

IBM SPSS Missing Values 29

IBM

Poznámka

Před použitím těchto informací a produktu, který podporují, si přečtěte informace v tématu [“Upozornění” na stránce 19](#).

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

Kapitola 1. Chybějící hodnoty.....	1
Úvod do chybějících hodnot.....	1
Analýza chybějící hodnoty.....	1
Zobrazení vzorů chybějících hodnot.....	3
Zobrazení popisné statistiky pro chybějící hodnoty.....	4
Odhadování statistiky a chybějící hodnoty.....	4
Dodatečné funkce příkazu MVA.....	6
Vícenásobné imputace.....	7
Analyzovat vzory.....	8
Impute chybějící datové hodnoty.....	9
Další funkce příkazu MULTIPLE IMPUTATION.....	12
Práce s více imputací daty.....	13
Analýza více imputací dat.....	13
Více voleb imputace.....	17
Upozornění.....	19
Ochranné známky.....	20
Rejstřík.....	23

Kapitola 1. Chybějící hodnoty

Následující funkce chybějících hodnot jsou obsaženy v části SPSS Statistics Premium Edition nebo v rámci volby Chybějící hodnoty.

Úvod do chybějících hodnot

Případy s chybějícími hodnotami představují důležitou výzvu, protože typické modelovací postupy jednoduše tyto případy zlikvidují z analýzy. Pokud existuje jen málo chybějících hodnot (velmi zhruba, méně než 5% z celkového počtu případů) a tyto hodnoty lze považovat za chybějící náhodně; tj. to, zda hodnota chybí, nezávisí na jiných hodnotách, pak je typická metoda odstranění listwise relativně "bezpečná". Volba Chybějící hodnoty vám může pomoci určit, zda je odstranění listwise dostatečné, a poskytuje metody pro zacházení s chybějícími hodnotami, když není.

Analýza chybějící hodnoty versus použití více imputací

Volba Chybějící hodnoty poskytuje dvě sady procedur pro obsluhu chybějících hodnot:

- Procedury Vícenásobné imputace poskytují analýzu vzorků chybějících dat, která směřuje k případné vícenásobné imputaci chybějících hodnot. To znamená, že se vytváří více verzí datové sady, každá obsahuje svou vlastní sadu imputovaných hodnot. Jsou-li prováděny statistické analýzy, jsou odhady parametrů pro všechny imputované datové soubory sloučeny a poskytují odhady, které jsou obecně přesnější, než by bylo možné pouze s jednou imputací.
- Analýza chybějící hodnoty poskytuje mírně odlišnou sadu popisných nástrojů pro analýzu chybějících dat (nejvíce zvláště Little's MCAR test) a zahrnuje řadu jednotlivých metod imputace. Všimněte si, že vícenásobná imputace je obecně považována za nadřazenou k jediné imputaci.

Úlohy chybějících hodnot

Můžete začít s analýzou chybějících hodnot podle následujících základních kroků:

1. **Proveďte nejednotnost.** Použijte analýzu chybějící hodnoty a analýzu vzorků, abyste prozkoumali vzorky chybějících hodnot ve vašich datech a určili, zda je zapotřebí více imputace.
2. **Chybějící hodnoty Impute.** Použít chybějící datové hodnoty Impute k násobení chybějících hodnot.
3. **Analýzujte "kompletní" data.** Použijte jakoukoli proceduru, která podporuje více dat imputace. Informace o analýze více datových sad imputace a seznam procedur, které podporují tato data, viz "Analýza více imputací dat" na stránce 13.

Analýza chybějící hodnoty

Procedura chybějící hodnoty analýzy provádí tři primární funkce:

- Popisuje vzorek chybějících dat. Kde se nacházejí chybějící hodnoty? Jak rozsáhlé jsou? Mají dvojice proměnných tendenci mít ve více případech chybějící hodnoty? Jsou datové hodnoty extrémní? Chybí hodnoty náhodně?
- Odhady střední hodnoty, směrodatné odchylky, odchylky a korelace pro různé chybějící metody hodnoty: listwise, pairwise, regression nebo EM (expectation-maximization). Metoda pairwise také zobrazuje počet dokončených případů po dvojicích.
- Filly v (imputes) chybí hodnoty s odhadovanými hodnotami pomocí regrese nebo EM metod; nicméně, více imputace je obecně považován za poskytnutí přesnější výsledky.

Analýza chybějící hodnoty pomáhá řešit několik problémů způsobených neúplnými daty. Pokud se případy s chybějícími hodnotami systematicky odlišují od případů bez chybějících hodnot, výsledky mohou být zavádějící. Chybějící data také mohou snížit přesnost vypočtených statistik, protože je méně informací než původně plánováno. Další obavu je, že předpoklady v pozadí mnoha statistických postupů jsou založeny na úplných případech a chybějící hodnoty mohou komplikovat požadovanou teorii.

Příklad. Při hodnocení léčby leukemie se měří několik proměnných. Nicméně ne všechna měření jsou k dispozici pro každého pacienta. Vzorky chybějících dat jsou zobrazeny, tabelážděné a náhodným náhodným výběrem. Analýza EM se používá k odhadu prostředků, korelací a kovariance. Používá se také k určení toho, že data chybí zcela náhodně. Chybějící hodnoty se pak nahradí imputovanými hodnotami a uloží se do nového datového souboru pro další analýzu.

Statistika. Univariate statistics, including number of nonmissing values, mean, standard deviation, number of missing values, and number of extreme values. Odhadované prostředky, kovarianční matice a korelační matice pomocí metod listwise, pairwise, EM nebo regresních metod. Malý test MCAR s výsledky EM. Souhrn středních hodnot různých metod. Pro skupiny definované podle chybějících a nechybějících hodnot: *t* testů. Pro všechny proměnné: vzory s chybějícími hodnotami zobrazovaly případy podle proměnných.

Aspekty dat

Data. Data mohou být kategorická nebo kvantitativní (scale nebo continuous). Můžete však odhadovat statistiky a vkládat chybějící data pouze pro kvantitativní proměnné. Pro každou proměnnou chybí hodnoty, které nejsou kódované jako systémově-chybějící hodnoty musí být definovány jako uživatel-*chybí*. Pokud má například položka v dotazníku odpověď *Nevědět* kódovanou jako 5 a chcete-li ji považovat za chybějící, měla by mít položka 5 kódovanou jako hodnotu uživatele-*chybí*.

Frekvence frekvencí. Hodnoty frekvence (replikace) jsou tímto postupem uznány. Případy se zápornou nebo nulovou hodnotou váhy replikace jsou ignorovány. Neceločíselné váhy jsou oříznuty.

Předpoklady. Listwise, párovaný a regresní odhad závisí na předpokladu, že vzorek chybějících hodnot nezávisí na hodnotách dat. (Tato podmínka je známá jako **chybějící zcela náhodná** nebo MCAR.) Proto všechny metody (včetně metody EM) pro odhad poskytují konzistentní a nezaujaté odhady korelací a kovariance, když jsou data MCAR. Porušení hypotézy MCAR může vést k předpojatým odhadům vyvolaným metodami listwise, pairwise a regresními metodami. Nejsou-li data MCAR, musíte použít odhad EM.

Odhad EM závisí na předpokladu, že se vzorek chybějících dat vztahuje pouze k pozorovaným datům. (Tato podmínka se nazývá **chybí náhodně** nebo MAR.) Tento předpoklad umožňuje upravit odhady pomocí dostupných informací. Například ve studii o vzdělání a příjmu mohou mít subjekty s nízkým vzděláním další chybějící hodnoty příjmů. V tomto případě jsou data MAR, ne MCAR. Jinými slovy, pro MAR je pravděpodobnost, že příjem bude zaznamenán, závisí na úrovni vzdělání. Pravděpodobnost se může lišit podle vzdělání, ale ne podle příjmu *v rámci této úrovně vzdělávání*. Je-li pravděpodobnost, že příjmy se zaznamenávají, také liší podle hodnoty příjmu v každé úrovni vzdělání (například lidé s vysokými příjmy nehlásí je), pak data nejsou MCAR ani MAR. Tato situace není neobvyklá, a pokud se použije, žádná z metod není vhodná.

Související postupy. Mnoho procedur vám umožňuje použít lišice nebo porovnávání po dvojicích. Lineární regrese a analýza faktoru umožňují nahrazení chybějících hodnot průměrnými hodnotami. V přídavném modulu Prognózování je k dispozici několik metod, které nahradí chybějící hodnoty v časové řadě.

Jak získat analýzu chybějící hodnoty

1. Z nabídky vyberte:

Analyzovat > Analýza chybějící hodnoty ...

2. Vyberte alespoň jednu kvantitativní proměnnou (měřítka) pro odhad statistiky a volitelně vkládanou chybějící hodnoty.

Volitelně můžete:

- Vyberte kategoriální proměnné (číselné nebo řetězcové) a zadejte mezní hodnotu počtu kategorií (**Maximální kategorie**).
- Klepněte na volbu **Vzorky**, chcete-li zkrokovat vzorky chybějících dat. Další informace naleznete v tématu [“Zobrazení vzorů chybějících hodnot”](#) na stránce 3.
- Chcete-li zobrazit deskriptivní statistiky chybějících hodnot, klepněte na volbu **Deskriptivy**. Další informace naleznete v tématu [“Zobrazení popisné statistiky pro chybějící hodnoty”](#) na stránce 4.

- Vyberte metodu pro odhad statistiky (střední hodnoty, odchylky a korelace) a případně vkládanou chybějící hodnoty. Další informace naleznete v tématu [“Odhadování statistiky a chybějící hodnoty”](#) na stránce 4 .
- Vyberete-li volbu EM nebo Regrese, klepněte na volbu **Proměnné** a zadejte dílčí sadu, která má být použita pro odhad. Další informace naleznete v tématu [“Proměnné predicted a Predictor”](#) na stránce 6 .
- Vyberte proměnnou popisku případu. Tato proměnná se používá k označování případů ve vzorech tabulek, které zobrazují jednotlivé případy.

Zobrazení vzorů chybějících hodnot

Můžete zvolit zobrazení různých tabulek, které zobrazují vzory a rozsah chybějících dat. Tyto tabulky vám mohou pomoci identifikovat:

- Kde jsou chybějící hodnoty umístěny
- Určuje, zda dvojice proměnných mají tendenci mít chybějící hodnoty v jednotlivých případech
- Zda jsou datové hodnoty extrémní

Zobrazit

Pro zobrazení vzorů chybějících dat jsou k dispozici tři typy tabulek.

Tabulkové případy. Vzory chybějící hodnoty v proměnných analýzy jsou tabulované s frekvencí zobrazenými pro každý vzorek. Pomocí volby **Řadit proměnné podle chybějícího vzorku hodnoty** určete, zda mají být počty a proměnné řazeny podle podobnosti se vzory. Použijte **Vynechat vzorky s méně než n% případů** , abyste eliminovali vzorky, které se vyskytují zřídka.

Případy s chybějícími hodnotami. Každý případ s chybějící nebo extrémní hodnotou je tabulován pro každou analytickou proměnnou. Pomocí volby **Řadit proměnné podle chybějícího vzorku hodnoty** určete, zda mají být počty a proměnné řazeny podle podobnosti se vzory.

Všechny případy. Každý případ je tabulován a chybějící a extrémní hodnoty jsou indikovány pro každou proměnnou. Případy jsou vypsány v pořadí, v jakém se objevují v datovém souboru, pokud není proměnná uvedena v **Řadit podle**.

V tabulkách, které zobrazují jednotlivé případy, se používají následující symboly:

- + . Extrémně vysoká hodnota
- . Extrémně nízká hodnota
- S.** Systém-chybějící hodnota
- A.** První typ uživatele-chybějící hodnota
- B.** Druhý typ uživatele-chybějící hodnota
- C.** Třetí typ uživatele-chybějící hodnota

Proměnné

Můžete zobrazit další informace pro proměnné, které jsou zahrnuty v analýze. Proměnné, které přidáte do pole **Další informace pro** , se zobrazí jednotlivě v tabulce chybějících vzorů. Pro kvantitativní (měřítka) proměnné se zobrazí střední hodnota; pro kategoriální proměnné je zobrazen počet případů, které mají vzorek v každé kategorii.

- **Řadit podle.** Případy jsou vypsány na základě vzestupného nebo sestupného pořadí hodnot zadané proměnné. K dispozici pouze pro **Všechny případy**.

Chcete-li zobrazit vzory chybějících hodnot

1. V dialogovém okně Analýza chybějící hodnoty vyberte proměnnou, pro kterou chcete zobrazit chybějící vzorky hodnot.
2. Klepněte na volbu **Vzory**.
3. Vyberte tabulku (y) vzoru, které chcete zobrazit.

Zobrazení popisné statistiky pro chybějící hodnoty

Univariátní statistika

Univariate statistics can help you identify the general extent of missing data. Pro každou proměnnou se zobrazí následující:

- Počet nechybějících hodnot
- Počet a procento chybějících hodnot

Pro kvantitativní (měřítko) proměnné jsou také zobrazeny následující údaje:

- Střední
- Směrodatná odchylka
- Počet extrémně vysokých a nízkých hodnot

Statistika proměnné ukazatele

Pro každou proměnnou je vytvořena proměnná indikátoru. Tato kategorická proměnná označuje, zda je proměnná přítomná nebo chybí pro jednotlivý případ. Proměnné indikátorů se používají k vytvoření neshody, testu t a tabulek frekvence.

Neshoda v procentech. Pro každou dvojici proměnných se zobrazí procento případů, ve kterých jedna proměnná má chybějící hodnotu a druhá proměnná má nechybějící hodnotu. Každý diagonální prvek v tabulce obsahuje procentní část chybějících hodnot pro jednotlivou proměnnou.

t testy se skupinami formovanými indikátorových proměnných. Pro každou kvantitativní proměnnou jsou porovnávány prostředky dvou skupin s využitím statistiky t Studentova. Skupiny určují, zda je proměnná přítomná nebo chybí. Zobrazí se statistika t , stupně volnosti, počty chybějících a nechybějících hodnot a prostředky obou skupin. Můžete také zobrazit jakékoli dva-stoleté pravděpodobnosti přidružené k statistice t . Pokud výsledky analýzy vyústí ve více než jeden test, nepoužívejte tyto pravděpodobnosti pro testování významnosti. Součet pravděpodobností je vhodný pouze v případě, že je vypočten jediný test.

Křížové hodnoty kategoriálních a indikátorových proměnných. Pro každou kategorickou proměnnou se zobrazí tabulka. Pro každou kategorii tabulka zobrazuje četnost a procentní část nechybějících hodnot pro ostatní proměnné. Zobrazeny jsou také procentní části každého typu chybějící hodnoty.

Vynechání proměnných chybí méně než $n\%$ případů. Chcete-li snížit velikost tabulky, můžete vynechat statistiku, která se vypočítává pouze pro malý počet případů.

Zobrazit deskriptivní statistiky

1. V hlavní dialogové okně Analýza chybějící hodnoty vyberte proměnné, pro které chcete zobrazit deskriptivní statistiky chybějící hodnoty.
2. Klepněte na položku **Deskriptivy**.
3. Vyberte deskriptivní statistiky, které chcete zobrazit.

Odhadování statistiky a chybějící hodnoty

Můžete zvolit odhad středních hodnot, směrodatných odchylek, kovariance a korelací pomocí listwise (pouze úplné případy), po dvojicích, EM (expectation-maximization) a/nebo regresní metody. Můžete také zvolit, zda imputovat chybějící hodnoty (odhad náhradních hodnot). Všimněte si, že vícenásobný imputace je obecně považován za nadřazený k jediné imputaci za účelem vyřešení problému chybějících hodnot. Little's MCAR test je stále užitečný pro určení, zda imputace je nutná.

Metoda Listwise

Tato metoda používá pouze úplné případy. Pokud má některá z proměnných analýzy chybějící hodnoty, je tento případ vynechán z výpočtů.

Metoda po dvojicích

Tato metoda se dívá na dvojice proměnných analýzy a používá případ pouze v případě, že má chybějící hodnoty pro obě proměnné. Frekvence, průměry a směrodatné odchylky se vypočítávají zvlášť pro každou dvojici. Protože ostatní chybějící hodnoty v případě jsou ignorovány, korelace a kovariance pro dvě proměnné nezávisí na hodnotách, které chybí v jiných proměnných.

Metoda EM

Tato metoda předpokládá distribuci částečně chybějících dat a báží odvozuje z pravděpodobnosti pod touto distribucí. Každá iterace se skládá z kroku E a kroku M. Krok E vyhledá podmíněné očekávání "chybějících" dat vzhledem k pozorovaným hodnotám a aktuálním odhadům parametrů. Tato očekávání se poté nahrazují údaji o "chybějících". V kroku M se vypočtou maximální odhady pravděpodobnosti parametrů, jako kdyby chybějící data byla vyplněna. "Chybějící" je uzavřen v uvozovkách, protože chybějící hodnoty nejsou přímo vyplněny. Místo toho se funkce těchto funkcí používají v pravděpodobnosti protokolu.

Roderick J. A. Little's chi-square statistic for testing whether values are missing completed at random (MCAR) is printed as a patnote to the EM matrix. Pro tento test je nulová hypotéza to, že data chybí zcela náhodně a hodnota p je významná na úrovni 0,05. Je-li hodnota menší než 0,05, data se zcela náhodně nezmešká. Data mohou chybět náhodně (MAR) nebo nesmějí chybět náhodně (NMAR). Nemůžete předpokládat jeden nebo druhý a potřebujete analyzovat data, abyste určili, jak data chybí.

Regresní metoda

Tato metoda vypočítá více odhadů lineární regrese a má volby pro rozšíření odhadů s náhodnými komponentami. Pro každou předpovězenou hodnotu může procedura přidat reziduum z náhodně vybraného úplného případu, náhodné běžné odchylky nebo náhodné odchylky (měřítko druhou odmocninou zbytkového střední hodnoty) z rozdělení t .

Volby odhadu EM

Při použití iterativního procesu odhaduje metoda EM prostředky, kovarianční matici a korelaci kvantitativních (scale) proměnných s chybějícími hodnotami.

Distribuce. EM provádí inference na základě pravděpodobnosti v rámci uvedené distribuce. Předvoleně se předpokládá normální rozdělení. Pokud víte, že jsou ocasy rozdělení delší než běžné rozdělení, můžete požádat o to, aby procedura konstruoval funkci pravděpodobnosti ze studentova rozdělení t s n stupni volnosti. Smíchaná normální distribuce také poskytuje distribuci s delšími ocasy. Uveďte poměr směrodatné odchylky smíšené běžné distribuce a poměru směsi obou distribucí. Smíšená normální distribuce předpokládá, že se liší pouze směrodatné odchylky rozdělení. To znamená, že prostředky musí být stejné.

Maximum iterací. Nastaví maximální počet iterací pro odhad skutečné kovariance. Procedura se zastaví při dosažení tohoto počtu iterací, a to i v případě, že se odhady nekonvergují.

Uložit dokončená data. Datovou sadu s imputovanými hodnotami můžete uložit místo chybějících hodnot. Uvědomte si však, že statistiky založené na kovariance používající imputované hodnoty podhodnocují jejich příslušné hodnoty parametrů. Stupeň podhodnocení je úměrný počtu případů, které jsou společně nenápadné.

Chcete-li zadat volby EM

1. V hlavním dialogovém okně Analýza chybějící hodnoty vyberte proměnnou, pro kterou chcete odhadnout chybějící hodnoty pomocí metody EM.
2. Vyberte položku **EM** ve skupině Odhad.
3. Chcete-li určit proměnné předpovědi a prediktora, klepněte na volbu **Proměnné**. Další informace naleznete v tématu [“Proměnné predicted a Predictor”](#) na stránce 6 .
4. Klepněte na **EM**.
5. Vyberte volby EM, které chcete.

Volby odhadu regrese

Regresní metoda odhaduje chybějící hodnoty pomocí vícenásobného lineární regrese. Zobrazí se matice kovariance a korelační matice předpovídaných proměnných.

Úprava odhadu. Metoda regrese může přidat náhodnou komponentu k odhadům regrese. You can select residuals, normal variates, Student's *t* variates, or no adjustment.

- *Reziduy.* Chybové výrazy jsou vybírány náhodně z pozorovaných zbytkových chyb úplných případů, které mají být přidány do regresních odhadů.
- *Normální rozptyl.* Chybový výraz je náhodně vykreslen z rozdělení s očekávanou hodnotou 0 a směrodatnou odchylkou rovnající se druhé odmocnině střední kvadratická chybovo_chyby regrese.
- *Studentovo t Varianty.* Chybové výrazy jsou náhodně vykresleny z *t* distribuce se zadanými stupni volnosti a jsou zmenšeny střední kvadratická chyba (RMSE).

Maximální počet prediktorů. Nastavuje maximální limit počtu prediktorů (nezávislých) proměnných použitých v procesu odhadu.

Uložit dokončená data. Zapíše datovou sadu v aktuální relaci nebo externí datový soubor IBM SPSS Statistics s chybějícími hodnotami nahrazenými hodnotami odhadovanými regresní metodou.

Chcete-li určit volby regrese

1. V dialogovém okně Analýza chybějící hodnoty vyberte proměnnou, pro kterou chcete odhadnout chybějící hodnoty pomocí regresní metody.
2. Vyberte položku **Regrese** ve skupině Odhad.
3. Chcete-li určit proměnné předpovědi a prediktora, klepněte na volbu **Proměnné**. Další informace naleznete v tématu [“Proměnné predicted a Predictor”](#) na stránce 6 .
4. Klepněte na tlačítko **Regrese**.
5. Vyberte volby regrese, které chcete.

Proměnné predicted a Predictor

Ve výchozím nastavení jsou pro odhad EM a regrese použity všechny kvantitativní proměnné. V případě potřeby můžete zvolit specifické proměnné jako predikované a predikční proměnné v rámci odhadu (ú). Daná proměnná může být v obou seznamech, ale existují situace, ve kterých byste mohli chtít omezit použití proměnné. Někteří analytici jsou například nepohodlní odhad hodnot proměnných výsledku. Můžete také chtít použít různé proměnné pro různé odhady a spustit proceduru vícekrát. Například, pokud máte sadu položek, které jsou hodnocení zdravotních sester a jiné sady, které jsou hodnocení lékařů, můžete chtít provést jednu práci pomocí položky sester pro odhad chybějících položek ošetřovatelů a další spuštění pro odhady položek lékařů.

Další úvaha vzniká při použití regresní metody. Ve vícenásobné regresi může použití velké části nezávislých proměnných přinést horší predikované hodnoty než menší dílčí sada. Proto musí proměnné dosáhnout hodnoty *Faž do-zadejte* mezní hodnotu 4.0, která má být použita. Tento limit lze změnit pomocí syntaxe.

Chcete-li určit predikované a predikční proměnné

1. V dialogovém okně Analýza chybějící hodnoty vyberte proměnnou, pro kterou chcete odhadnout chybějící hodnoty pomocí regresní metody.
2. Vyberte **EM** nebo **Regrese** ve skupině Odhadce.
3. Klepněte na volbu **Proměnné**.
4. Chcete-li jako predikované a predikční proměnné použít spíše specifické než všechny proměnné, vyberte volbu **Vybrat proměnné** a přesuňte proměnné do příslušného seznamu (ú).

Dodatečné funkce příkazu MVA

Jazyk syntaxe příkazu vám také umožňuje:

- Určete samostatné popisné proměnné pro vzory chybějících hodnot, vzory dat a objemné vzory pomocí klíčového slova DESCRIBE na dílčích příkazech MPATTERN, DPATTERN nebo TPATTERN .
- Pro tabulku datových vzorů zadejte více než jednu proměnnou řazení pomocí dílčího příkazu DPATTERN .
- Určete více než jednu proměnnou řazení pro datové vzory pomocí dílčího příkazu DPATTERN .
- Uvedte toleranci a konvergenci pomocí dílčího příkazu EM .
- Uvedte toleranci a F-to-enter, pomocí dílčího příkazu REGRESSION .
- Uvedte různé seznamy proměnných pro EM a Regrese, pomocí dílčích příkazů EM a REGRESSION .
- Určete různé procentní části pro potlačení zobrazených případů, pro každý z produktů TTESTS, TABULATE a MISMATCH.

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

Vícenásobné imputace

Účelem vícenásobné imputace je generovat možné hodnoty pro chybějící hodnoty, a tím vytvořit několik "kompletní" sady dat. Analytické procedury, které pracují s více datovými sadami imputace, produkuje výstup pro každou "úplnou" datovou sadu a spolu se sloučenými výstupy, které odhadují, jaké výsledky by byly, kdyby původní datová sada neměla žádné chybějící hodnoty. Tyto společné výsledky jsou obecně přesnější než výsledky jednotlivých metod imputace.

Aspekty použití více neputacích dat












Proměnné analýzy. Proměnné analýzy mohou být:

- *Nominální.* Proměnná může být považována za nominální, když její hodnoty představují kategorie bez vlastního hodnocení (například oddělení společnosti, v níž zaměstnanec pracuje). Příklady nominálních proměnných zahrnují region, poštovní směrovací číslo a náboženskou příslušnost.
- *Pořadové.* Proměnnou lze považovat za ordinální, když její hodnoty představují kategorie s nějakým vlastním hodnocením (například úroveň spokojenosti služeb od vysoce nespokojené s vysoce spokojené). Příklady ordinálních proměnných zahrnují skóre přístupu představující stupeň spokojenosti nebo důvěryhodnost a skóre hodnocení preferencí.
- *Měřítko.* Proměnná může být považována za měřítko (souvislá), když její hodnoty reprezentují uspořádané kategorie se smysluplnou metrikou, takže porovnání vzdálenosti mezi hodnotami je vhodné. Příklady proměnných měřítka zahrnují věk v rocích a příjem v tisících dolarů.

Procedura předpokládá, že byla přiřazena vhodná úroveň měření ke všem proměnným; nicméně můžete dočasně změnit úroveň měření proměnné tak, že pravým tlačítkem myši klepnete na proměnnou ze seznamu zdrojových proměnných a vyberete úroveň měření z rozevírací nabídky. Chcete-li trvale změnit úroveň měření proměnné,

Ikona vedle každé proměnné v seznamu proměnných identifikuje úroveň měření a datový typ:

Tabulka 1. Ikony úrovně měření

	Číselné	Řetězec	Datum	Čas
Měřítko (spojité)		není k dispozici		
Pořadové				
Nominální				

Frekvence frekvencí. Hodnoty frekvence (replikace) jsou tímto postupem uznány. Případy se zápornou nebo nulovou hodnotou váhy replikace jsou ignorovány. Neceločíselné váhy jsou zaokrouhleny na nejbližší celé číslo.

Váha analýzy. Váhy analýzy (regrese nebo vzorkování) jsou začleněny do souhrnů chybějících hodnot a do modelů imputace. Případy s negativní nebo nulovou hmotností analýzy jsou vyloučeny.

Složené ukázky. Procedura vícenásobného imputace explicitně neobsahuje vrstvy, klastry nebo jiné komplexní struktury odběru vzorků, i když může přijmout konečné váhové závaží ve formě proměnné váhy analýzy. Všimněte si také, že procedury Complex Sampling momentálně automaticky neanalyzují imputované datové sady. Úplný seznam procedur, které podporují fondy, naleznete v tématu [“Analýza více imputací dat”](#) na stránce 13.

Chybějící hodnoty. Jak uživatel-tak i systémem-chybějící hodnoty jsou považovány za neplatné hodnoty; to znamená, že oba typy chybějících hodnot jsou nahrazeny, když jsou hodnoty imputovány a s oběma hodnotami se zachází jako s neplatnými hodnotami proměnných, které se používají jako prediktory v modelech imputace. Hodnoty uživatelských a systémových chybějících hodnot jsou v analýzách chybějících hodnot také považovány za chybějící.

Replikované výsledky (chybějící datové hodnoty Impute). Chcete-li replikovat výsledky imputace přesně, použijte stejnou inicializační hodnotu pro generátor náhodných čísel, stejné datové pořadí a stejné pořadí proměnných, kromě použití stejných nastavení procedury.

- **Generování náhodných čísel.** Procedura používá náhodné generování čísel během výpočtu imputovaných hodnot. Chcete-li v budoucnu reprodukovat stejné náhodné výsledky, použijte pro generátor náhodných čísel stejnou inicializační hodnotu před každým spuštěním procedury Chybějící hodnoty dat.
- **Pořadí případů.** Hodnoty jsou imputovány v pořadí případů.
- **Pořadí proměnných.** Metoda imputace plně podmíněné specifikace (FCS) imputuje hodnoty v pořadí uvedeném v seznamu Proměnné analýzy.

Existují dvě dialogová okna, která jsou vyhrazena pro více imputace.

- Volba [Analyzovat vzory](#) poskytuje popisná měřítka vzorů chybějících hodnot v datech a může být užitečná jako průzkumný krok před imputací.
- Volba [Imputovat chybějící datové hodnoty](#) se používá ke generování více imputací. Úplné datové sady lze analyzovat pomocí procedur, které podporují více datových sad imputace. Informace o analýze více datových sad imputace a seznam procedur, které podporují tato data, viz [“Analýza více imputací dat”](#) na stránce 13 .

Analýzovat vzory

Analýza vzorků poskytuje popisná měřítka vzorů chybějících hodnot v datech a může být užitečná jako průzkumný krok před imputací.

Příklad. Telekomunikační poskytovatel chce lépe porozumět vzorům užívání služeb v její databázi zákazníků. Mají kompletní údaje o službách používaných jejich zákazníky, ale demografické informace, které společnost shromáždila, mají řadu chybějících hodnot. Analýza vzorů chybějících hodnot může pomoci určit další kroky pro imputaci.

Z nabídky vyberte:

Analýzovat > Více imputace > Analyzovat vzory ...

1. Vyberte alespoň dvě proměnné analýzy. Procedura analyzuje vzory chybějících dat pro tyto proměnné.

Volitelná nastavení

Váha analýzy. Tato proměnná obsahuje váhy analýzy (regrese nebo vzorkování). Tento postup zahrnuje váhy analýzy v souhrnech chybějících hodnot. Případy s negativní nebo nulovou hmotností analýzy jsou vyloučeny.

Výstup. K dispozici je následující volitelný výstup:

- **Souhrn chybějících hodnot.** Tím se zobrazí výšečový graf s výšečovým grafem ukazující počet a procentní část proměnných analýzy, případů nebo jednotlivých datových hodnot, které mají jednu nebo více chybějících hodnot.

- **Vzorky chybějících hodnot.** Tím se zobrazí roztahlé vzory chybějících hodnot. Každý vzorek odpovídá skupině případů se stejným vzorkem neúplných a úplných dat o proměnných analýzy. Tento výstup můžete použít k určení toho, zda lze použít metodu monotónové imputace pro vaše data, nebo pokud ne, jak blízce jsou data přibližná monotónové vzorky. Procedura rozdává analýzy proměnných pro odhalení nebo přibližování monotónní struktury. Pokud po změně pořadí neexistuje žádný monotónový vzor, můžete vyvodit závěr, že data mají monotónní vzorec, když jsou proměnné analýzy uspořádány jako takové.
- **Proměnné s nejvyšší frekvencí chybějících hodnot.** Zobrazí tabulku proměnných analýzy seřazených podle procenta chybějících hodnot v sestupném pořadí. Tabulka obsahuje deskriptivní statistiky (střední a směrodatné odchylky) pro proměnné měřítka.

Můžete určit maximální počet proměnných, které mají být zobrazeny, a minimální procentní podíl chybějící proměnné, která má být zahrnuta do zobrazení. Zobrazí se sada proměnných, které splňují obě kritéria. Například při nastavení maximálního počtu proměnných na 50 a minimální procentní části chybí na 25 požadavků, které tabulka zobrazuje až 50 proměnných, které mají nejméně 25% chybějících hodnot. Existuje-li 60 analytických proměnných, ale pouze 15 má 25% nebo více chybějících hodnot, výstup obsahuje pouze 15 proměnných.

Impute chybějící datové hodnoty

K vygenerování více imputací se používají chybějící datové hodnoty. Úplné datové sady lze analyzovat pomocí procedur, které podporují více datových sad imputace. Informace o analýze více datových sad imputace a seznam procedur, které podporují tato data, viz [“Analýza více imputací dat”](#) na stránce 13 .

Příklad. Telekomunikační poskytovatel chce lépe porozumět vzorům užívání služeb v její databázi zákazníků. Mají kompletní údaje o službách používaných jejich zákazníky, ale demografické informace, které společnost shromáždila, mají řadu chybějících hodnot. Kromě toho tyto hodnoty nejsou zcela náhodné, takže k dokončení datové sady bude použita vícenásobná imputace.

Z nabídky vyberte:

Analyzovat > Více imputace > Nechybějící datové hodnoty Impute ...

1. Vyberte alespoň dvě proměnné v modelu imputace. Procedura impuuje více hodnot pro chybějící data pro tyto proměnné.
2. Uveďte počet imputací, které se mají vypočítat. Při výchozím nastavení je tato hodnota 5.
3. Uveďte datovou sadu nebo datový soubor formátu IBM SPSS Statistics, do kterého mají být zapsána imputovaná data.

Výstupní datová sada se skládá z původních dat případu s chybějícími daty a sadou případů s imputovanými hodnotami pro každou imputaci. Pokud má například původní datová sada 100 případů a vy máte pět imputací, výstupní datová sada bude mít 600 případů. Všechny proměnné ve vstupní datové sadě jsou zahrnuty do datové sady výstupu. Vlastnosti slovníku (názvy, štítky atd.) existujících proměnných se zkopírují do nové datové sady. Soubor také obsahuje novou proměnnou, *Imputation_*, číselnou proměnnou, která označuje imputaci (0 pro původní data, nebo 1 ..n pro případy s imputovanými hodnotami).

Procedura automaticky definuje proměnnou *Imputation_* jako proměnnou rozdělení , když se vytvoří výstupní datová sada. Jsou-li rozdělení v platnosti, když se procedura provádí, výstupní datová sada obsahuje jednu sadu imputací pro každou kombinaci hodnot rozdělených proměnných.

Volitelná nastavení

Váha analýzy. Tato proměnná obsahuje váhy analýzy (regrese nebo vzorkování). Tento postup zahrnuje analytické váhy v regresním a klasifikačních modelech používaných k vkládání chybějících hodnot. Váhy analýzy se také používají v souhrnech imputovaných hodnot; například střední hodnota, směrodatná odchylka a směrodatná chyba. Případy s negativní nebo nulovou hmotností analýzy jsou vyloučeny.

Pole s neznámou úrovní měření

Výstraha na úrovni měření se zobrazí, když je úroveň měření pro jednu nebo více proměnných (polí) v datové sadě neznámá. Jelikož úroveň měření ovlivňuje výpočet výsledků pro tuto proceduru, všechny proměnné musí mít definovanou úroveň měření.

Data skenování. Přečte data v aktivní datové sadě a přiřadí výchozí úroveň měření k jakýmkoli polím s momentálně neznámou úrovní měření. Je-li datová sada velká, může to nějakou dobu trvat.

Přiřadit ručně. Otevře dialogové okno se seznamem všech polí s neznámou úrovní měření. Toto dialogové okno můžete použít k přiřazení úrovně měření k těmto polím. Úroveň měření můžete také přiřadit v pohledu Proměnné v editoru dat.

Vzhledem k tomu, že úroveň měření je pro tuto proceduru důležitá, nemůžete přistupovat k dialogovému oknu pro spuštění této procedury, dokud nebude mít všechna pole definovanou úroveň měření.

Metoda

Ouško Metoda uvádí, jak budou přisouzena chybějící hodnoty, včetně typů použitých modelů. Kategorické prediktory jsou označeny indikátorem (fiktivní).

Metoda imputace

Metoda **Automaticky** skenuje data a používá metodu monotónové barvy, pokud data zobrazují neproporcionální vzorek chybějících hodnot; jinak se použije plně podmíněná specifikace. Pokud jste si jisti, kterou metodu chcete použít, můžete ji zadat jako metodu **Vlastní**.

Plně podmíněná specifikace

Toto je iterační metoda Markov řetězce Monte Carlo (MCMC), která může být použita, pokud je vzorek chybějících dat je libovolný (monotónní nebo non-monotónní).

Pro každou iteraci a pro každou proměnnou v pořadí uvedeném v seznamu proměnných platí, že metoda plně podmíněné specifikace (FCS) odpovídá modelu (single dependent variable), který používá všechny ostatní dostupné proměnné v modelu jako prediktory, a pak imputuje chybějící hodnoty pro proměnnou, která je vektována. Metoda pokračuje, dokud není dosaženo maximálního počtu iterací a imputované hodnoty na maximální iteraci se uloží do imputované datové sady.

Maximum iterací

Uvádí počet iterací, neboli "kroky", které přijal řetězec Markov použitý metodou FCS. Pokud byla metoda FCS vybrána automaticky, použije výchozí počet 10 iterací. Když výslovně vyberete FCS, můžete zadat vlastní počet iterací. Možná budete muset zvýšit počet iterací, pokud se Markov řetězec nekonvergoval. Na kartě Výstup můžete uložit data historie iterace FCS a zakreslit ji pro posouzení konvergence.

Monotónní

Jedná se o neiterativní metodu, která může být použita pouze v případě, že data mají neproporcionální vzorek chybějících hodnot. Existuje monotónový vzorec, který umožňuje seřadit proměnné tak, aby v případě, že proměnná má nechybějící hodnotu, měly všechny předchozí proměnné také nechybějící hodnoty. Při zadávání této hodnoty jako metody **Vlastní** nezapomeňte určit proměnné v seznamu v pořadí, v němž je uveden vzor monotónní šablony.

Pro každou proměnnou v monotónové objednávce se monotónní metoda hodí k modelu univariate (single dependent variable), který používá všechny předchozí proměnné v modelu jako prediktory, a pak imputuje chybějící hodnoty pro proměnnou, která se vešla. Tyto imputované hodnoty se ukládají do imputované datové sady.

Zahrnout dvousměrné interakce

Je-li metoda imputace vybrána automaticky, imputační model pro každou proměnnou obsahuje konstantní termín a hlavní efekty pro predikční proměnné. Při výběru specifické metody můžete volitelně zahrnout všechny možné dvousměrné interakce mezi kategoriálními prediktory prediktorů.

Typ modelu pro proměnné měřítka

Lineární regrese

Je-li metoda imputace vybrána automaticky, použije se lineární regrese jako univariální model pro proměnné měřítka.

Prediktivní střední shoda (PMM)

Při výběru specifické metody můžete alternativně zvolit prediktivní střední shodu (PMM) jako model pro proměnné měřítka. PMM je variantou lineární regrese, která zajišťuje, že imputované hodnoty jsou věrohodné. Pro PMM je imputovaná hodnota založena na hodnotě definované pro **Náhodně vybrat celý případ od nejbližší (k) předpovědi**, kde (*k*) je kladné celé číslo s výchozí hodnotou 5.

Logistická regrese se vždy používá jako univariate model pro kategoriální proměnné. Bez ohledu na typ modelu se kategorické prediktory zpracovávají pomocí kódování indikátoru (fiktivní).

Singulární tolerance

Samotné matrice (nebo neměnitelné) mají lineárně závislé sloupce, které mohou způsobit vážné problémy s algoritmem odhadování. Dokonce blízko-singulární matice může vést k špatné výsledky, takže procedura bude léčit matici, jejíž určujícím faktorem je menší než tolerance jako singulární. Zadejte kladnou hodnotu.

Omezení

Ouško Omezení vám umožňuje omezit roli proměnné během imputace a omezit rozsah imputovaných hodnot proměnné měřítka tak, aby byly věrohodné. Kromě toho můžete omezit analýzu na proměnné s méně než maximálním procentem chybějících hodnot.

Skenování dat pro souhrn proměnných. Klepnutím na volbu **Data skenování** způsobí, že seznam zobrazí proměnné analýzy a zjištěné procento chybějících, minimálních a maximálních hodnot pro každou z nich. Souhrny mohou být založeny na všech případech nebo mohou být omezeny na skenování prvních *n* případů, jak je uvedeno v textovém poli Případy. Klepnutím na volbu **Znovu skenovat data** můžete aktualizovat souhrny distribuce.

Definovat omezení

- **Funkce.** To vám umožňuje přizpůsobit sadu proměnných, které mají být imputovány a/nebo považovány za prediktory. Typicky je každá proměnná analýza považována za závislou a prediktorem v modelu imputace. Volbu **Role** lze použít k vypnutí imputace pro proměnné, které chcete použít jako **Pouze prediktor**, nebo k vyloučení proměnných z použití jako prediktorů (**Pouze Impute**) a tím modelu předpovědi je kompaktnější. Toto je jediné omezení, které lze zadat pro kategoriální proměnné, nebo pro proměnné, které se používají pouze jako prediktory.
- **Min a Max.** Tyto sloupce umožňují určit minimální a maximální přípustné imputované hodnoty pro proměnné měřítka. Pokud přičtená hodnota spadá mimo tento rozsah, procedura kreslí jinou hodnotu, dokud nenajde jednu v rozsahu nebo pokud nedosáhne maximálního počtu tahů (viz **Maximální počet tahů** níže). Tyto sloupce jsou k dispozici pouze v případě, že je vybrána volba **Lineární regrese** jako typ modelu proměnné měřítka na kartě Metoda.
- **Zaokrouhlení.** Některé proměnné mohou být použity jako měřítka, ale mají hodnoty, které jsou samozřejmě dále omezené; například počet lidí v domácnosti musí být celé číslo a částka vynaložená během návštěvy do obchodu s potravinami nesmí mít zlomkové centy. Tento sloupec umožňuje určit nejmenší hodnotu pro přijetí. Chcete-li například získat celočíselné hodnoty, zadejte hodnotu 1 jako hodnotu zaokrouhlení; chcete-li získat hodnoty zaokrouhlené na nejbližší cent, zadejte hodnotu 0,01. Obecně jsou hodnoty zaokrouhleny na nejbližší celé číslo násobek zaokrouhlení. Následující tabulka ukazuje, jak různé hodnoty zaokrouhlení působí na základě imputované hodnoty 6.64823 (před zaokrouhlením).

Ohraničující hodnota	Hodnota, na kterou je hodnota 6.64832 zaokrouhlena
10	10
1	7
0.25	6.75
0.1	6.6

Tabulka 2. Výsledky zaokrouhlení (pokračování)	
Ohraničující hodnota	Hodnota, na kterou je hodnota 6.64832 zaokrouhlena
0.01	6.65

Vyloučit proměnné s velkými objemy chybějících dat. Typicky jsou proměnné analýzy imputovány a používány jako prediktory bez ohledu na to, kolik chybějících hodnot mají, za předpokladu, že mají dostatečné údaje pro odhad imputace modelu. Můžete zvolit vyloučení proměnných, které mají vysoký procentní podíl chybějících hodnot. Pokud například uvedete hodnotu 50 jako **Maximální procentní podíl**, nebudou proměnné analýzy, které mají více než 50% chybějících hodnot, imputovány, ani se nepoužijí jako prediktory v modelech imputace.

Maximální počet tahů. Jsou-li zadány minimální nebo maximální hodnoty pro imputované hodnoty proměnných měřítka (viz **Min a Max** výše), pokusí se procedura o vykreslení hodnot pro případ, dokud nenajde sadu hodnot, které jsou v uvedených rozsazích. Pokud není sada hodnot získána v rámci zadaného počtu tahů, procedura nakreslí jinou sadu parametrů modelu a opakuje proces vykreslování případu. Pokud sada hodnot v rámci rozsahů není získána v rámci zadaného počtu velkých a malých písmen, dojde k chybě.

Všimněte si, že zvýšení těchto hodnot může zvýšit dobu zpracování. Pokud procedura trvá delší dobu nebo není schopna nalézt vhodné tahy, zkontrolujte minimální a maximální hodnoty, které jsou uvedené, aby se zajistilo, že jsou vhodné.

Výstup

Obrazovka. Řídí zobrazení výstupu. Celkový souhrn imputace je vždy zobrazen, což zahrnuje tabulky týkající se specifikací imputace, iterace (pro metodu plně podmíněné specifikace), imputované proměnné imputované proměnné, závislé proměnné vyloučené z imputace a sekvence imputace. Jsou-li zadány, zobrazí se také konstatty pro analýzu proměnných.

- **Imputační model.** To zobrazuje model imputace pro závislé proměnné a prediktory a zahrnuje univariální typ modelu, efekty modelu a počet přisouzaných hodnot.
- **Deskriptivní statistiky.** To zobrazí deskriptivní statistiky pro závislé proměnné, pro které jsou hodnoty imputovány. Pro proměnné měřítka jsou popisnými statistikami střední hodnota, počet, směrodatná odchylka, minimální a maximální odchylka původních vstupních údajů (před imputací), imputované hodnoty (imputace) a úplné údaje (původní a imputované hodnoty dohromady-imputace). Pro kategoriální proměnné popisná statistika zahrnuje počet a procento podle kategorie pro původní vstupní data (před imputací), imputované hodnoty (imputace) a úplné údaje (původní a imputované hodnoty společně-imputace).

Historie iterací. Je-li použita metoda imputace plně podmíněné specifikace, můžete si vyžádat datovou sadu, která obsahuje data historie iterace pro imputaci FCS. Datová sada obsahuje prostředky a směrodatné odchylky podle iterace a imputace pro každé závislé měřítko, pro které jsou hodnoty imputovány. Můžete zakreslit data, abyste mohli posoudit konvergenci modelu.

Další funkce příkazu MULTIPLE IMPUTATION

Jazyk syntaxe příkazu vám také umožňuje:

- Určete podmnožinu proměnných, pro které jsou zobrazeny deskriptivní statistiky (dílní příkaz `IMPUTATIONSUMMARIES`).
- Uveďte jak analýzu chybějících vzorků, tak imputaci v jednom spuštění procedury.
- Určete maximální povolený počet parametrů modelu při imputování libovolné proměnné (klíčové slovo `MAXMODELPARAM`).

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

Práce s více imputací daty

Je-li vytvořena datová sada s více imputací (MI), přidá se proměnná s názvem *Imputation_* s popisem proměnné *Imputation Numbera* datová sada se seřadí ve vzestupném pořadí. Případy z původní datové sady mají hodnotu 0. Případy pro imputované hodnoty jsou číslovány od 1 do *M*, kde *M* je počet imputací.

Když otevřete datovou sadu, přítomnost prvku *Imputation_* identifikuje datovou sadu jako možnou datovou sadu MI.

Aktivace vícenásobné datové sady imputace pro analýzu

Datovou sadu musíte rozdělit pomocí volby **Porovnat skupiny** s proměnnou seskupení *Imputation_*, aby bylo možné s analýzou zacházet jako s datovou sadou MI. Můžete také definovat rozdělení na jiných proměnných.

Z nabídky vyberte:

Data > Rozdělit soubor ...

1. Vyberte volbu **Porovnat skupiny**.
2. Vyberte volbu *Imputační číslo [Imputation_]* jako proměnnou, na které se budou seskupovat případy.

Případně, když zapnete značení (viz níže), je soubor rozdělen na *Imputační číslo [Imputation_]*.

Rozlišování hodnot získaných z pozorovaných hodnot

Můžete rozlišit imputované hodnoty od sledovaných hodnot podle barvy pozadí buňky, písma a tučného písma (pro imputované hodnoty). Při vytváření nové datové sady v aktuální relaci s chybějícími hodnotami Impute jsou značení standardně zapnuty. Při otevření uloženého datového souboru, který obsahuje imputace, jsou značení vypnuta.

Chcete-li zapnout značení, vyberte z nabídky Editoru dat následující volby:

Zobrazit > Označit Impuovaná data ...

Případně můžete zapnout značení klepnutím na tlačítko značení imputace na pravém okraji panelu úprav v zobrazení dat Editoru dat.

Přesun mezi imputacemi

1. Z nabídky vyberte:

Upravit > Přejít na imputaci ...

2. Vyberte imputaci (nebo původní data) z rozevíracího seznamu.

Případně můžete vybrat imputaci z rozevíracího seznamu na panelu úprav v zobrazení dat editoru dat.

Relativní pozice případu je při výběru imputace zachována. Pokud například v původní datové sadě existuje 1000 případů, případ 1034, případ 34th v první imputaci, se zobrazí v horní části mřížky. Pokud v rozevírací nabídce vyberete imputaci **2**, bude v horní části mřížky zobrazena velikost písmen 2034, 34th v imputaci 2. Vyberete-li v rozevírací nabídce **Původní data**, bude se v horní části mřížky zobrazovat velká a malá písmena. Pozice sloupce je také zachována při přechodu mezi imputacemi, takže je snadné porovnávat hodnoty mezi imputacemi.

Transformace a úpravy imputovaných hodnot

Někdy budete muset provést transformace na imputovaná data. Můžete například chtít vzít protokol všech hodnot proměnné platu a uložit výsledek do nové proměnné. Hodnota vypočtená pomocí imputovaných dat bude považována za imputovanou, pokud se liší od hodnoty vypočtené s použitím původních dat.

Upravíte-li imputovanou hodnotu v buňce Editoru dat, bude tato buňka stále považována za přisouzena. Nedoporučuje se upravovat imputované hodnoty tímto způsobem.

Analýza více imputací dat

Mnoho postupů podporuje sdružování výsledků z analýzy násobení imputovaných datových sad. Je-li zapnuto značení imputace, zobrazí se vedle procedur, které podporují sdružování, speciální ikona.

V podnabídce Analyzovat statistiku nabídky Analyzovat, například Frekvence, deskriptivní ukazatele, Prozkoumat a Kontingenční tabulky, které podporují sdružování podpory, zatímco poměr, P-P Plots a Q-Q Plots nikoli.

Tabulkový výstup i model PMML lze uložit do fondu. Neexistuje žádný nový postup pro požadavek na souhrnný výstup. Místo toho vám nová karta v dialogovém okně Volby dává globální kontrolu nad vícenásobným výstupem imputace.

- **Sdružování tabulkového výstupu.** Standardně jsou při spuštění podporované procedury na vícenásobné datové sadě imputace (MI) automaticky vytvářeny výsledky pro každou imputaci, původní (neimputovaná) data a sloučené (konečné) výsledky, které berou v úvahu odchylky mezi imputací. Statistika ve fondu se liší podle postupu.
- **Sloučení produktu PMML.** Z podporovaných procedur, které exportují PMML, můžete také získat společnou PMML. PMML ve fondu je požadován stejným způsobem jako, a je uložen místo nesloučené PMML.

Nepodporované procedury nevygenerují z fondu výstupu ani sdílené soubory PMML.

Úrovně použití fondu

Výstup je sloučený s použitím jedné ze dvou úrovní:

- **Dosud neaktuální kombinace.** K dispozici je pouze parametr ve fondu.
- **Jedinečná kombinace.** Směsný parametr, jeho standardní chyba, zkušební statistický test a efektivní stupně volnosti, p -hodnota, interval spolehlivosti a sdružování diagnostiky (zlomek chybějících informací, relativní účinnost, relativní nárůst rozptylu) jsou zobrazeny, jsou-li k dispozici.

Koeficienty (regrese a korelace), prostředky (a střední rozdíly) a počty jsou obvykle sdružené. Je-li k dispozici standardní chyba statistiky, použije se univariate pooling; jinak se použije nepředčítaný fond.

Procedury, které podporují rozdělování do oblastí

Následující procedury podporují datové sady MI, na úrovni fondů určených pro každý kus výstupu.

Frekvence. Podporovány jsou následující funkce:

- Statistická tabulka podporuje prostředky v Univariate pooling (pokud S.E. Je také požadováno) a platné N a chybějící N v Naivním pooling.
- Tabulka Frekvence podporuje frekvenci v Naivním sdružování.

Popisovače. Podporovány jsou následující funkce:

- Tabulka Deskriptivní statistiky podporuje prostředky ve fondu fondů (pokud je S.E. Je také požadováno) a N v Naivním sdružování.

Kontingenční tabulky. Podporovány jsou následující funkce:

- Tabulka Crosstaculation podporuje počet v Naivním sdružování.

Prostředky. Podporovány jsou následující funkce:

- Tabulka sestav podporuje střední hodnotu Univariate pooling (pokud S.E. Je také požadováno) a N v Naivním sdružování.

Jednovzorkový test T. Podporovány jsou následující funkce:

- Statistická tabulka podporuje střední hodnotu Univariate pooling a N v Naivním pooling.
- Testovací tabulka podporuje Mean Difference u fondů Univariate.

Test Independent-Samples T Test. Podporovány jsou následující funkce:

- Tabulka Statistika skupiny podporuje prostředky v rámci fondů Univariate a N v Naisei se sdružováním.
- Testovací tabulka podporuje Mean Difference u fondů Univariate.

Paired-test T Test. Podporovány jsou následující funkce:

- Statistická tabulka podporuje prostředky v oblasti Univariate pooling a N v Naivním sdružování do společné oblasti.
- The Correlations table supports Correlations and N at Naiving pooling.
- Testovací tabulka podporuje střední hodnotu ve fondu Univariate.

Jednofaktorová ANOVA. Podporovány jsou následující funkce:

- Tabulka Deskriptivní statistiky podporuje střední hodnotu Univariate pooling a N v Naivním sdružování.
- Tabulka Contrast Tests podporuje hodnotu Contrast u fondů Univariate.

GLM jedinečnost a vícerozměrný. Podporovány jsou následující funkce:

- Tabulka Between-Subjects Factors podporuje N v Naive pooling.
- Tabulka Deskriptivní statistiky podporuje střední a N v Naive sdružování.
- Tabulka Odhady parametrů podporuje koeficient B při Univariate pooling.
- The Estimated Marginal Means: Odhady table supports Mean at Univariate pooling.
- The Estimated Marginal Means: Pairwise Comparisons table supports Mean Difference at Univariate pooling

Lineární smíšené modely. Podporovány jsou následující funkce:

- Tabulka Deskriptivní statistiky podporuje střední a N v Naivní sdružování.
- Tabulka Odhady tabulky s pevnou efektivací podporuje odhad na použití fondu Univariate.
- Odhady tabulky parametrů Covariance podporují odhad na použití fondu Univariate.
- The Estimated Marginal Means: Odhady table supports Mean at Univariate pooling.
- The Estimated Marginal Means: Pairwise Comparisons table supports Mean Difference at Univariate pooling

Zobecně lineární modely a zobecně rovnice odhadování. Tyto postupy podporují sdružený PMML.

- Tabulka Informace o kategoriickém Proměnném podporuje N a Procenta v Najevněm sdružování.
- Tabulka Souvislé informace o proměnných podporuje N a střední hodnotu v nejichrovaném fondu.
- Tabulka Odhady parametrů podporuje koeficient B při Univariate pooling.
- Odhadovaná mezní hodnota: Odhad srovnávací tabulky podporuje střední hodnotu v Naivní oblasti.
- The Estimated Marginal Means: Odhady table supports Mean at Univariate pooling.
- The Estimated Marginal Means: Pairwise Comparisons table supports Mean Difference at Univariate pooling

Bibivariate Correlations. Podporovány jsou následující funkce:

- Tabulka Deskriptivní statistiky podporuje střední a N v Naivní sdružování.
- The Correlations table supports Correlations and N at Univariate pooling. Všimněte si, že korelace se transformují pomocí Fisherovy transformace z před sdílením a poté zpětně transformovány po sloučení.

Částečné korelace. Podporovány jsou následující funkce:

- Tabulka Deskriptivní statistiky podporuje střední a N v Naivní sdružování.
- Tabulka Correlations podporuje vzájemné vztahy Correlations v Naiving poolingu.

Lineární regrese. Tento postup podporuje sdružené PMML.

- Tabulka Deskriptivní statistiky podporuje střední a N v Naivní sdružování.
- The Correlations table supports Correlations and N at Naiving pooling.
- Tabulka Koeficienty podporuje B při Univariate pooling a Correlations v Naivním sdružování.
- Srovnávací tabulka Correlations podporuje vztahy Correlations v Naivním sdružování.
- Tabulka Statistika reziduí podporuje střední hodnoty a N v Naivním poolingu.

Binární logistická regrese. Tento postup podporuje sdružené PMML.

- Proměnné v tabulce Rovnice podporují B v Univariate pooling.

Multinomická Logistická regrese. Tento postup podporuje sdružené PMML.

- Tabulka Odhady parametrů podporuje koeficient B při Univariate pooling.

Ordinální regrese. Podporovány jsou následující funkce:

- Tabulka Odhady parametrů podporuje koeficient B při Univariate pooling.

Analýza diskriminantů. Tato procedura podporuje fond XML ve fondu.

- Tabulka Statistiky skupiny podporuje Střední a Platné hodnoty N v Naitom. fondu.
- Tabulka Matrices v oblasti v rámci skupin podporuje korelace ve fondu Najeven.
- Tabulka Cononical Discriminant Function Covariační tabulka podporuje nestandardizované koeficienty v Nanegovaných pooling.
- Funkce v tabulce Centroids skupiny podporuje nestandardizované koeficienty ve fondu Naiving pooling.
- Tabulka Koeficienty klasifikačních funkcí podporuje koeficienty v Najevpřípadě pooling.

Test chí-kvadrát. Podporovány jsou následující funkce:

- Deskriptivní tabulka podporuje průměr a N v Najektivním sdružování.
- Tabulka s frekvencí podporuje Pozorování N v Naivní oblasti.

Binomický test. Podporovány jsou následující funkce:

- Deskriptivní tabulka podporuje prostředky a N v Najektivním sdružování.
- Tabulka Test podporuje N, Pozorovaný proporci a Test Proportion v Naisoltě pooling.

Spouští test. Podporovány jsou následující funkce:

- Deskriptivní tabulka podporuje prostředky a N v Najektivním sdružování.

One-Sample Kolmogorovan-Smirnov Test. Podporovány jsou následující funkce:

- Deskriptivní tabulka podporuje prostředky a N v Najektivním sdružování.

Testy Two-Independent-Samples. Podporovány jsou následující funkce:

- Tabulka ranks podporuje střední pozici a N v Naivní sdružování.
- Tabulka Frekvence podporuje N v Naivním sdružování.

Testy pro několik nezávislých ukázek. Podporovány jsou následující funkce:

- Tabulka ranks podporuje střední pozici a N v Naivní sdružování.
- Tabulka Frekvence podporuje počty v Naivním sdružování.

Testy dvou souvisejících ukázek. Podporovány jsou následující funkce:

- Tabulka ranks podporuje střední pozici a N v Naivní sdružování.
- Tabulka Frekvence podporuje N v Naivním sdružování.

Testy pro několik souvisejících ukázek. Podporovány jsou následující funkce:

- Tabulka rank podporuje Mean Rank v Naivní sdružování.

Regrese systému Cox. Tento postup podporuje sdružené PMML.

- Proměnné v tabulce Rovnice podporují B v Univariate pooling.
- Tabulka COVariate znamená, že tabulka podporuje průměr v Naivním pooling.

Jednovzorkový proporci. Podporovány jsou následující funkce:

- Tabulka One-Sample Proportions Confiseit Intervals supports Proportion at Univariate pooling.
- Tabulka One-Sample Proportions Test podporuje Sledované-Test Value v Univariate pooling.

Proporce nezávislých vzorků. Podporovány jsou následující funkce:

- Tabulka Statistické údaje skupiny Proportions Independent-Samples Group podporuje část Proportion v Univariate pooling.
- Tabulka intervalů spolehlivosti Independent-Samples Defidence (Independent-Samples Defidence Intervals) podporuje
- Testovací tabulka Proportions Independent-Samples Test podporuje rozdíl v části Proportions v Univariate pooling.

Paired-Samples Proportions. Podporovány jsou následující funkce:

- Tabulka Statistické údaje o porcích vysílaných ukázek podporuje část Proportion v Univariate pooling.
- Tabulka Intervaly spolehlivosti vzorných vzorků podporuje rozdíl ve fondu Naiving pools.
- Zkušební tabulka Paired-Samples Proportions Test podporuje rozdíl v porcích ve fondu Univariate.

Více voleb imputace

Karta Vícenásobné imputace řídí dva druhy předvoleb souvisejících s vícenásobnými imputacemi:

Označování imputovaných dat. Ve výchozím nastavení budou buňky obsahující imputovaná data mít jinou barvu pozadí než buňky obsahující neimputovaná data. Charakteristickým vzhledem k imputovaným datům by mělo být snadné se posouvat přes datovou sadu a lokalizovat tyto buňky. Můžete změnit výchozí barvu pozadí buňky, písmo a nastavení imputovaných dat tučným písmem.

Výstup analýzy. Tato skupina řídí typ výstupu Prohlížeče produkovaný při každé analýze imputované datové sady. Standardně se výstup bude produkovat pro původní datovou sadu (pre-imputation) a pro každou z předaných datových sad. Kromě toho budou pro tyto postupy, které podporují sdružování imputovaných dat, vytvořeny konečné sloučené výsledky. Je-li prováděno jednorozměrné sdružování, zobrazí se také sdružování diagnostiky. Můžete však potlačit jakýkoli výstup, který nechcete vidět.

Nastavení více možností imputace

V nabídce vyberte:

Upravit > Volby

Klepněte na kartu Vícenásobné imputace.

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 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.

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 chybějící hodnoty
 Deskriptivní statistika [4](#)
 em [5](#)
 impurování chybějících hodnot [4](#)
 metody [4](#)
 očekávání-maximalizace [6](#)
 odhad statistiky [4](#)
 příkaz additional features [6](#)
 Regrese [6](#)
 Test MCAR [4](#)
 Vzorky [3](#)
Analyzovat vzory [8](#)

C

COVARIANCE
 v analýze chybějící hodnoty [5](#), [6](#)

E

em
 v analýze chybějící hodnoty [5](#)

F

frekvenční tabulky
 v analýze chybějící hodnoty [4](#)

H

historie iterace
 ve více imputaci [12](#)

CH

chybějící hodnoty
 univariate statistiky [4](#)
chybějící proměnné indikátoru
 v analýze chybějící hodnoty [4](#)

I

Impunit chybějící datové hodnoty
 metoda imputace [10](#)
 Omezení [11](#)
 výstup [12](#)

K

kategorie tabulek
 v analýze chybějící hodnoty [4](#)
Korelace
 v analýze chybějící hodnoty [5](#), [6](#)

M

Malinův test MCAR
 v analýze chybějící hodnoty [1](#)
monotónová imputace
 ve více imputaci [10](#)

N

neshoda
 v analýze chybějící hodnoty [4](#)
Neúplná data
 viz Analýza chybějící hodnoty [1](#)
normální rozptyly
 v analýze chybějící hodnoty [6](#)

O

odstranění litwise
 v analýze chybějící hodnoty [1](#)
odstranění po dvojicích
 v analýze chybějící hodnoty [1](#)

P

plně podmíněná specifikace
 ve více imputaci [10](#)
počty extrémních hodnot
 v analýze chybějící hodnoty [4](#)
proměnné indikátoru
 v analýze chybějící hodnoty [4](#)
případy roztakování
 v analýze chybějící hodnoty [3](#)

R

Regrese
 v analýze chybějící hodnoty [6](#)

Ř

řazení případů
 v analýze chybějící hodnoty [3](#)

S

směrodatná odchylka
 v analýze chybějící hodnoty [4](#)
Střední
 v analýze chybějící hodnoty [4–6](#)
Studentův t test
 v analýze chybějící hodnoty [6](#)

T

t test

v analýze chybějící hodnoty [4](#)

Test MCAR

v analýze chybějící hodnoty [1](#)

V

více imputace

analyzovat vzory [8](#)

impute chybějící datové hodnoty [9](#)

Vícenásobné imputace [13](#)

Z

zbytkové chyby

v analýze chybějící hodnoty [6](#)

