

Manuál pro analýzu dat v softwaru STATISTICA

Software STATISTICA je produkt StatSoft, Inc. (www.statsoft.com, www.statsoft.cz). STATISTICA je dostupná v rámci MU z <https://inet.muni.cz/auth/login> (login stejný jako do www.is.muni.cz, seznam dostupných softwarů lze najít v oddílu Provozní služby).

Načtení datového souboru

Soubor -> Otevřít -> *vybrat datový soubor* -> Otevřít -> Importovat vybraný list do tabulky -> *vybrat list Excelovského souboru* -> OK -> *nechat zatržené* 1. řádek jako názvy proměnných -> OK -> Importovat jako textové popisky

Uložení datového souboru

Soubor -> Uložit -> *zadáme název souboru* -> Uložit

Zapnutí automatického filtru

Označit všechny sloupce (např. pomocí CTRL+A nebo kliknutím do levého horního rohu tabulky) -> Data -> Automatický filtr -> Automatický filtr

1. Vizualizace dat

Vytváření grafů pomocí záložky Grafy.

Koláčový graf

Grafy -> 2D grafy -> Výšečové grafy -> *zvolit proměnnou (např. Gender) (v záložce Detaily je možné zvolit, jakou legendu, typ a tvar grafu chceme (Legenda, Typ, Tvar) -> OK*
Po dvojným kliknutím na graf se nám ukáže okno Možnosti grafu, kde lze libovolně měnit barvu grafu i typ a tvar grafu a další parametry

Sloupcový graf (na ose y počty lidí)

Grafy -> Histogramy -> Proměnné -> *zvolit proměnnou (např. Group) -> OK -> zrušit zatržení Typ proložení: Normální -> na záložce Detaily zatrhnout Mezery mezi sloupci -> OK*

Sloupcový graf (na ose y procenta)

Grafy -> Histogramy -> Proměnné -> *zvolit proměnnou (např. Group) -> OK -> zrušit zatržení Typ proložení: Normální -> na záložce Detaily zatrhnout Mezery mezi sloupci -> na záložce Detaily změnit u Osa Y hodnotu N na % -> OK*

Histogram (na ose y procenta)

Grafy -> Histogramy -> Proměnné -> *zvolit proměnnou (např. Age) -> OK -> na záložce Detaily změnit u Osa Y hodnotu N na % (lze např. si vypsat i základní popisnou statistiku zatržením Popisné statistiky) -> OK*

Krabicový graf (s vykreslením odlehých hodnot)

Grafy -> 2D grafy -> Krabicové grafy... -> Proměnné -> *zvolit proměnnou (např. Age) jako Závislé prom. -> OK -> OK*

Krabicový graf (s minimem a maximem)

Grafy -> 2D Grafy -> Krabicové grafy... -> Proměnné -> *zvolit proměnnou (např. Age) jako Závislé prom. -> OK -> na záložce Detaily -> u Svorka zvolit Min-Max -> u Odl. hodn. & extrémy zvolit Vyp. -> OK*

2. Příprava dat pro analýzu

Nastavení formátu u MMSE na double

Dvakrát kliknout na šedé políčko s názvem proměnné -> nastavit Typ na Double -> nastavit Formát zobrazení na Číslo -> OK

Nastavení formátu u scan_date na datum

Dvakrát kliknout na šedé políčko s názvem proměnné -> nastavit Formát zobrazení na Datum -> vybrat formát 17/03/10 -> OK

Identifikace a odstranění duplikací

Data -> Filtrování dat/Překódování -> Filtrovat duplicitní případy -> Vstup: Proměnné -> ID -> OK -> u Output *zatrhnout* Vytvořit tabulku duplicit -> OK

*Je patrné, že se vždy zachová první záznam a druhý záznam je vyřazen bez ohledu na datum pořízení skenu. Pokud chceme, aby byl vždy odstraněn záznam se starším datem, je nejprve nutné data seřadit podle data pořízení skenu (sestupně) pomocí: Data -> Seřadit -> označit 1-ID -> Přidat prom. -> označit 30-scan_date -> Přidat prom. -> Sestupně -> OK -> Zahrnout formátování
Nový datový soubor bez duplikací uložit.*

Odstranění chybějících a chybných hodnot

Data -> Podmnožina -> Případy -> *zatrhnout* Povolit podmínky výběru – zapnout filtr -> do Zadané výrazem napsat `v4="" OR v4>110 OR v7=""` -> OK -> OK

Nový datový soubor bez chybějících a chybných hodnot uložit.

Rekódování proměnné Gender, aby obsahovala pouze hodnoty F a M

1. způsob – ručně: *Vyfiltrovat si řádek s hodnotou FF a hodnotu FF přepsat na F*

2. způsob – vytvořením nové proměnné: *Označit proměnnou za proměnnou Gender -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Gender_rek) -> do kolonky Dlouhé jméno napsat `=iif(v3="FF","F",v3)` -> OK*

Rekódování proměnné Group, aby obsahovala pouze hodnoty 1 (CN), 2 (MCI) a 3 (AD)

Označit proměnnou za proměnnou Group -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Group_3kat) -> do kolonky Dlouhé jméno napsat `=iif(v2=3;2;iif(v2=4;3;v2))` -> OK

Jiný způsob pomocí Data -> Překódovat...

Vytvoření textových popisků u kvalitativní proměnné

Dvakrát kliknout na šedé políčko s názvem proměnné -> Textové hodnoty -> zadat textové popisky a jejich příslušné číselné hodnoty -> OK -> OK

3. Popisná sumarizace dat

Popisná sumarizace dat pomocí Statistiky -> Základní statistiky/tabulky. Obecný popis dialogového okna pro sumarizaci dat, vizualizace a další analýzy je uveden na Obr. 1.

Popisná sumarizace kvalitativních dat – frekvenční tabulka

Statistiky -> Základní statistiky/tabulky -> Tabulky četností -> Proměnné -> zvolit proměnnou (např. Group) -> OK -> Výpočet

Popisná sumarizace kvantitativních dat

Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Age) -> OK -> na záložce Detailní výsledky *zatrhnout* Medián, Variační koeficient, Dolní & horní kvartily -> Výpočet

Výběr dat pro analýzu/graf

Obr. 1. Popis dialogového okna sloužícího pro sumarizaci, vizualizaci a další analýzy dat.

Popisná sumarizace kvantitativních dat – zapnutí filtru (vyfiltrování posledních 20 pacientů)

Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Height a Height_cor) -> OK -> Select Cases -> Zapnout filtr -> některé, vybrané: -> do "nebo číslý případů" zadat 814-833 -> OK -> na záložce Detailní výsledky zatrhnout Medián -> Výpočet

4. Transformace dat

Logaritmická transformace

Označit proměnnou za proměnnou, kterou chceme logaritmovat -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Weight_log) -> do kolonky Dlouhé jméno napsat =Log(v9) (Pozor, v softwaru STATISTICA je přirozený logaritmus označen jako Log(x) místo Ln(x)!) -> OK

Standardizace dat

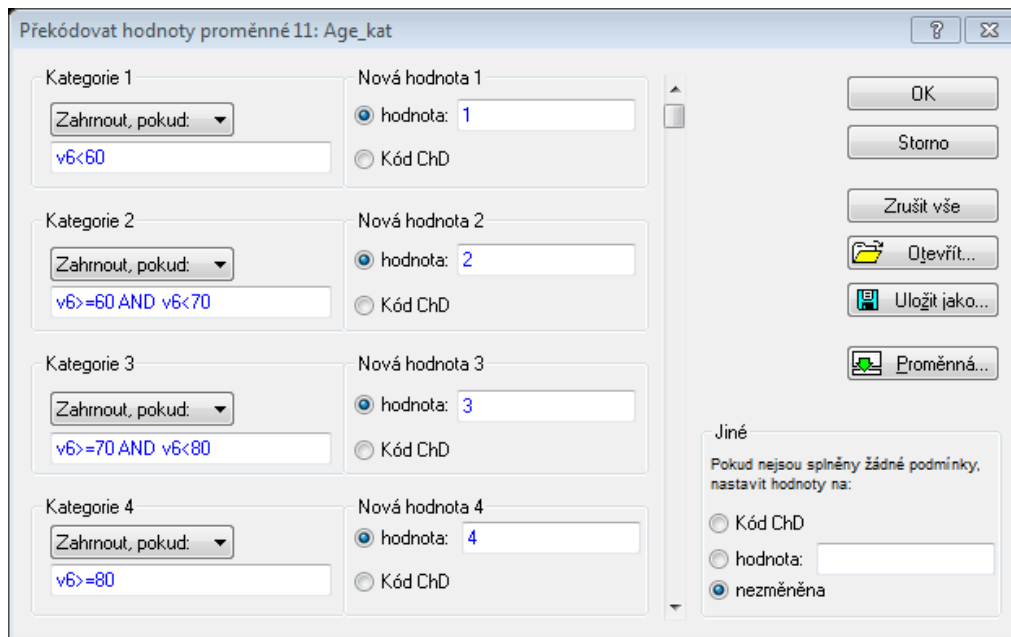
Označit proměnnou za proměnnou, kterou chceme standardizovat -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Age_st) -> do kolonky Dlouhé jméno napsat =v6 -> OK -> Data -> Standardizovat... -> OK

Centrování dat

Označit proměnnou za proměnnou, kterou chceme centrovat -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Height_centr) -> do kolonky Dlouhé jméno napsat =v9-174.15 (průměr vypočítaný pomocí Popisné statistiky) -> OK

Kategorizace

Označit proměnnou za proměnnou, kterou chceme kategorizovat -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Age_kat) -> OK -> Data -> Překódovat... (zkontrolovat si, že v záhlaví je správný název proměnné, jinak vybrat správnou proměnnou pomocí tlačítka Proměnná...) -> zadat podmínky a nové hodnoty (viz Obr. 2) -> OK



The screenshot shows a dialog box titled "Překódovat hodnoty proměnné 11: Age_kat". It contains four sections for defining categories:

- Kategorie 1:** "Zahnout, pokud:" dropdown is set to "v6<60". "Nová hodnota 1" is set to "hodnota: 1".
- Kategorie 2:** "Zahnout, pokud:" dropdown is set to "v6>=60 AND v6<70". "Nová hodnota 2" is set to "hodnota: 2".
- Kategorie 3:** "Zahnout, pokud:" dropdown is set to "v6>=70 AND v6<80". "Nová hodnota 3" is set to "hodnota: 3".
- Kategorie 4:** "Zahnout, pokud:" dropdown is set to "v6>=80". "Nová hodnota 4" is set to "hodnota: 4".

Each section has radio buttons for "hodnota:" (selected) and "Kód ChD". On the right side, there are buttons for "OK", "Storno", "Zrušit vše", "Otevřít...", "Uložit jako...", and "Proměnná...". At the bottom right, there is a "Jiné" section with radio buttons for "Kód ChD", "hodnota:", and "nezměněna" (selected).

Obr. 2. Ukázka kategorizace věku.

5. Intervaly spolehlivosti

Výpočet intervalu spolehlivosti a střední chyby průměru (standard error)

Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Age) -> OK -> na záložce Detailní výsledky zatrhnout Meze spolehl. prům. a Sm. chyba průměru -> Výpočet

Výpočet kvantilů Studentova rozložení

Statistiky -> Pravděpodobnostní kalkulátor -> Rozdělení... -> t (Studentovo) -> zatrhnout Inverze -> jako p zadat 0.975 -> jako sv (počet stupňů volnosti) zadat 832 -> Výpočet (vypočítá nám to hodnotu t)

6. Užitečná nastavení

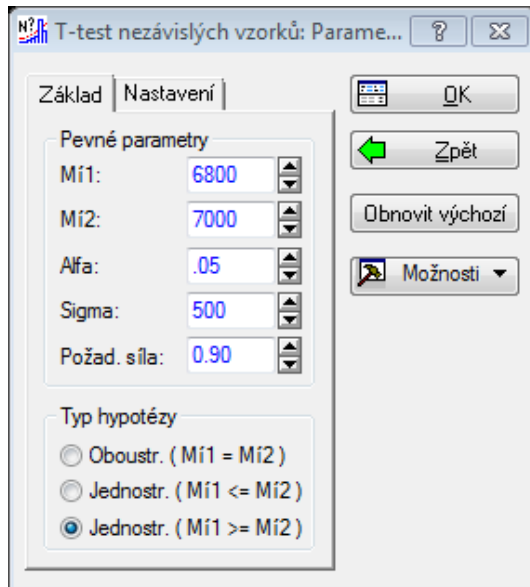
Vypnutí automatického překreslování grafů

Soubor -> Správce výstupů -> Grafy -> Nastavení -> Aktualizace dat přepnout na Uzamčen -> zrušit zatržení u Aktualizace stavů případů v tabulce -> OK

7. Výpočet velikosti vzorku a power analýza

Výpočet velikosti vzorku

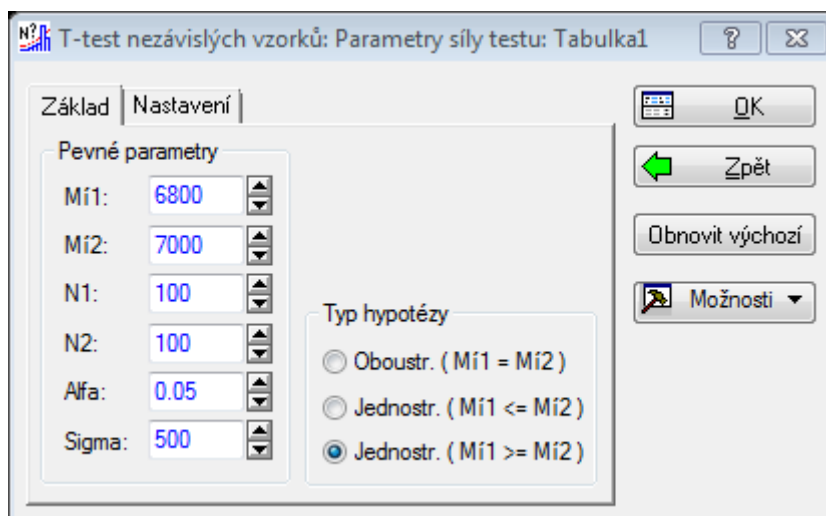
Statistiky -> Analýza síly testu -> Výpočet velikosti vzorku -> zvolit typ výpočtu podle typu našeho experimentu, který budeme chtít provést (např. Dva průměry, t-test, nezáv. vzorky) -> OK -> nastavit parametry a zvolit typ hypotézy (např. viz Obr. 3 – Pozor! Zadáváme typ nulové hypotézy, tedy zadáváme opak toho, co chceme prokázat!) -> OK -> Vypočítat N



Obr. 3. Ukázka nastavení parametrů u výpočtu velikosti vzorku.

Power analýza

Statistiky -> Analýza síly testu -> Výpočet síly testu -> zvolit typ výpočtu podle typu našeho experimentu, který budeme chtít provést (např. Dva průměry, t-test, nezáv. vzorky) -> OK -> nastavit parametry a zvolit typ hypotézy (např. viz Obr. 4 – Pozor! Zadáváme typ nulové hypotézy, tedy zadáváme opak toho, co chceme prokázat!) -> OK -> Vypočítat sílu



Obr. 4. Ukázka nastavení parametrů u power analýzy.

8. Statistické testy pro kvantitativní data – parametrické testy

Jednovýběrový t-test

Statistiky -> Základní statistiky/tabulky -> t-test, samost. vzorek -> OK -> zvolit *proměnnou* (např. Hippocampus_volume (mm³)) -> OK -> Test všech průměrů vůči: 6575 -> na záložce Možnosti *zatrhnout* Výpočet mezi spolehl. -> Výpočet

Párový t-test

Statistiky -> Základní statistiky/tabulky -> t-test, závislé vzorky -> OK -> zvolit *proměnné* (např. Hippocampus_volume (mm³) jako 1. seznam proměnných a Hippocampus_volume_24 (mm³) jako 2. seznam proměnných) -> OK -> Výpočet

Dvouvýběrový t-test

Statistiky -> Základní statistiky/tabulky -> t-test, nezávislé, dle skupin -> OK -> zvolit *proměnné* (např. Putamen_volume (mm³) jako Závislé proměnné a Gender_rek jako Grupovací proměnná) -> na záložce Možnosti lze zvolit Levenův test (test homogenity rozptylů) a Meze spol. pro odhady -> Výpočet

Analýza rozptylu (ANOVA) a post-hoc testy – 1. způsob

Statistiky -> Základní statistiky/tabulky -> Rozklad & jednofakt. ANOVA -> OK -> zvolit *proměnné* (např. Hippocampus_volume (mm³) jako Závislé proměnné a Group_3kat jako Grupovací proměnné) -> OK -> OK -> na záložce ANOVA & testy *kliknout* na Analýza rozptylu (vypíše ANOVA tabulku); dále lze vypsát i výsledky testů homogenity rozptylů: Leveneovy testy, Brown-Forsytheho testy -> na záložce Post-hoc *kliknout* na Tukeyův HSD (v případě vyrovnaných počtů subjektů ve skupinách), Tukey HSD - nestejná N (v případě nestejných počtů subjektů ve skupinách) nebo Schefféův test (pro stejné i nestejně počty subjektů ve skupinách)

Analýza rozptylu (ANOVA) a post-hoc testy – 2. způsob

Statistiky -> ANOVA -> Jednofaktorová ANOVA -> OK -> zvolit *proměnné* (např. Hippocampus_volume (mm³) jako Seznam závislých proměnných a Group_3kat jako Kategor. nezávislá proměnná (faktor)) -> OK -> OK -> Všechny efekty (vypíše ANOVA tabulku; prvního řádku s interceptem si nevěšíme) -> Více výsledků -> na záložce Post-hoc *kliknout* na Tukeyův HSD (pro stejné počty subjektů ve skupinách), HSD nestejně N (pro nestejně počty subjektů ve skupinách) nebo Schefféův (pro stejné i nestejně počty subjektů)

9. Statistické testy pro kvantitativní data – neparametrické testy

Wilcoxonův test – jednovýběrový

STATISTICA neumožňuje počítat jednovýběrový Wilcoxonův test přímo. Je nutné nejprve vytvořit novou proměnnou, která bude mít ve všech řádcích hodnotu, se kterou chceme srovnávat naše data: Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. mmse_konst) -> do kolony Dlouhé jméno napsat =27,5 (hodnota konstanty, se kterou chceme srovnávat) -> OK
Poté můžeme použít pro výpočet párový Wilcoxonův test: Statistiky -> Neparametrická statistika -> Porovnání dvou závislých vzorků (proměnné) -> OK -> zvolit *proměnné* (např. MMSE jako 1. seznam proměnných a mmse_konst jako 2. seznam proměnných) -> OK -> Wilcoxonův párový test (Je možné vypočítat i Znaménkový test, který je též neparametrickou alternativou párového t-testu.)

Wilcoxonův test – párový

Statistiky -> Neparametrická statistika -> Porovnání dvou závislých vzorků (proměnné) -> OK -> zvolit *proměnné* (např. MMSE jako 1. seznam proměnných a MMSE_24 jako 2. seznam proměnných) -> OK -> Wilcoxonův párový test

(Je možné vypočítat i Znaménkový test, který je též neparametrickou alternativou párového t-testu.)

Mannův-Whitneyův test

Statistiky -> Neparametrická statistika -> Porovnání dvou nezávislých vzorků (skupiny) -> OK -> zvolit proměnné (např. Hippocampus_volume (mm3) jako Seznam závislých proměnných a Gender_rek jako Nezáv. (groupov.) proměnná) -> OK -> M-W U test

Kruskalův-Wallisův test

Statistiky -> Neparametrická statistika -> Porovnání více nezávislých vzorků (skupiny) -> OK -> zvolit proměnné (např. MMSE jako Seznam závislých proměnných a Group_3kat jako Nezáv. (groupov.) proměnná) -> OK -> Výpočet (vypíše výsledky Kruskalova-Wallisova testu) -> Vícenás. porovnání průměrného pořadí pro vš. sk. (vypíše výsledky post hoc analýzy)

10. Ověření předpokladů statistických testů

Vykreslení Q-Q grafu pro jednotlivé skupiny

Grafy -> 2D grafy -> Normální pravděpodobnostní grafy... -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Výstup do jediné složky -> přepnout Uspořádání skupin na Sestupně -> OK -> lze zatrhnout Shapiro-Wilksův test (test normality dat) -> OK

Ověření normality pomocí Základní statistiky/tabulky pro jednotlivé skupiny

Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> OK -> na záložce Detailní výsledky zatrhnout Medián (to teď není nutné) -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Výstup do jediné složky a Sloučit tabulkové výsledky v jedné tabulce -> OK -> přepnout Uspořádání skupin na Sestupně -> OK -> na záložce Normalita zatrhnout Shapiro-Wilkův W test -> Tabulky četností -> na záložce Pravd. & bod. grafy lze nechat vykreslit Normální pravděpod. graf (tzn. Q-Q graf) -> na záložce Normalita lze nechat vykreslit Histogramy

Ověření normality pomocí Histogramy... pro jednotlivé skupiny

Grafy -> Histogramy -> Proměnné -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> OK -> na záložce Detaily změnit u Osa Y hodnotu N na %, zatrhnout Shapiro-Wilksův test a Kolmogorov-Smirnovův test -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Výstup do jediné složky a Sloučit tabulkové výsledky v jedné tabulce -> OK -> přepnout Uspořádání skupin na Sestupně -> OK -> OK

11. Další užitečné příkazy

Vykreslení tečkového grafu

Grafy -> Bodové grafy -> zvolit proměnné (např. Hippocampus_volume (mm3) jako X a Hippocampus_volume_24 (mm3) jako Y -> na záložce Detaily můžeme zatrhnout Korelace a p (lin. prolož.) -> OK

Popisná sumarizace kvantitativní proměnné podle kategorií kvalitativní proměnné

Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> OK -> na záložce Detailní výsledky zatrhnout Medián -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Výstup do jediné složky a Sloučit tabulkové výsledky v jedné tabulce -> OK -> přepnout Uspořádání skupin na Sestupně -> OK -> Výpočet

Vytvoření šablony grafů

Upravit si graf do finální podoby -> 2x kliknout do grafu -> Styly... -> Více -> 2x kliknout na Graf (aby se celý ten strom zavřel) -> kliknout na tlačítko s třemi tečkami -> Uložit jako -> pojmenovat graf -> Uložit -> Zavřít -> OK

Úprava grafu podle šablony

2x kliknout do grafu -> Styly... -> Více -> 2x kliknout na Graf (aby se celý ten strom zavřel) -> otevřít rozbalovací nabídku s typy grafů -> zvolit šablonu -> Upravit -> Zavřít -> OK

12. Analýza kontingenčních tabulek

Kontingenční tabulka absolutních četností

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group_3kat do List 1 a Age_kat do List 2)-> OK -> OK -> Výpočet

Kontingenční tabulka procent

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group_3kat do List 1 a Age_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Procenta z počtu v řádku, Procenta z počtu ve sloupci nebo Procenta celkového počtu -> Výpočet

Kontingenční tabulka – očekávané četnosti

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group_3kat do List 1 a Age_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Očekávané četnosti -> Výpočet

Pearsonův chí-kvadrát test

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group_3kat do List 1 a Age_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Pearsonův & M-V chí-kvadrát -> na záložce Detailní výsledky kliknout na Detailní 2-rozm. tabulky

Fisherův exaktní test

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Gender do List 1 a mmse_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Fisher exakt., Yates, McNemar (2 x 2) -> na záložce Detailní výsledky kliknout na Detailní 2-rozm. tabulky

McNemarův test

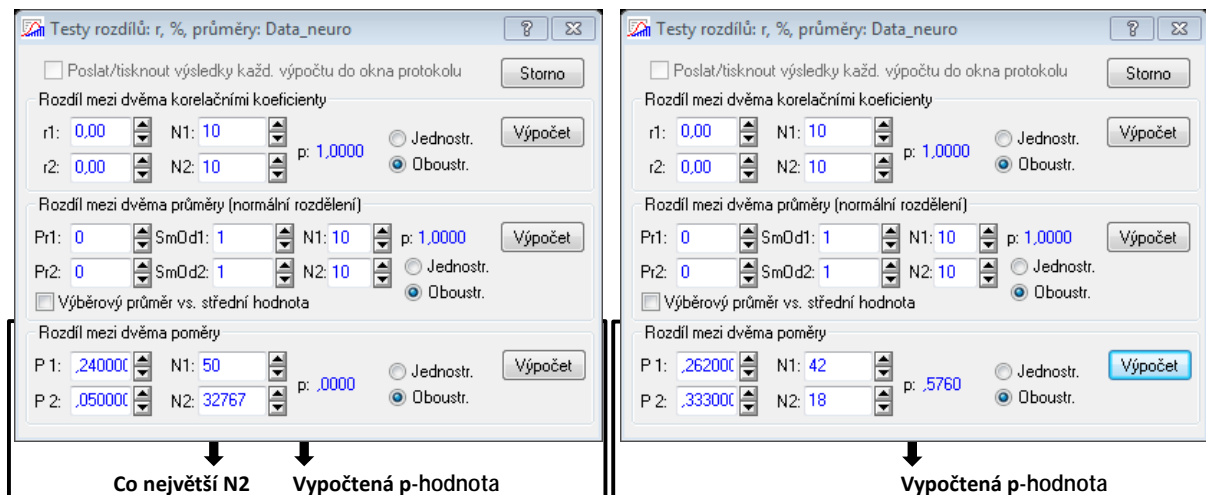
Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. mmse_kat do List 1 a mmse24_kat do List 2) -> OK -> OK -> na záložce Možnosti zatrhnout Fisher exact, Yates, McNemar (2 x 2) -> na záložce Detailní výsledky kliknout na Detailní 2-rozm. tabulky

13. Testy binomických dat

Jednovýběrový binomický test

Příklad: Mezi 50 pacientů s Alzheimerovou chorobou je 12 pacientů s MMSE skóre nižším než daná hranice. Ověřte, zda podíl pacientů s nižším skóre je stejný jako v běžné populaci (v běžné populaci uvažujte, že 5% lidí má hodnoty nižší než hranice).

Statistiky -> Základní statistiky/tabulky -> Testy rozdílů: r, %, průměry -> Rozdíl mezi dvěma poměry -> *zadat pravděpodobnosti a počty subjektů (viz Obr. 5 vlevo)* -> Oboustr. -> Výpočet (*dostaneme p-hodnotu*)



Obr. 5. Ukázka zadávání parametrů v jednovýběrovém (vlevo) a dvouvýběrovém binomickém testu (vpravo).

Dvouvýběrový binomický test

Příklad: Mezi 42 pacienty s Alzheimerovou chorobou (AD) je 11 pacientů s MMSE skóre nižším než daná hranice. Mezi 18 pacienty s mírnou kognitivní poruchou (MCI) je 6 pacientů s MMSE skóre nižším než daná hranice. Ověřte, zda se podíly pacientů s nižším skóre u pacientů s AD a MCI liší.

$$p_1 = 11/42 = 0,262 \text{ a } p_2 = 6/18 = 0,333$$

Statistiky -> Základní statistiky/tabulky -> Testy rozdílů: r, %, průměry -> Rozdíl mezi dvěma poměry -> *zadat pravděpodobnosti a počty subjektů (viz Obr. 5 vpravo)* -> Oboustr. -> Výpočet (*dostaneme p-hodnotu*)

14. ROC analýza

- I. Vytvoření proměnné `mmse_neg` (`mmse_neg = -mmse`), protože je nutné, aby hodnoty skóre pro kontrolní subjekty byly menší než hodnoty pro pacienty.
- II. Statistiky -> Pokročilé lineární/nelineární modely -> Zobecněné lineární/nelineární modely -> Logitový model -> OK -> *zvolit proměnné* (`group_01_CnMci` jako Závislá proměnná a `mmse_neg` jako Spojité nezáv. prom. -> *zvolit Odezv. kódy (zadat hodnoty: 1 0* (je nutné to mít v tomto pořadí, aby byla správně vypočítaná senzitivita a specifita; 1 značí rizikovou skupinu, 0 kontrolní skupinu)) -> OK -> OK -> na záložce Rezid. 1 kliknout na ROC křivka (vykreslí se graf s ROC křivkou a AUC)
- III. Pro zjištění cut-off (nejlepšího dělicího bodu) je nutné nejdříve vypsát frekvenční tabulku hodnot spojité proměnné: Statistiky -> Základní statistiky/tabulky -> Tabulky četností -> zvolit proměnnou (např. MMSE) -> Výpočet
- IV. Výslednou tabulku je nutné zkopírovat do Excelu: kliknout na levý horní roh tabulky (tím se celá tabulka označí) -> kliknout do tabulky pravým tlačítkem myši -> Kopírovat se záhlavími -> vložit do Excelu

- V. Do Excelu za tuto tabulku přikopírovat tabulku, která byla vytvořena jako výsledek ROC analýzy
- VI. V Excelu spočítat specifitu (tzn. udělat 1-(sloupeček s 1-Specifita)) a pak spočítat součet senzitivity a specifity -> vybrat řádek s největší hodnotou součtu senzitivity a specifity
- VII. Vytvoření kategorizovaného MMSE skóre s využitím cut-off: Vložit -> Přidat proměnné -> pojmenovat novou proměnnou (např. mmse_kat2) a do kolonky Dlouhé jméno napsat =iif(MMSE<=28;1;0) -> OK
- VIII. Ověření vypočítané senzitivity a specifity: Statistika -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. mmse_kat2 do List 1 a group_01_CnMci do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Procenta z počtu ve sloupci -> Výpočet (senzitivita je tady procento v pravém dolním rohu, specifita je procento v levém horním rohu – záleží ale na nakódování dat a v Listu 1 musí být výsledek diagnostického testu a v Listu 2 skutečnost)

15. Korelační analýza

Bodový graf ("Scatterplot")

Grafy -> Bodové grafy... -> zvolit proměnné (např. Putamen_volume (mm3) jako X a Amygdala_volume (mm3) jako Y -> OK -> na záložce Detaily lze zatrhnout Korelace a p (lin. prolož.) (vypočítá Pearsonův korelační koeficient a p-hodnotu) a R kvadrát -> OK

Výpočet Pearsonova korelačního koeficientu

Statistika -> Základní statistiky/tabulky -> Korelační matice -> OK -> 1 seznam proměn. -> zvolit proměnné (např. Amygdala_volume (mm3) a Putamen_volume (mm3)) -> na záložce Možnosti zvolit Zobrazit r, p-hodnoty a N -> Výpočet

Srovnání dvou korelačních koeficientů

Statistika -> Základní statistiky/tabulky -> Testy rozdílů: r, %, průměry -> OK -> zadáme hodnoty korelačního koeficientu a počet subjektů obou výběrů -> Výpočet (spočítá nám to p-hodnotu)

Výpočet Spearmanova korelačního koeficientu (výpočet čtvercové korelační matice)

Statistika -> Neparametrická statistika -> Korelace (Spearman, Kendallovo tau, gama) -> OK -> zvolit proměnné (např. MMSE a Hippocampus_volume (mm3)) -> OK -> Spearmanův koef. R

Výpočet Spearmanova korelačního koeficientu (výpočet detailní tabulky)

Statistika -> Neparametrická statistika -> Korelace (Spearman, Kendallovo tau, gama) -> OK -> zvolit Detailní report (místo Čtvercová matice) -> zvolit proměnné (např. MMSE do 1. seznam proměnných a Hippocampus_volume (mm3) do 2. seznam proměnných) -> OK -> Spearmanův koef. R

16. Regresní analýza

Lineární regrese a odstranění vlivu kovariát

Statistika -> Vícenásobná regrese -> zvolit proměnné (např. Nucl_caud_volume (mm3) jako Závislá prom., Age a gender_01,... jako Seznam nezáv. proměnných) -> OK -> OK -> Výpočet: Výsledky regrese (vypíše regresní koeficienty a p-hodnoty) -> OK

- Vykreslení Q-Q grafu pro rezidua: kliknout na Normální p-graf reziduí
- Vykreslení histogramu reziduí: na záložce Rezidua kliknout na Histogram reziduí (vpravo lze zvolit, zda chceme vykreslit histogram reziduí či standardizovaných reziduí)
- Vykreslení bodového grafu predikovaných hodnot a reziduí: na záložce Bodové grafy kliknout na Předpovědi vs. rezidua
- Uložení reziduí: na záložce Uložit kliknout na Uložit rezidua & předpovědi -> zvolit proměnné, které bude nově vytvořená tabulka dále obsahovat -> OK

17. Analýza přežití

Kaplanův-Meierův odhad funkce přežití pro jeden výběr

Statistiky -> Pokročilé lineární/nelineární modely -> Analýza přežívání -> Kaplan-Meierova metoda -> OK -> zvolit proměnné (např. čas jako Časy přežívání a umrti jako Indikátor cenzorov.) -> OK -> Kódy pro ukončené: 1 -> Kód pro cenzorované: 0 -> OK

- Vykreslení křivky přežití: *kliknout na Časy přežívání vs. kum. podíly přeživ.*
- Vypsání tabulky pro výpočet x-letého přežití: *kliknout na Výsledky: analýza přežívání*
- Vypsání tabulky s mediánem přežití: *na záložce Detaily kliknout na Kvantily funkce přežívání*

Kaplanův-Meierův odhad funkce přežití pro dva výběry

Statistiky -> Pokročilé lineární/nelineární modely -> Analýza přežívání -> Porovnání dvou vzorků -> OK -> zvolit proměnné (např. čas jako Přežívání, umrti jako Cenzor. prom., Skupina jako Grupovací prom.) -> OK -> Kódy pro ukončené: 1 -> Kód pro cenzorované: 0 -> zkontrolovat, jestli kategorie grupovací proměnné jsou v pořadí, jaké chceme -> OK

- Vykreslení křivek přežití: *na záložce Grafy funkcí kliknout na Kum. podíl přeživ. dle skupin (Kaplan Meier)*
- Výpočet testu na srovnání přežití ve skupinách: *na záložce Základní výsledky kliknout na Gehanův Wilcoxonův test nebo Ln-pořadový test (p-hodnota je v záhlaví tabulky)*
- Vypsání tabulky pro výpočet x-letého přežití: *na záložce Dvouvýběrové testy kliknout na Podíly přeživ. dle skupin (nevypisuje se ale střední chyba nutná pro výpočet intervalů spolehlivosti a nevypočítá se medián přežití – je nutné tyto údaje počítat pro každou křivku zvlášť pomocí Kaplanova-Meierova odhadu funkce přežití pro jeden výběr (postup viz výše))*

Kaplanův-Meierův odhad funkce přežití pro tři a více výběrů

Statistiky -> Pokročilé lineární/nelineární modely -> Analýza přežívání -> Porovnání více vzorků -> OK -> zvolit proměnné (např. čas jako Přežívání, umrti jako Cenzor. prom., Skupina jako Grupovací prom.) -> OK -> Kódy pro ukončené: 1 -> Kód pro cenzorované: 0 -> Kódy (skupin) -> Vše (kdyžtak změnit pořadí kategorií podle toho, jak potřebujeme) -> OK -> OK

- Vykreslení křivek přežití: *kliknout na Kumul. podíl přeživ. (Kaplan-Meier) dle skupin*
- Výpočet testu na srovnání přežití ve skupinách: *kliknout na Výpočet: časy přežívání & skóre (p-hodnota je v záhlaví tabulky)*
- Vypsání tabulky pro výpočet x-letého přežití: *na záložce Detaily kliknout na Procenta přeživ. dle skup. (nevypisuje se ale střední chyba nutná pro výpočet intervalů spolehlivosti a nevypočítá se medián přežití – je nutné tyto údaje počítat pro každou křivku zvlášť pomocí Kaplanova-Meierova odhadu funkce přežití pro jeden výběr (postup viz výše))*

Coxův model proporcionálních rizik

Statistiky -> Pokročilé lineární/nelineární modely -> Coxovy modely proporcionálního rizika -> zvolit proměnné (např. OS_doba_mesice jako Doby přežití, ECOG1 jako Kovariáty, Umrti jako Cenzorovací proměnná) -> OK -> Kód pro výskyt události: 1 -> Kód pro cenzorovaná data: 0 -> OK -> Parametrické odhady

Poděkování

Příprava výukových materiálů předmětu „DSAN01 Analýza dat pro Neurovědy“ byla finančně podporována prostředky projektu FRVŠ č. 942/2013 „Inovace materiálů pro interaktivní výuku a samostudium předmětu Analýza dat pro Neurovědy“.