

# IBM SPSS COMPLEX SAMPLES

## Analyzujte výsledky komplexních výběrových šetření korektním způsobem

Korektní zpracování výzkumů založených na komplexních výběrových plánech není snadné. Statistické odhady a testy hypotéz jsou závislé na designu výběru, výpočtové vzorce pro odhady i pro jejich směrodatné chyby je třeba aplikovat podle situace a podle pravděpodobnostního modelu vzniku dat. Správné postupy a výsledky zajistí speciální statistický produkt pro takovou situaci.

Modul IBM® SPSS® Complex Samples vám korektně a snadno vypočítá odhady s jejich standardními chybami i v komplikovaných designech komplexních výběrů. Modul také zrealizuje váš výběrový design a připraví pravidla pro zpracování.

Modul můžete využít pro:

- **výběrová šetření** – popisné a inferenční statistiky z výběrových dat
- **výzkum trhu** – analýzy spokojenosti zákazníků, značky, komplexní studie trhu
- **zdravotnický výzkum** – analýzy velkých populací podle vybraných ukazatelů, jako například ukazatelů výživy, dopravních nehod, konzumace alkoholu apod.
- **sociální a demografický výzkum** – komplexní výzkumy o společnosti a sociálních problémech
- **výzkum veřejného mínění** – výzkum názorů na politická témata

## Nejdůležitější:

- Získejte korektní odhady prodava z komplexních výběrů
- Zvyšte přesnost svých analýz a zajistěte reprezentativnost výběru stratifikací
- Vybírejte skupinky jednotek
- Pracujte s víceúrovňovými výběry

## IBM SPSS Complex Samples nabízí všechny potřebné nástroje pro práci s komplexními výběry.

### Zahrnuje:

- intuitivního názorného průvodce (*Sampling Wizard*), který vás provede krok za krokem procesem přípravy výběrového plánu,
- uživatelsky příjemného průvodce pro analýzu dat (*Analysis Preparation Wizard*), který pomáhá s přípravou datových souborů z rozsáhlých složitých šetření,
- predikce numerických výstupů pomocí obecných lineárních modelů modulu IBM SPSS Complex Samples (*CSGLM*),
- predikce ordinálních výstupů pomocí ordinální regrese modulu IBM SPSS Complex Samples (*CSORDINAL*),
- predikce kategorizovaných výstupů pomocí logistické regrese modulu IBM SPSS Complex Samples (*CSLOGISTIC*),
- predikce délky života pomocí Coxovy regrese modulu IBM SPSS Complex Samples (*CSCOXREG*).

### Modul IBM SPSS Complex Samples pracuje s následujícími typy výběrů:

- **stratifikovaný výběr** – zajistíte si přesnější závěry a vyšší reprezentativnost rozdělením populace do klíčových skupin – oblastí/strat. Oblastmi mohou být regiony, pohlaví, věkové skupiny, povolání i jiné charakteristiky nebo kombinace více proměnných.
- **skupinkový výběr** – využijte skupinky výběrových jednotek. Skupinky jsou města, školy, třídy ve školách, nemocnice, rodiny, obyvatelé domu, jiné geografické jednotky apod. Výběrovými objekty jsou pak studenti, pacienti či obyvatelé. Skupinky zavádíme do výběrů obvykle z důvodů snížení nákladů a zjednodušení organizace výzkumu nebo za účelem podchytit skupinky jako jednotky analýzy.
- **vícestupňový výběr** – výběr postupně vytváříte ve stupních hierarchicky vložených úrovní skupinek. Nejprve zavedete první stupeň výběru podle určitých populačních kritérií a vybíráte jednotky prvního stupně (např. okresy). Druhý stupeň bude realizován na již vybraných jednotkách z prvního stupně (jednotkami druhého stupně mohou být např. obce). Opakováním těchto kroků lze docílit vícestupňového výběru (dále vybíráme postupně například sčítací okrsky, domy, domácnosti, z nichž nakonec vybereme osobu).

### Dosáhnete spolehlivějších závěrů

Každý výzkumník by si měl být jist spolehlivostí svých výsledků. Většina konvenčních statistických programů pracuje pouze se vzorci založenými na předpokladu prostého náhodného výběru. Provést prostý náhodný výběr v praxi je ale prakticky možné jen velmi zřídka a není to ani efektivní

z hlediska přesnosti odhadů a finančních nákladů pro velká šetření. Analýza výběrů konvenčními statistickými postupy proto vede k chybám z porušení předpokladů odhadových a testovacích postupů. Standardní chyby statistik bývají často podhodnoceny a zkreslují tak dosahovanou přesnost. Modul IBM SPSS Complex Samples poskytuje statisticky hodnotnější závěry o populacích z vašich výběrových dat, neboť zahrne vliv designu všech aspektů komplexního přístupu k výběrovému šetření.

### Pracujte efektivně a bez problémů

Modul IBM SPSS Complex Samples pracuje s výběrovými daty velmi snadno. Díky intuitivnímu uživatelskému rozhraní bez odkladů analyzujete data a ihned interpretujete výsledky. Po skončení práce datové soubory publikujete včetně výběrových a analytických plánů. Plány slouží také jako šablony pro budoucí aplikace. Tím nejen zvyšujete spolehlivost a šetříte svůj čas díky jednoduché aplikaci hotového zadání, ale vaši práci zužitkují i jiní, kteří chtějí postupy rekapitulovat, zkontrolovat nebo na vaši práci navázat.

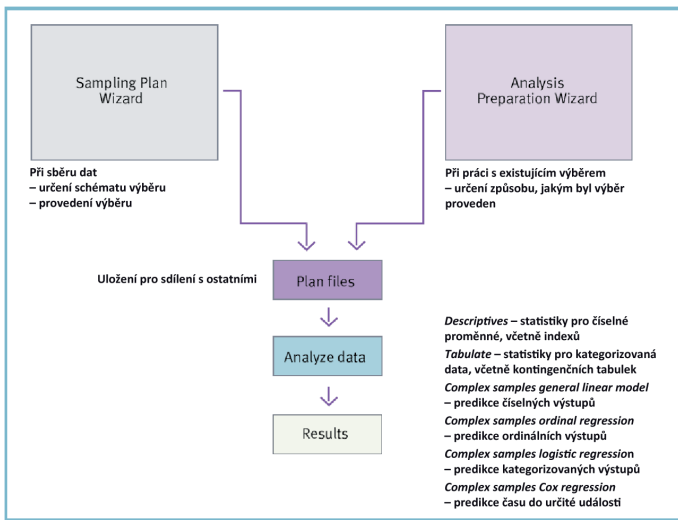
Práci v modulu IBM SPSS Complex Samples zahájíte pomocí průvodců, kteří vás navedou, a tak postupně určíte všechny parametry nutné pro plánování výběrových šetření. Vytváříte-li vlastní výběry, využijte průvodce pro zadání výběrového designu (*Sampling Wizard*). Jestliže analyzujete převzatá výběrová data, je k dispozici průvodce pro přípravu analýz *Analysis Preparation Wizard*, pomocí něhož definujete, jakým způsobem byl výběr získán a tím určíte, jak korektně odhadovat standardní chyby. Poté co jste vytvořili výběr nebo specifikovali způsob výpočtu standardních chyb a uložili tyto informace do analytického plánu, můžete začít analyzovat data a získáte korektní výsledky.

Modul IBM SPSS Complex Samples si osvojíte rychle a bezprostředně s ním začnete pracovat. Pro seznámení s celým postupem můžete také využít online help systém. K dispozici máte interaktivní nápovědu a případové studie či výukový program. Tato podpora zefektivní využití softwaru pro vaše konkrétní data. Práce bude rychlejší a ušetříte tak čas a náklady na svůj projekt.

Modul IBM SPSS Complex Samples nabízí:

- výpočet korektních bodových odhadů statistik jako jsou součty, průměry a poměrové ukazatele
- výpočet standardních chyb těchto statistik
- korektní intervaly spolehlivosti a testy hypotéz
- predikce číselných proměnných
- predikce ordinálních proměnných
- predikce kategorizovaných proměnných
- predikce délky života (*survival analysis*)

## Přesná analýza výzkumných dat



IBM SPSS Complex Samples je k dispozici pro instalaci v režimu lokální desktopové aplikace, avšak při požadavku na větší výkon a škálovatelnost lze provést také instalaci v režimu klient/server.

## SPECIFIKACE

## Complex Sample Plan (CSPLAN)

Procedura nabízí prostředí pro specifikaci výběrového designu a analýzy; využívají ji další procedury modulu IBM SPSS Complex Samples. CSPLAN neprovádí samostatně ani výběr ani neanalyzuje data. Pro realizaci výběru slouží procedura CSSELECT (popsaná níže) se vstupy, které vzniknou v CSPLAN. Analýzu se vstupy z CSPLAN pak realizují procedury CSDESCRIPTIVES, CSTABULATE, CSGLM, CSLOGISTIC nebo CSORDINAL (popsané níže).

- vytvoření designu výběru: využití k extrakci výběrových jednotek z aktivního datového souboru (výběrové opory)
- vytvoření analytického designu: užití k analýze výběrových dat komplexních výběrů
- při přípravě designu výběru se příslušný analytický design automaticky ukládá do zvláštního souboru, který slouží pro následnou realizaci výběru i pro provedení analýzy
- zobrazení výběrového i analytického plánu
- specifikace designu v externím souboru
- určení proměnných charakterizujících plán výběru, které se automaticky vytvoří při realizaci výběru nebo se využívají v procesu selekce, při odhadech nebo v analýzách (specifikace PLANVARS)
  - určení výsledných výběrových vah pro každý případ a pro použití v dalších analytických procedurách modulu IBM SPSS Complex Samples
  - indikace celkových vah, které se automaticky generují při realizaci výběru procedurou CSSELECT
  - výběr vah, ze kterých se počítají výsledné váhy ve vícestupňovém výběru

- řízení výstupu procedury CSPLAN specifikací PRINT
  - zobrazení souhrnu specifikací každé fáze výběrového plánu
  - zobrazení tabulky specifikací MATRIX
- specifikace DESIGN určuje stupně, stratifikační proměnné, skupinkové proměnné a tvoří popisná označení pro jednotlivé stupně
- specifikace METHOD nabízí výběr z řady metod přiřazujících jednotkám shodnou nebo různou pravděpodobnost včetně prostého a systematického náhodného výběru. K dispozici jsou též metody pro výběry s pravděpodobnostmi zahrnutí úměrnými velikosti skupiny (PPS). Případy mohou být vybírány s vrácením (WR) nebo bez vrácení (WOR).
  - SIMPLE\_WOR: výběr případů se stejnou pravděpodobností bez vrácení
  - SIMPLE\_WR: výběr případů se stejnou pravděpodobností s vrácením
  - SIMPLE\_SYSTEMATIC: výběr případů s pevným krokem z celé opory nebo oblasti; uvnitř prvního intervalu je zvolen náhodný začátek
  - SIMPLE\_CHROMY: výběr jednotek postupně se stejnou pravděpodobností bez vrácení
  - PPS\_WOR: výběr případů s pravděpodobností úměrnou velikosti bez vrácení
  - PPS\_WR: výběr případů s pravděpodobností úměrnou velikosti s vrácením
  - PPS\_SYSTEMATIC: systematický výběr s pravděpodobností úměrnou velikosti bez vrácení
  - PPS\_CHROMY: výběr jednotek postupně s pravděpodobností úměrnou velikosti bez vrácení
  - PPS\_BREWER: výběr dvou případů z každé oblasti s pravděpodobností úměrnou velikosti bez vrácení
  - PPS\_MURTHY: výběr dvou případů z každé oblasti s pravděpodobností úměrnou velikosti bez vrácení
  - PPS\_SAMPFORD: zobecnění Brewerovy metody pro výběr více než dvou případů z každé oblasti s pravděpodobností úměrnou velikosti bez vrácení
  - volba procenta jednotek nebo počtu vybraných případů na každém stupni výběru. Specifikace výstupních proměnných, které se generují při realizaci výběru (například váhy), pro každý stupeň výběru
  - metody odhadu: s vrácením, stejná pravděpodobnost bez vrácení na prvním stupni, nestejná pravděpodobnost bez vrácení
  - možnost zahrnutí korekce na konečnou populaci při odhadu rozptylu u prostého náhodného výběru
  - odhad pro nestejnou pravděpodobnost bez vrácení: pouze pro první stupeň
  - specifikace proměnných: určení vstupních proměnných výběrového procesu včetně vah a pravděpodobností zahrnutí

- *specifikace počtu vybraných případů na daném stupni příkazem SIZE*
- *specifikace procenta vybraných případů na daném stupni, např. výběrový podíl specifikačním příkazem RATE*
- *specifikace minimálního počtu vybraných případů při použití RATE; vhodné pro situace, kdy podíl vybraných případů je díky zaokrouhlování v některé z oblastí malý*
- *specifikace maximálního počtu vybraných případů při použití RATE; vhodné pro situace, kdy podíl vybraných případů je díky zaokrouhlování v některé oblasti větší než požadovaný*
- *specifikace míry velikosti (Measure of Size – MoS) v designu s pravděpodobností úměrnou velikosti (PPS); určení proměnných, které obsahují MoS pro proceduru CSSELECT a její aplikaci pro výběrovou oporu*
- *vytvoření informačních proměnných pro každý stupeň výběru specifikací STAGEVARS;*

můžete obdržet:

- *podíl jednotek vybraných z populace na každém stupni na základě zadaných pravděpodobností zahrnutí*
- *kumulativní výběrové váhy pro daný stupeň*
- *identifikace případů, které byly vybrány více než jednou pro výběry s vrácením; duplikační index pro případy vybrané na daném stupni*
- *velikost populace na daném stupni*
- *počet vybraných jednotek na daném stupni*
- *výběrový poměr vybraných jednotek*
- *váha pro daný stupeň*
- *výběr metody odhadu na daném stupni pomocí specifikace ESTIMATOR*
  - *stejně pravděpodobnosti výběru bez vrácení*
  - *nestejně pravděpodobnosti výběru bez vrácení*
  - *výběr s vrácením*
- *specifikace velikosti populace každého elementu výběru pomocí specifikace POPSIZE*
- *volba podílu případů vybraných z populace na daném stupni pomocí specifikace INCLPROB*

### Complex Samples Selection (CSSELECT)

Realizace komplexních pravděpodobnostních výběrů z populace. CSSELECT vybírá případy podle designu definovaného procedurou CSPLAN.

- *volba výchozího nastavení generátoru pseudonáhodných čísel specifikací CRITERIA*
- *volba, zda uživatelem definované chybějící hodnoty klasifikačních proměnných (oblasti, skupinky) budou či nebudou považovány za validní případy specifikací CLASSMISSING*
- *užití novějšího generátoru náhodných čísel Mersenne Twister pro výběr případů*
- *volba vstupních a výstupních souborů specifikací DATA*

- *změna názvů proměnných při zápisu výběrových vah procedurou CSSELECT; obdobně i dalších informačních proměnných (např. pravděpodobnosti zahrnutí)*
- *zápis vybraných případů do externího souboru s možností výběru specifikovaných proměnných*
- *automatické uložení pravděpodobností zahrnutí dvojic jednotek pro první stupeň výběru do externího souboru, pokud výběrový plán obsahuje specifikaci PPS\_WR*
- *generování textového souboru obsahujícího pravidlo, které popisuje charakteristiky vybraných jednotek*
- *výstupy specifikované příkazem PRINT*
  - *sumarizace distribucí vybraných případů v jednotlivých oblastech; informace je členěna podle výběrových stupňů designu*
  - *výpis sumarizace případů*

### Complex Samples Descriptives (CSDSCRIPTIVES)

Odhaduje průměry, součty, podíly a počítá jejich standardní chyby, efekty designu, intervaly spolehlivosti a provádí testování hypotéz na základě výběrových dat. Procedura bere při odhadu variability v úvahu design, podle kterého byl výběr proveden (stratifikace, stupně, metody se stejnou pravděpodobností zahrnutí případu, pravděpodobnost úměrná velikosti, výběry s vrácením, výběry bez vrácení). Uživatel může zvolit i analýzy dílčích populací.

- *volba názvu souboru s designem výběru vytvořeného procedurou CSPLAN; soubor obsahuje specifikace designu analýz se specifikací PLAN*
- *volba názvu souboru pro uložení pravděpodobností zahrnutí dvojic jednotek (pokud výběrový plán obsahuje specifikaci PPS\_WR)*
- *výběr analyzovaných proměnných užívaných specifikacemi MEAN, SUM a SUMMARY*
- *odhad průměrů a součtů proměnných specifikacemi MEANS, SUM a SUMMARY*
  - *požadován je Studentův t test populačních průměrů a součtů; testová hodnota se zadává klíčovým slovem TTEST; pro dílčí populace určené v SUBPOP je test proveden samostatně i pro celý základní soubor*
- *odhad podílů proměnných uvedených ve specifikaci SUMMARY pomocí specifikace RATIO*
  - *požadován je Studentův t test populačních podílů; testová hodnota se zadává klíčovým slovem TTEST*
  - *asociace syntaxe s odhadem průměru, součtu či poměru, což zahrnuje:*
    - *počet platných pozorovaných případů v datovém souboru pro každý odhad průměru, součtu či podílu*
    - *velikost populace pro odhad průměru, součtu či podílu*
    - *standardní chybu pro odhad průměru, součtu či podílu*
    - *koeficient variace*
    - *efekty designu*



- odmocninu z efektu designu
- interval spolehlivosti
- výběr subpopulací, pro které se provede analýza s použitím specifikace SUBPOP
  - zobrazení výsledků pro všechny subpopulace v jediné tabulce nebo v samostatných tabulkách
- způsob zacházení s vynechanými hodnotami
  - výpočet každé statistiky na základě všech platných dat analyzované proměnné (proměnných); výpočet podílů vychází ze všech případů s platnými hodnotami obou zvolených proměnných; při této volbě mohou statistiky vycházet z výběrů různé velikosti
  - do výpočtů vstupují pouze případy, které mají platné hodnoty u všech analyzovaných proměnných; všechny statistiky pro různé proměnné vycházejí ze stejné velké souboru
  - vyloučení uživatelem definovaných vynechaných hodnot v oblastech, ve skupinkách a u proměnných definujících subpopulace
  - zahrnutí uživatelem definovaných vynechaných hodnot v oblastech, ve skupinkách a u subpopulačních proměnných; uživatelem definované chybějící hodnoty jsou zpracovávány jako platná data

### Complex Sample Tabulate

Vytváří tabulky četností (třídění prvního stupně) a kontingenční tabulky (třídění druhého stupně) včetně standardních chyb, efektů designu, intervalů spolehlivosti a testů hypotéz pro komplexní výběry. Procedura odhaduje variabilitu na základě použitého designu a metod výběru (stejně pravděpodobnosti zahrnutí, PPS metody, výběrové procedury WR a WOR) CSTABULATE vytváří tabulky také pro dílčí populace.

- určení XML souboru s designem výběru vytvořeného procedurou CSPLAN příkazem PLAN
- volba souboru s uloženými pravděpodobnostmi zahrnutí dvojic jednotek (pokud výběrový plán obsahuje specifikaci PPS\_WR)
- tabulky mohou obsahovat následující statistiky:
  - velikost populace: odhad velikosti populace v jednotlivých buňkách tabulky a pro marginálie
  - standardní chyba: výpočet standardních chyb pro všechny odhady velikosti populace
  - řádková a sloupcová procenta: vyjadřují odhad velikosti populace v každé buňce jako procenta z odhadu velikosti populace pro celý řádek resp. sloupec; statistika je k dispozici pro kontingenční tabulky se dvěma vstupy
  - tabulková procenta: odhad velikosti populace v každé buňce jako procenta z odhadu velikosti populace pro celou tabulku

- koeficient variace
- efekt designu
- odmocnina z efektu designu
- intervaly spolehlivosti: výběr hladiny spolehlivosti mezi 0 % a 100 %
- nevážené četnosti: nevážené četnosti vyjadřují počet platných pozorování v datovém souboru pro odhad velikosti populace
- kumulativní odhady velikosti populace: kumulativní odhady velikosti populace jsou k dispozici pouze v tabulkách četností s jedním vstupem
- kumulativní procenta: kumulativními procenty lze odhadovat velikost populace pouze v tabulkách četností s jedním vstupem
- odhady očekávané velikosti populace: odhady očekávané velikosti populace pro každé pole kontingenční tabulky za předpokladu nezávislosti proměnných; k dispozici pouze pro kontingenční tabulky se dvěma vstupy
- rezidua: zobrazení rozdílů mezi odhady pozorovaných a očekávaných velikostí populace v každé buňce; k dispozici pouze pro kontingenční tabulky se dvěma vstupy
- Pearsonova rezidua: jsou k dispozici pouze pro kontingenční tabulky se dvěma vstupy
- adjustovaná Pearsonova rezidua: jsou k dispozici pouze pro kontingenční tabulky se dvěma vstupy
- pro celou tabulku lze využít následující statistiky a testy
  - test shody četností
  - test nezávislosti
  - podíl šancí
  - relativní riziko
  - rozdíly rizik
- výběr analyzovaných subpopulací specifikací SUBPOP
  - výsledky pro všechny subpopulace v jediné tabulce nebo v samostatných tabulkách
- způsob zacházení s vynechanými hodnotami
  - každá tabulka vychází ze všech platných dat pro tabelované proměnné; rozsahy výběrů se nemusí rovnat
  - každá tabulka vychází z případů, kde jsou platné hodnoty pro všechny analyzované proměnné; rozsahy výběrů se musí rovnat
  - uživatelem definované vynechané hodnoty uvnitř oblastí, skupinek a pro subpopulační proměnné nejsou zahrnuty do výpočtu
  - uživatelem definované vynechané hodnoty pro oblasti, skupinky a subpopulační proměnné jsou zahrnuty do výpočtu jako platné hodnoty

### Complex Samples General Linear Models (CSGLM)

Procedura nabízí modely lineární regrese, analýzy rozptylu

(ANOVA) a analýzy kovariance (ANCOVA) pro výběry provedené pomocí komplexních výběrových metod. Procedura odhaduje variabilitu na základě použitého designu a metod výběrů (stejně pravděpodobnosti zahrnutí, PPS metody, výběrové procedury WR a WOR). Procedura CSGLM volitelně poskytuje analýzy pro subpopulace.

- *modely*
  - hlavní efekty
  - všechny interakce  $n$ -tého řádu
  - plně kombinované faktory
  - vlastní modely, včetně hierarchických uspořádání
  - IBM Software Business Analytics IBM SPSS Statistics
- *statistiky*
  - odhady parametrů modelu: odhady koeficientů, standardní chyby,  $t$  testy, intervaly spolehlivosti, efekt designu a odmocnina z efektu designu pro každý odhad koeficientu
  - populační průměry závislých proměnných a kovariát
  - ověření modelu
  - informace o designu výběru
- *testy hypotéz*
  - testové statistiky: Waldův  $F$  test, upravený Waldův  $F$  test, Waldův chí-kvadrát test, upravený Waldův chí-kvadrát test
  - úprava pro vícenásobná porovnávání: nejmenší signifikantní rozdíl, Bonferroniho nerovnost, sekvenční Bonferroniho nerovnost, Šidákův test, sekvenční Šidákův test
  - výběrové stupně volnosti: založené na designu výběru nebo definované uživatelem
- *odhady průměrů: používají se odhady marginálních průměrů pro faktory a interakce v modelu*
  - kontrasty: jednoduchý, kontrast odchylek, Helmertův, opakovaný, polynomický
- *proměnné modelu mohou být uloženy do aktivního datového souboru a/nebo exportovány do externího souboru obsahujícího parametrické matice*
- *proměnné: predikované hodnoty a rezidua*
- *kovarianční a korelační matice parametrů a jejich další statistiky lze exportovat jako datový soubor IBM SPSS Statistics*
- *odhady parametrů a/nebo kovarianční matice parametrů lze exportovat také do XML souboru*
- *výstupy:*
  - informace o designu výběru (strata, PSU apod.)
  - odhady regresních koeficientů a  $t$  testy
  - sumární informace o závislé proměnné, kovariátách a faktorech
  - sumární informace o výběru, včetně nevážených počtů a velikosti populace
  - intervaly spolehlivosti pro odhady parametrů a uživatelem specifikované hladiny spolehlivosti

- Waldův  $F$  test pro efekty modelu
- efekty designu
- vícenásobný koeficient determinace
- matice koeficientů kontrastů ( $L$ )
- kovarianční matice odhadů regresních koeficientů
- odmocnina ze střední kvadratické chyby (mean square error)
- kovarianční a korelační matice regresních koeficientů
- *zpracování chybějících hodnot*
  - vyloučení chybějících hodnot metodou listwise
- *další možnosti*
  - uživatelem definovaný jmenovatel a stupně volnosti použité při výpočtu signifikancí ( $p$ -hodnot) u všech testových statistik
  - diagnostika kolinearity
  - vhodnost modelu pro subpopulace

### Complex Samples Ordinal (CSORDINAL)

Provádí regresní analýzu binární nebo ordinální závislé proměnné s využitím zvolené kumulativní funkce spojení (link function) pro komplikované designy komplexních výběrů. Procedura odhaduje variabilitu na základě použitého designu výběru (stejných pravděpodobností zahrnutí, PPS metody, výběrové procedury WR a WOR). Volitelně poskytuje procedura CSORDINAL analýzy pro subpopulace.

- *modely*
  - hlavní efekty
  - všechny interakce  $n$ -tého řádu
  - plně kombinované faktory
  - volba modelu, včetně hierarchických uspořádání
- *statistiky*
  - odhady parametrů modelu: odhady koeficientů, exponenciální odhady, standardní chyby všech odhadnutých koeficientů,  $t$  testy, intervaly spolehlivosti, efekt designu a odmocnina z efektu designu pro každý odhadnutý koeficient, kovariance a korelace odhadnutých parametrů
  - ověření modelu: Pseudo  $R^2$  a klasifikační tabulka
  - test rovnoběžnosti přímek: Waldův test rovnosti směrnic, odhady parametrů pro zobecněný model (nestejně směrnice), kovariance odhadů parametrů pro zobecněný model
  - sumární statistiky pro proměnných modelu
  - informace o designu výběru
- *testy hypotéz*
  - testové statistiky: Waldův  $F$  test, upravený Waldův  $F$  test, Waldův chí-kvadrát test, upravený Waldův chí-kvadrát test
  - úprava pro vícenásobná porovnávání: nejmenší signifikantní rozdíl, Bonferroniho nerovnost, sekvenční Bonferroniho nerovnost, Šidákův test, sekvenční Šidákův test

- výběrové stupně volnosti: na základě designu výběru nebo pevně definované uživatelem
- proměnné modelu mohou být uloženy v aktivním datovém souboru nebo exportovány do externího souboru obsahujícího parametrické matice
  - proměnné: predikovaná kategorie a pravděpodobnost predikované kategorie, pravděpodobnost pozorované kategorie, kumulativní pravděpodobnosti (jedna proměnná na kategorii), předpovídané pravděpodobnosti (jedna proměnná na kategorii)
  - kovarianční a korelační matice parametrů a další statistiky, mohou být exportovány do souboru IBM SPSS Statistics
  - odhady parametrů a/nebo parametrická kovarianční matice mohou být exportovány do XML souboru
- tři metody odhadu: Newton-Raphsonova metoda, Fisherovo skórování a Fisherovo skórování s následnou Newton-Raphsonovou metodou
- specifikace modelu pomocí kumulativní funkce spojení (link function): Cauchit, komplementární log-log, logit, negativní log-log a probit
- kumulativní podíly šancí pro specifikované faktory a kovariáty (k dispozici pouze pro logitovou funkci spojení)
- výstupy:
  - informace o designu výběru (strata, PSU apod.)
  - sumární informace o závislé proměnné, kovariátách a faktorech
  - sumární informace o výběru, včetně nevážených počtů a velikosti populace
  - intervaly spolehlivosti pro odhady parametrů a uživatelem specifikované hladiny spolehlivosti
  - sumární statistiky modelu
  - Waldův F test pro efekty modelu, upravený Waldův F test, Waldův chí-kvadrát test, upravený Waldův chí kvadrát test
  - efekty designu
  - klasifikační tabulka
  - matice (L) koeficientů kontrastů
  - kovarianční matice odhadů regresních koeficientů
  - tabulka obecné odhadnutelné funkce
  - korelační matice regresních koeficientů
- zpracování chybějících hodnot
  - vyloučení chybějících hodnot metodou listwise
- další možnosti
  - uživatelem definovaný jmenovatel, stupně volnosti, použité pro výpočet signifikancí (p-hodnot) všech testových statistik
  - diagnostika kolinearity
  - vhodnost modelu pro subpopulace

### Complex Samples Logistic Regression (CSLOGISTIC)

Provádí binární i multinomické (MLR) logistické regresní analýzy pro data získaná komplexními výběrovými metodami.

Procedura odhaduje variabilitu na základě použitého designu a metod (stejně pravděpodobnosti zahrnutí, PPS metody, výběrové procedury WR a WOR). Volitelně poskytuje procedura CSLOGISTIC analýzy v subpopulacích.

- modely
  - hlavní efekty
  - všechny interakce n-tého řádu
  - plně kombinované faktory
  - volba modelu, včetně hierarchických uspořádání
- statistiky
  - odhady parametrů modelu: odhady koeficientů, exponenciální odhady, standardní chyby, t testy, intervaly spolehlivosti, efekt designu a odmocnina z efektu designu pro každý odhad koeficientu, kovariance a korelace odhadnutých parametrů
  - ověření modelu: Pseudo R2 a klasifikační tabulka
  - sumární statistiky pro proměnné modelu
  - informace o designu výběru
- testy hypotéz
  - testové statistiky: Waldův F test, upravený Waldův F test, Waldův chí-kvadrát test, upravený Waldův chí-kvadrát test
- úprava pro vícenásobná porovnávání: nejmenší signifikantní rozdíl, Bonferroniho nerovnost, sekvenční Bonferroniho nerovnosti, Šidákův test, sekvenční Šidákův test
- výběrové stupně volnosti: založené na designu výběru nebo definované uživatelem
- proměnné modelu lze uložit do aktivního datového souboru a/nebo exportovat do externího souboru obsahujícího parametrické matice
  - proměnné: predikované kategorie a pravděpodobnosti predikcí
  - kovarianční a korelační matice parametrů a jejich další statistiky, lze exportovat do souboru IBM SPSS Statistics
  - odhady parametrů a/nebo parametrická kovarianční matice mohou být exportovány do XML souboru
- výstupy:
  - informace o designu výběru (strata, PSU apod.)
  - sumární informace o závislé proměnné, kovariátách a faktorech
  - sumární informace o výběru, včetně nevážených počtů a velikosti populace
  - intervaly spolehlivosti pro odhady parametrů a uživatelem specifikované hladiny spolehlivosti
  - sumární statistiky modelu
  - Waldův F test pro efekty modelu
  - efekty designu
  - klasifikační tabulka
  - matice koeficientů kontrastů (L)
  - kovarianční matice odhadů regresních koeficientů

- *odmocnina z průměrné kvadratické chyby (mean square error)*
- *kovarianční a korelační matice regresních koeficientů*
- *zpracování chybějících hodnot*
  - *vyločení chybějících hodnot metodou listwise*
- *další možnosti*
  - *uživatелеm definovaný jmenovatel, stupně volnosti, použité pro výpočet signifikancí (p-hodnot) všech testových statistik*
  - *diagnostika kolinearity*
  - *vhodnost modelu pro subpopulace*

## Complex Samples Cox Regression (CSCOXREG)

Aplikuje Coxovu regresi proporcionálního rizika na analýzu přežití, tj. analýzu doby do určité události, pro data sebraná na základě komplexního výběrového designu. CSCOXREG podporuje číselné i kategorizované prediktory, které mohou být závislé na čase. CSCOXREG představuje jednoduchý způsob analýzy rozdílů v podskupinách a také analýzy vlivu určité skupiny prediktorů. Procedura dokáže pracovat s daty zahrnujícími opakovaná pozorování, jako například několik návštěv pacienta u lékaře, opakované rozhovory nebo pozorování.

- *čas a událost: určení proměnné s časem trvání a hodnot indukujících, že událost, kterou sledujeme, již nastala*
  - *čas přežití*
    - *začátek intervalu (počátek rizika)*
      - *čas 0*
      - *změny u subjektu*
    - *konec intervalu*
    - *událost jako individuální hodnoty či rozpětí hodnot*
- *prediktory*
  - *faktory*
  - *kovariáty*
  - *časově závislé prediktory*
- *podskupiny: stratifikace analýz a/nebo jejich omezení na konkrétní subpopulace*
- *modely*
  - *hlavní efekty*
  - *všechny interakce n-tého řádu*
  - *volba modelu, včetně hierarchických uspořádání*
- *statistiky*
  - *informace o designu výběru*
  - *shrnutí událostí a cenzorování*
  - *nastavení rizika v čase události*
  - *odhady parametrů modelu: odhady koeficientů, exponenciální odhady, standardní chyby, t testy, intervaly spolehlivosti, efekt designu a odmocnina z efektu designu pro každý odhad koeficientu, kovariance a korelace odhadnutých parametrů*
  - *předpoklady modelu*
    - *test proporcionálních rizik*

- *odhady parametrů pro alternativní model*
- *kovarianční matice pro alternativní model*
- *základní funkce přežití a kumulativní funkce rizika*
- *testy hypotéz*
  - *testové statistiky: F test, upravený F test, chí-kvadrát test, upravený chí-kvadrát test*
  - *úprava pro vícenásobná porovnávání: nejmenší signifikantní rozdíl, Bonferroniho nerovnost, sekvenční Bonferroniho nerovnost, Šidákův test, sekvenční Šidákův test*
  - *výběrové stupně volnosti: založené na designu výběru nebo definované uživatelem*
- *proměnné modelu lze uložit do aktivního datového souboru a/nebo exportovat do externího souboru obsahujícího parametrické matice*
- *proměnné: funkce přežití, dolní a horní mez intervalu spolehlivosti pro funkci přežití, kumulativní funkce rizika, horní a horní mez intervalu spolehlivosti pro kumulativní funkci rizika, predikované hodnoty lineárního prediktoru, Schoenfeldova rezidua (jedna proměnná pro každý parametr modelu), Martingalova rezidua, devianční rezidua, Coxova-Snellova rezidua, skórová rezidua (jedna proměnná pro každý parametr modelu), DFBeta rezidua (jedna proměnná pro každý parametr modelu) agregovaná Martingalova rezidua, agregovaná devianční rezidua, agregovaná Coxova-Snellova rezidua, agregovaná skórová rezidua (jedna proměnná pro každý parametr modelu) a agregovaná DFBeta rezidua*
- *(jedna proměnná pro každý parametr modelu) parametr modelu a agregovaná DFBeta rezidua (jedna proměnná pro každý parametr modelu)*
- *export modelu a/nebo funkce přežití*
  - *ve formátu datového souboru IBM SPSS Statistics*
  - *export funkce přežití ve formátu datového souboru IBM SPSS Statistics*
  - *export modelu ve formátu XML*
- *možnosti specifikace kritérií pro odhady, metod pro výpočet funkcí přežití a intervalů spolehlivosti a práce s uživatelem definovanými chybějícími hodnotami*
  - *odhady: maximální počet iterací, maximální počet kroků půlení, omezení počtu iterací založené na změně odhadů parametrů, omezení počtu iterací založené na změně logaritmické věrohodnosti, zobrazení historie iterací a tie-breaking metoda odhadu parametrů (Efronova nebo Breslowa)*
  - *funkce přežití: metoda odhadu základních funkcí přežití (Efronova, Breslowa nebo product-limit) a intervaly spolehlivosti pro funkce přežití (transformované nebo původní jednotky)*
    - *určení hladiny pro intervaly spolehlivosti*
    - *ošetření chybějících hodnot (jejich zpracování jako platných nebo neplatných)*