

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

File -> Open -> vybrat datový soubor -> Open -> Import selected sheet to a Spreadsheet -> vybrat list Excelovského souboru -> OK -> nechat zatržené Get variable names from first row -> OK -> Import as Text Labels

Uložení datového souboru

File -> Save -> zadáme název souboru -> Save

Zapnutí automatického filtru

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

1. Vizualizace dat

Vytváření grafů pomocí záložky Graphs.

Koláčový graf

Graphs -> 2D Graphs -> Pie Charts -> zvolit proměnnou (např. Gender) (v záložce Advanced je možné zvolit, jakou legendu, typ a tvar grafu chceme (Pie Legend, Type, Shape) -> OK
Po dvojitým kliknutím na graf se nám ukáže okno Graph Options, kde lze libovolně měnit barvu grafu i typ a tvar grafu a další parametry

Sloupcový graf (na ose y počty lidí)

Graphs -> Histograms -> Variables -> zvolit proměnnou (např. Group) -> OK -> zrušit zatržení Fit type: Normal -> na záložce Advanced zatrhnout Breaks between columns -> OK

Sloupcový graf (na ose y procenta)

Graphs -> Histograms -> Variables -> zvolit proměnnou (např. Group) -> OK -> zrušit zatržení Fit type: Normal -> na záložce Advanced zatrhnout Breaks between columns -> na záložce Advanced změnit u Y axis hodnotu N na % -> OK

Histogram (na ose y procenta)

Graphs -> Histograms -> Variables -> zvolit proměnnou (např. Age) -> OK -> na záložce Advanced změnit u Y axis hodnotu N na % (lze např. si vypsat i základní popisnou statistiku zatržením Descriptive statistics) -> OK

Krabicový graf (s vykreslením odlehých hodnot)

Graphs -> 2D Graphs -> Box Plots... -> Variables -> zvolit proměnnou (např. Age) jako Dependent variable -> OK -> OK

Krabicový graf (s minimem a maximem)

Graphs -> 2D Graphs -> Box Plots... -> Variables -> zvolit proměnnou (např. Age) jako Dependent variable -> OK -> na záložce Advanced -> u Whisker zvolit Min-Max -> u Outliers zvolit Off -> OK

2. Příprava dat pro analýzu

Nastavení formátu u MMSE na double

Dvakrát kliknout na šedé políčko s názvem proměnné -> nastavit Type na Double -> nastavit Display format na Number -> OK

Nastavení formátu u scan_date na datum

Dvakrát kliknout na šedé políčko s názvem proměnné -> nastavit Display format na Date -> vybrat formát 17/03/10 -> OK

Identifikace a odstranění duplikací

Data -> Data Filtering/Recoding -> Filter Duplicate Cases -> Input: Variables -> ID -> OK -> u Output zatrhnout Create duplicates spreadsheet -> 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 -> Sort -> označit 1-ID -> Add Var(s) -> označit 30-scan_date -> Add Vars(s) -> Descending -> OK -> Include Formatting

Nový datový soubor bez duplikací uložit.

Odstranění chybějících a chybných hodnot

Data -> Subset -> Cases -> zatrhnout Enable Selection Condition -> do By expression 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 -> Insert -> Add Variables -> Name -> zadat název nové proměnné (např. Gender_rek) -> do Long name 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 -> Insert -> Add Variables -> Name -> zadat název nové proměnné (např. Group_3kat) -> do Long name napsat $=iif(v2=3;2;iif(v2=4;3;v2))$ -> OK

Jiný způsob pomocí Data -> Recode...

Vytvoření textových popisků u kvalitativní proměnné

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

3. Popisná sumarizace dat

Popisná sumarizace dat pomocí Statistics -> Basic Statistics/Tables. Obecný popis dialogového okna pro sumarizaci dat, vizualizace a další analýzy je uveden na Obr. 1.

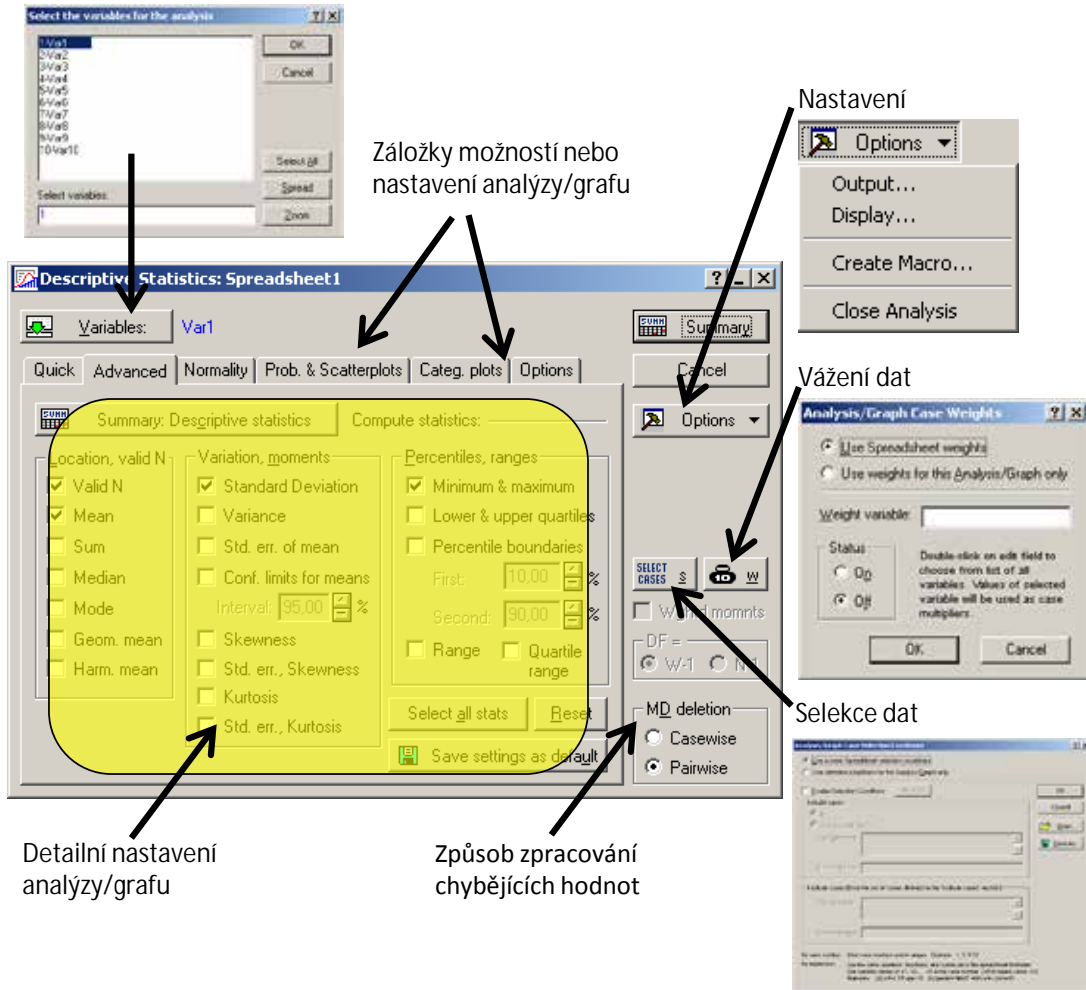
Popisná sumarizace kvalitativních dat – frekvenční tabulka

Statistics -> Basic Statistics/Tables -> Frequency tables -> Variables -> zvolit proměnnou (např. Group) -> OK -> Summary

Popisná sumarizace kvantitativních dat

Statistics -> Basic Statistics/Tables -> Descriptive statistics -> Variables -> zvolit proměnnou (např. Age) -> OK -> na záložce Advanced zatrhnout Median, Coefficient of variation, Lower & upper quartiles -> Summary

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)

Statistics -> Basic Statistics/Tables -> Descriptive statistics -> Variables -> zvolit proměnnou (např. Height a Height_cor) -> OK -> Select Cases -> Enable Selection Conditions -> Specific, selected by: -> do "or case number" zadat 814-833 -> OK -> na záložce Advanced zatrhnout Median -> Summary

4. Transformace dat

Logaritmická transformace

Označit proměnnou za proměnnou, kterou chceme logaritmovat -> Insert -> Add Variables -> Name -> zadat název nové proměnné (např. Weight_log) -> do Long name napsat =Log(v9) (Pozor, v softwaru STATISTICA je přirozený logaritmus označen jako Log(x) místo Ln(x)!) -> OK

Standardizace dat

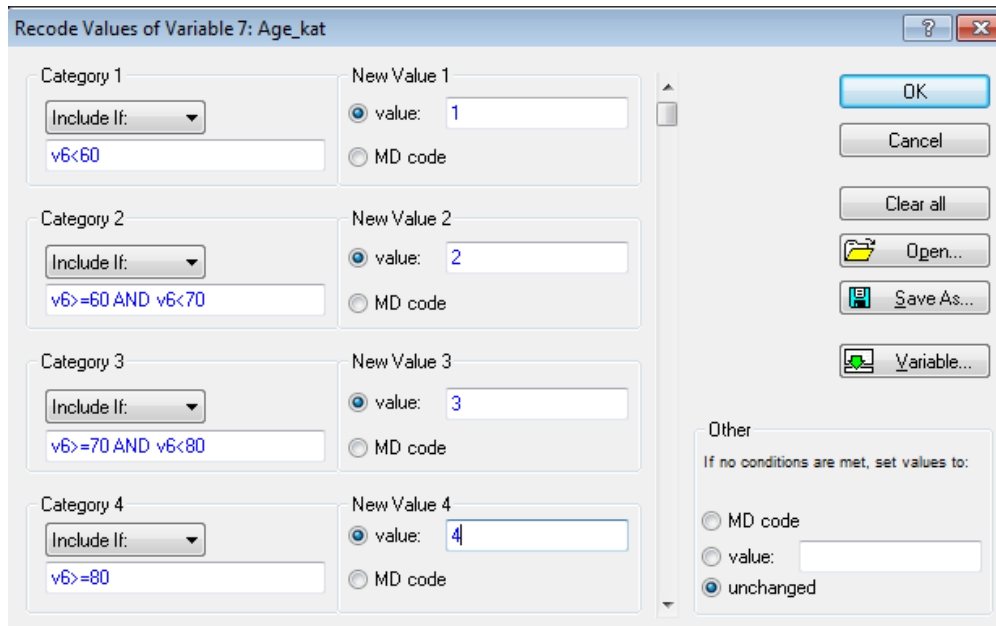
Označit proměnnou za proměnnou, kterou chceme standardizovat -> Insert -> Add Variables -> Name -> zadat název nové proměnné (např. Age_st) -> do Long name napsat =v6 -> OK -> Data -> Standardize... -> OK

Centrování dat

Označit proměnnou za proměnnou, kterou chceme centrovat -> Insert -> Add Variables -> Name -> zadat název nové proměnné (např. Height_centr) -> do Long name napsat =v9-174.15 (průměr vypočítaný pomocí Descriptive statistics) -> OK

Kategorizace

Označit proměnnou za proměnnou, kterou chceme kategorizovat -> Insert -> Add Variables -> Name -> zadat název nové proměnné (např. Age_kat) -> OK -> Data -> Recode... (zkontrolovat si, že v záhlaví je správný název proměnné, jinak vybrat správnou proměnnou pomocí tlačítka Variable...) -> zadat podmínky a nové hodnoty (viz Obr. 2) -> OK



Obr. 2. Ukázka kategorizace věku.

5. Intervaly spolehlivosti

Výpočet intervalu spolehlivosti a střední chyby průměru (standard error)

Statistics -> Basic Statistics/Tables -> Descriptive statistics -> Variables -> zvolit proměnnou (např. Age) -> OK -> na záložce Advanced zatrhnout Conf. limits for means a Std. err. of mean -> Summary

Výpočet kvantilů Studentova rozložení

Statistics -> Probability Calculator -> Distributions... -> t (Student) -> zatrhnout Inverse -> jako p zadat 0.975 -> jako df (degrees of freedom – počet stupňů volnosti) zadat 832 -> Compute (vypočítá nám to hodnotu t)

6. Užitečná nastavení

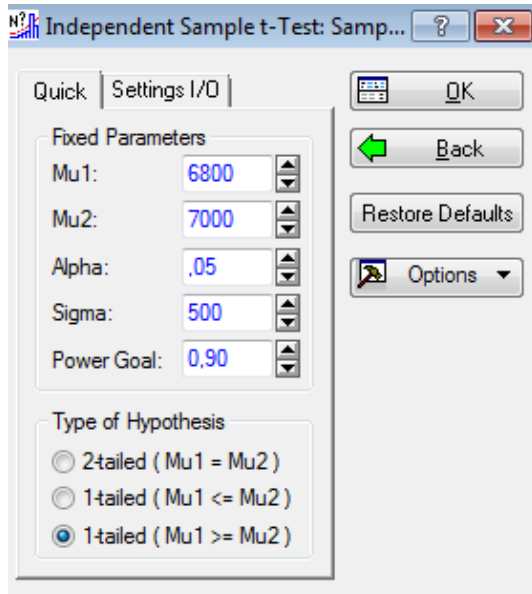
Vypnutí automatického překreslování grafů

File -> Output Manager -> Graphs -> Settings -> Data Update přepnout na Locked -> zrušit zatržení u Update spreadsheet case states -> OK

7. Výpočet velikosti vzorku a power analýza

Výpočet velikosti vzorku

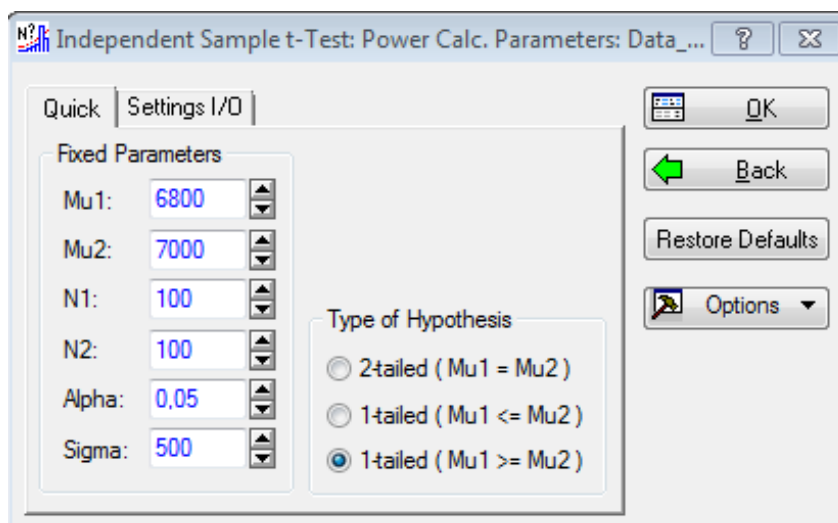
Statistics -> Power Analysis -> Sample Size Calculation -> zvolit typ výpočtu podle typu našeho experimentu, který budeme chtít provést (např. Two Means, t-Test, Ind. Samples) -> OK -> nastavit parametry a zvolit typ hypotézy (např. viz Obr. 3 – Pozor! Zadáváme typ nulové hypotézy, tedy zadáváme opak toho, co chceme prokázat!) -> OK -> Calculate N



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

Power analýza

Statistics -> Power Analysis -> Power Calculation -> zvolit typ výpočtu podle typu našeho experimentu, který budeme chtít provést (např. Two Means, t-Test, Ind. Samples) -> OK -> nastavit parametry a zvolit typ hypotézy (např. viz Obr. 4 – Pozor! Zadáváme typ nulové hypotézy, tedy zadáváme opak toho, co chceme prokázat!) -> OK -> Calculate Power



Obr. 4. Ukázka nastavení parametrů u power analýzy.

8. Statistické testy pro kvantitativní data – parametrické testy

Jednovýběrový t-test

Statistics -> Basic Statistics/Tables -> t-test, single sample -> OK -> zvolit proměnnou (např. Hippocampus_volume (mm³)) -> OK -> Test all means against: 6575 -> na záložce Options zatrhnout Compute conf. limits -> Summary

Párový t-test

Statistics -> Basic Statistics/Tables -> t-test, dependent samples -> OK -> zvolit proměnné (např. Hippocampus_volume (mm³) jako First variable a Hippocampus_volume_24 (mm³) jako Second variable) -> OK -> Summary

Dvouvýběrový t-test

Statistics -> Basic Statistics/Tables -> t-test, independent, by groups -> OK -> zvolit proměnné (např. Putamen_volume (mm³) jako Dependent variable a Gender_rek jako Grouping variable) -> na záložce Options lze zvolit Levene's test (test homogeneity rozptylů) a CI for estimates -> Summary

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

Statistics -> Basic Statistics/Tables -> Breakdown & one-way ANOVA -> OK -> zvolit proměnné (např. Hippocampus_volume (mm³) jako Dependent variable a Group_3kat jako Grouping variable) -> OK -> OK -> na záložce ANOVA & tests kliknout na Analysis of Variance (vypíše ANOVA tabulku); dále lze vypsat i výsledky testů homogeneity rozptylů: Levene tests, Brown-Forsythe test -> na záložce Post-hoc kliknout na Tukey honest significant difference (HSD) (v případě vyrovnaných počtů subjektů ve skupinách), Tukey HSD for unequal N (v případě nestejných počtů subjektů ve skupinách) nebo Scheffé test (pro stejné i nestejně počty subjektů ve skupinách)

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

Statistics -> ANOVA -> One-way ANOVA -> OK -> zvolit proměnné (např. Hippocampus_volume (mm³) jako Dependent variable a Group_3kat jako Categorical predictor (factor)) -> OK -> OK -> All effects (vypíše ANOVA tabulku; prvního řádku s interceptem si nevšímáme) -> More results -> na záložce Post-hoc kliknout na Tukey HSD (pro stejné počty subjektů ve skupinách), Unequal N HSD (pro nestejně počty subjektů ve skupinách) nebo Scheffé (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: Insert -> Add Variables -> Name -> zadat název nové proměnné (např. mmse_konst) -> do Long name napsat =27,5 (hodnota konstanty, se kterou chceme srovnávat) -> OK

Poté můžeme použít pro výpočet párový Wilcoxonův test: Statistics -> Nonparametric -> Comparing two dependent samples (variables) -> OK -> zvolit proměnné (např. MMSE jako First variable a mmse_konst jako Second variable) -> OK -> Wilcoxon matched pairs test

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

Wilcoxonův test – párový

Statistics -> Nonparametric -> Comparing two dependent samples (variables) -> OK -> zvolit proměnné (např. MMSE jako First variable a MMSE_24 jako Second variable) -> OK -> Wilcoxon matched pairs test

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

Mannův-Whitneyův test

Statistics -> Nonparametric -> Comparing two independent samples (groups) -> OK -> zvolit proměnné (např. Hippocampus_volume (mm3) jako Dependent variable list a Gender_rek jako Indep. (grouping) variable) -> OK -> M-W U test

Kruskalův-Wallisův test

Statistics -> Nonparametrics -> Comparing multiple indep. samples (groups) -> OK -> zvolit proměnné (např. MMSE jako Dependent variable list a Group_3kat jako Indep. (grouping) variable) -> OK -> Summary (vypíše výsledky Kruskalova-Wallisova testu) -> Multiple comparisons of mean ranks for all groups (vypíše výsledky post hoc analýzy)

10. Ověření předpokladů statistických testů

Vykreslení Q-Q grafu pro jednotlivé skupiny

Graphs -> 2D Graphs -> Normal Probability Plots... -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> By Group -> Grouping Variable(s) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Output to a single folder -> přepnout Sorting of Groups na Ascending -> OK -> lze zatrhnout Shapiro-Wilk test (test normality dat) -> OK

Ověření normality pomocí Basic Statistics/Tables pro jednotlivé skupiny

Statistics -> Basic Statistics/Tables -> Descriptive statistics -> Variables -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> OK -> na záložce Advanced zatrhnout Median (to teď není nutné) -> By Group -> Grouping Variable(s) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Output to a single folder a Accumulate tabular results in a single spreadsheet -> OK -> přepnout Sorting of Groups na Ascending -> OK -> na záložce Normality zatrhnout Shapiro-Wilk's W test -> Frequency tables -> na záložce Prob. & Scatterplots lze nechat vykreslit Normal probability plot (tzn. Q-Q graf)

Ověření normality pomocí Histograms... pro jednotlivé skupiny

Graphs -> Histograms -> Variables -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> OK -> na záložce Advanced změnit u Y axis hodnotu N na %, zatrhnout Shapiro-Wilk test a Kolmogorov-Smirnov test -> By Group -> Grouping Variable(s) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Output to a single folder a Accumulate tabular results in a single spreadsheet -> OK -> přepnout Sorting of Groups na Ascending -> OK -> OK

11. Další užitečné příkazy

Vykreslení tečkového grafu

Graphs -> Scatterplots -> zvolit proměnné (např. Hippocampus_volume (mm3) jako X a Hippocampus_volume_24 (mm3) jako Y -> na záložce Advanced můžeme zatrhnout Corr. and p (linear fit) -> OK

Popisná sumarizace kvantitativní proměnné podle kategorií kvalitativní proměnné

Statistics -> Basic Statistics/Tables -> Descriptive statistics -> Variables -> zvolit proměnnou (např. Hippocampus_volume (mm3)) -> OK -> na záložce Advanced zatrhnout Median -> By Group -> Grouping Variable(s) -> vybrat proměnnou (např. Gender_rek) -> OK -> zatrhnout Output to a single folder a Accumulate tabular results in a single spreadsheet -> OK -> přepnout Sorting of Groups na Ascending -> OK -> Summary

Vytvoření šablony grafů

Upravit si graf do finální podoby -> 2x kliknout do grafu -> Styles... -> More -> 2x kliknout na Graph (aby se celý ten strom zavřel) -> kliknout na tlačítko s třemi tečkami -> Save as -> pojmenovat graf -> Save -> Close -> OK

Úprava grafu podle šablony

2x kliknout do grafu -> Styles... -> More -> 2x kliknout na Graph (aby se celý ten strom zavřel) -> otevřít rozbalovací nabídku s typy grafů -> zvolit šablonu -> Edit -> Close -> OK

12. Analýza kontingenčních tabulek

Kontingenční tabulka absolutních četností

Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. *Group_3kat* do List 1 a *Age_kat* do List 2)-> OK -> OK -> Summary

Kontingenční tabulka procent

Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. *Group_3kat* do List 1 a *Age_kat* do List 2)-> OK -> OK -> *na záložce* Options *zatrhnout* Percentages of row counts, Percentages of column counts *nebo* Percentages of total count -> Summary

Kontingenční tabulka – očekávané četnosti

Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. *Group_3kat* do List 1 a *Age_kat* do List 2)-> OK -> OK -> *na záložce* Options *zatrhnout* Expected frequencies-> Summary

Pearsonův chí-kvadrát test

Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. *Group_3kat* do List 1 a *Age_kat* do List 2)-> OK -> OK -> *na záložce* Options *zatrhnout* Pearson & M-L Chi-square -> *na záložce* Advanced *kliknout* na Detailed two-way tables

Fisherův exaktní test

Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. *Gender* do List 1 a *mmse_kat* do List 2)-> OK -> OK -> *na záložce* Options *zatrhnout* Fisher exact, Yates, McNemar -> *na záložce* Advanced *kliknout* na Detailed two-way tables

McNemarův test

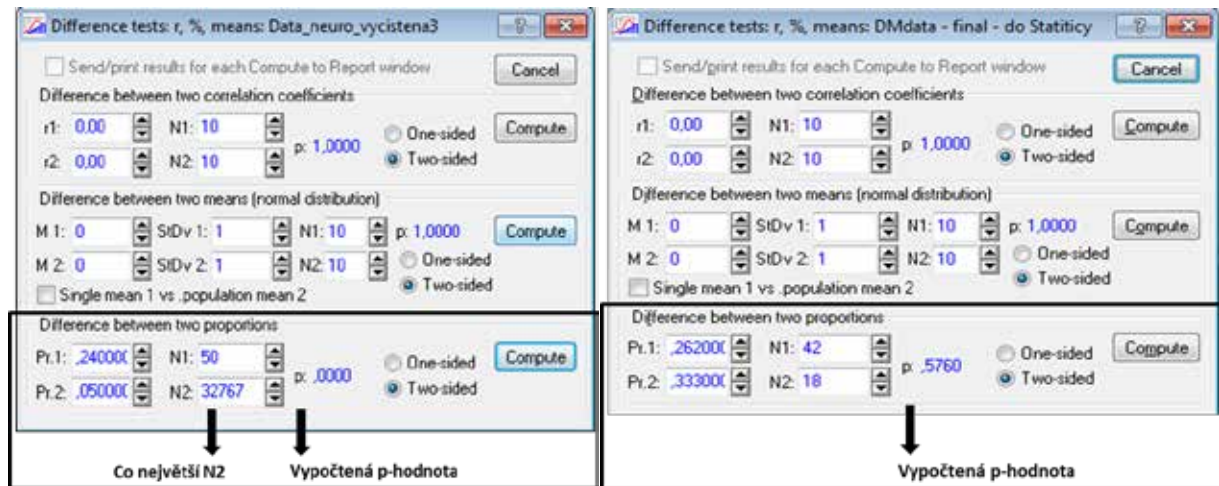
Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. *mmse_kat* do List 1 a *mmse24_kat* do List 2)-> OK -> OK -> *na záložce* Options *zatrhnout* Fisher exact, Yates, McNemar -> *na záložce* Advanced *kliknout* na Detailed two-way tables

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).

Statistics -> Basic Statistics/Tables -> Difference tests: r, %, means -> Difference between two proportions -> *zadat pravděpodobnosti a počty subjektů* (viz Obr. 5 vlevo) -> Two-sided -> Compute (*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.

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

Statistics -> Basic Statistics/Tables -> Difference tests: r, %, means -> Difference between two proportions -> *zadat pravděpodobnosti a počty subjektů* (viz Obr. 5 vpravo) -> Two-sided -> Compute (*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. Statistics -> Advanced Linear/Nonlinear Models -> Generalized Linear/Nonlinear Models -> Logit model -> OK -> *zvolit proměnné* (group_01_CnMci jako Dependent variable a mmse_neg jako Continuous predictor -> *zvolit Response codes* (*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 Resid. 1 kliknout na ROC Curve (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: Statistics -> Basic Statistics/Tables -> Frequency Tables -> zvolit proměnnou (např. MMSE) -> Summary
- 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 -> Copy with headers -> 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-Specificity)) 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: Insert -> Add Variables -> pojmenovat novou proměnnou (např. mmse_kat2) a do Long name napsat =iif(MMSE<=28;1;0) -> OK
- VIII. Ověření vypočítané senzitivity a specifity: Statistics -> Basic Statistics/Tables -> Tables and banners -> OK -> Specify tables (select variables) (např. mmse_kat2 do List 1 a group_01_CnMci do List 2)-> OK -> OK -> na záložce Options zatrhnout Percentage of column counts -> Summary (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")

Graphs -> Scatterplots... -> zvolit proměnné (např. Putamen_volume (mm3) jako X a Amygdala_volume (mm3) jako Y -> OK -> na záložce Advanced lze zatrhnout Corr. and p (linear fit) (vypočítá Pearsonův korelační koeficient a p-hodnotu) a R square (linear fit) -> OK

Výpočet Pearsonova korelačního koeficientu

Statistics -> Basic Statistics/Tables -> Correlation matrices -> OK -> One variable list -> zvolit proměnné (např. Amygdala_volume (mm3) a Putamen_volume (mm3)) -> na záložce Options zvolit Display r, p-values, and N's -> Summary

Srovnání dvou korelačních koeficientů

Statistics -> Basic Statistics/Tables -> Difference tests: r, %, means -> OK -> zadáme hodnoty korelačního koeficientu a počet subjektů obou výběrů -> Compute (spočítá nám to p-hodnotu)

Výpočet Spearmanova korelačního koeficientu (výpočet čtvercové korelační matice)

Statistics -> Nonparametrics -> Correlation (Spearman, Kendall tau, gamma) -> OK -> zvolit proměnné (např. MMSE a Hippocampus_volume (mm3)) -> OK -> Spearman rank R

Výpočet Spearmanova korelačního koeficientu (výpočet detailní tabulky)

Statistics -> Nonparametrics -> Correlation (Spearman, Kendall tau, gamma) -> OK -> zvolit Detailed report (místo Square matrix) -> zvolit proměnné (např. MMSE do First variable list a Hippocampus_volume (mm3) do Second variable list) -> OK -> Spearman rank R

16. Regresní analýza

Lineární regrese a odstranění vlivu kovariát

Statistics -> Multiple Regression -> zvolit proměnné (např. Nucl_caud_volume (mm3) jako Dependent var., Age a gender_01,... jako Independent variables) -> OK -> OK -> Summary: Regression results (vypíše regresní koeficienty a p-hodnoty) -> OK

- Vykreslení Q-Q grafu pro rezidua: kliknout na Normal plot of residuals
- Vykreslení histogramu reziduí: na záložce Residuals kliknout na Histogram of residuals (vpravo lze zvolit, zda chceme vykreslit histogram reziduí či standardizovaných reziduí)
- Vykreslení bodového grafu predikovaných hodnot a reziduí: na záložce Scatterplots kliknout na Predicted vs. residuals
- Uložení reziduí: na záložce Save kliknout na Save residuals & predicted -> zvolit proměnné, které bude nově vytvořená tabulka dále obsahovat -> OK

17. Analýza přežití

Kaplanův-Meierův odhad funkce přežití pro jeden výběr

Statistics -> Advanced Linear/Nonlinear Models -> Survival Analysis -> Kaplan & Meier product-limit method -> OK -> zvolit proměnné (např. cas jako Survival times a umrti jako Censoring indicator) -> OK -> Code for complete responses: 1 -> Code for censored responses: 0 -> OK

- Vykreslení křivky přežití: *kliknout na Survival times vs. cum. proportion surviving*
- Vypsání tabulky pro výpočet x-letého přežití: *kliknout na Summary: Product-limit survival analysis*
- Vypsání tabulky s mediánem přežití: *na záložce Advanced kliknout na Percentiles of survival function*

Kaplanův-Meierův odhad funkce přežití pro dva výběry

Statistics -> Advanced Linear/Nonlinear Models -> Survival Analysis -> Comparing two samples -> OK -> zvolit proměnné (např. cas jako Survival, umrti jako Censoring var, Skupina jako Grouping variable) -> OK -> Code for complete responses: 1 -> Code for censored responses: 0 -> zkontrolovat, jestli kategorie grupovací proměnné jsou v pořadí, jaké chceme -> OK

- Vykreslení křivek přežití: *na záložce Function plots kliknout na Cum. prop. surviving by group (Kaplan Meier)*
- Výpočet testu na srovnání přežití ve skupinách: *na záložce Quick kliknout na Gehan's Wilcoxon test nebo Log-rank test (p-hodnota je v záhlaví tabulky)*
- Vypsání tabulky pro výpočet x-letého přežití: *na záložce Two-sample tests kliknout na Proportion surviving by group (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

Statistics -> Advanced Linear/Nonlinear Models -> Survival Analysis -> Comparing multiple samples -> OK -> zvolit proměnné (např. cas jako Survival, umrti jako Censoring var, Skupina jako Grouping variable) -> OK -> Code for complete responses: 1 -> Code for censored responses: 0 -> Codes (for groups) -> All (kdyžtak změnit pořadí kategorií podle toho, jak potřebujeme) -> OK -> OK

- Vykreslení křivek přežití: *kliknout na Cumulative proportion surviving (Kaplan-Meier) by group*
- Výpočet testu na srovnání přežití ve skupinách: *kliknout na Summary: Survival times & scores (p-hodnota je v záhlaví tabulky)*
- Vypsání tabulky pro výpočet x-letého přežití: *na záložce Advanced kliknout na Percent surviving by group (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

Statistics -> Advanced Linear/Nonlinear Models -> Cox Proportional Hazards Model -> zvolit proměnné (např. OS_doba_mesice jako Survival, ECOG1 jako Covariates, Umrti jako Censoring variable) -> OK -> Code for complete responses: 1 -> Code for censored responses: 0 -> OK -> Parameter estimates

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“.