearman-vysoká spolehlivost  
v Analytické spolehlivosti [216](#)  
split-half reliability

split-half reliability (pokračování)  
v Analytické spolehlivosti [215](#), [216](#)  
Spouští test  
chybějící hodnoty [200](#)  
oříznuté body [199](#), [200](#)  
příkaz additional features [200](#)  
statistika [200](#)  
volby [200](#)  
spustí test  
Jedna-ukázka neparametrických testů [189](#), [190](#)  
standardizace  
v analýze TwoStep Cluster [178](#)  
standardizované hodnoty  
v deskriptivních [84](#)  
standardizované zbytkové chyby  
v GLM [126](#)  
v lineární regresi [142](#)  
standardní chyba  
v analýze ROC [233](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v GLM [126](#)  
v ROC křivce [235](#)  
ve Prozkoumat [88](#)  
standardní chyba šikmosti  
in Means [97](#)  
v krychlích OLAP [99](#)  
v souhrnu [95](#)  
standardní chyba špičatosti  
in Means [97](#)  
v krychlích OLAP [99](#)  
v souhrnu [95](#)  
Statistika Mantel-Haenszel  
v kontingenčních tabulkách [91](#)  
statistika podílů sloupců  
v kontingenčních tabulkách [92](#)  
Střední  
in Means [97](#)  
podskupina [96](#), [98](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v jednofaktorové ANOVA [117](#)  
v krychlích OLAP [99](#)  
v měrových statistikách [223](#), [225](#)  
v souhrnech sestav v řádcích [212](#)  
v souhrnu [95](#)  
ve Prozkoumat [88](#)  
ve Souhrnech sestav ve sloupcích [213](#)  
více sloupců sestavy [214](#)  
Student-Newman-Keuls  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)  
Studentizovaná rezidua  
v lineární regresi [142](#)  
Studentův t test [109](#)  
studie s odpovídajícími dvojicemi  
v paired-test vzorků T [111](#)  
Svařová statistika  
v jednofaktorové ANOVA [117](#)

## Š

šikmost  
in Means [97](#)

šikmost (*pokračování*)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v krychlích OLAP [99](#)  
v souhrnech sestav v řádcích [212](#)  
v souhrnu [95](#)  
ve Prozkoumat [88](#)  
ve Souhrnech sestav ve sloupcích [213](#)

špičatost  
in Means [97](#)  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v krychlích OLAP [99](#)  
v souhrnech sestav v řádcích [212](#)  
v souhrnu [95](#)  
ve Prozkoumat [88](#)  
ve Souhrnech sestav ve sloupcích [213](#)

Švestková  
v ordinální regresi [145](#)

## T

t test  
in Independent-Samples T Test [109](#)  
v jednovzorkového testu T [112](#)  
v paired-test vzorků T [111](#)

tabulka klasifikace  
v Nearest Sousední analýza [168](#)

tabulky kmenového a koncového listu  
ve Prozkoumat [88](#)

Tamhaneův T2  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

tau-b  
v kontingenčních tabulkách [91](#)

tau-c  
v kontingenčních tabulkách [91](#)

Test her a Howellův porovnávání po dvojicích  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

test chí-kvadrát  
Jedna-ukázka neparametrických testů [189](#), [190](#)

Test Kolmogorov-Smirnov  
Jedna-ukázka neparametrických testů [189](#), [190](#)  
Test Liliefors [190](#)

Test Liliefors [190](#), [200](#), [201](#)

Test na aditivitu Tukeyho  
v Analytické spolehlivosti [215](#), [216](#)

Test Scheffé  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

Testy dvou souvisejících ukázek  
chybějící hodnoty [204](#)  
příkaz additional features [205](#)  
statistika [204](#)  
typy testů [204](#)  
volby [204](#)

testy homogenity rozptylu  
v jednofaktorové ANOVA [117](#)

testy nezávislosti  
chí kvadrát [91](#)

testy normality  
ve Prozkoumat [88](#)

Testy pro několik nezávislých vzorků  
definování rozsahu [205](#)  
chybějící hodnoty [205](#)  
příkaz additional features [206](#)  
seskupování proměnných [205](#)  
statistika [205](#)  
typy testů [205](#)  
volby [205](#)

Testy pro několik souvisejících ukázek  
příkaz additional features [207](#)  
statistika [206](#)  
typy testů [206](#)

Testy Two-Independent-Samples  
definice skupin [203](#)  
chybějící hodnoty [203](#)  
příkaz additional features [203](#)  
seskupování proměnných [203](#)  
statistika [203](#)  
typy testů [203](#)  
volby [203](#)

tolerance  
v lineární regresi [143](#)

Tolerance  
Lineární regrese [144](#)

tornádo, grafy  
v simulaci [247](#)

Transformační matice  
Analýza faktoru [172](#)

Trychtýř [73](#)

Tukey je opravdu významný rozdíl  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

Tukeyův odhad váhy  
ve Prozkoumat [88](#)

Tukeyův test b  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)

Tvůrce simulace [238](#)

## U

Ukazatel Lance a Williams dispoďobnosti  
ve Vzdálenosti [133](#)

ukazatel rozdílu vzorků  
ve Vzdálenosti [133](#)

Ukazatele disperze  
v deskriptivních [85](#)  
v frekvencích [83](#)  
v měrových statistikách [223](#), [225](#)  
ve Prozkoumat [88](#)

ukázka školení  
v Nearest Sousední analýza [164](#)

Uložit dialogové okno  
Binární meta-analýza [60](#)  
Regrese meta-analýzy [72](#)  
Souvislá meta-analýza [44](#)  
Velikost binárního efektu metaanalýzy [68](#)  
Velikost souvislé analýzy meta-analýzy [51](#)

úplné faktoriální modely  
v GLM [119](#)

Upravená hodnota R2  
v lineární regresi [143](#)

Upravené Binomické střední hodnoty

Upravené Binomické střední hodnoty (*pokračování*)  
v paired-rozměry vzorků [105](#)  
Upravený R-čtverec  
v lineárních modelech [136](#)

## V

V  
v kontingenčních tabulkách [91](#)  
variační koeficient (COV)  
v měrových statistikách [223](#), [225](#)  
Vážené Kappy  
Podmínky [219](#)  
příklad [218](#)  
statistika [218](#)  
statistiky [219](#)  
tisk [220](#)  
zkřížžení [220](#)  
Vážené kappy na Cohen [218–220](#)  
vážené předpovězené hodnoty  
v GLM [126](#)  
vážený průměr  
v měrových statistikách [223](#), [225](#)  
váženými nejmenšími  
v lineární regresi [140](#)  
Velikost binárního efektu metaanalýzy  
Bublina-karta [75](#)  
Dialogové okno Analýza [63](#)  
Dialogové okno Bias [65](#)  
Dialogové okno Inference [64](#)  
Dialogové okno kontrastu [65](#)  
Dialogové okno Kritéria [62](#)  
Dialogové okno Oříznout a Vyplnit [66](#)  
Dialogové okno Tisk [67](#)  
Dialogové okno vykreslení [69](#)  
Karta " L' Abb'e Plot [78](#)  
Karta Výkres funklu [76](#)  
Karta Výkres lesa [73](#)  
Kumulativní karta Zákresový strom [74](#)  
Ouško Galbraitní [77](#)  
Uložit dialogové okno [68](#)  
Velikost souvislé analýzy meta-analýzy  
Bublina-karta [75](#)  
Dialogové okno Analýza [46](#)  
Dialogové okno Bias [49](#)  
Dialogové okno Inference [47](#)  
Dialogové okno kontrastu [48](#)  
Dialogové okno Oříznout a Vyplnit [49](#)  
Dialogové okno Tisk [51](#)  
Dialogové okno vykreslení [52](#)  
Karta " L' Abb'e Plot [78](#)  
Karta Výkres funklu [76](#)  
Karta Výkres lesa [73](#)  
Kumulativní karta Zákresový strom [74](#)  
Ouško Galbraitní [77](#)  
Uložit dialogové okno [51](#)  
Více odpovědí  
příkaz additional features [210](#)  
více porovnání  
v jednofaktorové ANOVA [115](#)  
více R  
v lineární regresi [143](#)  
více sad odpovědí  
Kódová kniha [79](#)

vícenásobná regrese  
v lineární regresi [140](#)  
vizualizace  
modely klastrování [180](#)  
vlastní modely  
v GLM [119](#)  
vliv velikosti  
in Independent-Samples T Test  
[109](#)  
V Před-T-test vzorků [111](#)  
vodorovné pruhové grafy  
v frekvencích [84](#)  
vrstvy  
v kontingenčních tabulkách [90](#)  
Výběr funkcí  
v Nearest Sousední analýza [168](#)  
Výkres lesa [73](#)  
výpis případů [94](#)  
Vyplnit na stránku  
v lineární regresi [142](#)  
výrazy interakce [119](#), [147](#), [148](#)  
výsečové grafy  
v frekvencích [84](#)  
vzdálenost bloku  
ve Vzdálenosti [133](#)  
Vzdálenosti  
dispodobnost ukazatele [133](#)  
opatření podobnosti [133](#)  
příkaz additional features [134](#)  
příklad [132](#)  
statistika [132](#)  
transformace hodnot [133](#)  
transformace ukazatelů [133](#)  
výpočet vzdáleností mezi proměnnými [132](#)  
výpočet vzdálenosti mezi případy [132](#)  
Vzorek pro odchvatu  
v Nearest Sousední analýza [164](#)

## W

Wald  
in Independent-Samples Proportions  
[108](#)  
v jednovzorkových porcích [102](#)  
v paired-rozměry vzorků [105](#)  
Wald (korigovaná kontinuita)  
in Independent-Samples Proportions  
[108](#)  
v jednovzorkových porcích [102](#)  
v paired-rozměry vzorků [105](#)  
Wald (Vodivost nápravného zařízení)  
in Independent-Samples Proportions  
[108](#)  
v paired-rozměry vzorků [105](#)  
Wald H0  
in Independent-Samples Proportions  
[108](#)  
Wald H0 (Kontinuita opravena)  
in Independent-Samples Proportions  
[108](#)  
Wald-Wolfowitz běží  
v testech Two-Independent-Samples  
[203](#)  
Waller-Duncan t test



Waller-Duncan t test (*pokračování*)  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)  
Wilks " lambda  
v analýze diskriminant [171](#)  
Wilsonovo skóre (opraveno kontinuity)  
v jednovzorkových porcích [102](#)

## Y

Yates ' korekce kontinuity  
v kontingenčních tabulkách [91](#)

## Z

z skóre  
uložení jako proměnné [84](#)  
v deskriptivních [84](#)  
Zákresový  
Regrese meta-analýzy [75](#)  
zaliční grafy  
v Hierarchické analýze klastru [185](#)  
Zátěž  
ve vícerozměrném škálování [220](#)  
zbytkové chyby  
uložení v lineární regresí [142](#)  
uložení v odhadu křivky [159](#)  
v kontingenčních tabulkách [92](#)  
zesílení  
v lineárních modelech [134](#)  
Zkouška Levenem  
v jednofaktorové ANOVA [117](#)  
ve Prozkoumat [88](#)  
zkouška paralelních lovných šňůr  
v ordinální regresí [146](#)  
Zkouška porovnání Gabriela po dvojicích  
v GLM [123](#)  
v jednofaktorové ANOVA [115](#)  
Zkouška skříní skříní  
v analýze diskriminant [170](#)  
zkoušky linearity  
in Means [97](#)  
zkřížžení  
v kontingenčních tabulkách [89](#)  
více odpovědí [209](#)  
Zobrazení modelu  
v Nearest Sousední analýza [166](#)  
zpětná eliminace  
v lineární regresí [141](#)  
zpracování hluku  
v analýze TwoStep Cluster [178](#)





