

*IBM SPSS Regression 29*

**IBM**

**Poznámka**

Před použitím těchto informací a produktu, který podporují, si přečtěte informace v tématu [“Upozornění” na stránce 49](#).

**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. Regrese.....</b>	<b>1</b>
Výběr procedury pro binární logistickou regresi.....	1
Logistická regrese.....	1
Pravidlo pro Logistické regresní sady.....	3
Metody pro výběr proměnných regresní regrese.....	3
Logistická regrese s definováním kategorických proměnných.....	3
Logistický regresní operace Uložit nové proměnné.....	4
Logistické volby regrese.....	5
Další funkce příkazu LOGISTIC REGRESSION.....	5
Multinomická Logistická regrese.....	5
Multinomická Logistická regrese.....	6
Multinomická Logistic Regression Reference Category.....	7
Multinomická Logistická regresní statistika.....	7
Multinomický Logistický regresní kritéria.....	8
Multinomická Logistic Regresní Volby.....	8
Multinomický Logistic Regression Uložit.....	9
Další funkce příkazu NOMREG.....	9
Probit regrese.....	9
Rozsah analýzy probitového analýzy.....	10
Volby analýzy Probit.....	10
Další funkce příkazu PROBIT.....	11
Regrese kvantilů.....	11
Měření kvantile: Kritéria.....	12
Množstevní regrese: Model.....	13
Quantile Regression: Zobrazit.....	14
Množstevní regrese: Uložit.....	16
Množstevní regrese: Export.....	16
Nelineární regrese.....	17
Podmíněná logika (nelineární regrese).....	18
Nelineární parametry regrese.....	18
Nelineární regresní běžné modely.....	19
Funkce nelineárních regresních ztrát.....	19
Omezení nelineárních regresních parametrů.....	20
Nelineární regresní operace Uložit nové proměnné.....	20
Nelineární regresní volby.....	20
Interpretace nelineárních regresních výsledků.....	20
Další funkce příkazu NLR.....	21
Odhad váhy.....	21
Volby odhadu váhy.....	22
Další funkce příkazu WLS.....	22
Regrese dvoufázového nejmenších čtverců.....	23
Dvoufázové volby regrese nejmenších čtverců.....	24
Doplňující funkce příkazu 2SLS.....	24
Kategorické schéma kódování proměnných.....	24
Odchylka.....	24
Jednoduchá.....	24
Helmert.....	25
Rozdíl.....	25
Polynomní.....	25
Opakované.....	26
Jiný.....	26

Indikátor.....	27
Regrese jádra jádra.....	27
parametry jádra.....	29
Regrese jádra jádra: Volby.....	29
Modely s parametrickým akceorickým časem.....	30
Parametrický akcelerovaný model času selhání: Kritéria.....	31
Modely s parametrickým akceorickým časem: Model.....	31
Modely s parametrickým akceorickým časem selhání: Odhad.....	32
Parametrické akcelerované modely času selhání: Tisk.....	33
Parametrické akcelerované modely času selhání: Predict.....	33
Modely s parametrickým akceorickým selháním: Plot.....	34
Parametrické akcelerované modely času selhání: Export.....	35
Přežití AFT definování událostí pro proměnné stavu.....	35
Parametrické akcelerované modely času selhání: Vyberte kategorii.....	35
Parametrické sdílené chatrné modely.....	36
Parametrické sdílené chatrné modely: Kritéria.....	37
Parametrické sdílené chatrné modely: model.....	38
Parametrické sdílené chatrné modely: Odhad.....	38
Parametrické sdílené chatrné modely: Tisk.....	40
Parametrické sdílené chatrné modely: Predict.....	40
Parametrické sdílené chatrné modely: Zakreslit.....	41
Parametrické sdílené chatrné modely: Export.....	42
Parametrické sdílené chatrné modely: Definování událostí.....	42
Parametrické sdílené chatrné modely-Příklady.....	43
Parametrické sdílené chatrné modely-Případová studie pro opakovací data.....	45
<b>Upozornění.....</b>	<b>49</b>
Ochranné známky.....	50
<b>Rejstřík.....</b>	<b>53</b>

---

# Kapitola 1. Regrese

Následující funkce regrese jsou zahrnuty v části SPSS Statistics Standard Edition nebo Regresní volba.

## Výběr procedury pro binární logistickou regresi

---

Binární logistické regresní modely lze instalovat pomocí Logistic regresních postupů a Multinomial Logistic Regression procedure. Každý postup má možnosti, které nejsou dostupné v druhé. Důležitou teoretickou odlišností je to, že Logistic Regression provádí všechny předpovědi, zbytkové chyby, vliv statistiky a dobroty-testů pomocí dat na úrovni jednotlivých případů, bez ohledu na to, jak jsou data zadávána a zda je počet proměnných kovariátů menší než celkový počet případů, zatímco Multinomický Logistický regresní procedura interně agreguje případy k vytvoření subpopulací s identickými kovariátovými vzory pro prediktory, produkující předpovědi, zbytkové chyby a dobroty proložení testů založených na těchto subpopulacích. Jsou-li všechny prediktory kategorické nebo všechny spojité prediktory mají pouze omezený počet hodnot-takže existuje několik případů na každém odlišném kovariátu-přístup subpopulace může produkovat platné testy dobrotivosti a informativních zbytkových chyb, zatímco individuální přístup na úrovni případu nemůže.

### Logistická regrese

Poskytuje následující jedinečné funkce:

- Hommer-Lemeshow test dobroty pro model
- Postupné analýzy
- Kontrastky, které definují parametrizaci modelu
- Alternativní mezní body pro klasifikaci
- Klasifikační grafy
- Model osazen na jednu sadu případů k zadrhovací sadě případů
- Uloží předpovědi, zbytkové chyby a ovlivňují statistiky.

### Multinomická Logistická regrese

Poskytuje následující jedinečné funkce:

- Pearson a deviance chi-kvadrát testy pro dobroty fit na model
- Specifikace subpopulací pro seskupování údajů pro zkoušky na dobré kondicích
- Výpis počtů, předpokládaných počtů a zbytkových chyb podle dílčích populací
- Úprava rozptylu rozptylu pro více než disperzi
- Matice rozptylu odhadů parametrů
- Zkoušky lineárních kombinací parametrů
- Explicitní specifikace vnořených modelů
- Fit 1-1 odpovídá podmíněným modelům pro logistickou regresi s použitím různých proměnných

### Poznámky:

- Oba tyto procedury odpovídají modelu pro binární data, která jsou generalizovaným lineárním modelem s binomickou distribucí a linkovou funkcí. Je-li pro vaše data vhodnější jiná spojivá funkce, měli byste použít proceduru Generalized Linear Models.
- Pokud jste opakovali měření binárních dat nebo záznamů, které jsou jinak souvztažné, měli byste zvážit procedury Generalized Linear Mixed Models nebo Generalized Estimation Equations Equations.

---

## Logistická regrese

Logistická regrese je užitečná pro situace, ve kterých chcete být schopni předpovídat přítomnost nebo nepřítomnost charakteristiky nebo výsledku na základě hodnot sady proměnných prediktorů. Je

podobný modelu lineární regrese, ale je vhodný pro modely, kde závislá proměnná je dichotomová. Logistické regresní koeficienty lze použít k odhadu poměru pravděpodobnosti pro každou z nezávislých proměnných v modelu. Logistická regrese se vztahuje na širší spektrum výzkumných situací než při analýze diskriminačních.

**Příklad.** Jaké charakteristiky životního stylu jsou rizikové faktory pro ischemickou chorobu srdce (CHD)? Vzhledem k vzorku pacientů změřených ke stavu kouření, dietě, cvičení, užívání alkoholu a statusu CHD můžete vytvořit model pomocí čtyř proměnných životního stylu, které předpovídají přítomnost nebo nepřítomnost CHD u vzorku pacientů. Model lze poté použít k odvození odhadů poměrů pravděpodobnosti pro každý faktor, který vám například řekne, jak daleko pravděpodobnější kuřáci mají vyvinout CHD než nekuřáci.

**Statistika.** Pro každou analýzu: celkové případy, vybrané případy, platné případy. Pro každou kategorickou proměnnou: kódování parametru. Pro každý krok: proměnné (y) zadané nebo odebrané, historie iterace, -2 protokolu-pravděpodobnost, dobrota přizpůsobení, Hommer-Lemhow dobroty-proložení, model chí kvadrát, zlepšení chí kvadrát, klasifikační tabulka, korelace mezi proměnnými, sledované skupiny a předpovězené pravděpodobnosti grafu, reziduální chí kvadrát. Pro každou proměnnou v rovnici: koeficient ( $B$ ), standardní chyba  $B$ , statistika Wald, odhadovaný poměr šancí ( $\exp(B)$ ), interval spolehlivosti pro  $\exp(B)$ , protokol pravděpodobnosti, pokud je výraz odebrán z modelu. Pro každou proměnnou, která není v rovnici: skóre statistiky. Pro každý případ: pozorovaná skupina, předpokládaná pravděpodobnost, předpokládaná skupina, zbytková, standardizovaná reziduum.

**Metody.** Modely můžete odhadnout pomocí položek bloku proměnných nebo libovolné z následujících krokových metod: dopředu podmíněnou, vpřed LR, dopředu Wald, zpětné podmíněné, zpětné LR nebo vzad Wald.

## Informace o Logistickém regresních datech

**Data.** Závislá proměnná by měla být dichotomová. Nezávislé proměnné mohou být buď úrovně intervalu nebo kategorické; jsou-li kategorické, měly by být fiktivní nebo indikátory kódované (v proceduře se dá automaticky odkódovat kategoriální proměnné automaticky).

**Předpoklady.** Logistická regrese se neopírá o distribuční předpoklady v tom samém smyslu, že analýza diskriminant dělá. Vaše řešení však může být stabilnější, pokud prediktory mají normální rozdělení. Navíc, podobně jako u jiných forem regrese může multilinéarita mezi prediktory vést k předpojatým odhadům a nafukovací standardní chyby. Postup je neefektivnější, když členství ve skupině je skutečně kategorickou proměnnou; je-li členství ve skupině založeno na hodnotách souvislé proměnné (například "vysoké IQ" versus "nízké IQ"), měli byste zvážit použití lineární regrese, abyste využili výhod bohatších informací nabízených nepřetržitou proměnnou samotnou.

**Související postupy.** Použijte proceduru Scatterplot k zobrazení dat pro vícelinearitu. Jsou-li splněny předpoklady vícehodnotového normality a kovariance-kovarianční matice, můžete být schopni získat rychlejší řešení pomocí procedury analýzy diskriminací. Pokud jsou všechny vaše prediktorové proměnné kategorické, můžete také použít lineární proceduru. Je-li vaše závislá proměnná souvislá, použijte lineární regresní proceduru. Můžete použít proceduru ROC Curve k vynesení pravděpodobností uložených pomocí procedury Logistic Regression.

## Získání Logistické regresní analýzy

1. Z nabídky vyberte:

**Analyzovat > Regrese > Binární Logistic ...**

2. Vyberte jednu dichotomickou závislou proměnnou. Tato proměnná může být číselná nebo řetězcová.

3. Vyberte jednu nebo více kovariány. Chcete-li zahrnout podmínky interakce, vyberte všechny proměnné zahrnuté do interakce a poté vyberte **> a\*b>**.

Chcete-li zadat proměnné do skupin (**blocks**), vyberte proměnné pro blok a klepněte na tlačítko **Další** a zadejte nový blok. Opakujte, dokud nebudou zadány všechny bloky.

Volitelně můžete vybrat případy pro analýzu. Zvolte proměnnou výběru a zadejte kritéria pravidla.

## Pravidlo pro Logistické regresní sady

Případy definované pravidlem výběru jsou zahrnuty do odhadu modelu. Například, pokud jste vybrali proměnnou a **rovná se** a uvedli jste hodnotu 5, pak pouze ty případy, pro které má zvolená proměnná hodnotu 5, jsou zahrnuty do odhadu modelu.

Výsledky statistiky a klasifikace jsou generovány pro vybrané i nevybrané případy. Poskytuje mechanismus pro klasifikaci nových případů na základě dříve existujících dat nebo pro rozdělení dat na logické části na části školení a testování dílčích sad, aby bylo možné provést ověření platnosti na generovaném modelu.

## Metody pro výběr proměnných regresní regrese

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.

- **Zadat.** Procedura pro výběr proměnné, v níž jsou všechny proměnné v bloku zadány v jednom kroku.
- **Postoupit výběr (podmíněný).** Kroková metoda výběru s testováním položek na základě významnosti statistiky se skóre a testování odebrání založené na pravděpodobnosti statistiky pravděpodobnosti založené na odhadech podmíněného parametru.
- **Postoupit výběr (poměr pravděpodobnosti).** Kroková metoda výběru s testováním položek na základě významnosti statistiky skóre a testování odebrání založené na pravděpodobnosti statistiky pravděpodobnosti založené na odhadech maximální dílčí pravděpodobnosti.
- **Postoupit výběr (Wald).** Kroková metoda výběru se testováním položek na základě významnosti statistiky skóre a testování odebrání založené na pravděpodobnosti statistiky Wald.
- **Zpětné odstranění (podmíněné).** Výběr zpětného kroku zpět. Testování odebrání je založeno na pravděpodobnosti statistiky pravděpodobnosti založené na odhadech podmíněného parametru.
- **Zpětná eliminace (poměr pravděpodobnosti).** Výběr zpětného kroku zpět. Testování odebrání je založeno na pravděpodobnosti statistiky pravděpodobnosti založené na odhadech maximální dílčí pravděpodobnosti.
- **Zpětné odstranění (Wald).** Výběr zpětného kroku zpět. Testování odebrání je založeno na pravděpodobnosti statistiky Wald.

Hodnoty významnosti ve vašem výstupu jsou založeny na montáži jednoho modelu. Proto jsou hodnoty významnosti obecně neplatné, když se použije metoda po krocích.

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**.

## Logistická regrese s definováním kategorických proměnných

Můžete uvést podrobnosti o tom, jak bude Logistický regresní procedura zpracovávat kategoriální proměnné:

**Kovariáty.** Obsahuje seznam všech proměnných, které jsou uvedeny v hlavním dialogovém okně, buď samostatně, nebo jako součást interakce, v libovolné vrstvě. Pokud jsou některé z těchto řetězcových proměnných nebo jsou kategorické, můžete je použít pouze jako kategoriální kovarianty.

**Kategorické proměnné Cvariety.** Zobrazí seznam proměnných identifikovaných jako kategoriální. Každá proměnná zahrnuje notaci v závorkách označující kontrastní kódování, které se má použít. Řetězcové proměnné (označené symbolem < za jejich názvy) jsou již přítomny v seznamu Categorical Covariates. Vyberte libovolné jiné kategoriální proměnné ze seznamu Cvariates a přesuňte je do seznamu Kategorieckých proměnných.

**Změnit kontrast.** Umožňuje změnit kontrastní metodu. Dostupné kontrastní metody jsou:

- **Indikátor.** Kontrastní látky označují přítomnost nebo nepřítomnost členství v kategorii. Referenční kategorie je znázorněna v matici kontrastu jako řádek nul.

- **Jednoduchý.** Každá kategorie predikční proměnné (kromě referenční kategorie) se porovnává s referenční kategorií.
- **Rozdíl.** Každá kategorie proměnné prediktoru s výjimkou první kategorie je porovnána s průměrným účinkem předchozích kategorií. Známe také jako obrácené Helmmert kontrasts.
- **Helmert.** Každá kategorie proměnné prediktoru s výjimkou poslední kategorie je porovnána s průměrným efektem následných kategorií.
- **Opakováno.** Každá kategorie proměnné prediktoru (s výjimkou poslední kategorie) se porovná s další kategorií.
- **Polynomiální.** Ortogonální polynom kontrasty. Předpokládá se, že kategorie jsou rovnoměrně rozloženy. Polynomiální kontrasty jsou k dispozici pouze pro číselné proměnné.
- **Odchylka.** Každá kategorie proměnné prediktoru s výjimkou referenční kategorie je porovnána s celkovým výsledkem.

Vyberete-li volbu **Odchylka, Jednoduchá** nebo **Indikátor**, vyberte jako referenční kategorii buď hodnotu **První**, nebo **Poslední**. Všimněte si, že metoda se ve skutečnosti nezmění, dokud neklepnete na tlačítko **Změnit**.

Řetězcové proměnné musí být kategorické proměnné. Chcete-li odebrat řetězcovou proměnnou ze seznamu Categorical Covariates, musíte odebrat všechny výrazy obsahující proměnnou ze seznamu Cvariates v hlavním dialogovém okně.

## Logistický regresní operace Uložit nové proměnné

Výsledky logistické regrese můžete uložit jako nové proměnné v aktivní datové sadě:

**Předpovězené hodnoty.** Uloží hodnoty předpovídané modelem. Dostupné možnosti jsou Pravděpodobnosti a členství ve skupinách.

- *Pravděpodobnosti.* Pro každý případ se ukládá očekávaná pravděpodobnost výskytu události. V tabulce ve výstupu se zobrazí název a obsah všech nových proměnných. Událost "event" je kategorie závislé proměnné s vyšší hodnotou; například pokud závislá proměnná nabývá hodnot 0 a 1, je předpovězená pravděpodobnost uložení kategorie 1 uložena.
- *Předpovězené členství ve skupinách.* Skupina s největší hodnotou posterior založenou na skóre diskriminačních skóre. Skupina, do níž model předpovídá, kam daný případ patří.

**Ovlivnění.** Uloží hodnoty ze statistik, které měří vliv případů na předpovězenou hodnotu. Dostupné možnosti jsou Cook's, Fevers values a DfBeta (s).

- *Cookovy.* Logistické regresní analogový analogový Cook vliv statistiky. Ukazatel toho, jak dlouho se změní zbytkové chyby všech případů, pokud byl konkrétní případ vyloučen z výpočtu regresních koeficientů.
- *Hodnota pákového efektu.* Relativní vliv každého pozorování na přizpůsobení modelu.
- *DfBetasy.* Rozdíl v hodnotě beta je změna regresního koeficientu, která vyplývá z vyloučení konkrétního případu. Hodnota se vypočítává pro každý termín v modelu, včetně konstanty.

**Residuals.** Uloží zbytkové chyby. Dostupné volby jsou nestandardizované, Logit, Studentized, Standardized a Deviance.

- *Nestandardizované reziduální reziduuum.* Rozdíl mezi pozorovanou hodnotou a hodnotou předpovězenou modelem.
- *Logit Reziduální.* Zbytková hodnota pro případ, je-li předpovězena v měřítku logit. Protokol zbytková je zbytková hodnota rozdělená podle předpokládané doby pravděpodobnosti 1 minus očekávaná pravděpodobnost.
- *Stentizované reziduuum.* Změna odchylky modelu, je-li případ vyloučen.
- *Standardizované reziduální reziduuum.* Zbytek vydělený odhadem jeho směrodatné odchylky. Standardizované zbytkové chyby, které jsou také známe jako Pearsonové zbytkové chyby, mají střední hodnotu 0 a směrodatnou odchylku 1.
- *Deviance.* Residuals based on the model deviance.



**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.

## Logistické volby regrese

Můžete určit volby pro analýzu logistické regrese:

**Statistika a položky lots.** Umožňuje vám vyžádat statistiky a grafy. Dostupné volby jsou Classification plots, Hosmer-Lemeshow goodness-of-fit, Casewise listing of residuals, Correlations of estimates, Iteration history, and CI for exp ( $B$ ). Vyberte jednu z alternativ ve skupině Zobrazit pro zobrazení statistiky a zakreslených grafů buď na každém kroku, nebo pouze pro poslední model, v posledním kroku.

- *Hosmer-Lemeshow goodness-of-fit statistic.* Tento dobrotivý statistický údaj je robustnější než tradiční dobrotivá vhodná pro logistickou regresi, zvláště pro modely s kontinuální kovariáty a studie s malými velikostmi vzorku. Je založen na seskupování případů do snížení rizika a porovnání pozorovaných pravděpodobností s očekávanou pravděpodobností v každém decilu.

**Pravděpodobnost pro Stepwise.** Umožňuje vám řídit kritéria, podle kterých jsou proměnné zadávány a odebírány z rovnice. Můžete určit kritéria pro položku nebo odebrání proměnných.

- *Pravděpodobnost krokování.* Proměnná se zadává do modelu, pokud je pravděpodobnost její statistiky skóre menší než hodnota vstupní hodnoty a je odebrána, pokud je pravděpodobnost větší než hodnota Odebrání. Chcete-li přepsat výchozí nastavení, zadejte kladné hodnoty pro položku a Odebrání. Položka musí být menší než odebrání.

**Uzavření klasifikace.** Umožňuje určit bod řezu pro klasifikaci případů. Případy s předpovězenými hodnotami, které překračují uzavření klasifikace, jsou klasifikovány jako kladné, zatímco ty s předpovězenými hodnotami nižšími, než je uzavření objektu, jsou klasifikovány jako záporné. Chcete-li změnit výchozí nastavení, zadejte hodnotu mezi 0.01 a 0.99.

**Maximum iterací.** Umožňuje vám změnit maximální počet iterací, které model iteruje před ukončením.

**Zahrnout do modelu konstantu.** Umožňuje označit, zda má model obsahovat konstantní termín. Je-li tato volba zakázána, bude konstantní výraz roven 0.

## Další funkce příkazu LOGISTIC REGRESSION

Jazyk syntaxe příkazu vám také umožňuje:

- Identifikovat výstup s hodnotami podle hodnot nebo proměnných popisků proměnné.
- Řízení rozteče mezi iteracemi iterace. Místo tisku odhadů parametrů po každé iteraci můžete požádat o odhad parametrů po každé  $n$ . iteraci.
- Změňte kritéria pro ukončení iterace a kontrolu CRC.
- Zadejte seznam proměnných pro výpisy z případu.
- Skonává paměť tím, že zadržuje data pro každou skupinu rozdělených souborů v externím pracovním souboru během zpracování.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Multinomická Logistická regrese

Multinomická Logistická regrese je užitečná pro situace, ve kterých chcete být schopni klasifikovat předměty na základě hodnot sady proměnných prediktoru. Tento typ regrese je podobný logistickým regresi, ale je obecnější, protože závislá proměnná není omezena na dvě kategorie.

**Příklad.** Pro efektivnější tržní filmy chtějí filmová studia předpovídat, jaký typ filmu se bude pravděpodobně vidět filmový moviegoer. Prostřednictvím Multinomického logistické regrese může studio určit sílu vlivu na věk, pohlaví a seznamovací status člověka na typ filmu, který preferují. Studio může pak posluhovat reklamní kampaň určitého filmu směrem ke skupině lidí, kteří se pravděpodobně budou vidět.

**Statistika.** Historie iterací, koeficienty parametru, asymptotické kovariance a korelační matice, testy pravděpodobnosti poměrů pro model a dílčí účinky, -2 protokolu-pravděpodobnost. Pearson a deviance chi-square dobroty fit. Cox a Snell, Nagelkerke, a McFadden  $R^2$ . Klasifikace: pozorované versus předpokládané frekvence podle kategorie odpovědí. Sražení: pozorované a předpovězené frekvence (se zbytkových chyb) a poroporce podle kovariátové šablony a kategorie odezvy.

**Metody.** Model mnohomoření je vhodný pro celý faktoriální model nebo pro uživatelem specifikovaný model. Odhad parametru se provádí pomocí iterativního algoritmu maximální pravděpodobnosti.

## Aspekty multimatických Logistických regresních dat

**Data.** Závislá proměnná by měla být kategorická. Nezávislé proměnné mohou být faktory nebo kovariáty. Obecně platí, že faktory by měly být kategoriálními proměnnými a proměnné by měly být spojitými proměnnými.

**Předpoklady.** Předpokládá se, že poměr šancí každé ze dvou kategorií je nezávislý na všech ostatních kategoriích odpovědí. Je-li například na trh zaveden nový výrobek, tento předpoklad stanoví, že tržní podíly všech ostatních výrobků jsou úměrně rovným dílem ovlivněny. Také vzhledem k proměnné kovariát se předpokládá, že odpovědi jsou nezávislé mnohoromické proměnné.

## Získání multinomické logistické regrese

1. Z nabídky vyberte:

**Analyzovat > Regrese > Multinomial Logistic Regression ...**

2. Vyberte jednu závislou proměnnou.

3. Faktory jsou volitelné a mohou být buď číselné, nebo kategorické.

4. CVariates jsou volitelné, ale musí být numerické, pokud jsou uvedeny.

## Multinomická Logistická regrese

Standardně procedura Multinomial Logistic Regression produkuje model s faktorem a kovariuje hlavní efekty, ale můžete zadat vlastní model nebo si vyžádat postupný výběr modelu s tímto dialogovým oknem.

**Určete model.** Model s hlavními účinky obsahuje kovariantní a hlavní účinky, ale bez interakčních efektů. Úplný faktoriální model obsahuje všechny hlavní účinky a všechny interakce mezi faktory a faktory. Neobsahuje interakce covariate. Můžete vytvořit vlastní model pro uvedení dílčích sad interakcí faktoru nebo kovariát interakcí, nebo požadujte postupný výběr modelových výrazů.

**Faktory & proměnné.** Jsou vypsány faktory a kovariány.

**Vynucené vstupní podmínky.** Podmínky přidané do seznamu vynuceného záznamu jsou vždy obsaženy v modelu.

**Krokové podmínky.** Podmínky přidané do postupného seznamu jsou zahrnuty v modelu podle jednoho z následujících uživatelem vybraných metod stemwise:

- **Předat položku.** Tato metoda začíná bez nekrokových termínů v modelu. V každém kroku se do modelu přidá nejvýznamnější termín, dokud se do modelu nepřidají žádné z nevýrazových termínů z modelu, které by byly statisticky významným přínosem.
- **Zpětné odstranění.** Tato metoda začíná zadáním všech termínů uvedených ve stepném seznamu do modelu. V každém kroku je z modelu odebrán nejméně významný krokový výraz, dokud všechny zbývající nepostupné podmínky nemají statisticky významný příspěvek na daný model.
- **Předat krokové.** Tato metoda začíná modelem, který by měl být vybrán metodou dopředného zadávání. Od té doby se tento algoritmus střídá mezi zpětným odstraňováním po krocích v modelu a forwardové položky za podmínek, které vynechal z modelu. To pokračuje, dokud žádné podmínky nespĺní kritéria zadání nebo odebrání.
- **Zpětná kroková.** Tato metoda začíná modelem, který by byl vybrán metodou zpětné eliminace. Od té doby se algoritmus mezi forwardovým záznamem zaměřenými na termíny vynechte z modelu a zpětné

je eliminován po krocích v rámci modelu. To pokračuje, dokud žádné podmínky nesplní kritéria zadání nebo odebrání.

**Zahrnout zachycení do modelu.** Umožňuje zahrnout nebo vyloučit výraz zachycení pro model.

## Podmínky sestavení

Pro vybrané faktory a kovariáty:

**Interakce.** Vytvoří interakční období všech vybraných proměnných na úrovni nejvyšší úrovně.

**Hlavní účinky.** Vytvoří výraz main-effects pro každou vybranou proměnnou.

**Všechny 2-way.** Vytvoří všechny možné dvousměrné interakce mezi vybranými proměnnými.

**Vše 3-way.** Vytvoří všechny možné tři interakce mezi vybranými proměnnými.

**Všechny 4-way.** Vytvoří všechny možné čtyřcestné interakce mezi vybranými proměnnými.

**Všechny 5-way.** Vytvoří všechny možné vzájemné interakce mezi vybranými proměnnými.

## Multinomial Logistic Regression Reference Category

Ve výchozím nastavení je u kategorie odkazů v Multinomial Logistickém regresí poslední kategorie kategorie. Toto dialogové okno vám umožňuje řídit odkazové kategorie a způsob, jakým jsou kategorie seřazeny.

**Referenční kategorie.** Určete první, poslední nebo vlastní kategorii.

**Kategorie Order.** Ve vzestupném pořadí definuje nejnižší hodnota první kategorii a nejvyšší hodnota definuje poslední. V sestupném pořadí nejvyšší hodnota definuje první kategorii a nejnižší hodnota definuje poslední.

## Multinomial Logistic Regression Statistics

Můžete určit následující statistiky pro Multinomial Logistic Regression:

**Souhrn zpracování případů.** Tato tabulka obsahuje informace o uvedených kategoriálních proměnných.

**Model.** Statistiky pro celkový model.

- **Pseudokvadrát.** Vytiskne statistiky Cox a Snell, Nagelkerke a McFadden  $R^2$ .
- **Souhrn kroků.** Tato tabulka shrnuje účinky zadané nebo odebrané v každém kroku v metodě krokové metody. Neprodukuje se, dokud není v dialogovém okně Model zadán nepostupný model.
- **Informace o montáži modelu.** Tato tabulka porovnává instalované a zachycovací modely nebo modely s hodnotou null.
- **Kritéria informací.** Tato tabulka tiskne informační kritérium Akake (AIC) a společnosti Schwarz z Bayesova informačního kritéria (BIC).
- **Součet pravděpodobností buněk.** Tiskne tabulku pozorovaných a očekávaných frekvencí (se zbytkovou frekvencí) a proporcemi podle kategorie vzorků a odezev kovariate.
- **Tabulka klasifikace.** Tiskne tabulku pozorovaných versus předpokládaných odpovědí.
- **Goodness of fit chi-square statistics.** Vypíše statistické údaje o Pearsonové a pravděpodobnosti-kvadrantní statistiky. Statistika se vypočítává pro vzorce kovariate určené všemi faktory a kovariáty nebo uživatelem definovanou podmnožinou faktorů a kovariany.
- **Monotónická opatření.** Zobrazí tabulku s informacemi o počtu shodných dvojic, dekordant dvojic a vázaných dvojic. V této tabulce jsou také zobrazeny Somers' D, Goodman a Kruskal's Gamma, Kendall's tau-a a Concordance Index C.

**Parametry.** Statistiky související s parametry modelu.

- **Odhady.** Vypíše odhady parametrů modelu s uživatelem určenou úrovní důvěry.

- **Test poměru pravděpodobnosti.** Vypíše testy pravděpodobnosti poměru pravděpodobnosti k částečným účinkům modelu. Test celkového modelu se vytiskne automaticky.
- **Asymptotické korelace.** Vypíše matici korelací odhadu parametrů.
- **Asymptotické kovariance.** Tiskne matici parametrů odhadu kovariance.

**Definujte podskupiny.** Umožňuje vybrat podmnožinu faktorů a kovariátů za účelem definování vzorků kovariátů používaných pravděpodobností buněk a testů dobrotivosti.

## Multinomický Logistický regresní kritéria

Můžete uvést následující kritéria pro Multinomická Logistic Regression:

**Iterace.** Umožňuje vám uvést maximální počet opakování cyklu pomocí algoritmu, maximálního počtu kroků v kroku-halving, tolerance konvergence pro změny v protokolu-pravděpodobnosti a parametry, jak často se bude tisknout průběh iterativního algoritmu, a v jaké iteraci by měl procedura zahájit kontrolu úplného nebo kvaziúplného oddělení dat.

- **Sbližování pravděpodobnosti protokolu.** Předpokládá se konvergence, pokud je absolutní změna ve funkci protokolu pravděpodobnosti menší než uvedená hodnota. Kritérium se nepoužije, je-li hodnota 0. Uvedte nezápornou hodnotu.
- **Konvergenční parametr.** Konvergence se předpokládá, pokud je absolutní změna v odhadu parametrů menší než tato hodnota. Kritérium se nepoužije, je-li hodnota 0.

**Rozdílová data.** Umožňuje vám uvést nezápornou hodnotu menší než 1. Tato hodnota se přidá do každé prázdné buňky křížové konfigurace kategorie odezvy podle vzorce kovariate. To pomáhá stabilizovat algoritmus a zabránit předpojatosti v odhadech.

**Tolerance tolerance.** Umožňuje vám uvést toleranci použitou při kontrole jedinečnosti.

## Multinomická Logistic Regresní Volby

Můžete zadat následující volby pro Multinomická Logistic Regression:

**Měřítka rozptylu.** Umožňuje vám uvést hodnotu měřítka rozptylu, která se použije pro opravu odhadu matice kovariance parametru. **Deviance** odhaduje hodnotu měřítka pomocí statistiky funkce deviance (pravděpodobnosti-ratio chí-square). **Pearson** odhaduje hodnotu měřítka pomocí statistiky Pearson chí-square. Můžete také zadat vlastní hodnotu měřítka. Musí se jednat o kladnou číselnou hodnotu.

**Volby Stepwise.** Tyto volby vám poskytují řízení statistických kritérií, když se k sestavení modelu používají metody krokové metody. Ignorují se, pokud není v dialogovém okně Model zadán nepostupný model.

- **Pravděpodobnost záznamu.** Toto je pravděpodobnost statistiky pravděpodobnosti-poměr pravděpodobnosti pro záznam proměnné. Čím větší je uvedená pravděpodobnost, tím jednodušší je, že proměnná vstoupí do modelu. Toto kritérium je ignorováno, pokud není zvolena dopředná položka, dopředu kroková metoda nebo zpětná kroková metoda.
- **Test položky.** Toto je metoda pro zadání výrazů v metodách krokových metod. Vyberte si mezi zkouškou pravděpodobnosti a skóre testu. Toto kritérium je ignorováno, pokud není zvolena dopředná položka, dopředu kroková metoda nebo zpětná kroková metoda.
- **Pravděpodobnost odebrání.** Toto je pravděpodobnost statistického poměru pravděpodobnosti k odebrání proměnné. Čím větší je uvedená pravděpodobnost, tím jednodušší je, že proměnná zůstane v modelu. Toto kritérium je ignorováno, pokud není vybrána zpětná eliminace, dopředný krok vpřed, nebo zpětná kroková metoda.
- **Test odebrání.** Jedná se o metodu odebrání výrazů v metodách krokových metod. Vyberte si mezi testem pravděpodobnosti pravděpodobnosti a Wald testem. Toto kritérium je ignorováno, pokud není vybrána zpětná eliminace, dopředný krok vpřed, nebo zpětná kroková metoda.
- **Minimální krokové efekty v modelu.** Při použití zpětné eliminace nebo zpětné postupné metody udává tento parametr minimální počet termínů, které mají být zahrnuty do modelu. Zachycení se nezapočítává jako modelový výraz.

- **Maximální počet efektů v modelu.** Při použití dopředné nebo dopředné krokové metody to určuje maximální počet výrazů, které mají být zahrnuty do modelu. Zachycení se nezapočítává jako modelový výraz.
- **Hierarchicky omezte položku a odebrání výrazů.** Tato volba umožňuje zvolit, zda mají být zavedena omezení pro zahrnutí modelových výrazů. Hierarchie vyžaduje, aby všechny podmínky nižšího řádu, které jsou součástí výrazu, aby byly zahrnuty, musí být nejprve v modelu, aby byly zahrnuty všechny termíny, které mají být zahrnuty. Pokud je například požadavek hierarchie v platnosti, faktory *Mariální stav* a *Pohlaví* musí být oba v modelu před přidáním interakce *Mariální stav \* Pohlaví*. Tři volby přepínače určují roli proměnných v určování hierarchie.

## Multinomický Logistic Regression Uložit

Dialogové okno Uložit vám umožňuje uložit proměnné do pracovního souboru a exportovat informace o modelu do externího souboru.

**Uložené proměnné.** Mohou být uloženy následující proměnné:

- **Odhadnutá pravděpodobnost odezvy.** Toto jsou odhadované pravděpodobnosti klasifikace vzoru faktorů/kovariátů do kategorií odezvy. Existuje tolik pravděpodobného pravděpodobnosti, protože existují kategorie proměnné odezvy; až 25 budou uloženy.
- **Kategorie Predicted.** Jedná se o kategorii odezvy s největší očekávanou pravděpodobností pro vzorek součinitel/kovariát.
- **Předpokládané pravděpodobnosti kategorie.** Jedná se o maximum odhadovaných pravděpodobností odezvy.
- **Pravděpodobnost skutečné kategorie.** Toto je odhadovaná pravděpodobnost zařazení faktorů/kovariátového vzorku do sledované kategorie.

**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.

## Další funkce příkazu NOMREG

Jazyk syntaxe příkazu vám také umožňuje:

- Uvedte referenční kategorii závislé proměnné.
- Zahrnout případy s uživateli chybějícími hodnotami.
- Přizpůsobte testy hypotézy tak, že uvedete hypotézy s hodnotou null jako lineární kombinace parametrů.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Probit regrese

Tento postup měří vztah mezi silou stimulu a poměrným podílem případů, které na stimulu vykazují určitou odpověď. To je užitečné pro situace, kdy máte dichotomous výstup, který je myšlenka být ovlivněna nebo způsobeno úrovní některé nezávislé proměnné (s) a je zvláště dobře hodí pro experimentální údaje. Tento postup vám umožní odhadnout sílu stimulu potřebného k vyvolání určité části odpovědi, jako je medián efektivní dávky.

**Příklad.** Jak efektivní je nový pesticid při usmrcování mravenců a jaká je vhodná koncentrace k použití? Můžete provést experiment, ve kterém vystavujete vzorky mravenců různým koncentracím pesticidu a poté zaznamenáte počet usmrčených mravenců a počet exponovaných mravenců. Použití probit regrese na tyto údaje, můžete určit sílu vztahu mezi koncentrací a zabíjením, a můžete určit, jaká vhodná koncentrace pesticidu by bylo, kdyby jste chtěli být jisti zabít, řekněme 95% exponovaných mravenců.

**Statistika.** Regresní koeficienty a standardní chyby, zachycení a standardní chyba, Pearson dobra-vhodné čchi-náměstí, pozorované a očekávané frekvence, a intervaly spolehlivosti pro efektivní úrovně nezávislé proměnné (s). Plot: transformované grafy odezvy.

Tato procedura používá algoritmy navržené a implementované v prostředí NPSOL<sup>®</sup> od Gill, Murray, Saunders & Wright, aby bylo možné odhadnout parametry modelu.

## Aspekty probitových regresních dat

**Data.** Pro každou hodnotu nezávislé proměnné (nebo pro každou kombinaci hodnot pro více nezávislých proměnných) by proměnná odezvy měla být součtem počtu případů s těmi hodnotami, které ukazují odezvu zájmu, a celková pozorovaná proměnná by měla být součtem celkového počtu případů s těmito hodnotami pro nezávislou proměnnou. Proměnná faktorů by měla být kategorická, kódovaná jako celá čísla.

**Předpoklady.** Pozorování by měla být nezávislá. Pokud máte velký počet hodnot pro nezávislé proměnné vzhledem k počtu pozorování, jak byste mohli v observační studii, chi-square a cchi-of-fit statistické údaje nemusí být platné.

**Související postupy.** Analýza Probit je úzce spjata s logistickou regresí; ve skutečnosti, pokud si zvolíte transformaci logit, tento postup v podstatě spočítá logistickou regresí. Obecně platí, že analýza probit je vhodná pro určené experimenty, zatímco logistická regrese je vhodnější pro observační studie. Rozdíly ve výstupu odrážejí tyto různé emphases. Zkušební postup analýzy uvádí odhady efektivních hodnot pro různé míry odezvy (včetně střední účinné dávky), zatímco logistické regresní procedury hlásí odhady poměrů pravděpodobnosti pro nezávislé proměnné.

## Získání probového regresní analýzy

1. Z nabídky vyberte:

**Analyzovat > Regrese > Probit ...**

2. Vyberte proměnnou frekvence odezvy. Tato proměnná označuje počet případů, které vykazují odezvu na testovací stimul. Hodnoty této proměnné nemohou být záporné.
3. Vyberte celkovou pozorovanou proměnnou. Tato proměnná označuje počet případů, na které byl podnět použit. Hodnoty této proměnné nemohou být záporné a nemohou být menší než hodnoty proměnné frekvence odezvy pro každý případ.

Volitelně můžete vybrat proměnnou faktorů. Pokud ano, definujte rozsah pro skupiny pomocí volby **Definovat rozsah**.

4. Vyberte jednu nebo více kovariací (s). Tato proměnná obsahuje úroveň stimulu aplikovaného na každé pozorování. Chcete-li změnit kovariate, vyberte transformaci z rozevřacího seznamu **Transformace**. Pokud není použita žádná transformace a existuje řídicí skupina, pak je řídicí skupina zahrnuta do analýzy.
5. Vyberte buď model **Probit**, nebo **Logit**.

### **Probitový model**

Použije probitovou transformaci (inverzní funkci součtové standardní standardní distribuční funkce) k proporcím odezvy.

### **Model protokolu**

Použije transformaci logit (log odds) na poměry odezvy.

## Rozsah analýzy probitového analýzy

To vám umožní uvést úroveň proměnné faktorů, které budou analyzovány. Úroveň faktorů musí být kódovány jako následná celá čísla a všechny úrovně v rozsahu, který uvedete, budou analyzovány.

## Volby analýzy Probit

Můžete určit volby pro analýzu probit:

**Statistika.** Umožňuje vám vyžádat si následující nepovinné statistiky: Frekvence, relativní medián potence, zkouška paralelismu a intervaly spolehlivosti.

- *Relativní medián Potencisy*. Zobrazuje poměr mediánu potencií pro každou dvojici úrovní faktoru. Také ukazuje 95% intervaly spolehlivosti pro každou relativní střední účinnost. Relativní střední potenciace nejsou k dispozici, pokud nemáte proměnnou faktoru nebo pokud máte více než jednu kovariaci.
- *Test paralelizmu*. Test hypotézy, že všechny úrovně faktoru mají společný sklon.
- *Intervaly úředního vědomí*. Intervaly spolehlivosti pro dávkování agenta potřebného k vytvoření určité pravděpodobnosti odezvy.

V případě, že jste zvolili více než jednu kovariaci, nejsou k dispozici žádné intervaly spolehlivosti a relativní medián potence. Relativní střední účinnost a zkuška paralelizmu jsou k dispozici pouze tehdy, pokud jste vybrali proměnnou faktoru.

**Přirozená rychlost odezvy.** Umožňuje vám označit přirozenou rychlost odezvy i v případě absence stimulu. Dostupné alternativy jsou Žádné, Vypočítat z dat nebo Hodnota.

- *Vypočítat z dat*. Odhadněte přirozenou odezvu odezvy z ukázkových dat. Vaše data by měla obsahovat případ představující řídicí úroveň, pro kterou se hodnota proměnné liší od 0. Probit odhaduje přirozený poměr odpovědí na základě podílu odpovědí pro řídicí úroveň jako počáteční hodnotu.
- *Hodnota*. Nastavuje přirozenou rychlost odezvy v modelu (vyberte tuto položku, když znáte přirozenou rychlost odezvy v předstihu). Uvedte poměr přirozené odezvy (poměr musí být menší než 1). Například, pokud se odezva vyskytne 10% času, kdy je stimul 0, zadejte 0.10.

**Kritéria.** Umožňuje vám řídit parametry iterativního algoritmu odhadu parametrů. Můžete přepsat výchozí hodnoty pro Maximum iterací, Limit kroku a Toleranční toleranci.

## Další funkce příkazu PROBIT

Jazyk syntaxe příkazu vám také umožňuje:

- Vyžádejte si analýzu jak na probit, tak i u modelů logit.
- Řídit zacházení s chybějícími hodnotami.
- Transformujte kovariany bázi jinými než 10 nebo přirozenými logaritmy.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Regrese kvantilů

Regrese je statistická metoda široce používaná v kvantitativním modelování. Multiple lineární regrese je základní a standardní přístup, ve kterém výzkumníci používají hodnoty několika veličin k vysvětlení nebo predikci středních hodnot výsledku stupnice. Za mnoha okolností se však více zajímáme o medián nebo o libovolný kvantil výsledku měřítka.

Kvantitaní regrese modeluje vztah mezi sadou prediktorů (nezávislých) proměnných a konkrétními percentily (nebo "kvantily") cílové (závislé) proměnné, nejčastěji střední hodnotou. Má dvě hlavní výhody oproti běžné nejméně čtvercové regresi:

- Kvantitaová regrese nečiní žádné předpoklady ohledně distribuce cílové proměnné.
- Kvantitalová regrese má tendenci odolat vlivu odlehlých pozorování

Kvantitaní regrese se široce používá pro výzkum v průmyslových odvětvích, jako je ekologie, zdravotní péče a finanční ekonomie.

### Příklad

Jaký je vztah mezi celkovým příjmem domácnosti a podílem příjmu, který je vynaložen na potraviny? Engel zákon je postřeh v ekonomii, ve kterém se uvádí, že podle vzestupu příjmů klesá podíl příjmů vynaložených na potraviny, a to i v případě, že absolutní výdaje na potraviny rostou. Při použití kvantilu regrese na tyto údaje můžete určit, které výdaje na potraviny se mohou týkat 90% rodin (pro 100 rodin s daným příjmem), pokud se nezajímají o střední výdaje na potraviny.

### Statistika

Quantile Regression, simplex approach, Frisch-Newton interior-point non-linear optimization algorithm, Barrodale and Roberts, Bofinger, Hall Sheather, šířka frekvenčního pásma, maticové

manipulace, konvergenční kritéria, regresní váhy, predikovaný cíl, predikovaný cíl, predikční rezidua, tabulka, odhady rozptylu, odhady parametrů, kovarianční matice, korelační matrice, sledované hodnoty, interval spolehlivosti.

Tato procedura používá algoritmy navržené Koenker, R. W. a Bassett, G. W. (1978). Regresní kvantily, *Ekonometrika*, 46, 33-50.

## Posouzení kvantilových regresních dat

### Data

Je vyžadována jedna číselná závislá proměnná. Cílová proměnná musí být spojitá proměnná. Prediktory mohou být spojitě proměnné nebo fiktivní proměnné pro kategoriální prediktory. Pro spuštění analýzy je nezbytný buď výraz zachycení, nebo alespoň jeden prediktor.

### Předpoklady

Kvantilová regrese nečiní předpoklady o distribuci cílové proměnné a vymění vliv na výstupky, které mají vliv na přelhaní pozorování.

### Související postupy

Kvantilová analýza souvisí s regresí běžného nejméně mocniny.

## Získání analýzy kvantilové regrese

1. Z nabídky vyberte:

**Analyzovat > Regrese > Kvantil ...**

Dialogové okno vám umožňuje uvést cíl, faktor, kovariate a proměnné váhy, které se použijí pro analýzu regrese kvantilu. Dialogové okno také poskytuje volbu uchování paměti pro komplexní analýzu nebo velké datové sady.

2. Vyberte číselnou cílovou proměnnou. Pro spuštění analýzy je vyžadována pouze jedna cílová proměnná. Povoleny jsou pouze číselné proměnné.
3. Volitelně můžete vybrat jednu nebo více proměnných faktorů. Proměnné měřítka nejsou povoleny.
4. Volitelně můžete vybrat jednu nebo více proměnných covariate. Řetězcové proměnné nejsou povoleny.

**Poznámka:** Pokud jsou v dialogovém okně Model vybrány oba seznamy **Faktor (s)** i **Covariate (s)** a v dialogovém okně Model je vybrána volba **Zahrnout zachycení v modelu**, zobrazí se následující zpráva:

No effects have been specified. Therefore, an intercept only model will be fit.  
Do you want to fit an intercept-only model?

5. Volitelně můžete vybrat proměnnou regresní váhy. Řetězcové proměnné nejsou povoleny.
6. Volitelně můžete vybrat volbu **Conserve memory for complex analysis or large datasets**. Toto nastavení řídí, zda jsou data během zpracování držena v externím souboru. Povolení nastavení může pomoci šetřit paměťové prostředky při spouštění složitých analýz nebo analýz s velkými datovými sadami.

## Měření kvantile: Kritéria

Dialogové okno Kritéria poskytuje volby pro

### Kvantil

Poskytuje volby pro určení kvantil (ů).

#### Určit jednotlivé kvantily

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 tlačítek **Přidat**, **Změnit** a **Odebrat** můžete pracovat s hodnotami v seznamu hodnot kvantil.

Všechny hodnoty musí být jedinečné (duplicitní hodnoty nejsou povoleny). Standardní hodnota je 0.5.



### Určit kvantily mřížky

When selected, a grid of quantiles 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.

### Metoda odhadu

Poskytuje volby pro určení metody odhadu modelu.

#### Automaticky zvolený programem

Umožňuje proceduře automaticky vybrat příslušnou metodu odhadu. Toto je výchozí nastavení.

#### Algoritmus Simplex

Volá algoritmus simplex, který byl vyvinut společností Barrodale a Roberts.

#### Frisch-Newton interior-bod nelineární optimalizace

Volání na algoritmus Frisch-Newton vnitřní-bod nelineární optimalizační algoritmus.

### Následný odhad

Poskytuje volby pro post-odhad kovariance-kovariance odhadů parametru a intervaly spolehlivosti pro predikované cílové hodnoty.

#### Předpokládejme, že případy jsou IID

Je-li tato volba vybrána, předpokládá se, že chybové výrazy jsou nezávislé a shodně distribuované. Není-li toto nastavení vybráno, čas výpočtu může výrazně vzrůst u velkých modelů. Nastavení je ve výchozím nastavení vybráno.

#### Typ šířky pásma

Určuje, která metoda šířky pásma se používá k odhadu rozptylu odchylky rozptylu odhadů parametru (**Bofinger** nebo **Hall-Sheather**). Výchozím nastavením je **Bofinger**.

### Číselná metoda

Poskytuje následující volby:

#### Singulární tolerance

Uvádí hodnotu tolerance pro maticové manipulace ve vnitřní-bodové metodě. Uvedená hodnota musí být jedna, dvojitá hodnota v  $(0, 10^{-3})$ , s předvoleným nastavením  $10^{-12}$ .

#### Konvergence

Určuje kritérium konvergence pro číselnou metodu. Uvedená hodnota musí být jedna, dvojitá hodnota v  $(0, 10^{-3})$ , při výchozím nastavení  $10^{-6}$ .

#### Maximum iterací

Určuje maximální počet iterací. Uvedená hodnota musí být jedno kladné celé číslo. Výchozí hodnota je 2000.

### Chybějící hodnoty

Poskytuje volby pro určení, jak se zachází s chybějícími hodnotami.

#### Vyloučit jak uživatelské, tak systémové chybějící hodnoty

Je-li tato volba vybrána, jsou vyloučeny chybějící hodnoty jak pro uživatele, tak i pro systém.

#### Uživatel-chybějící hodnoty jsou považovány za platné

Je-li tato volba vybrána, jsou hodnoty uživatelských chybějících hodnot považovány za platné.

### Interval spolehlivosti (%)

Určuje úroveň významnosti. Je-li zadána, hodnota musí být jednoduchá hodnota typu double v rozsahu 0 až 100. Výchozí hodnota je 95.

## Množstevní regrese: Model

Dialogové okno Model poskytuje volby pro uvedení efektů a vah použitých v modelu. Je-li vynechán nebo zadán sám, bude model obsahovat výraz zachycení a všechny hlavní efekty s kovariáty v seznamu proměnných a faktory v seznamu faktorů.

## Určit efekty modelu

Výchozí model je pouze pro zachycení, takže musíte explicitně zadat jiné efekty modelu. Eventuálně můžete sestavovat vnořené nebo nevnořené podmínky. Je-li vybrána volba **Výrazy sestavení**, jsou k dispozici následující volby efektu a interakce pro nevnořené podmínky.

### Hlavní účinky

Vytvoří výraz main-effects pro každou vybranou proměnnou.

### Interakce

Vytvoří funkční období interakce nejvyšší úrovně pro všechny vybrané proměnné.

### Faktoriál

Vytvoří všechny možné interakce a hlavní efekty vybraných proměnných.

### 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.

Je-li vybrána volba **Sestavit vnořené podmínky**, můžete sestavovat vnořené podmínky. Vnořené termíny jsou užitečné pro modelování efektu faktoru nebo proměnné, jejichž hodnoty nemají interakci s úrovněmi jiného faktoru. Například řetězec obchodu s potravinami může dodržovat výdajové návyky svých zákazníků v několika lokalitách skladu. Vzhledem k tomu, že každý zákazník často navštívuje pouze jedno z těchto umístění, lze říci, že efekt *Zákazník* může být **vnořen v** efektu *Umístění úložiště*.

Kromě toho můžete zahrnout efekty interakce, jako je polynomiální výrazy zahrnující stejné proměnné, nebo přidat více úrovní vnoření do vnořené výrazu.

### Vnoření poznámek:

- Chcete-li zahrnout efekt pro interakci mezi dvěma faktory, použijte klíčové slovo BY nebo hvězdičku (\*) k připojení k faktorům, které se podílejí na interakci.
- Faktory uvnitř účinku interakce musí být odlišné.
- Chcete-li zahrnout efekt pro vnoření jednoho výrazu do jiného, použijte dvojice závorek.
- Je-li přítomno více než jeden pár závorek, každá dvojice závorek musí být uzavřena nebo vnořena do jiné dvojice závorek.
- Vícenásobné vnoření je povoleno.
- Interakce mezi vnořenými efekty nejsou podporovány.

**Omezení:** Vnořené termíny mají následující omezení:

- Všechny faktory v rámci interakce musí být jedinečné. Je-li tedy A faktor, pak uvedení  $A^*$  A je neplatné.
- Všechny faktory ve vnořeném efektu musí být jedinečné. Je-li tedy A faktor, pak zadání A (A) je neplatné.
- Žádný vliv nemůže být vnořen v rámci proměnné. Pokud tedy A je součinitel a X je proměnná, pak zadání A (X) je neplatné.

### Zahrnout zachycení do modelu

Je-li tato volba vybrána, je do modelu zahrnut výraz zachycení. Není-li tato volba vybrána, je pro spuštění analýzy vyžadován alespoň jeden prediktor. Nastavení je ve výchozím nastavení povoleno.

## Quantile Regression: Zobrazit

Dialogové okno Zobrazit poskytuje výstup a nastavení pro ovládací prvek grafu.

## Tisk

K dispozici jsou následující volby výstupu.

### Odhady parametrů

Zobrazuje odhady parametrů a odpovídající statistiky testu a intervaly spolehlivosti. Volitelně můžete volitelně zobrazit odhady parametrů exponované na základě prvotního odhadu parametrů.

### Matice Cvariance pro odhady parametrů

Zobrazí odhadnutou matici kovariance parametru.

### Matice korelace pro odhad parametrů

Zobrazí odhadovanou matici korelace parametru.

## Zakreslit a tabulky

K dispozici jsou následující možnosti vykreslení:

### Vykreslit odhady parametrů pro

Můžete zvolit vykreslení odhadů parametrů pro specifický počet nejvyšších efektů nebo pro všechny efekty v modelu. Nastavení **Prvních xx efektů** řídí počet kategorií nebo kombinace kategorií ve smíšeném efektu, které jsou vykresleny v rámci interakce jedné proměnné a jednoho nebo dvou faktorů. Hodnota musí být jediné kladné celé číslo (50 je výchozí nastavení).

#### Poznámky:

- Grafy předpovědi se vytvářejí pro všechny efekty, je-li zadaná celočíselná hodnota větší než počet kategorií nebo kombinací.
- Toto nastavení je platné pouze tehdy, je-li v dialogovém okně Kritéria určeno více hodnot pro nastavení **Hodnoty kvantilu**. Když je uvedena jedna hodnota kvantilu, nevytvářejí se žádné grafy.

### Zobrazit předpověď pozorovaný zkusaný graf

Řídí tvorbu predikovaného a pozorovaného grafu hodnot. Je-li tato volba povolena, vytvoří se jediný výkres obsahující body (s různými barevnými tečkami znázorňujícími různé kvantily). Nastavení je ve výchozím nastavení vypnuto.

### Předpověďte efekty v modelu

Je-li povolena tato volba, jsou k dispozici následující volby:

#### Graf zakreslení nebo tabulace horních x efektů

Uvedte počet nejlepších efektů, jejichž vykreslovací nebo predikční tabulka bude vytvořena. 3 je výchozí hodnota.

**Poznámka:** Vyhodnotové grafy nebo tabulky předpovědi se vytvářejí pro všechny účinky, je-li uvedená hodnota větší než počet platných efektů v modelu.

#### Graf zakreslení nebo tabuláčkovat uživatelem zadané efekty

Platné pokyny pro účinek jsou:

- Effect with one covariate (including a high power of the covariate itself): Create a single plot containing the lines predicted by different quantiles.
- Efekt s jedním faktorem: Tabulovat predikce pro kategorie faktoru podle různých kvantilů.
- Vliv s interakcí dvou faktorů: Pro každý kvantil, tabulate predikce pro kategorie dvou faktorů.
- Účinek na interakci jedné proměnné a jednoho nebo dvou faktorů: Pro každý kvantil vytvořte graf obsahující řádky pro každou kategorii nebo kombinaci kategorií v rámci interakce interakce.
- Maximální počet kombinací, které mají být zakresleny, je řízen hodnotou uvedenou pro **kategorie maximální kategorie xx kombinací kombinací kategorií ve smíšeném efektu**.

Efekty, které jsou přesunuty ze seznamu **Efekty modelu** na seznam **Řádky předpovědi** se používají pro vykreslení. Vykreslování není vytvořeno v případech, kdy jsou uvedené efekty konstantní (byly odebrány ze sestavení modelu).

Efekty, které jsou přesunuty z **Modelové efekty** do seznamu **Tabulky předpovědí**, se používají pro tabelování. Tabulky se nevytvorí v případech, kdy jsou uvedené efekty konstantní (odebrané ze sestavení modelu).

#### **Vykreslit maximum xx kategorií nebo kombinací kategorií ve smíšeném účinku**

Řídí maximální počet kombinací kategorií, které mají být zakreslené. Předvolená hodnota je 10.

## **Množstevní regrese: Uložit**

Dialogové okno Uložit poskytuje volby pro přidělení skóre modelu.

#### **Předpovězená hodnota odpovědi**

Vyberete-li tuto volbu, bude předpovězena předpokládaná cílová hodnota.

#### **Gumová**

Je-li vybrána tato volba, jsou skóre zbytkové chyby skóre.

#### **Dolní hranice intervalu předpovědi**

Je-li vybrána tato volba, budou skórování dolních hranic intervalů předpovědi.

#### **Horní hranice intervalu předpovědi**

Je-li vybrána tato volba, budou skórování horních hranic intervalů předpovědi.

**Poznámka:** Název proměnné může být zadán pro každou volbu uložení. Je-li zadán název kořenového adresáře, musí to být platný název proměnné. Je použit kořenový název následovaný podtržítkem "\_" a smysluplným příponou kvantilu, pokud je v dialogovém okně Kritéria určeno více hodnot pro nastavení **Hodnoty kvantilu**.

## **Množstevní regrese: Export**

Dialogové okno Exportovat poskytuje volby pro uvedení, které statistiky jsou exportovány, jak jsou exportovány statistiky (externí datové soubory nebo datové sady) a řídí, jak se zachází s daty během zpracování (proces normálně, nebo je držen v externím pracovním souboru během zpracování).

#### **Matice rozptylu odhadů parametrů**

Je-li tato volba vybrána, volby pro zápis kovarianční matice odhadu parametru na externí datový soubor, nebo dříve deklarovaná data, jsou zapnuty.

#### **Korelační matice odhadů parametrů**

Je-li tato volba vybrána, jsou povoleny volby pro zápis matice korelace z odhadů parametrů do externího datového souboru nebo dříve deklarovaných datových sad.

#### **Matice kovariance/korelace bude uložena do jedné datové sady nebo do externího souboru za přítomnosti více regresních kvantilů.**

Je-li přítomno více kvantilů, tato volba přepíná ukládání kovariance/korelačních matic na jednu nebo více datových sad nebo externích datových souborů. Když není povoleno, matice se uloží do jednoho, externího datového souboru nebo datové sady. Když je povoleno, matice se uloží do více externích datových souborů nebo datových sad. Toto nastavení se projeví pouze v případě, že je v dialogovém okně Kritéria určeno více hodnot pro nastavení **Hodnoty kvantilu**.

**Poznámka:** Tato volba je k dispozici pouze v případě, že je vybrána volba **Kovarianční matice odhadů parametrů** nebo **Korelační matice odhadů parametrů**.

#### **Exportovat informace o modelu do souboru XML**

Je-li tato volba vybrána, poskytuje volby pro export informací o modelu do specifického názvu souboru XML a umístění.

#### **Exportovat ve formátu XML**

Je-li vybrána volba **Exportovat informace o modelu do souboru XML**, můžete exportovat buď odhady parametrů a kovarianční matice, nebo pouze odhady parametrů. **Odhady parametrů a kovarianční matice** jsou výchozí nastavení.

## Konvence pro názvy souborů

- Je-li zadána jedna hodnota pro nastavení **Hodnoty kvantilu** v dialogovém okně Kritéria, použijí se `savefile` a `dataset` k pojmenování externích datových souborů nebo datové sady.
- Je-li pro nastavení **Hodnota kvantil** v dialogovém okně Kritéria určeno více hodnot, je každý kvantil uložen do externího datového souboru nebo datové sady.
- Znak podtržítka "\_" následovaný smysluplným příponou kvantil je automaticky připojen k datovému souboru nebo k názvu datové sady. Například, když je 0.25, 0.50 a 0.75 uvedeno jako **Quantile values**, jsou přípony `_25`, `_50a` `_75` připojeny k názvům datových souborů (před příponou `.sav`).
- Další číslice lze zadat pro každou příponu kvantile (je-li to nutné).
- Úvodní nula a desetinná čárka **Quantile values** se nepoužívají v příponě.
- Je-li pro **Hodnoty kvantilu** uvedena exponenciální notace, je při zobrazení v příponě převedena na desítkovou hodnotu.

## Nelineární regrese

Nelineární regrese je metoda vyhledání nelineárního modelu vztahu mezi závislou proměnnou a sadou nezávislých proměnných. Na rozdíl od tradiční lineární regrese, která je omezena na odhad lineárních modelů, může nelineární regrese odhadnout modely s libovolnými vztahy mezi nezávislými a závislými proměnnými. To lze provést pomocí algoritmů iterativního odhadu. Všimněte si, že tento postup není nutný pro jednoduché polynomiální modely ve tvaru  $Y = A + BX^{**2}$ . Definováním  $W = X^{**2}$  se dostaneme jednoduchý lineární model  $Y = A + BW$ , který lze odhadnout pomocí tradičních metod, jako je například lineární regrese.

**Příklad.** Může se předpovídat populace na základě času? Scatterplot ukazuje, že se zdá být silný vztah mezi obyvatelstvem a časem, ale vztah je nelineární, takže to vyžaduje speciální metody odhadu z Nonlinear regresního postupu. Vytvořením vhodné rovnice, jako je například model logistického populačního růstu, můžeme dosáhnout dobrého odhadu modelu, což nám umožňuje vytvářet předpovědi o počtu obyvatel, které se ve skutečnosti neměřili.

**Statistika.** Pro každou iteraci: odhad parametrů a zbytkový součet čtverců. Pro každý model: součet čtverců pro regresi, reziduální, nekorigovaný celkový a korigovaný celkový, parametr odhadů, asymptotické standardní chyby a asymptotické korelační matice parametrů odhadů.

**Poznámka:** Omezená nelineární regrese používá algoritmy navržené a implementované ve funkci NPSOL<sup>®</sup> od Gill, Murray, Saunders a Wright k odhadu parametrů modelu.

### Pokyny k nelineární regresní datům

**Data.** Závislé a nezávislé proměnné by měly být kvantitativní. Kategorické proměnné, jako např. náboženství, hlavní nebo region trvalého pobytu, je třeba rekódovat do binárních (fiktivních) proměnných nebo jiných typů kontrastních proměnných.

**Předpoklady.** Výsledky jsou platné pouze v případě, že jste zadali funkci, která přesně popisuje vztah mezi závislými a nezávislými proměnnými. Kromě toho je velmi důležitá volba dobrých startovních hodnot. I v případě, že jste zadali správný funkční tvar modelu, pokud používáte špatné počáteční hodnoty, může se váš model nepřibližovat, nebo můžete získat lokálně optimální řešení, a ne tak, aby bylo optimální.

**Související postupy.** Mnoho modelů, které se objevují nelineární v první řadě, lze transformovat na lineární model, který lze analyzovat pomocí procedury lineární regrese. Pokud si nejste jisti, jaký správný model by měl být správný, procedura odhadu křivky vám může pomoci identifikovat užitečné funkční vztahy ve vašich datech.

### Získání analýzy nelineárních regresních analýz

1. Z nabídky vyberte:

**Analyzovat > Regrese > Nelineární ...**

2. Vyberte jednu numerickou závislou proměnnou ze seznamu proměnných ve své aktivní datové sadě.

3. Chcete-li sestavit výraz modelu, zadejte výraz do pole **Modelovat výraz** nebo vložte komponenty (proměnné, parametry, funkce) do pole.
  4. Označte parametry ve svém modelu klepnutím na volbu **Parametry**.
- Segmentovaný model (jeden, který má různé formy v různých částech své domény) musí být zadán s použitím podmíněné logiky v rámci jediného příkazu modelu.

## Podmíněná logika (nelineární regrese)

Pomocí podmíněné logiky můžete určit segmentovaný model. Chcete-li použít podmíněnou logiku ve výrazu modelu nebo funkci ztráty, vytvoříte součet řady výrazů, z nichž jedna pro každou podmínku. Každý člen se skládá z logického výrazu (v závorkách) násobený výrazem, který má být výsledkem toho, že logický výraz má hodnotu true.

Vezměme si například segmentovaný model, který se rovná 0 pro  $X \leq 0$ ,  $X$  pro  $0 < X < 1$  a 1 pro  $X \geq 1$ . Výraz pro toto je:

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

Logické výrazy v závorkách se všechny vyhodnocují na hodnotu 1 (true) nebo 0 (false). Proto:

Pokud  $X \leq 0$ , hodnota se sníží na  $1 * 0 + 0 * X + 0 * 1 = 0$ .

Pokud je  $0 < X < 1$ , sníží se na  $0 * 0 + 1 * X + 0 * 1 = X$ .

Pokud  $X \geq 1$ , sníží se na  $0 * 0 + 0 * X + 1 * 1 = 1$ .

Složitější příklady lze snadno sestavit nahrazením různých logických výrazů a výsledných výrazů. Nezapomeňte, že dvojité nerovnosti, jako například  $0 < X < 1$ , musí být napsány jako složené výrazy, jako například  $(X > 0 \ \& \ X < 1)$ .

Řetězcové proměnné lze použít v logických výrazech:

$$(\text{city} = \text{'New York'}) * \text{costliv} + (\text{city} = \text{'Des Moines'}) * 0.59 * \text{costliv}$$

To přináší jeden výraz (hodnota proměnné *costliv*) pro New Yorkery a další (59% z této hodnoty) pro obyvatele Des Moines. Řetězcové konstanty musí být uzavřeny v uvozovkách nebo apostrofech, jak je zobrazeno zde.

## Nelineární parametry regrese

Parametry jsou části vašeho modelu, které používají nelineární regresní procedury. Parametry mohou být aditivní konstanty, multiplikativní koeficienty, exponenty nebo hodnoty použité při vyhodnocování funkcí. Všechny vámi definované parametry se zobrazí (se svými počátečními hodnotami) na seznamu Parametry v hlavním dialogovém okně.

**Název.** Pro každý parametr musíte zadat název. Tento název musí být platným názvem proměnné a musí se jednat o název použitý ve výrazu modelu v hlavním dialogovém okně.

**Počáteční hodnota.** Umožňuje vám uvést počáteční hodnotu pro parametr, pokud možno co nejbližší k očekávanému konečnému řešení. Špatné startovací hodnoty mohou vést k selhání sbíhavost nebo k konvergenci na řešení, které je lokální (spíše než globální), nebo je fyzicky nemožné.

**Použit počáteční hodnoty z předchozí analýzy.** Pokud jste již spustili nelineární regresi z tohoto dialogového okna, můžete vybrat tuto volbu pro získání počátečních hodnot parametrů z jejich hodnot v předchozím spuštění. To vám umožní pokračovat v hledání, když se algoritmus konverguje pomalu. (Počáteční počáteční hodnoty se budou i nadále zobrazovat v seznamu Parametry v hlavním dialogovém okně.)

*Poznámka:* Tento výběr bude zachován v tomto dialogovém okně pro zbytek vaší relace. Pokud model změníte, ujistěte se, že jste zrušili výběr.

## Nelineární regresní běžné modely

Níže uvedená tabulka poskytuje vzorovou syntaxi pro mnoho publikovaných nelineárních regresních modelů. Vybraný model náhodně nevejde do vašich dat dobře. Vhodné počáteční hodnoty parametrů jsou nezbytné a některé modely vyžadují omezení, aby se sbíhaly.

Tabulka 1. Příklad syntaxe modelu

Název	Modelový výraz
Asymptotická regrese	$b_1 + b_2 * \text{výraz} (b_3 * x)$
Asymptotická regrese	$b_1 - (b_2 * (b_3 ** x))$
Hustota	$(b_1 + b_2 * x) ** (-1/ b_3)$
Gauss	$b_1 * (1 - b_3 * \exp(-b_2 * x ** 2))$
Gompertzová	$b_1 * \exp(-b_2 * \exp(-b_3 * x))$
Johnson-Schumacher	$b_1 * \exp(-b_2 / (x + b_3))$
Protokol-upraveno	$(b_1 + b_3 * x) ** b_2$
Protokol-Logistický	$b_1 - \ln(1 + b_2 * \exp(-b_3 * x))$
Metcherlich Returns of Diminshing Returns	$b_1 + b_2 * \text{výraz} (-b_3 * x)$
Michaelis Menten	$b_1 * x / (x + b_2)$
Morgan-Mercer-Florin	$(b_1 * b_2 + b_3 * x ** b_4) / (b_2 + x ** b_4)$
Peál-Reed	$b_1 / (1 + b_2 * \exp(-(b_3 * x + b_4 * x ** 2 + b_5 * x ** 3)))$
Poměr krychlí Cubics	$(b_1 + b_2 * x + b_3 * x ** 2 + b_4 * x ** 3) / (b_5 * x ** 3)$
Poměr kvadratických hodnot	$(b_1 + b_2 * x + b_3 * x ** 2) / (b_4 * x ** 2)$
Richards	$b_1 / ((1 + b_3 * \exp(-b_2 * x)) ** (1/ b_4))$
Verhulst	$b_1 / (1 + b_3 * \exp(-b_2 * x))$
Von Bertalando	$(b_1 ** (1 - b_4) - b_2 * \exp(-b_3 * x)) ** (1 / (1 - b_4))$
WEIBULL	$b_1 - b_2 * \exp(-b_3 * x ** b_4)$
Hustota výnosu	$(b_1 + b_2 * x + b_3 * x ** 2) ** (-1)$

## Funkce nelineárních regresních ztrát

**Funkce ztráty** v nelineární regresi je funkce, která je minimalizována algoritmem. Vyberte buď **Součet obdélníkových zbytkových chyb** k minimalizaci součtu druhých mocnín zbytkových chyb nebo **Funkce ztráty definované uživatelem**, aby se minimalizovala jiná funkce.

Vyberete-li volbu **Funkce ztráty definované uživatelem**, je třeba definovat funkci ztráty, jejíž součet (ve všech případech) by měl být minimalizován výběrem hodnot parametrů.

- Většina funkcí ztráty zahrnuje speciální proměnnou **RESID\_**, která představuje zbytkovou hodnotu. (Výchozí součet funkce ztráty druhých mocnín zbytkových chyb může být zadán explicitně jako **RESID\_\*\*2**.) Pokud potřebujete použít předpokládanou hodnotu ve své funkci ztráty, bude se rovnat závislé proměnné minus zbytková hodnota.
- Je možné zadat funkci podmíněné ztráty pomocí podmíněné logiky.

Můžete buď zadat výraz do pole Funkce uživatelem definované ztráty, nebo vložit komponenty výrazu do pole. Řetězcové konstanty musí být uzavřeny v uvozovkách nebo apostrofech a numerické konstanty musí být zapsány v americkém formátu, s tečkou jako desetinným oddělovačem.

## Omezení nelineárních regresních parametrů

**Omezení** je omezení přípustných hodnot pro parametr během iterativního vyhledávání řešení. Lineární výrazy se vyhodnocují před přijetím kroku, takže můžete použít lineární omezení k zabránění krokům, které mohou způsobit přetečení. Nelineární výrazy jsou vyhodnoceny po provedení kroku.

Každá rovnice nebo nerovnost vyžaduje následující prvky:

- Výraz zahrnující alespoň jeden parametr v modelu. Zadejte výraz nebo použijte pomocnou klávesnici, která vám umožní vložit čísla, operátory nebo závorky do výrazu. Můžete zadat buď požadovaný parametr (y) spolu se zbytkem výrazu, nebo vložit ze seznamu parametrů vlevo. Běžné proměnné nelze použít v omezení.
- Jeden ze tří logických operátorů  $<$ ,  $=$ , nebo  $>$ .
- Numerická konstanta, na kterou je výraz porovnáván s použitím logického operátoru. Zadejte konstantu. Numerické konstanty musí být zapsány v americkém formátu, s tečkou jako desetinný oddělovač.

## Nelineární regresní operace Uložit nové proměnné

Do aktivního datového souboru můžete uložit počet nových proměnných. Dostupné volby jsou Residuals, Predicted values, Derivatives a funkce Loss function. Tyto proměnné lze použít v následných analýzách k testování vhodnosti modelu nebo k identifikaci problémových případů.

- *Reziduy*. Uloží zbytkové chyby s residencí s názvem proměnné.
- *Předpovězené hodnoty*. Uloží predikované hodnoty s názvem proměnné pred\_.
- *Deriváty*. Jeden derivát se ukládá pro každý parametr modelu. Derivátové názvy jsou vytvářeny prefixem 'd.' na prvních šest znaků názvů parametrů.
- *Hodnoty funkce ztráty*. Tato volba je k dispozici, pokud zadáte vlastní funkci ztráty. Proměnná ztráta názvu proměnné \_ je přiřazena k hodnotám funkce ztráty.

## Nelineární regresní volby

Volby vám umožňují řídit různé aspekty vaší nelineární regresní analýzy:

*Odhady zaváděcího programu*. Metoda odhadu standardní chyby statistiky za použití opakovaných vzorků z původní datové sady. To se provádí pomocí vzorkování (s náhradou), aby bylo možné získat mnoho vzorků stejné velikosti jako původní datové sady. Nelineární rovnice se odhaduje pro každý z těchto vzorků. Směrodatná chyba každého odhadu každého parametru se pak vypočítá jako směrodatná odchylka odhadnutých odhadnutých na zaváděcí program. Hodnoty parametrů z původních dat se použijí jako počáteční hodnoty pro každý vzorek zaváděcího programu. To vyžaduje sekvenční kvadratický programovací algoritmus.

**Metoda odhadu**. Umožňuje vybrat metodu odhadu, je-li to možné. (Určité volby v tomto nebo jiném dialogovém okně vyžadují sekvenční kvadratický programovací algoritmus.) Dostupné alternativy zahrnují Sekvenční kvadratické programování a Levenberg-Marquardt.

- *Sekvenční kvadratické programování*. Tato metoda je k dispozici pro omezené a neomezené modely. Sekvenční kvadratické programování se používá automaticky, pokud zadáte omezený model, uživatelsky definovanou ztrátu funkce nebo bootstrapping. Můžete zadat nové hodnoty pro Maximum iterací a Limit kroku a můžete změnit výběr v rozevírácích seznamech pro optimální toleranci, přesnost funkce a nekonečnou velikost kroku.
- *Levenberg-Marquardt*. Jedná se o výchozí algoritmus pro modely bez omezení. Metoda Levenberg-Marquardt není k dispozici, pokud zadáte omezený model, uživatelsky definovanou ztrátu funkce nebo bootstrap. Můžete zadat nové hodnoty pro maximální počet iterací a můžete změnit výběr v rozevírácích seznamech pro konvergenci a konvergenci parametru Sum-of-squay.

## Interpretace nelineárních regresních výsledků

Nelineární regresní problémy často představují výpočetní obtíže:



- Volba počátečních hodnot pro parametry ovlivňuje konvergenci. Snažte se zvolit počáteční hodnoty, které jsou přiměřené a pokud možno i blízké očekávanému konečnému řešení.
- Někdy se jeden algoritmus chová lépe než ten druhý na konkrétním problému. V dialogovém okně Volby vyberte druhý algoritmus, je-li k dispozici. (Pokud uvedete funkci ztráty nebo určité typy omezení, nemůžete použít algoritmus Levenberg-Marquardt.)
- Když se iterace zastaví pouze z toho důvodu, že se vyskytl maximální počet iterací, model "final" pravděpodobně není dobrým řešením. Vyberte volbu **Použít počáteční hodnoty z předchozí analýzy** v dialogovém okně Parametry, chcete-li pokračovat v iteraci, nebo ještě lépe zvolit jiné počáteční hodnoty.
- Modely, které vyžadují umocňování hodnot nebo rozsáhlých datových hodnot, mohou způsobit přetečení nebo podtoky (čísla jsou příliš velká nebo příliš malá, aby je počítač mohl reprezentovat). Někdy se jim můžete vyhnout vhodnou volbou výchozích hodnot nebo uložením omezení na parametry.

## Další funkce příkazu NLR

Jazyk syntaxe příkazu vám také umožňuje:

- Nazvěte soubor, ze kterého se čtou počáteční hodnoty pro odhady parametrů.
- Uvedte více než jeden příkaz modelu a funkci ztráty. Tím je usnadněno určení segmentovaného modelu.
- Dodatek své vlastní deriváty spíše než používat ty, které jsou vypočteny programem.
- Určete počet vzorků zaváděcího programu, které mají být vygenerovány.
- Určete další kritéria iterace, včetně nastavení kritické hodnoty pro kontrolu odvození a definování konvergenčního kritéria pro korelaci mezi zbytkovými rezidui a odvozenými hodnotami.

Další kritéria pro příkaz CNLR (omezená nelineární regrese) umožňují:

- Zadejte maximální povolený počet menších iterací v rámci každé hlavní iterace.
- Nastavte kritickou hodnotu pro kontrolu odvození.
- Nastavte omezení kroku.
- Uvedte toleranci havárie, abyste určili, zda jsou počáteční hodnoty ve svých uvedených mezích.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Odhad váhy

Standardní lineární regresní modely předpokládají, že odchylka je konstantou v rámci sledované populace. Když to není tento případ (například, když případy, které jsou vysoké na některých atributech ukazují více variability než případy, které jsou nízké na tomto atributu) lineární regrese pomocí obyčejných nejmenších čtverců (OLS) již neposkytuje optimální model odhadů. Pokud rozdíly v variabilitě lze předpovědět z jiné proměnné, procedura odhadu váhy může vypočítat koeficienty lineární regresní modelu s použitím vážených nejmenších čtverců (WLS), tak aby přesnější pozorování (tj. ty s menší variabilitou) měly větší váhu při určování regresní koeficienty). Procedura odhadu váhy testuje rozsah transformací váhy a udává, které z nich budou nejlépe vyhovovat datům.

**Příklad.** Jaké jsou dopady inflace a nezaměstnanosti na změny cen akcií? Vzhledem k tomu, že akcie s vyšším podílem akcií často vykazují větší variabilitu než akcie s nízkými hodnotami akcií, nejmenších čtverců nepřinese optimální odhady. Odhad hmotnosti vám umožňuje účtovat za účinek ceny akcií na variabilitě změn cen při výpočtu lineárního modelu.

**Statistika.** Hodnoty pravděpodobnosti protokolu pro každou mocninu zdrojové proměnné váhy testovány, více  $R$ ,  $R$ na druhou, upravené  $R$ na druhou mocninu, tabulku ANOVA pro model WLS, nestandardizované a standardizované odhady parametrů a pravděpodobnosti protokolu pro model WLS.

### Aspekty dat odhadu váhy

**Data.** Závislé a nezávislé proměnné by měly být kvantitativní. Kategorické proměnné, jako např. náboženství, hlavní nebo region trvalého pobytu, je třeba rekódovat do binárních (fiktivních) proměnných

nebo jiných typů kontrastních proměnných. Proměnná váhy by měla být kvantitativní a měla by souviset s variabilitou v závislé proměnné.

**Předpoklady.** Pro každou hodnotu nezávislé proměnné musí být distribuce závislé proměnné normální. Vztah mezi závislou proměnnou a každou nezávislou proměnnou by měl být lineární a všechny pozorování by měly být nezávislé. Rozdíl závislé proměnné se může lišit v různých úrovních nezávislé proměnné, ale rozdíly musí být předpověditelné na základě proměnné váhy.

**Související postupy.** K zobrazení dat lze použít proceduru prozkoumání. Průzkum poskytuje testy pro normality a homogenitu rozptylu a také grafické zobrazení. Pokud se zdá, že vaše závislá proměnná má stejný rozptyl napříč úrovněmi nezávislých proměnných, můžete použít lineární regresní proceduru. Pokud se zdá, že vaše data porušují předpoklad (jako je normalnost), pokuste se je transformovat. Pokud vaše data nejsou ve spojení lineárně a transformace nepomáháme, použijte alternativní model v proceduře odhadu křivky. Je-li vaše závislá proměnná dichotomózní (například, zda je konkrétní prodej dokončen nebo zda je položka vadná), použijte Logistický regresní postup. Pokud je vaše závislá proměnná cenzurovaná (například doba přežití po operaci), použijte Tabulky Life Tables, Kaplan-Meier nebo Cox Regression, které jsou dostupné ve vlastních tabulkách a pokročilých statistikách. Nejsou-li vaše data nezávislá (například, pokud sledujete stejnou osobu za několika podmínek), použijte proceduru Opakované míry, která je k dispozici ve vlastních tabulkách a v rozšířené statistice.

## Získání analýzy odhadu váhy

1. Z nabídky vyberte:

**Analyzovat > Regrese > Odhad váhy ...**

2. Vyberte jednu závislou proměnnou.

3. Vyberte jednu nebo více nezávislých proměnných.

4. Vyberte proměnnou, která je zdrojem heterocedasticity jako proměnné váhy.

### **Proměnná váhy**

Data jsou váženy recipročním z této proměnné umocené na mocnina. Regresní rovnice se vypočítává pro každý zadaný rozsah hodnot výkonu a udává sílu, která maximalizuje funkci pravděpodobného logaritmicko-pravděpodobnosti.

### **Rozsah výkonu**

Tento parametr se používá ve spojení s proměnnou váhy k výpočtu vah. Několik regresních rovnic bude vhodné, jeden pro každou hodnotu v rozsahu výkonu. Hodnoty zadané do pole Test rozsahu Power a přes textové pole musí být mezi -6.5 a 7.5 včetně. Hodnoty napájení se pohybují od nízké až vysoké hodnoty, v přírůstcích určených hodnotou uvedenou. Celkový počet hodnot v rozsahu výkonu je omezen na 150.

## Volby odhadu váhy

Můžete určit volby pro analýzu odhadu váhy:

**Uložit nejlepší váhu jako novou proměnnou.** Přidává váhovou proměnnou do aktivního souboru. Tato proměnná se nazývá *WGT\_n*, kde *n* je číslo, které je zvoleno tak, aby proměnná poskytla jedinečný název.

**Zobrazení ANOVA a odhadů.** Umožňuje vám řídit, jak jsou statistiky zobrazeny ve výstupu. Dostupné alternativy jsou pro nejlepší napájení a pro každou hodnotu napájení.

## Další funkce příkazu WLS

Jazyk syntaxe příkazu vám také umožňuje:

- Poskytněte jednu hodnotu pro napájení.
- Uveďte seznam hodnot výkonu, nebo smíchejte rozsah hodnot se seznamem hodnot pro napájení.

Úplné informace o syntaxi najdete v příručce *Command Syntax Reference*.

## Regrese dvoufázového nejmenších čtverců

Standardní lineární regresní modely předpokládají, že chyby v závislé proměnné jsou nekorelované s nezávislými proměnnými ( $s$ ). Když se nejedná o tento případ (například když jsou vztahy mezi proměnnými obousměrné), lineární regrese pomocí obyčejných nejmenších čtverců (OLS) již neposkytuje optimální odhad modelu. Regrese druhé fáze po druhé fázi používá instrumentální proměnné, které nejsou korelovány s chybovými podmínkami pro výpočet odhadnutých hodnot problematických prediktorů (první fáze), a pak tyto vypočtené hodnoty používá k odhadu lineárního regresního modelu závislé proměnné (druhá fáze). Vzhledem k tomu, že vypočtené hodnoty jsou založeny na proměnných, které nejsou korelovány s chybami, jsou výsledky dvoustupňového modelu optimální.

**Příklad.** Je poptávka po komoditě spojená s jeho cenou a příjmy spotřebitelů? Potíž v tomto modelu spočívá v tom, že cena a poptávka mají vzájemný vliv na sebe navzájem. To znamená, že cena může ovlivnit poptávku a poptávku může také ovlivnit cenu. Dvoufázový regresní model s nejnižší hodnotou by mohl použít příjmy spotřebitelů a nejednotkovou cenu k výpočtu serveru proxy za cenu, která není korelována s chybami měření v poptávce. Tento server proxy je nahrazen za cenu sám v původně zadaném modelu, který se poté odhadne.

**Statistika.** Pro každý model: standardizované a nestandardizované regresní koeficienty, vícenásobné  $R$ ,  $R^2$ , upravené  $R^2$ , směrodatná chyba tabulky odhadu, analýzy rozptylového systému, předpokládané hodnoty a zbytkové chyby. Také, 95% intervaly spolehlivosti pro každý regresní koeficient, a matice korelací a kovariance odhadů parametrů.

### Pokyny pro regresní data dvou fází nejmenších čtverců

**Data.** Závislé a nezávislé proměnné by měly být kvantitativní. Kategorické proměnné, jako např. náboženství, hlavní nebo region trvalého pobytu, je třeba rekódovat do binárních (fiktivních) proměnných nebo jiných typů kontrastních proměnných. Vysvětlující proměnné *Endogenous* by měly být kvantitativní (ne kategorické).

**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.

**Související postupy.** Pokud se domníváte, že žádná z vašich proměnných prediktorů není korelována s chybami ve vaší závislé proměnné, můžete použít lineární regresní proceduru. Pokud se zdá, že vaše data poruší jeden z předpokladů (například normalita nebo konstanta rozptylu), zkuste je transformovat. Pokud vaše data nejsou ve spojení lineárně a transformace nepomáháme, použijte alternativní model v proceduře odhadu křivky. Je-li vaše závislá proměnná dichotomózní, jako např. zda je určitý prodej dokončen či nikoli, použijte Logistický regresní postup. Pokud vaše data nejsou nezávislá, například, pokud sledujete stejnou osobu za několika podmínek -- použijte proceduru Opakované měření.

### Získání dvoufázové analýzy regrese s nejmenších čtverečků

1. Z nabídky vyberte:

**Analyzovat > Regrese > 2-Fáze Nejméně čtverců ...**

2. Vyberte jednu závislou proměnnou.

3. Vyberte jednu nebo více vysvětlujících proměnných (prediktorů).

4. Vyberte jednu nebo více instrumentálních proměnných.

- *Instrumentální.* Jedná se o proměnné použité k výpočtu předpovídaných hodnot pro endogenní proměnné v první fázi dvoustupňové analýzy nejmenších čtverců. Stejně proměnné se mohou objevit jak ve vysvětlivkách, tak v seznamu Instrumental. Počet instrumentálních proměnných musí být přinejmenším stejný jako počet vysvětlujících proměnných. Jsou-li všechny uvedené vysvětlující a instrumentální proměnné stejné, výsledky jsou stejné jako výsledky z lineární regrese.

Vysvětlující proměnné nejsou uvedeny jako instrumentální jsou považovány za endogenní. Obvykle jsou všechny exogenní proměnné ve Vysvětlivčím seznamu uvedeny také jako pomocné proměnné.

## Dvoufázové volby regrese nejmenších čtverců

Pro svou analýzu můžete vybrat následující volby:

**Uložit nové proměnné.** Umožňuje vám přidat nové proměnné do aktivního souboru. Dostupné možnosti jsou Predicted a Residuals.

**Zobrazte kovarianci parametrů.** Umožňuje vám tisknout kovarianční matici odhadů parametrů.

## Doplňující funkce příkazu 2SLS

Jazyk syntaxe příkazu vám také umožňuje odhad více rovnic najednou. Úplné informace o syntaxi naleznete v příručce *Odkaz na syntaxi příkazu*.

## Kategorické schéma kódování proměnných

V mnoha procedurách můžete požádat o automatické nahrazení kategorické nezávislé proměnné sadou kontrastních proměnných, které pak budou zadány nebo odebrány z rovnice jako blok. Můžete uvést, jak se má sada kontrastních proměnných kódovat, obvykle na dílčím příkazu CONTRAST. Tato příloha vysvětluje a ilustruje, jak různé typy kontrastu požadované na CONTRAST skutečně fungují.

## Odchylka

**Odchylka od střední hodnoty.** V maticových termínech mají tyto kontrasty formu:

```
mean ( 1/k  1/k  ...  1/k  1/k)
df(1) (1-1/k -1/k  ... -1/k -1/k)
df(2) ( -1/k 1-1/k ... -1/k -1/k)
.
.
df(k-1) ( -1/k -1/k  ... 1-1/k -1/k)
```

kde  $k$  je počet kategorií pro nezávislou proměnnou a poslední kategorie je při výchozím nastavení vynechána. Například odchylka kontrastuje pro nezávislou proměnnou se třemi kategoriemi:

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3  2/3 -1/3)
```

Chcete-li vynechat kategorii jinou než poslední, uveďte číslo vynechané kategorie v závorkách za klíčovým slovem DEVIATION. Následující dílčí příkaz například získá odchylky pro první a třetí kategorii a vynechá sekundu:

```
/CONTRAST (FACTOR)=DEVIATION(2)
```

Předpokládejme, že *factor* má tři kategorie. Výsledná matice kontrastu bude

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3  2/3)
```

## Jednoduchá

**Jednoduché kontrasty.** Porovná každou úroveň faktoru  $k$  poslednímu. Obecný maticový formulář je

```
mean (1/k  1/k  ...  1/k  1/k)
df(1) ( 1  0  ...  0  -1)
df(2) ( 0  1  ...  0  -1)
.
.
df(k-1) ( 0  0  ...  1  -1)
```

kde  $k$  je počet kategorií pro nezávislou proměnnou. Příklad: Jednoduché kontrasty pro nezávislou proměnnou se čtyřmi kategoriemi jsou následující:

```
(1/4  1/4  1/4  1/4)
( 1  0  0  -1)
( 0  1  0  -1)
( 0  0  1  -1)
```

Chcete-li místo poslední kategorie použít jinou kategorii, určete v závorkách za klíčovým slovem SIMPLE pořadové číslo referenční kategorie, což nemusí být nutně hodnota přidružená k dané kategorii. Například následující dílčí příkaz CONTRAST získá kontrastní matici, která vynechá druhou kategorii:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Předpokládejme, že *faktor* má čtyři kategorie. Výsledná matice kontrastu bude

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 -1 1 0)
( 0 -1 0 1)
```

## Helmert

**Pomoci kontrastů.** Porovnává kategorie nezávislé proměnné se střední hodnotou následných kategorií. Obecný maticový formulář je

```
mean (1/k 1/k ... 1/k 1/k 1/k)
df(1) ( 1 -1/(k-1) ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) ( 0 1 ... -1/(k-2) -1/(k-2) -1/(k-2))
.
.
df(k-2) ( 0 0 ... 1 -1/2 -1/2)
df(k-1) ( 0 0 ... 0 1 -1)
```

kde  $k$  je počet kategorií nezávislé proměnné. Např. nezávislá proměnná se čtyřmi kategoriemi má matici kontrastu Helmert v následujícím tvaru:

```
(1/4 1/4 1/4 1/4)
( 1 -1/3 -1/3 -1/3)
( 0 1 -1/2 -1/2)
( 0 0 1 -1)
```

## Rozdíl

**Rozdíl nebo inverzní funkce helmert kontrastů.** Porovnává kategorie nezávislé proměnné se střední hodnotou předchozích kategorií proměnné. Obecný maticový formulář je

```
mean ( 1/k 1/k 1/k ... 1/k)
df(1) ( -1 1 0 ... 0)
df(2) ( -1/2 -1/2 1 ... 0)
.
.
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

kde  $k$  je počet kategorií pro nezávislou proměnnou. Rozdíl mezi nezávislými proměnnými a čtyřmi kategoriemi je například následující:

```
( 1/4 1/4 1/4 1/4)
( -1 1 0 0)
( -1/2 -1/2 1 0)
( -1/3 -1/3 -1/3 1)
```

## Polynomiální

**Ortogonalní polynomiální kontrasty.** První stupeň svobody obsahuje lineární účinek napříč všemi kategoriemi; druhý stupeň svobody, kvadratický efekt, třetí stupeň svobody, kubické atd. a tak dále, pro účinky na vyšší úrovni.

Můžete zadat velikost mezer mezi úrovněmi zpracování měřenou danou kategorickou proměnnou. Rovnoměrné řádkování, které je výchozí, pokud vynecháte metriku, může být uvedeno jako následná celá čísla od 1 do  $k$ , kde  $k$  je počet kategorií. Má-li proměnná *drug* tři kategorie, dílčí příkaz

```
/CONTRAST(DRUG)=POLYNOMIAL
```

je stejné jako

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Rovnoměrné řádkování však není vždy nutné. Předpokládejme například, že *droga* představuje různé dávky drogy podávané ve třech skupinách. Pokud je dávka podaná druhé skupině je dvojnásobek, který byl podáván první skupině a dávka podaná ke třetí skupině je trojnásobek, který byl podáván první skupině, kategorie ošetření jsou rovnoměrně rozloženy, a odpovídající metrický pro tuto situaci se skládá z po sobě jdoucích celá čísla:

```
/CONTRAST (DRUG)=POLYNOMIAL (1, 2, 3)
```

Je-li však dávka podávaná druhé skupině čtyřikrát vyšší než dávka podaná do první skupiny a dávka podaná ke třetí skupině je sedmkrát vyšší než dávka podaná k první skupině, je vhodnou metrikou následující:

```
/CONTRAST (DRUG)=POLYNOMIAL (1, 4, 7)
```

V obou případech je výsledkem kontrastní specifikace to, že první stupeň svobody pro *látku* obsahuje lineární účinek úrovně dávkování a druhý stupeň svobody obsahuje kvadratický účinek.

Polynomiální kontrasty jsou zvláště užitečné při zkouškách trendů a pro zkoumání charakteru povrchů odpovědí. Můžete také použít polynomiální kontrasty k provedení nelineární prokládání křivky, jako např. curvilinear regrese.

## Opakované

**Porovná přílehlé úrovně nezávislé proměnné.** Obecný maticový formulář je

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      :
      :
df(k-1) ( 0 0 0 ... 1 -1)
```

kde  $k$  je počet kategorií pro nezávislou proměnnou. Např. opakované kontrasty pro nezávislou proměnnou se čtyřmi kategoriemi jsou následující:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Tyto kontrasty jsou užitečné při analýze profilu a všude tam, kde je třeba rozlišovat skóre.

## Jiný

**Kontrast definovaný uživatelem.** Umožňuje vstup speciálních kontrastů ve formě čtvercových matic s tolika řádky a sloupci, jak jsou tam kategorie dané nezávislé proměnné. Pro MANOVA a LOGLINEAR je první zadaný řádek vždy střední nebo konstantní, efekt a představuje sadu vah ukazujících, jak v dané proměnné průměrné jiné nezávislé proměnné, jsou-li nějaká. Tento kontrast je obvykle vektor jedniček.

Zbývající řádky matice obsahují speciální kontrasty označující porovnání mezi kategoriemi proměnné. Obvykle jsou ortogonální kontrasty nejužitečnější. Ortogonální kontrasty jsou statisticky nezávislé a jsou neredundantní. Kontrasty jsou ortogonální, pokud:

- Pro každý řádek, kontrastní koeficienty součet na 0.
- Produkty odpovídajících koeficientů pro všechny páry disjunktní řádky také součet na 0.

Například předpokládejme, že léčba má čtyři úrovně a že chcete porovnat různé úrovně léčby s sebou navzájem. Vhodný speciální kontrast je

```
(1 1 1 1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0 2 -1 -1) compare 2nd with 3rd and 4th
(0 0 1 -1) compare 3rd with 4th
```

keré určíte pomocí následujícího dílčího příkazu CONTRAST pro MANOVA, LOGISTIC REGRESSION a COXREG:

```
/CONTRAST (TREATMNT)=SPECIAL ( 1 1 1 1
                                3 -1 -1 -1
                                0 2 -1 -1
                                0 0 1 -1 )
```

Pro LOGLINEAR musíte uvést:

```
/CONTRAST (TREATMNT)=BASIS SPECIAL ( 1 1 1 1
                                       3 -1 -1 -1
                                       0 2 -1 -1
                                       0 0 1 -1 )
```

Každý řádek kromě řádků znamená, že součet řádků je roven 0. Produkty každého páru disjunktní řádky součtu 0 stejně:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

Speciální kontrasty nemusí být ortogonální. Nesmí však být lineární kombinací navzájem. Pokud tomu tak je, procedura nahlásí lineární závislost a ukončí zpracování. Helmert, rozdíl, a polynomiální kontrasty jsou všechny kolmé kontrasty.

## Indikátor

**Kódování proměnné indikátoru.** Také známý jako fiktivní kódování, to není k dispozici v LOGLINEAR nebo MANOVA. Počet nových kódovaných proměnných je  $k-1$ . Případy v referenční kategorii jsou kódovány 0 pro všechny proměnné  $k-1$ . Případ v kategorii  $n^n$  je kódován 0 pro všechny proměnné indikátoru kromě  $n^n$ , které je kódováno 1.

## Regrese jádra

Metoda Kernel Ridge Regression je procedura rozšíření, která používá třídu Python **sklearn.kernel\_ridge.KernelRidge** k odhadu regresních modelů jádra. Modely regrese mostu jádra jsou neparametrické regresní modely, které jsou schopné modelovat lineární a nelineární vztahy mezi proměnnými prediktoru a výsledky. Výsledky mohou být velmi citlivé na volby hyperparametrů modelu. Metoda Kernel Ridge Regression usnadňuje volbu hodnot hyperparametru k ověření křížových ověření platnosti zadaných mřížek hodnot s použitím třídy **skarin.model\_selection.GridSearchCV**.

### Příklad

#### Statistika

Additive\_CHI2, CHI2, Coline, Laplacian, Linear, Polynomial, RBF, Sigmoid, Alpha, Gamma, Coef0, Degrey, crossvalidation, pozorovaných versus predicted, dual als coefficients, dual weight coefficients, kernel space, efficients coefficients, varial coefficients, coultily versus predicted

### Aspekty dat

#### Data

- Můžete zadat libovolnou nebo všechny z osmi různých funkcí jádra.
- Vybraná funkce jádra určuje, které hyperparametry jsou aktivní.
- Hyperparametry zahrnují alfa pro regularizaci ridge, které jsou společné pro všechna jádra a tři další hyperparametry pro každou specifickou funkci jádra.
- Je-li určeno více dílčích příkazů jádra nebo je zadána více než jedna hodnota některého parametru, provede se prohledávání mřížky s křížovým ověřením pro vyhodnocení modelů a výběr nejvhodnějšího modelu, který je založen na zadržených datech.
- Rozšíření přijímá dělené proměnné z procedury Rozdělit soubor a vah pomocí procedury Váhy s váhou.

- Jsou-li zahrnuty váhy, používají se při vytváření hodnot naložené ve všech analýzách. Kvůli omezením v metodě score ve třídě **skurn.model\_selection.GridSearchCV** nejsou vyhodnocení vzájemné validace, která se používají pro výběr modelu, vážena.

## Předpoklady

### Získání regresní regrese jádra

1. Z nabídky vyberte:

**Analýza > Regrese > Kernel Ridge ...**

2. Vyberte proměnnou **Dependent** .

3. Vyberte jednu nebo více proměnných **Nezávislé (é)** .

4. Výchozí nastavení **Jeden model** se použije tehdy, když je uvedena pouze jedna hodnota pro každý parametr funkce jádra. Když je zvoleno nastavení **Jeden model** , nemůžete uvést další funkce **Kernel (s)** a váhové váhy jsou plně uplatňovány v rámci analýzy, vyhodnocení a hodnocení výsledků. Chcete-li změnit uspořádání funkcí jádra, můžete také použít ovládací prvky se šipkami nahoru a dolů.

Volitelně můžete vybrat volbu **Výběr modelu** ze seznamu **Režim** .

Je-li vybrána volba **Výběr modelu** ze seznamu **Režim** , můžete do seznamu **Kernel (s)** přidat více funkcí jádra.

a. Chcete-li zahrnout další funkce jádra, klepněte na ovládací prvek pro přidání (+).

b. Klepněte na prázdnou buňku ve sloupci **Kernel** a vyberte funkci jádra.

c. Poklepáním na libovolnou buňku funkce jádra můžete zadat hodnoty parametrů funkce jádra pro příslušný sloupec (**Alfa, Gamma, Coef0, stupeň**). Více informací naleznete v části “parametry jádra” na stránce 29. Niže jsou uvedeny výchozí parametry pro vyladění funkcí jádra.

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

#### **Kosinus**

ALPHA=1

#### **Laplaciština**

ALPHA=1 GAMMA=1/p

#### **Lineární**

Výchozí funkce jádra. ALPHA=1

#### **Polynomní**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

#### **RBF**

ALPHA=1 GAMMA=1/p

#### **ID podpisu**

ALPHA=1 GAMMA=1/p COEF0=1

**Poznámka:** Je-li pro každý parametr funkce jádra zadána více než jedna hodnota, provede se vyhledávání v mřížce s křížovým ověřením pro vyhodnocení modelů a je vybrán ten nejlepší model, který je založen na zadržených datech.

5. Volitelně klepněte na tlačítko **Volby** , abyste uvedli počet převaděče, volby zobrazení, nastavení výkresu a položky k uložení. Více informací naleznete v části “Regrese jádra jádra: Volby” na stránce 29.

6. Klepněte na tlačítko **OK**.



## parametry jádra

Dialog **Parametry jádra** poskytuje volby pro určení jednotlivých hodnot parametrů funkce jádra a pro určení, že výběr modelu se provádí pomocí vyhledávání v mřížce nad kombinací jader a uvedených hodnot parametrů mřížky.

### Zadejte jednotlivé parametry

Povolte nastavení pro uvedení hodnot pro vybraný parametr funkce jádra.

- Zadejte hodnotu a klepněte na **Přidat**, abyste zahrnuli hodnotu do parametru funkce jádra.
- Vyberte hodnotu parametru a klepněte na tlačítko **Změnit**, chcete-li aktualizovat hodnotu.
- Vyberte hodnotu parametru a klepněte na tlačítko **Odebrat**, chcete-li odstranit hodnotu.

### Uřčit parametry mřížky

Povolením nastavení určete, že výběr modelu se provádí pomocí vyhledávání v mřížce nad kombinací jader a zadanými hodnotami parametrů mřížky.

## Regrese jádra jádra: Volby

Dialogové okno **Plot** poskytuje volby pro uvedení počtu přehybů, voleb zobrazení, nastavení výkresu a položek k uložení.

### Počet záhyby křížové validace

Počet rozdělení nebo záhyby v křížové validaci s výběrovým mřížkou pro výběr modelu. Zadejte celočíselnou hodnotu větší než 1. Výchozí hodnota je 5. Nastavení je k dispozici pouze v případě, že je vybrána volba **Výběr modelu** jako **Režim** v primárním dialogovém okně **Regrese v jádru jádra**.

### Zobrazit

Poskytuje volby pro určení, který výstup se má zobrazit, když je v platnosti křížové ověření platnosti.

#### Nejlepší

Výchozí nastavení zobrazí pouze základní výsledky pro vybraný nejlepší model.

#### Porovnat

Zobrazí základní výsledky pro všechny vyhodnocené modely.

#### Porovnat modely a záhyby

Zobrazí úplné výsledky pro každé rozdělení nebo rozdělení pro každý hodnocený model.

### bodová

Poskytuje volby pro uvedení zakreslených nebo zbytkových hodnot versus předpovězené hodnoty.

#### Sledováno vs. předpovězeno

Zobrazí bodový graf pozorovaných versus předpokládaných hodnot pro uvedený nebo nejlepší model.

#### Reziduy vs. predicted

Zobrazí bodový graf zbytkových chyb oproti předpovězeným hodnotám pro uvedený nebo nejlepší model.

### Uložit

Tabulka poskytuje volby pro uvedení proměnných, které se mají uložit do aktivní datové sady.

#### Předpovězené hodnoty

Uloží předpokládané hodnoty z uvedeného nebo nejlepšího modelu do aktivní datové sady. Může být zahrnut volitelný název proměnné.

#### Reziduum

Uloží zbytkové chyby z uvedených nebo nejlepších predikcí modelu do aktivní datové sady. Může být zahrnut volitelný název proměnné.

#### Duální koeficienty

Uloží duální nebo jádrový koeficient prostoru jádra z uvedeného modelu do aktivní datové sady. Může být zahrnut volitelný název proměnné. Nastavení není k dispozici, je-li vybrána volba **Výběr modelu** jako **Režim** v primárním dialogovém okně **Regrese v jádru jádra**.

## Modely s parametrickým akceorickým časem

Analýza modelu AFT (Parametric Accelerated Failure Time) vyvolává parametrický postup modelového přežití s neopakujícími se daty doby životnosti. Modely s parametrickým přežitím předpokládají, že doba přežívání následuje po známém rozdělení, a tato analýza se hodí k urychlenému modelům doby selhání s jejich modelové účinky úměrné času přežití.

### Získání analýzy parametrického modelu urychlených poruch

1. Z nabídky vyberte:

**Analyzovat > Přežití > Modely parametrického akcelerované doby selhání (AFT)**

2. Vyberte zdrojovou proměnnou.

#### Čas

##### Přežití

Jedna číselná proměnná, která udává délku doby přežití.

##### Spustit/Ukončit

Numerické proměnné označující **Počáteční čas** a **Koncový čas**.

#### Stav

Jednotlivý volitelný řetězec nebo číselná proměnná, která určuje jedno z následujících nastavení stavu:

##### Selhání/událost

Mapuje záznam do kategorie Selhání/události. Výchozí hodnota pro řetězcovou stavovou proměnnou je F.

##### Pravé cenzura

Mapuje záznam na správnou cenzurovanou kategorii. Výchozí hodnota pro řetězcovou stavovou proměnnou je R.

##### Levé cenzura

Mapuje záznam na kategorii vlevo cenzurování. Výchozí hodnota pro řetězcovou stavovou proměnnou je L.

##### Cenzření intervalu

Mapuje záznam do kategorie cenzurování intervalu. Pouze pro **Spustit/Ukončit**. Výchozí hodnota pro proměnnou stavu řetězce je I.

##### Zpracování nemapovaných hodnot

Ovládá kategorii, na kterou se mají mapovat nemapované záznamy. Chcete-li odstranit záznamy, které se nezdařilo namapovat, vyberte volbu **Vyloučit je z analýzy**.

Pro **Survival** je výchozí stav pro všechny případy **Selhání/Událost**. Pro **Spustit/Ukončit** je výchozí stav **Interval centrování intervalu**. Klepněte na tlačítko **Definovat událost**, chcete-li definovat událost pro stavovou proměnnou.

#### Kovariáta (proměnné)

Jedna nebo více volitelných číselných proměnných, které mají být považovány za kovariány. Všimněte si, že proměnná nemůže být zadána **Cvariate (s)** a **Fixed Factor (s)**.

#### Pevný faktor (y)

Jedna nebo více volitelných proměnných, které mají být považovány za faktory. Proměnná nemůže být zadána pomocí **Pevný faktor (y)** a **Kovariáta (proměnné)**.

#### Oříznutí vlevo

Jediná volitelná číselná proměnná pro oříznutí vlevo pouze pro **přežití**.

## Parametrický akcelerovaný model času selhání: Kritéria

### Podmínky

Volitelný panel k určení obecných kritérií.

### Interval spolehlivosti

Nepovinná procentní část k určení úrovně pro intervaly spolehlivosti regresních parametrů. Musí se jednat o jedinou číselnou hodnotu v rozsahu 0 až 100. Předvolba je 95.

### Chybějící hodnoty

Možnost řídit, jak se zachází s uživatelem s chybějícími hodnotami:

#### Vyloučit jak uživatelské, tak systémové chybějící hodnoty

Zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami. Jedná se o výchozí nastavení.

#### Uživatel-chybějící hodnoty jsou považovány za platné

Ignoruje uživatele, který postrádá hodnotu, a považuje je za platné hodnoty.

### Ošetření stavu

Pouze pro **Spustit/Ukončit**. Volba pro řízení způsobu práce se záznamy s chybnými stavovým polím:

#### Vyřadit konfliktní záznam

Uvolní konfliktní záznamy. Toto je výchozí nastavení.

#### Získat informace o čase v závislosti na stavu

Získává informace o čase podle stavu.

#### Odvodit stav podle informací o čase

Změní stav podle informací o čase.

## Modely s parametrickým akceleorickým časem: Model

### Model

Volitelný panel k určení voleb a nastavení modelu.

### Doba přežívání

Volba pro uvedení distribuce doby přežití.

#### WEIBULL

Uvádí Weibullovo rozdělení. Toto je výchozí nastavení.

#### Exponenciální

Určuje exponenciální rozdělení.

#### Protokol-normální

Určuje protokol-normální rozdělení.

#### Protokol-Logistický

Určuje logistickou distribuci protokolu.

### Nastavení proměnné Covariate

Určete proměnné kovariate.

### Nastavení faktoru

Určete proměnné faktoru.

### Počáteční hodnota zachycení

Volba pro určení počáteční hodnoty výrazu zachycení. Je-li tento parametr zadán, musí se jednat o jedinou číselnou hodnotu a nesmí mít hodnotu 0.

### Počáteční hodnota parametru stupnice

Volba pro řízení nastavení parametru scale.

### Standardní chyba odpovídající regrese OLS

Použije standardní chybu odpovídající regresní přímky nejmenších čtverců jako počáteční hodnotu.

### Neverše standardní chyba odpovídající regrese OLS

Používá převrácená hodnota směrodatné chyby.

### Uživatелеm zadaná hodnota

Je-li zadána jedna číselná hodnota, bude jako počáteční hodnota použita hodnota. Je-li tento parametr zadán, musí být větší než 0.

## Modely s parametrickým akcelerickým časem selhání: Odhad

### Odhad

Volitelný panel k určení nastavení pro kontrolu odhadu zrychlených modelů času selhání a volitelného procesu výběru funkcí.

### Střídavá metoda nebo vícenásobné kleště (ADMM)

#### Rychlý

Použije se rychle střídavý směr násobičů (ADMM). Jedná se o výchozí nastavení.

#### Tradiční

Použije tradiční algoritmus ADMM.

### Použití regularizace L-1

Ukonává proces k řízení výběru funkcí. Pole **Parametr penále** uvádí parametr penalty, který řídí regularizační proces. Musí se jednat o jedinou hodnotu větší než 0. Předvolené nastavení je 0.001.

### Vzorová kritéria konvergence

#### Přeměna parametru

Uvádí konvergenční kritéria pro parametr. Musí se jednat o jedinou číselnou hodnotu patřící do [ 0, 1). Předvolené nastavení je 0.000001. Pro **Typ** můžete vybrat buď **ABSOLUTE** pro použití absolutní konvergence na vnitřní optimalizaci, nebo **RELATIVE**, chcete-li použít relativní konvergenci na vnitřní optimalizaci. Volitelné pole **Hodnota** určuje klíčové slovo.

#### Cíl "Konvergence"

Určuje kritéria konvergence pro funkci cíle. Musí se jednat o jedinou číselnou hodnotu patřící do [ 0, 1). Výchozí nastavení je 0, které neuplatňuje kritéria konvergence. Pro **Typ** můžete vybrat buď **ABSOLUTE** pro použití absolutní konvergence na vnitřní optimalizaci, nebo **RELATIVE**, chcete-li použít relativní konvergenci na vnitřní optimalizaci. Volitelné pole **Hodnota** určuje klíčové slovo.

#### Hesská konvergence

Uvádí konvergenční kritéria pro Hessian matrix. Musí se jednat o jedinou číselnou hodnotu patřící do [ 0, 1). Výchozí nastavení je 0, které neuplatňuje kritéria konvergence. Pro **Typ** můžete vybrat buď **ABSOLUTE** pro použití absolutní konvergence na vnitřní optimalizaci, nebo **RELATIVE**, chcete-li použít relativní konvergenci na vnitřní optimalizaci. Volitelné pole **Hodnota** určuje klíčové slovo.

### Zbývající kritéria konvergence

Volba pro řízení procesu optimalizace.

#### Jak primitivní, tak dvojitý reziduum

Použije se jak primitivní, tak i dvojitý reziduální kritérium konvergence. Toto je výchozí nastavení.

#### Pouze zbytková zbytková hodnota

Použije se reziduální kritérium zbytkové konvergence.

#### Pouze dvojitý reziduum

Použije se zdvojené zbývající kritérium konvergence.

### Metoda

Nepovinný parametr k určení metody odhadu.

#### Automatický

Automaticky zvolí metodu založenou na ukázkové datové sadě. Jedná se o výchozí nastavení. Pole **Prahová hodnota prediktorů** určuje prahovou hodnotu počtu prediktorů a musí se jednat o jediné celé číslo větší než 1. Výchozí hodnota je 1000.

### **Newton-Raphson**

Použije se Newtona-Raphsonův způsob.

### **L-BFGS**

Použije se algoritmus BFGS s omezenými paměti. Pole **Aktualizovat** určuje počet minulých aktualizací spravovaných algoritmem BFGS s omezenou pamětí a musí být jediné celé číslo větší než nebo rovné 1. Výchozí hodnota je 5.

## **Iterace**

### **Maximum iterací**

Určuje maximální počet iterací. Musí se jednat o jediné celé číslo náležící uživateli [ 1, 100].  
Předvolené nastavení je 20.

### **Maximální krok-halving**

Určuje maximální počet kroků-halving. Musí se jednat o jediné celé číslo náležící uživateli [ 1, 20].  
Výchozí nastavení je 5.

### **Maximální počet hledání řádků**

Určuje maximální počet hledání řádku. Musí se jednat o jediné celé číslo náležící uživateli [ 1, 100].  
Předvolené nastavení je 20.

### **Absolutní konvergence pro proces iterace**

Určuje absolutní konvergenci pro proces vnější iterace. Musí se jednat o jedinou číselnou hodnotu, která patří (0, 1). Výchozí nastavení je 0.0001.

### **Relativní konvergence pro proces iterace**

Určuje relativní konvergenci pro proces vnější iterace. Musí se jednat o jedinou číselnou hodnotu, která patří (0, 1). Výchozí nastavení je 0.01.

## **Parametrické akcelerované modely času selhání: Tisk**

### **Tisk**

Volitelný panel pro ovládání výstupů tabulky.

### **Podrobnosti kódování faktoru**

Je-li vybráno, zobrazí se a vytiskne podrobnosti o kódování faktorů. Pokud v platnosti nejsou žádné faktory, je proces ignorován.

### **Počáteční hodnoty přiřazené regresním parametrům**

Je-li vybráno, zobrazí se počáteční hodnoty použité v procesu odhadu.

### **Historie iterace modelu**

Je-li vybráno, zobrazí historii iterací analýzy přežití. V poli **Počet kroků** zadejte počet kroků mezi 1 a 99999999. Výchozím nastavením je hodnota 1.

### **Výsledky výběru obsahující**

Řídí zobrazení podrobností výběru funkcí.

### **Vybrané i nevybrané proměnné**

Zobrazit jak vybrané, tak nevybrané proměnné v tabulce.

### **Pouze vybrané proměnné**

Zobrazit pouze vybrané proměnné.

### **Pouze nevybrané proměnné**

Zobrazit pouze nevybrané proměnné. Pole **Maximální počet proměnných k zobrazení** uvádí maximální počet proměnných vytištěných v tabulce. Výchozí nastavení je 30.

## **Parametrické akcelerované modely času selhání: Predict**

### **Predict**

Volitelný panel pro skóre a uložení predikované statistiky do aktivní datové sady.

## Časové hodnoty pro přidělení skóre

### Časové hodnoty definované závislé proměnné (závislé proměnné)

Skóre **Predikce** založené na časové proměnné zadané pro parametrický model přežití.

### Pravidelné intervaly

Skóre **Předpovědi** je založena na budoucích hodnotách času. Pole **Časový interval** určuje časový interval a musí se jednat o jedinou číselnou hodnotu větší než 0. Pole **Počet časových období** uvádí počet časových období a musí se jednat o jediné číselné celé číslo mezi 2 a 100.

### Doba trvání

Skóre **Předpovědi** je založena na časovém trvání a definuje budoucí hodnoty času. Musí se jednat o jedinou číselnou proměnnou.

## Předpovědi

### Přežití

Skóre a ukládá předpovídanou statistiku přežití na aktivní datovou sadu. Výchozí název přizpůsobené proměnné (nebo název kořenového adresáře) je `PredSurvival`.

### Riziko

Skóre a uloží předpovězená rizika do aktivní datové sady. Výchozí název přizpůsobené proměnné (nebo kořenový název) je `PredHazard`.

### Kumulativní riziko

Skóre a uloží předpovězená kumulativní rizika do aktivní datové sady. Výchozí název přizpůsobené proměnné (nebo název kořenového adresáře) je `PredCumHazard`.

### Podmíněné přežití

Skóre a uloží predikované statistiky podmíněného přežití do aktivní datové sady. Výchozí název přizpůsobené proměnné (nebo kořenový název) je `PredConditionalSurvival`. Pokud není zadána hodnota `PASTTIME`, proces bude ignorován. Hodnota **Minulá doba přežití** je povinná a uvádí hodnoty v minulosti pro přidělení skóre. Musí se jednat o jedinou číselnou proměnnou.

## Modely s parametrickým akceleorickým selháním: Plot

### bodová

#### Cox-Snell reziduální pozemek

Vyberte volbu **Zobrazit zákresový graf**, chcete-li vytvořit reziduální graf Cox-Snell. V poli **Počet bodů pro počet sběrných míst** zadejte číslo od 1 do 10000. Předvolené nastavení je 100.

#### Funkce Plots

Volba pro řízení grafů funkcí.

#### Typ

##### Přežití

Vytvoří graf pro funkce přežití.

##### Riziko

Vytvoří zakreslení pro funkce rizika.

##### Hustota

Vytvoří zákresový graf pro funkce hustoty.

#### Počet zobrazovaných bodů

Určuje počet funkčních bodů mezi 1 a 200. Předvolené nastavení je 100.

#### Kovariovat hodnoty pro zákresový graf

Volitelný parametr k určení uživatelem zadaných hodnot a jejich přiřazení k predictorům. Ve výchozím nastavení budou určené vykreslované grafy vytvořeny na **Střed** každé proměnné v platnosti a frekvence kategorií každého faktoru v platnosti. Je-li zadán, budou určené grafy vytvořeny na základě nastavení vzoru. Při výskytu duplicitních proměnných se nejprve rozezná jedna zadaná a zbylé by se ignorovaly. Platná proměnná musí být obsažena ve vzorovém efektu. Pro proměnnou kovariate musí být uživatelem zadaná hodnota číselná. Omise proměnné v účinku označuje, že frekvence kategorií a **Střední hodnota** bude použita jako výchozí pro

faktor a kovariate, resp. Je-li k proměnné přiřazena neplatná hodnota, požadovaný vzorek nebude vykreslen.

### Hodnoty faktoru pro vykreslení

Volitelný parametr k určení uživatelem zadaných hodnot a jejich přiřazení k predictorům. Při výskytu duplicitních proměnných se nejprve rozezná jedna zadaná a zbylé by se ignorovaly. Platná proměnná musí být obsažena ve vzorovém efektu. Omise proměnné v účinku označuje, že frekvence kategorie a střední hodnota by byly použity jako výchozí pro faktor a kovariate, resp. Je-li k proměnné přiřazena neplatná hodnota, požadovaný vzorek nebude vykreslen.

### Samostatné řádky pro

Volba pro určení kategorické proměnné, podle které budou vykresleny spojnicové grafy.

### Maximální počet řádků v grafu

Určuje maximální počet řádků v grafu, pokud je zadán parametr **Oddělte řádky pro**. Předvolené nastavení je 10.

## Parametrické akcelerované modely času selhání: Export

### Exportovat

Vyberte volbu **Exportovat informace o modelu do souboru XML**, chcete-li zapsat informace o modelu a parametru do souboru PMML pro přidělení skóre. Musíte uvést adresář a název souboru PMML, který má být uložen.

## Přežití AFT definování událostí pro proměnné stavu

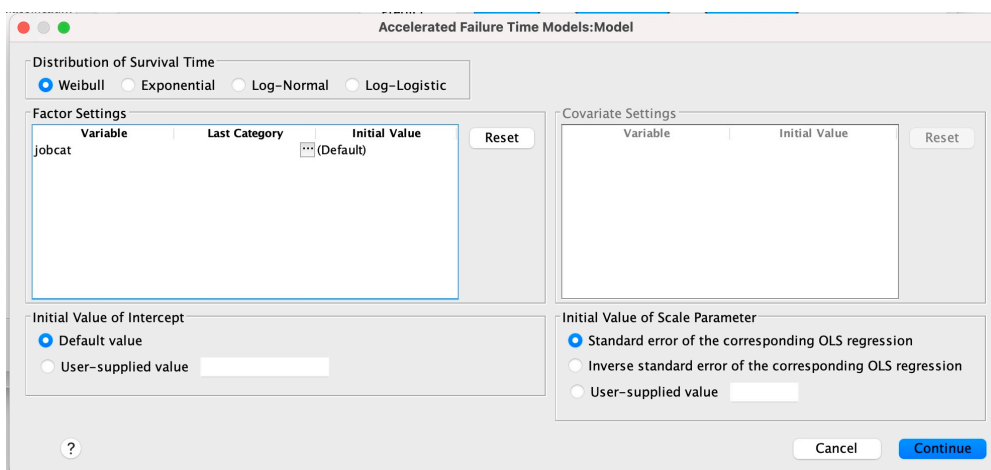
Výskyty zvolené hodnoty nebo hodnot pro stavovou proměnnou označují, že pro tyto případy došlo k události terminálu. Všechny ostatní případy se považují za cenzurované. Zadejte buď jednu hodnotu, nebo rozsah hodnot, které identifikují událost, o kterou se zajímáte.

## Parametrické akcelerované modely času selhání: Vyberte kategorii

Nastavení Výběr kategorie poskytuje volby pro výběr hodnoty, která označuje kategorii, která má být modelována jako základní úroveň pro porovnání.

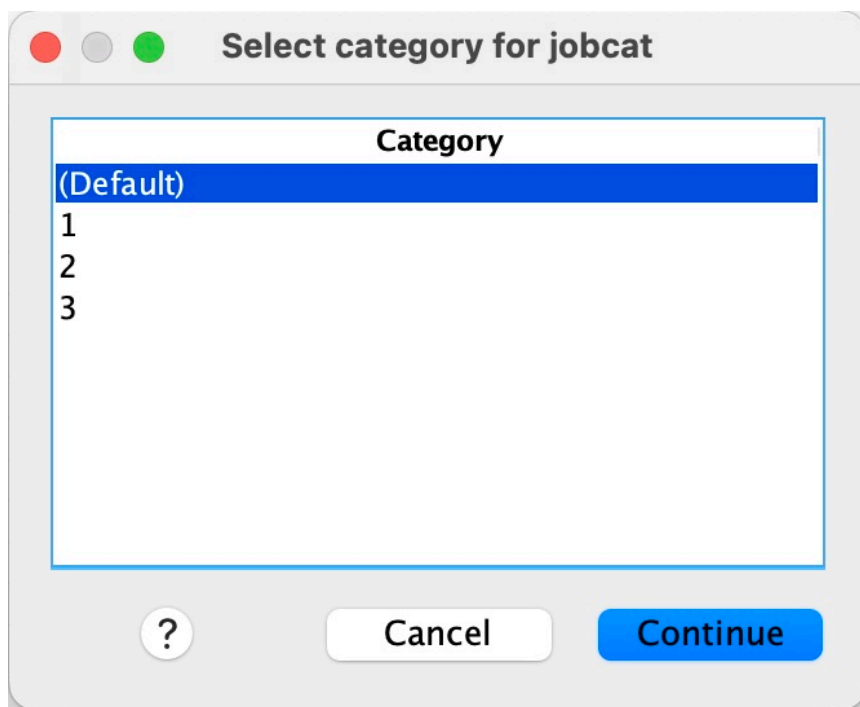
Výběr kategorie

Klepnutím na 'Poslední kategorie' otevřete dialogové okno 'Výběr kategorie'.



Obrázek 1. Model Accelerated Life Time Models-dialogové okno-Kategorie

Chcete-li označit kategorii jako úroveň baseline, vyberte hodnotu z dialogového okna 'Vybrat kategorii'.



Obrázek 2. Model Accelerated Life Time Models-Dialogové okno-Vybrat kategorii

Klepněte na Pokračovat.

## Parametrické sdílené chatrné modely

Parametrickou Sdílenou Frailty Models Survival analýza zahájí parametrický postup modelových modelů přežití s opakovaným vstupem do životního prostředí. Modely s parametrickým přežitím předpokládají, že doba přežití se řídí známým rozdělením, a tato analýza zahrnuje do parametrického modelu přežití s křehkou dobou trvání. S ohledem na individuální nebo skupinovou variabilitu se zachází jako s náhodnou komponentou kvůli nenáslednému účinku.

### Získání analýzy Parametrické sdílené modely zlotých modelů

1. Z nabídky vyberte:

**Analyzovat > Přežití > Parametrické sdílené modely-zlomky**

2. Vyberte zdrojovou proměnnou.

#### Čas

##### Přežití

Doba přežívání je znázorněna jednou proměnnou, která udává čas ukončení. Čas zahájení by byl nastaven na 0.

##### Spustit/Ukončit

Numerické proměnné, které označují hodnoty **Čas zahájení** a **Koncový čas**.

#### Předmět

Nezbytné pro spuštění procedury. Určuje jednu proměnnou pro ID subjektu.

#### Interval

Uvádí jednu a číselnou proměnnou pro číslo intervalu, které se používá k identifikaci různých rekurentních záznamů, které sdílejí stejné ID subjektu.



## Stav

Jednotlivý volitelný řetězec nebo číselná proměnná, která určuje jedno z následujících nastavení stavu:

### Selhání/událost

Mapuje záznam do kategorie Selhání/události. Výchozí hodnota pro řetězcovou stavovou proměnnou je F.

### Pravé cenzura

Mapuje záznam na správnou cenzurovanou kategorii. Výchozí hodnota pro řetězcovou stavovou proměnnou je R.

### Zpracování nemapovaných hodnot

Ovládá kategorii, na kterou se mají mapovat nemapované záznamy. Chcete-li odstranit záznamy, které se nezdařilo namapovat, vyberte volbu **Vyloučit je z analýzy**.

Klepněte na tlačítko **Definovat událost**, chcete-li definovat událost pro stavovou proměnnou.

## Kovariáta (proměnné)

Jedna nebo více volitelných číselných proměnných, které mají být považovány za kovariany. Všimněte si, že proměnná nemůže být zadána **Cvariate (s)** a **Fixed Factor (s)**.

## Pevný faktor (y)

Jedna nebo více volitelných proměnných, které mají být považovány za faktory. Proměnná nemůže být zadána pomocí **Pevný faktor (y)** a **Kovariáta (proměnné)**.

# Parametrické sdílené chatrné modely: Kritéria

## Podmínky

Volitelný panel k určení obecných kritérií.

### Interval spolehlivosti

Nepovinná procentní část k určení úrovně pro intervaly spolehlivosti regresních parametrů. Musí se jednat o jedinou číselnou hodnotu v rozsahu 0 až 100. Předvolba je 95.

### úroveň významnosti

Volba pro určení úrovně významnosti pro test poměru pravděpodobnosti pro komponentu frajnosti. Musí se jednat o jedinou číselnou hodnotu v rozsahu 0 až 1. Výchozí nastavení je 0.05.

### Chybějící hodnoty

Možnost řídit, jak se zachází s uživatelem s chybějícími hodnotami:

#### Vyloučit jak uživatelské, tak systémové chybějící hodnoty

Zachází s uživatelem s chybějícími hodnotami jako platnými hodnotami. Jedná se o výchozí nastavení.

#### Uživatel-chybějící hodnoty jsou považovány za platné

Ignoruje uživatele, který postrádá hodnotu, a považuje je za platné hodnoty.

## Ošetření intervalu

Volba pro řízení toho, jak se vypořádat se záznamy, jejichž interval je v konfliktu s počátečním a koncovým časem. To se projeví, pokud existují dvě časové proměnné s proměnnou Interval uvedenou v hlavním dialogu.

### Vyřadit konfliktní záznamy

Vyřadí všechny sériové záznamy předmětu, je-li hodnota intervalu v konfliktu s počátečním a koncovým časem. Toto je výchozí nastavení.

### Hodnoty intervalu zjišťování založené na času zahájení a ukončení

Vyhledá hodnotu intervalu od počátečního a koncového času.

## Parametrické sdílené chatrné modely: model

### Model

Volitelný panel k určení voleb a nastavení modelu.

#### Doba přežívání

Volba pro uvedení distribuce doby přežití.

##### **WEIBULL**

Uvádí Weibullovo rozdělení. Toto je výchozí nastavení.

##### **Exponenciální**

Uvádí Exponenciální rozdělení.

##### **Protokol-normální**

Určuje protokol-normální rozdělení.

##### **Protokol-Logistický**

Určuje logistickou distribuci protokolu.

#### Nastavení proměnné Covariate

Určete proměnné kovariate.

#### Nastavení faktoru

Určete proměnné faktoru.

#### Počáteční hodnota zachycení

Volba pro určení počáteční hodnoty výrazu zachycení. Je-li tento parametr zadán, musí se jednat o jedinou číselnou hodnotu a nesmí mít hodnotu 0.

#### Počáteční hodnota parametru stupnice

Volba pro řízení nastavení parametru scale.

##### **Standardní chyba odpovídající regrese OLS**

Použije standardní chybu odpovídající regresní přímky nejmenších čtverců jako počáteční hodnotu.

##### **Neverše standardní chyba odpovídající regrese OLS**

Používá převrácená hodnota směrodatné chyby.

##### **Uživatelé zadaná hodnota**

Je-li zadána jedna číselná hodnota, bude jako počáteční hodnota použita hodnota. Je-li tento parametr zadán, musí být větší než 0.

#### Komponenta frajty

Nepovinný parametr pro určení **Distribuce** komponenty frajnosti.

##### **Gama**

Určuje rozdělení gama. Toto je výchozí nastavení.

##### **Inverzní Gaussovo**

Určuje inverzní funkci Gaussovo rozdělení.

##### **Počáteční hodnota odchyly**

Určuje počáteční hodnotu rozptylu komponenty frajity. Musí se jednat o jedinou číselnou hodnotu větší než 0. Výchozí hodnota je 1.0 pro gama rozdělení a 0.1 pro inverzní-Gaussovo rozdělení.

## Parametrické sdílené chatrné modely: Odhad

### Odhad

Nepovinný panel k určení nastavení pro řízení odhadu sdílených modelů brailty a volitelného procesu výběru funkcí.

## Střídavá metoda nebo vícenásobné kleště (ADMM)

### Rychle

Použije se rychle střídavý směr násobičů (ADMM). Jedná se o výchozí nastavení.

### Tradiční

Použije tradiční algoritmus ADMM.

## Použití regularizace L-1

Ukonává proces k řízení výběru funkcí. Pole **Parametr penále** uvádí parametr penalty, který řídí regularizační proces. Musí se jednat o jedinou hodnotu větší než 0. Předvolené nastavení je 0.001.

## Vzorová kritéria konvergence

### Přeměna parametru

Uvádí konvergenční kritéria pro parametr. Musí se jednat o jedinou číselnou hodnotu patřící do  $[0, 1)$ . Výchozí nastavení je 0.000001. Pro **Typ** můžete vybrat buď **ABSOLUTE** pro použití absolutní konvergence na vnitřní optimalizaci, nebo **RELATIVE**, chcete-li použít relativní konvergenci na vnitřní optimalizaci. Volitelná položka **Hodnota** určuje číselnou prahovou hodnotu pro typ konvergence.

### Cíl "Konvergence"

Určuje kritéria konvergence pro funkci cíle. Musí se jednat o jedinou číselnou hodnotu, která patří do  $[0, 1)$ . Výchozí nastavení je 0, které neuplatňuje kritéria konvergence. Pro **Typ** můžete vybrat buď **ABSOLUTE** pro použití absolutní konvergence na vnitřní optimalizaci, nebo **RELATIVE**, chcete-li použít relativní konvergenci na vnitřní optimalizaci. Volitelná položka **Hodnota** určuje číselnou prahovou hodnotu pro typ konvergence.

### Hesská konvergence

Uvádí konvergenční kritéria pro Hessian matrix. Musí se jednat o jedinou číselnou hodnotu, která patří do  $[0, 1)$ . Výchozí nastavení je 0, které neuplatňuje kritéria konvergence. Pro **Typ** můžete vybrat buď **ABSOLUTE** pro použití absolutní konvergence na vnitřní optimalizaci, nebo **RELATIVE**, chcete-li použít relativní konvergenci na vnitřní optimalizaci. Volitelná položka **Hodnota** určuje číselnou prahovou hodnotu pro typ konvergence.

## Zbývající kritéria konvergence

Volba pro řízení procesu optimalizace.

### Jak primitivní, tak dvojitý reziduum

Použije se jak primitivní, tak i duální reziduální kritérium konvergence. Toto nastavení je standardně nastaveno.

### Pouze zbytková zbytková hodnota

Použije se reziduální kritérium zbytkové konvergence.

### Pouze duální reziduum

Použije se zdvojené zbývající kritérium konvergence.

## Metoda

Nepovinný parametr k určení metody odhadu.

### Automatický

Automaticky zvolí metodu založenou na ukázkové datové sadě. Tato metoda je při výchozím nastavení vybrána. Pole **Prahová hodnota prediktorů** určuje prahovou hodnotu počtu predikátů a musí se jednat o jediné celé číslo větší než 1. Výchozí hodnota je 1000.

### Newton-Raphson

Použije se Newtona-Raphsonův způsob.

### L-BFGS

Použije se algoritmus BFGS s omezenými pamětí. Pole **Aktualizovat** uvádí počet předchozích aktualizací, které jsou udržovány algoritmem BFGS s omezenou pamětí, a musí se jednat o jediné celé číslo větší než nebo rovné 1. Výchozí hodnota je 5.

## Iterace

### Maximum iterací

Určuje maximální počet iterací. Musí se jednat o jediné celé číslo, které patří [ 1, 300]. Předvolené nastavení je 20.

### Maximální krok-halving

Určuje maximální počet kroků-halving. Musí se jednat o jediné celé číslo, které patří do [ 1, 200]. Výchozí nastavení je 5.

### Maximální počet hledání řádků

Určuje maximální počet hledání řádku. Musí se jednat o jediné celé číslo, které patří [ 1, 300]. Předvolené nastavení je 20.

### Absolutní konvergence pro proces iterace

Určuje absolutní konvergenci pro proces vnější iterace. Musí se jednat o jedinou číselnou hodnotu, která patří do (0, 1). Výchozí nastavení je 0.0001.

### Relativní konvergence pro proces iterace

Určuje relativní konvergenci pro proces vnější iterace. Musí se jednat o jedinou číselnou hodnotu, která patří do (0, 1). Výchozí nastavení je 0.01.

## Parametrické sdílené chatrné modely: Tisk

### Tisk

Volitelný panel, který řídí výstupy tabulky.

#### Podrobnosti kódování faktoru

Je-li vybráno, zobrazí se a vytiskne podrobnosti o kódování faktorů. Pokud v platnosti nejsou žádné faktory, je proces ignorován.

#### Počáteční hodnoty, které jsou přiřazeny k regresním parametrům

Je-li vybráno, zobrazí se počáteční hodnoty, které se použijí v procesu odhadu.

#### Historie iterace modelu

Je-li vybráno, zobrazí historii iterací analýzy přežití. V poli **Počet kroků** zadejte počet kroků mezi 1 a 99999999. Výchozím nastavením je hodnota 1.

## Parametrické sdílené chatrné modely: Predict

### Predict

Volitelný panel pro skóre a uložení predikované statistiky do aktivní datové sady.

#### Časové hodnoty pro přidělení skóre

##### Časové hodnoty definované závislé proměnné (závislé proměnné)

Skóre **Predikce** založené na časové proměnné zadané pro parametrický model přežití.

##### Pravidelné intervaly

Skóre **Předpovědi** je založena na budoucích hodnotách času. Pole **Časový interval** určuje časový interval a musí se jednat o jedinou číselnou hodnotu větší než 0. Pole **Počet časových období** uvádí počet časových období a musí se jednat o jediné číselné celé číslo mezi 2 a 100.

##### Doba trvání

Skóre **Předpovědi** je založena na časovém trvání a definuje budoucí hodnoty času. Musí se jednat o jedinou číselnou proměnnou.

#### Předpovědi

##### Přežití

Skóre a ukládá předpovídanou statistiku přežití na aktivní datovou sadu. Standardní název přizpůsobené proměnné (nebo jméno uživatele root) je `PredSurvival`.

### **Riziko**

Skóre a uloží předpovězená rizika do aktivní datové sady. Standardní název přizpůsobené proměnné (nebo jméno uživatele root) je `PredHazard`.

### **Kumulativní riziko**

Skóre a uloží předpovězená kumulativní rizika do aktivní datové sady. Standardní název přizpůsobené proměnné (nebo jméno uživatele root) je `PredCumHazard`.

### **Podmíněné přežití**

Skóre a uloží predikované statistiky podmíněného přežití do aktivní datové sady. Standardní název přizpůsobené proměnné (nebo jméno uživatele root) je `PredConditionalSurvival`. Proces je ignorován, pokud není zadán `PASTTIME`. Hodnota **Minulá doba přežití** je povinná a uvádí hodnoty v minulosti pro přidělení skóre. Musí se jednat o jedinou číselnou proměnnou.

### **Nepodmíněné přežití**

Skóre a uloží předpovězené nepodmíněné statistiky o přežití do aktivní datové sady. Klíčové slovo je ve výchozím nastavení potlačeno. Je-li zadán, může být následován volitelným uživatelem zadaným názvem proměnné (nebo kořenovým názvem) zadaným v závorkách. Výchozí název je `PredUnCondSurvival`.

### **Nepodmíněné riziko**

Skóre a ukládá předpovězenou nepodmíněnou statistiku o riziku pro aktivní datovou sadu. Klíčové slovo je ve výchozím nastavení potlačeno. Je-li zadán, může být následován volitelným uživatelem zadaným názvem proměnné (nebo kořenovým názvem) zadaným v závorkách. Výchozí název je `PredUncondHazard`.

### **Nepodmíněné riziko cemu**

Skóre a uloží předpokládané nepodmíněné kumulativní statistiky o riziku do aktivní datové sady. Klíčové slovo je ve výchozím nastavení potlačeno. Je-li uveden, může být následován volitelným uživatelem zadaným názvem proměnné (nebo kořenovým názvem, který je uveden v závorkách). Výchozí název je `PredUncondCumHazard`.

## **Parametrické sdílené chatrné modely: Zakreslit**

### **Graf**

#### **Funkce Plots**

Volba pro řízení grafů funkcí.

#### **Typ**

##### **Přežití**

Vytvoří plot pro nepodmíněné funkce přežití.

##### **Riziko**

Vytvoří zakreslení pro nepodmíněné funkce rizika.

##### **Hustota**

Vytvoří zákresový graf pro funkce hustoty.

#### **Počet zobrazovaných bodů**

Určuje počet funkčních bodů mezi 1 a 200. Předvolené nastavení je 100.

#### **Kovariovat hodnoty pro zákresový graf**

Volba pro určení uživatelem zadaných hodnot a jejich přiřazení k predictorům. Ve výchozím nastavení budou určené vykreslované grafy vytvořeny ve střední části každé proměnné, která se v podstatě mění. Je-li zadán, budou určené grafy vytvořeny na základě nastavení vzoru. Při výskytu duplicitních proměnných se nejprve rozezná jedna zadaná a zbylé by se ignorovaly. Platná proměnná musí být obsažena ve vzorovém efektu. Pro proměnnou kovariate musí být uživatelem zadaná hodnota číselná. Vynechání proměnné v účinku znamená, že by se střední hodnota použila při výchozím nastavení pro kovariate. Je-li k proměnné přiřazena neplatná hodnota, požadovaný vzorek nebude vykreslen.

### Hodnoty faktoru pro vykreslení

Volba pro určení uživatelem zadaných hodnot a jejich přiřazení k predictorům. Ve výchozím nastavení budou určené grafy vytvořeny v četnosti kategorií každého faktoru v platnosti. Je-li zadán, budou určené grafy vytvořeny na základě nastavení vzoru. Při výskytu duplicitních proměnných se nejprve rozezná jedna zadaná a zbylé by se ignorovaly. Platná proměnná musí být obsažena ve vzorovém efektu. Vynechání proměnné v účinku označuje, že by se frekvence kategorií použila pro faktor jako výchozí. Je-li k proměnné přiřazena neplatná hodnota, požadovaný vzorek nebude vykreslen.

#### Samostatné řádky pro

Volba pro určení kategorické proměnné, podle které budou vykresleny spojnicové grafy.

#### Maximální počet řádků v grafu

Určuje maximální počet řádků v grafu, pokud je zadán parametr **Oddělte řádky pro** . Předvolené nastavení je 10.

## Parametrické sdílené chatrné modely: Export

### Exportovat

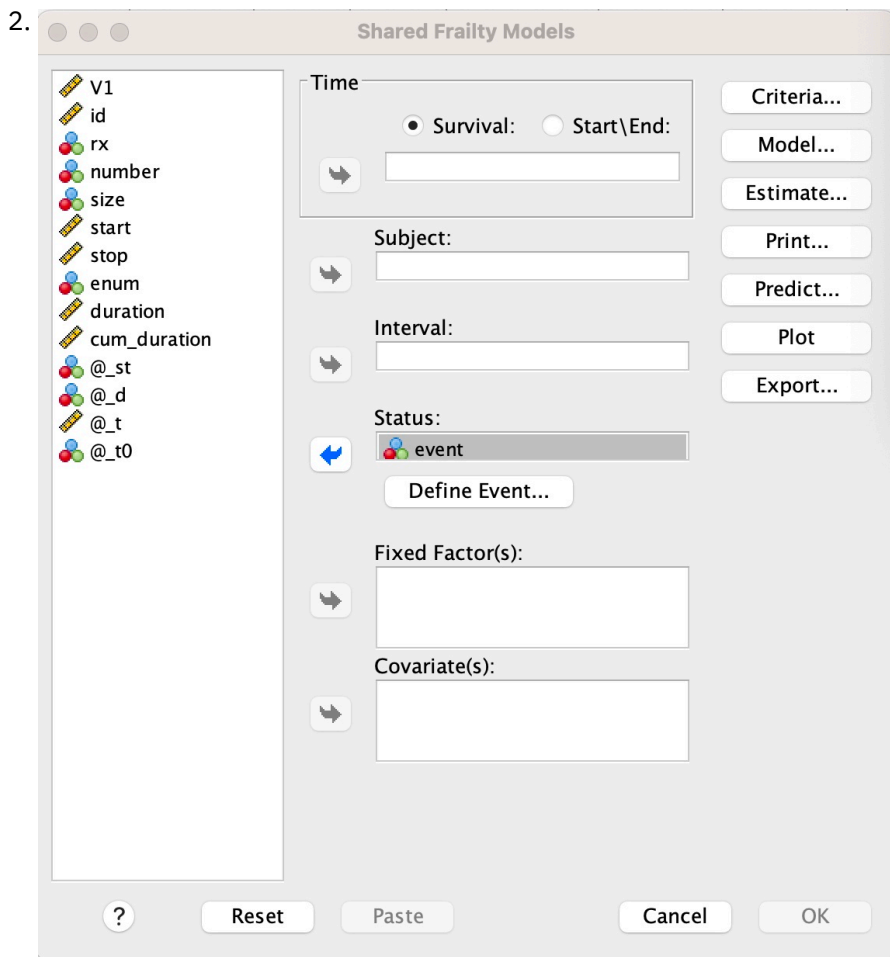
Vyberte volbu **Exportovat informace o modelu do souboru XML** , chcete-li zapsat informace o modelu a parametru do souboru PMML pro přidělení skóre. Musíte uvést adresář a název souboru PMML, který má být uložen.

## Parametrické sdílené chatrné modely: Definování události

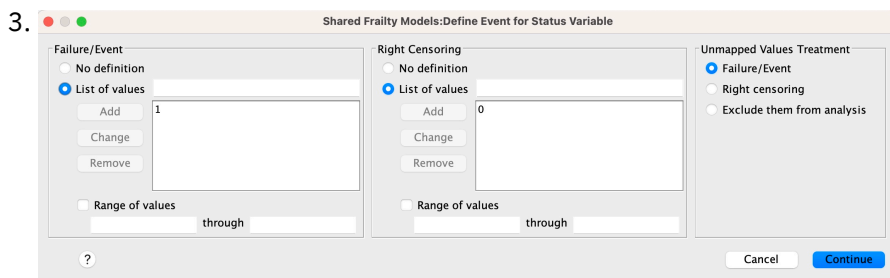
Volba pro definování stavu. Je-li proměnná stavu vynechána, stane se selhání nebo událost výchozím stavem pro všechny případy.

1. Z nabídky vyberte:

**Analyzovat > Přežití > Parametrické sdílené modely s použitím parametrického sdílení ...**



Obrázek 3. Dialogové okno Sdílené zlomené modely-dialogové okno-Stav



Obrázek 4. Dialogové okno Sdílené zlomené modely-dialogové okno Stav-Definovat událost

## Parametrické sdílené chatrné modely-Příklady

### Příklad 1

SURVREG RECURRENT y WITH x1 BY x2  
/MODEL PŘEDMĚT = ID ŽELEZNIČNÍ OBNOVENÍ=VÁŽENÍ.

U proměnných x1 a x2 je osazen parametrický model přežití se sdílenou frakcją pro přežití.

Survival čas je reprezentován jedinou proměnnou y.

Předměty jsou identifikovány pomocí ID proměnné.

Přežití se předpokládá, že sleduje Weibullovo rozdělení.

Odchylka od frailty se předpokládá, že bude následovat rozdělení gama.

Všechny platné záznamy se používají v analýze přežití.

### **Příklad 2**

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJECT = ID FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

U proměnných x1 a x2 je osazen parametrický model přežití se sdílenou frakou pro přežití.

Doba přežívání je znázorněna dvěma proměnnými y1 a y2 označenými za počáteční a koncový čas.

Předměty jsou identifikovány pomocí ID proměnné.

Předpokládá se, že doba přežívání se řídí běžnou distribucí protokolu.

Rozptyl palaty se předpokládá, že bude následovat inverzní Gaussovo rozdělení.

Časové intervaly jsou definovány proměnnou z. Pro každý předmět používá procedura pouze nekonfliktní záznamy a vyloučí z analýzy všechny záznamy po prvním stavu selhání.

### **Příklad 3**

```
SURVREG RECURRENT y1 y2 WITH x1 BY x2(1)
```

```
/MODEL SUBJECT = ID FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC
```

```
/STATUS VARIABLE=událost FAILURE=1 RIGHT=0.
```

U proměnných x1 a x2 je osazen parametrický model přežití se sdílenou frakou pro přežití. Přežití

čas je reprezentován dvěma proměnnými y1 a y2 označenými jako počáteční a koncový čas. Pro faktor x2 je kategorie "1" určena jako úroveň baseline, která má být modelována.

Předměty jsou identifikovány pomocí ID proměnné.

Předpokládá se, že doba přežívání je v souladu s logistickou distribucí protokolu.

Rozptyl palaty se předpokládá, že bude následovat inverzní Gaussovo rozdělení.

Událost proměnné je určena k definování stavu s 1 a 0 označováním selhání a pravého cenzurování.

### **Příklad 4**

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJE/ID
```

```
/STATUS VARIABLE=událost FAILURE=1 RIGHT=0
```

```
/PŘEDPOVÍDAT ZRUŠENÍ PODMÍNKY UNCONDHAZARD UNCONDC/HAZARD
```

```
/FUNCTIONPLOT CHIRÁL NEBEZPEČNOSTI NEBEZPEČÍ PLOTBY (x2).
```

U proměnných x1 a x2 je osazen parametrický model přežití se sdílenou frakou pro přežití. Survival čas je reprezentován jedinou proměnnou y.

Předměty jsou identifikovány pomocí ID proměnné.

Nepodmíněné přežití nebo riziko přežití založené na populacích, riziko a kumulativní riziko se zaznamenávají a ukládají se do aktivní datové sady.

Přežití bez podmíněného nebo popularizace a křivky rizik jsou vykresleny odděleně podle kategorií v x2.

### **Příklad 5**

```
SURVREG RECURRENT y WITH x1 BY x2
```



```
/MODEL PŘEDMĚT = ID ŽELEZNIČNÍ OBNOVENÍ=VÁŽENÍ
```

```
/STATUS VARIABLE=událost FAILURE=1 RIGHT=0
```

```
/ESTIMATION HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0SELECTFEATURES=TRUE  
PENALTY=0.01.
```

U proměnných x1 a x2 je osazen parametrický model přežití se sdílenou frajkou pro přežití. Survival čas je reprezentován jedinou proměnnou y.

Předměty jsou identifikovány pomocí ID proměnné.

Přežití se předpokládá, že sleduje Weibullovo rozdělení.

Odchylka od brailty se předpokládá, že bude následovat rozdělení gama.

Kritéria konvergence jsou založena na hessijské matici. Používá 1e-12 jako relativní konvergenci.

Model zahrnuje pokutový termín pro řízení regularizačního procesu. Parametr penále je nastaven na hodnotu 0.01.

### **Příklad 6**

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJE/ID
```

```
/STATUS VARIABLE=infect FAILURE=1 RIGHT=0
```

```
/ESTIMATION MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.
```

U proměnných x1 a x2 je osazen parametrický model přežití se sdílenou frajkou pro přežití. Survival čas je reprezentován jedinou proměnnou y.

Předměty jsou identifikovány pomocí ID proměnné.

Procedura uvádí maximální počet hledání řádku na 100, maximální počet iterací, které mají být 50, a maximální počet kroků-halving na 20.

## **Parametrické sdílené chatrné modely-Případová studie pro opakovací data**

Parametrické sdílené chatrné modely-Případová studie pro opakovací data

Příklad použití názvu případu-Vedlejší účinky léčby.

Herci-Vyšetřovatel veřejného zdraví a praktický lékař.

Předběžné podmínky-Vyčištěná datová sada dostupná na základě doby přežití, stavu a predikátů, které mají být upraveny.

Popis-Patrick, vyšetřovatel veřejného zdraví, vyšetřuje vzorek dat, který obsahuje 20 účastníků. Tito účastníci jsou zařazeni do studie o mírném vedlejší účinku, který je potenciálně způsoben novou léčbou. Projev projektantovi tvrdí, že mezi samci a samicemi nejsou žádné rozdíly, pokud jde o vedlejší účinek. Patrick by rád tuto hypotézu hodnotil. Proměnné, které jsou zahrnuty do vzorku dat, jsou vypsány níže uvedeným způsobem:

- patID: Identifikační číslo ID k identifikaci jedinečného účastníka.

- endTime: Doba přežívání (ve dnech) po ukončení léčby, která se měří od zahájení léčby až po nežádoucí účinek, který byl hlášen nebo cenzurován během 60 dnů.

- sideEffect: Stav bočního účinku, stav = 0, je-li cenzurován a stav = 1, je-li nahlášen mírný oboustranný účinek.

-věk: věk účastníka v období výzkumu.

-samice: samice = 0, je-li žena a samice = 1 samice.

Může se použít více procedur, což způsobí, že se více záznamů o časech opakování, které se měří pro určitého účastníka. Počáteční čas je vždy 0 pro každý záznam, který je vynechán ve vzorku dat. Patrick má zájem vizualizovat funkce přežití a nebezpečí, aby bylo možné srovnávat mezi mužem a ženou tím, že řídí jejich věk a křehkost. Je si vědom toho, že tato léčba, která jsou spravována stejnému účastníkovi, je více korelovaná. Za předpokladu, že doba přežití se řídí Weibullovo rozdělení, Patrick se rozhodne vytvořit parametrický model pro přežití se sdílenými podvody v produktu SPSS Statistics pro účely léčby závislosti na léčbě stejného účastníka.

Syntaxe-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILTY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

### Synopse:

Syntaxe, kterou určuje Patrick, označuje endTime jako jedinou závislou dobu. Procedura automaticky předpokládá, že počáteční čas je 0 pro každý záznam. Proměnné věk a ženy jsou modelovány jako kovariát a faktor, resp. Předpokládá se, že doba přežívání recidivy se předpokládá u Weibullova rozdělení. Předpokládá se, že nenápadný frajovaný termín sleduje rozdělení Gamma a modeluje se její komponenta rozptylu. Pokud jde o výstupy, tabulka Souhrn modelu poskytuje informace o proceduře a modelu. Tabulka Souhrn zpracování případů poskytuje souhrnný seznam stavu selhání/cenzurování a také ty případy, které jsou vyloučeny z analýzy.

Ve vzorku dat Patricka jsou všechny záznamy platné a jsou zahrnuty do analýzy. Srovnáním pravděpodobnosti protokolu s pravděpodobností odpovídajícího modelu bez komponenty frakity se nepodařilo dosáhnout významné úrovně ( $p$ -value = 0.168) modelu sdílené paměti. Patrick se ptá, zda je nezbytné zahrnout do modelu sdílené-frajný termín. Odhadovaný faktor akcelerace účastníka je 1.017, který se získává výpočtem exponentu odhadovaného koeficientu regrese 0.017 z [ ženy = 0.0]. Jeho přidružený 95% interval spolehlivosti je (.688, 1.504). Tyto výsledky naznačují, že mužský jedinec má téměř stejný akcelerace faktor jako samice jedince se stejným věkem a křehkou. Na úrovni počtu obyvatel

Patrick vykresluje nepodmíněné přežití a křivky nebezpečí odděleně pro samce a samice, které jsou hodnoceny ve stáří vzorku (37.45 let).

Patrick potvrzuje, že pro jakoukoli pevnou hodnotu doby přežití má muž a samice v průměru mít stejnou pravděpodobnost přežití. Zajímavé je, že navzdory neimodální formě, která je zobrazena v nepodmíněném měřidlovém grafu, Patrick zjistí, že během období 60 dnů se riziko obyvatelstva ve skutečnosti zvyšuje. Toto chování může znamenat, že se jedná o vliv frajnosti. K dalšímu zkoumání vedlejších účinků způsobených léčbou může Patrick pokračovat s modelem bez křehkého dílu a srovnávat chování samců a samic. Kromě toho může u účastníků za období delší než 60 dnů považovat více údajů za delší než 60 dnů.



## 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

analýza logistické regrese [1](#)  
Analýza Probit  
  definovat rozsah [10](#)  
  fiducial intervaly spolehlivosti [10](#)  
  Iterace [10](#)  
  Podmínky [10](#)  
  příkaz additional features [11](#)  
  přirozená míra odezvy [10](#)  
  relativní střední potence [10](#)  
  statistika [10](#)  
  test paralelismu [10](#)  
analýza přežití  
  v regresních regresí jádra [27](#)  
asymptotická regrese  
  v nelineární regresí [19](#)

## B

binární logistická regrese [1](#)  
buňky s nulovým pozorováním  
  in Multinomial Logistic Regression [8](#)

## C

Cookovy D  
  v Logistickém regresí [4](#)  
Cox a Snell R-náměstí  
  in Multinomial Logistic Regression [7](#)

## D

DfBeta  
  v Logistickém regresí [4](#)  
dobrota vložení  
  in Multinomial Logistic Regression [7](#)

## F

fiducial intervaly spolehlivosti  
  v Probit Analysis [10](#)  
funkce deviance  
  pro odhad hodnoty rozptylu rozptylu [8](#)

## G

Gauss model  
  v nelineární regresí [19](#)

## H

historie iterace  
  in Multinomial Logistic Regression [8](#)  
hodnota velikosti rozptylu  
  in Multinomial Logistic Regression [8](#)

Hosmer-Lemesyhow goodness-of-fit statistic  
  v Logistickém regresí [5](#)

## I

INTERCEPT  
  zahrnout nebo vyloučit [6](#)  
intervaly spolehlivosti  
  in Multinomial Logistic Regression [7](#)  
Iterace  
  in Multinomial Logistic Regression [8](#)  
  v Logistickém regresí [5](#)  
  v Probit Analysis [10](#)

## K

kategoriální proměnné [3](#)  
kategorie odkazů  
  in Multinomial Logistic Regression [7](#)  
Klasifikace  
  in Multinomial Logistic Regression [5](#)  
konstantní termín  
  v lineární regresí [5](#)  
kontrast  
  v Logistickém regresí [3](#)  
korelační matice  
  in Multinomial Logistic Regression [7](#)  
kovarianční matice  
  in Multinomial Logistic Regression [7](#)  
kovariace  
  v Logistickém regresí [3](#)  
kritérium konvergence  
  in Multinomial Logistic Regression [8](#)  
kroková volba  
  in Multinomial Logistic Regression [6](#)  
  v Logistickém regresí [3](#)

## L

Lineární regrese  
  odhad hmotnosti [21](#)  
  Regrese dvoufázového nejmenších čtverců [23](#)  
logistická regrese [1](#)  
Logistická regrese  
  binární [1](#)  
  definovat pravidlo výběru [3](#)  
  Hosmer-Lemesyhow goodness-of-fit statistic [5](#)  
  Iterace [5](#)  
  kategoriální proměnné [3](#)  
  Koeficienty [1](#)  
  konstantní termín [5](#)  
  kontrast [3](#)  
  metody výběru proměnných [3](#)  
  nastavit pravidlo [3](#)  
  ovlivňování ukazatelů [4](#)  
  pravděpodobnost nekrokových [5](#)

Logistická regrese (*pokračování*)  
predikované hodnoty [4](#)  
příkaz additional features [5](#)  
příklad [1](#)  
řetězcová proměnná\_řetězce [3](#)  
statistika [1](#)  
statistiky a grafy [5](#)  
ukládání nových proměnných [4](#)  
uzavření klasifikace [5](#)  
volby zobrazení [5](#)  
zbytkové chyby [4](#)  
logistická regrese SPSS [1](#)

## M

McFadden R-náměstí  
in Multinomial Logistic Regression [7](#)  
Metcherlich zákon o klesajících výnosech  
v nelineární regresi [19](#)  
Michaelis Menten model  
v nelineární regresi [19](#)  
Model Gompertz  
v nelineární regresi [19](#)  
model hustoty  
v nelineární regresi [19](#)  
model hustoty výnosu  
v nelineární regresi [19](#)  
Model Johnson-  
Schumacher  
v nelineární regresi [19](#)  
Model Richards  
v nelineární regresi [19](#)  
Model Verhulst  
v nelineární regresi [19](#)  
Model Von Bertalanffy  
v nelineární regresi [19](#)  
modely s hlavní účinky  
in Multinomial Logistic Regression [6](#)  
Modely s parametrickým akceleorickým časem  
analýza [30](#)  
bodová [34](#)  
exportovat [35](#)  
model [31](#)  
odhad [32](#)  
Podmínky [31](#)  
predikovat [33](#)  
tisk [33](#)  
Morgan-Mercer-Florenin model  
v nelineární regresi [19](#)  
Multinomická Logistická regrese  
export informací o modelu [9](#)  
kategorie odkazů [7](#)  
modely [6](#)  
Podmínky [8](#)  
příkaz additional features [9](#)  
statistika [7](#)  
uložit [9](#)

## N

Nagelkerke R-náměstí  
in Multinomial Logistic Regression [7](#)  
nelineární modely

nelineární modely (*pokračování*)  
v nelineární regresi [19](#)  
Nelineární regrese  
Algoritmus Levenberg-Marquardt [20](#)  
běžné nelineární modely [19](#)  
deriváty [20](#)  
funkce ztráty [19](#)  
interpretace výsledků [20](#)  
metody odhadu [20](#)  
odhady zaváděcího programu [20](#)  
omezení parametrů [20](#)  
parametry [18](#)  
počáteční hodnoty [18](#)  
podmíněná logika [18](#)  
predikované hodnoty [20](#)  
příkaz additional features [21](#)  
příklad [17](#)  
segmentovaný model [18](#)  
sekvenční kvadratické programování [20](#)  
statistika [17](#)  
ukládání nových proměnných [20](#)  
zbytkové chyby [20](#)

## O

Odhad váhy  
historie iterace [22](#)  
protokolování-pravděpodobnost [21](#)  
příkaz additional features [22](#)  
příklad [21](#)  
statistika [21](#)  
uložení nejlepších vah jako nové proměnné [22](#)  
zobrazení ANOVA a odhadů [22](#)  
odhady parametrů  
in Multinomial Logistic Regression [7](#)  
Okraj jádra  
Alpha [27](#)  
coef0 [27](#)  
DEGREE [27](#)  
Gamma [27](#)  
jednotlivý model [27](#)  
Výběr modelu [27](#)  
omezená regrese  
v nelineární regresi [20](#)  
omezení parametrů  
v nelineární regresi [20](#)

## P

pákové hodnoty  
v Logistickém regresi [4](#)  
Parametrické modely s metrikou  
stavové proměnné přežití [42](#)  
Parametrické sdílené chatrné modely  
analýza [36](#)  
exportovat [42](#)  
model [38](#)  
odhad [38](#)  
Podmínky [37](#)  
predikovat [40](#)  
tisk [40](#)  
zakreslit [41](#)  
Peal-Reed model

Peal-Reed model (*pokračování*)  
v nelineární regresi [19](#)

Pearson chi-square  
dobrota vložení [7](#)  
pro odhad hodnoty rozptylu rozptylu [8](#)

poměr cubics modelu  
v nelineární regresi [19](#)

poměr kvadratických modelů  
v nelineární regresi [19](#)

Poměr pravděpodobnosti  
dobrota vložení [7](#)  
pro odhad hodnoty rozptylu rozptylu [8](#)

Probit regrese  
příklad [9](#)  
statistika [9](#)

protokol-upravený model  
v nelineární regresi [19](#)

protokolování-pravděpodobnost  
in Multinomial Logistic Regression [7](#)  
v odhadu váhy [21](#)

předat výběr  
v Logistickém regresi [3](#)

Přežití AFT  
Dialogové okno přežití-Proměnné kategorie [35](#)

## R

Regrese dvoufázového nejmenších čtverců  
instrumentální proměnné [23](#)  
kovariance parametrů [24](#)  
příkaz additional features [24](#)  
příklad [23](#)  
statistika [23](#)  
ukládání nových proměnných [24](#)

Regrese jádra jádra  
parametry [29](#)  
parametry mřížky [29](#)  
Přehybné vody [29](#)  
uložit [29](#)  
výkresy [29](#)  
zobrazení [29](#)

Regrese kvantilů  
exportovat [16](#)  
model [13](#)  
Podmínky [12](#)  
příklad [11](#)  
statistika [11](#)  
uložit [16](#)  
zobrazení [14](#)

relativní střední potence  
v Probit Analysis [10](#)

rozdělení  
in Multinomial Logistic Regression [8](#)

Rozdíl  
jako korekce buněk s nulovými pozorováními [8](#)

## Ř

řetězcová proměnná\_řetězce  
v Logistickém regresi [3](#)

## S

singulární  
in Multinomial Logistic Regression [8](#)

step-halving  
in Multinomial Logistic Regression [8](#)

## T

tabulky klasifikace  
in Multinomial Logistic Regression [7](#)

tabulky pravděpodobnosti buněk  
in Multinomial Logistic Regression [7](#)

Tabulky životnosti  
stavové proměnné přežití [35](#)

test paralelismu  
v Probit Analysis [10](#)

## U

úplné faktoriální modely  
in Multinomial Logistic Regression [6](#)

## V

vlastní modely  
in Multinomial Logistic Regression [6](#)

## W

Weibullový model  
v nelineární regresi [19](#)

## Z

zpětná eliminace  
v Logistickém regresi [3](#)





