


Načtení a příprava dat pro analýzu

Načtení souboru: File – Open – Data (pokud načítáme Excelovský soubor, nezapomenout změnit Files of type na Excel (*.xls, *.xlsx, *.xlsm))

Olejblování hodnot: přepnout se do Variable View – kliknout na políčko Values a rozkliknout tlačítko se třemi tečkami (při zadávání textových popisků nedávat uvozovky, ty jsou tam doplněny automaticky); hodnoty a textové popisky pak lze přepínat v hlavním panelu pomocí tlačítka: 

Odstranění duplikací: Data - Identify duplicate cases...

Spojení souborů: Data – Merge Files...

Agregace (např. když pro každého pacienta máme víc řádků v tabulce (např. opakovaný rozbor krve) a chceme vytvořit novou tabulku, kde bude jeden pacient jen na jednom řádku (každá složka krve pak bude sumarizována mediánem či průměrem naměřených hodnot): Data - Aggregate

Vyfiltrování chybějících a chybných hodnot: Data - Select cases – If condition is satisfied ... (lze si vybrat druh výstupu: „Filter out unselected cases“, „Copy selected cases to a new dataset“ nebo „Delete unselected cases“)

Seřazení souboru podle nějakého parametru: Data – Sort Cases

Transformace proměnných

Logaritmická transformace: Transform – Compute Variable... – v aritmetických funkcích (Function group: Arithmetic) vybrat Ln

Standardizace: Analyze – Descriptive Statistics – Descriptives – *zatrhnout* Save standardized values as variables

Kategorizace: Transform – Visual Binning...

Překódování textových proměnných na proměnné s číselnými kategoriemi: Transform – Automatic Recode... (pokud jsou v datech chybějící hodnoty, je vhodné zatrhnout „Treat blank string values as user-missing“)

Popisná statistika

Popisná statistika kategoriálních dat: Analyze – Descriptive Statistics – Frequencies – vybrat proměnnou Gender – OK

Popisná statistika spojitých dat: Analyze – Descriptive Statistics – Descriptives – vybrat proměnnou Age – OK

Popisná statistika (včetně mediánu): Analyze – Tables – Custom Tables...

- přidávání proměnných přetažením do pravého okna
- Summary Statistics – zvolit Valid N, Mean, Median, Minimum, Maximum, Std. Deviation – Apply to Selection – OK
- Position: Columns nebo Rows
- Categories and Totals: zatrhnout Total (pokud chceme vypsát popisnou statistiku celkově, nejen podle kategorií)

Grafy

Vykreslení koláčového grafu: Graphs – Legacy Dialogs – Pie – Summaries for groups of cases – Define Slices by: Gender – OK

Vykreslení histogramu: Graphs – Legacy Dialogs – Histogram – Variable: Age – OK (dále lze zatrhnout Display normal curve a lze vykreslit histogramy podle kategoriální proměnné tím, že se kategoriální proměnná přetáhne do Rows nebo Columns)

Vykreslení krabicového grafu pro spojitě proměnné bez kategorizace: Graphs – Legacy Dialogs – Boxplot – Simple – Summaries of separate variables – Define – Boxes Represent: Age – OK

Vykreslení krabicového grafu pro spojitě proměnné s kategorizací: Graphs – Legacy Dialogs – Boxplot – Simple – Summaries for groups of cases – Define – Variable: Age; Category axis: Gender – OK

Vykreslení tečkového grafu (scatter plot): Graphs – Legacy Dialogs – Scatter/Dot – Simple Scatter – Define

Testování hypotéz – parametrické testy

Jednovýběrový t-test: Analyze – Compare Means – One-Sample T Test

Párový t-test: Analyze – Compare Means – Paired-Samples T Test

Dvouvýběrový t-test: Analyze – Compare Means – Independent-Samples T Test

ANOVA: Analyze – Compare Means – One-Way ANOVA (ověření homogenity rozptylů: na záložce Options zatrhnout Homogeneity of variance test; post-hoc testy: na záložce Post Hoc vybrat např. Tukey)

Výpočet Pearsonova korelačního koeficientu: Analyze – Correlate – Bivariate (nechat zatržené Correlation Coefficient: Pearson)

Testování hypotéz – neparametrické testy

Wilcoxonův test: Analyze – Nonparametric Tests – One Sample

Mannův-Whitneyův test: Analyze – Nonparametric Tests – Independent Samples... (nebo: Analyze - Nonparametric tests - Legacy Dialogs - 2 Independent Samples)

Kruskalův-Wallisův test: Analyze – Nonparametric Tests – Legacy Dialogs – K Independent Samples

Výpočet Spearmanova korelačního koeficientu: Analyze – Correlate – Bivariate (zatrhnout Correlation Coefficient: Spearman)

Analýza kontingenčních tabulek

Analyze – Descriptive Statistics – Crosstabs...

- Výpočet procent a očekávaných četností: záložka Cells
- Výpočet Pearsonova Chí-kvadrát testu: na záložce Statistics zatrhnout Chi-square
- Výpočet Fisherova exaktního testu pro tabulky větší než 2x2: na záložce Exact zatrhnout Exact (na záložce Statistics musí být zatržené Chi-square)
- Výpočet McNemarova testu: na záložce Statistics zatrhnout McNemar
- Výpočet relativního rizika (relative risk) a poměru šancí (odds ratio): na záložce Statistics zatrhnout Risk

Výpočet relativního rizika (relative risk) a poměru šancí (odds ratio) včetně p-hodnoty: Analyze – Regression – Binary Logistic – vybrat závisle proměnnou a kovariátu – Categorical (lze zvolit, zda je referenční kategorie Last nebo First; při změně kliknout na Change) – Options (zatrhnout CI for exp(B), aby se vypsaly intervaly spolehlivosti) – OK

ROC analýza

Analyze – ROC Curve – zadat Test Variable a State Variable (jako Value of State Variable zadat rizikovou kategorii) – na záložce Options lze zvolit, zda „Larger test result indicates more positive test“ nebo „Smaller test result indicates more positive test“ – Continue – zatržení „Standard error and confidence interval“ umožní k AUC vypočítat intervaly spolehlivosti a p-hodnotu – zatržení „Coordinate points of the ROC Curve“ umožní získat tabulku se senzitivitou a 1-specificitou pro jednotlivé cut-off body (po zkopírování této tabulky do Excelu je možno vypočítat specificitu a nalézt nejlepší cut-off)

Komentáře a další triky

Vypsání tabulky, kde jsou p-hodnoty pro mnoho kontingenčních tabulek dohromady: Utilities – OMS Control Panel – Output Types: Tables – Command Identifiers: Crosstabs – Table Subtypes for Selected Commands: Chi Square Tests – Output Destination: New dataset (nějak ten nový soubor pojmenovat) – Add – Paste; ve skriptu se po všech CROSSTABS musí napsat omsend.

Ověření normality dat: Analyze – Descriptive Statistics – Explore – na záložce Plots zatrhnout Normality plots with tests

Post hoc testy ke Kruskal-Wallis ANOVě v SPSS neexistují, řeší se to sadou Mann-Whitney testů (manuálně je pak nutno provést Bonferroniho korekci)

Výpočet p-hodnoty pro jednostrannou alternativu u testů v SPSS (SPSS totiž umožňuje počítat jen oboustranné alternativy):

- $p/2$ (pokud „směr“ v našich datech odpovídá jednostranné alternativě – např. pro jednostranný t-test, pokud H_1 : průměr > 25 a náš vypočtený průměr je opravdu > 25), kde p je p-hodnota vypočtená pro oboustrannou alternativu
- $1-p$ -hodnota/2 (pokud „směr“ v našich datech neodpovídá jednostranné alternativě – např. pro jednostranný t-test, pokud H_1 : průměr > 25 a náš vypočtený průměr je < 25)