

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 – Software – Nabídka softwaru).

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

Uložení datového souboru

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

Zapnutí automatického filtru

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

Vypnutí automatického překreslování grafů

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

1. Vizualizace dat

Vytváření grafů pomocí záložky Graphs.

Koláčový graf

Graphs -> 2D Graphs -> Pie Charts -> zvolit proměnnou (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 subjektů či objektů)

Graphs -> Histogram -> Variables -> zvolit proměnnou -> OK -> zrušit zatržení Fit type: Normal -> na záložce Advanced zatrhnout Breaks between columns -> OK

Sloupcový graf (na ose y procenta)

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

Krabicový graf (s vykreslením odlehlých hodnot)

Graphs -> 2D -> Box Plots... -> Variables -> zvolit proměnnou jako Dependent variable -> OK -> OK

Krabicový graf (s minimem a maximem)

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

Tečkový graf

Graphs -> Scatterplot -> zvolit *proměnné* -> na záložce *Advanced* můžeme *zatrhnout* *Corr. and p* (linear fit) -> OK

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

2. Příprava dat pro analýzu

Nastavení formátu proměnné na "double"

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

Nastavení formátu proměnné na datum

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

Identifikace a odstranění duplikací

Data -> *Filter/Recode* -> *Filter Duplicate Cases...* -> *Input: Variables* -> zvolit proměnnou či proměnné -> OK -> u *Output* *zatrhnout* *Create duplicates spreadsheet* -> OK

Odstranění chybějících a chybných hodnot

Data -> *Subset* -> *Cases* -> *zatrhnout* *Enable Selection Condition* -> do *By expression* napsat podmínku (např. $v4=""$ OR $v4 > 110$ OR $v7=""$) -> OK -> OK
Nový datový soubor bez chybějících a chybných hodnot uložit.

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*. 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* -> *Frequency tables* -> *Variables* -> zvolit proměnnou -> OK -> *Summary*

Popisná sumarizace kvantitativních dat

Statistics -> *Basic Statistics* -> *Descriptive statistics* -> *Variables* -> zvolit proměnnou -> OK -> na záložce *Advanced* *zatrhnout* *Median, Lower & upper quartiles* -> *Summary*

Výpočet intervalu spolehlivosti a střední chyby průměru (standard error)

Statistics -> *Basic Statistics* -> *Descriptive statistics* -> *Variables* -> zvolit proměnnou -> OK -> na záložce *Advanced* *zatrhnout* *Conf. limits for means a případně i Std. err. of mean* -> *Summary*

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

Statistics -> Basic Statistics -> Descriptive statistics -> Variables -> zvolit proměnnou -> OK -> Select Cases -> Enable Selection Conditions -> Specific, selected by: -> do "or case number" zadat 1-20 -> OK -> na záložce Advanced zatrhnout Median a případně další charakteristiky -> Summary

Popisná sumarizace kvantitativní proměnné podle kategorií kvalitativní proměnné

Statistics -> Basic Statistics -> Descriptive statistics -> Variables -> zvolit proměnnou -> OK -> na záložce Advanced zatrhnout Median a případně další charakteristiky -> By Group -> Grouping Variable(s) -> vybrat proměnnou -> OK -> zatrhnout Output to a single folder a Accumulate tabular results in a single spreadsheet a přepnout Sorting of Groups na Ascending či Descending -> OK -> Summary

Výběr dat pro analýzu/graf

The image shows a series of SPSS dialog boxes with arrows pointing to specific features:

- Select the variables for the analysis:** A list of variables (1-Var1 to 10-Var10) is shown. An arrow points to the 'Select variables' field.
- Descriptiv Statistics: Spreadsheet1:** The 'Summary: Descriptive statistics' tab is active. A yellow box highlights the 'Compute statistics' section, which includes 'Variation, moments' (Standard Deviation, Variance, etc.) and 'Percentiles, ranges' (Minimum & maximum, etc.). Arrows point to 'Standard Deviation' and 'Minimum & maximum'.
- Options:** A dropdown menu showing 'Options', 'Output...', 'Display...', 'Create Macro...', and 'Close Analysis'.
- Analysis/Graph Case Weights:** The 'Use spreadsheet weights' radio button is selected. An arrow points to the 'Weight variable' field.
- Analysis/Graph Case Selections Conditions:** The 'Enable Selection Conditions' checkbox is checked. An arrow points to the 'Specific, selected by' section.

Annotations in Czech:

- Nastavení:** Points to the 'Options' dropdown.
- Záložky možností nebo nastavení analýzy/grafu:** Points to the 'Options' tab in the main dialog.
- Vážení dat:** Points to the 'Analysis/Graph Case Weights' dialog.
- Selekce dat:** Points to the 'Analysis/Graph Case Selections Conditions' dialog.
- Detailní nastavení analýzy/grafu:** Points to the 'Compute statistics' section in the main dialog.
- Způsob zpracování chybějících hodnot:** Points to the 'MD deletion' section (Casewise, Pairwise).

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

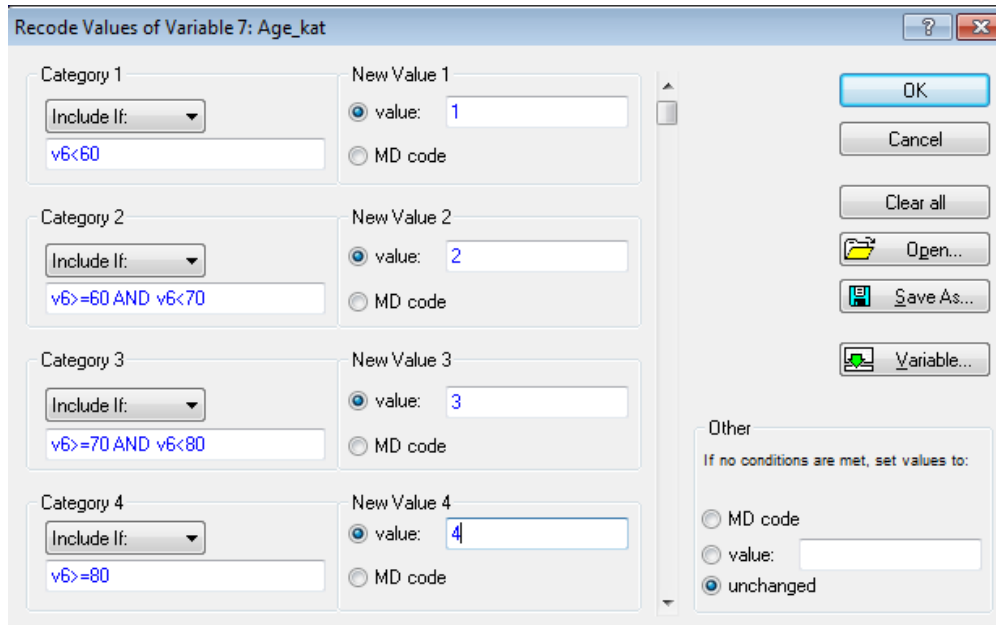
4. Transformace dat

Logaritmická transformace

Označit proměnnou za proměnnou, kterou chceme logaritmovat (např. v9) -> kliknout pravým tlačítkem myši -> Add Variables -> Name -> zadat název nové proměnné -> do Long name napsat =Log(v9) (Pozor, ve STATISTICe je přirozený logaritmus označen jako Log(x) místo Ln(x)!) -> OK

Kategorizace

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



Obr. 2. Ukázka kategorizace věku.

5. Statistické testy pro kvantitativní data – parametrické testy

Jednovýběrový t-test

Statistics -> Basic Statistics -> t-test, single sample -> OK -> zvolit proměnnou -> OK -> napsat hodnotu do Test all means against -> na záložce Options zatrhnout Compute conf. limits -> Summary

Párový t-test

Statistics -> Basic Statistics -> t-test, dependent samples -> OK -> zvolit proměnné -> OK -> Summary

Dvouvýběrový t-test

Statistics -> Basic Statistics -> t-test, independent, by groups -> OK -> zvolit proměnné (kvantitativní proměnnou jako Dependent variable a kategoriální proměnnou jako Grouping variable) -> na záložce Options lze zvolit Levene's test (test homogeneity rozptylů), CI for estimates a případně Test w/ separate variance estimates (pokud není splněn předpoklad homogeneity rozptylů) -> Summary

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

Statistics -> Basic Statistics -> Breakdown & one-way ANOVA -> OK -> zvolit proměnné (kvantitativní proměnnou jako Dependent variable a kategoriální proměnnou jako Grouping variable) -> OK -> OK -> na záložce ANOVA & tests kliknout na Analysis of Variance (vypíše ANOVA tabulku); dále lze vypsát i výsledky testů homogeneity rozptylů: Levene tests, Brown-Forsythe tests -> 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 nesterajných počtů subjektů ve skupinách) nebo Scheffé test (pro stejné i nesterajné počty subjektů ve skupinách)

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

Statistics -> ANOVA -> One-way ANOVA -> OK -> zvolit *proměnné (kvantitativní proměnnou jako Dependent variable a kategoriální proměnnou jako Categorical 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)

6. 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: Data -> Variables -> Add... -> Name -> zadat název nové proměnné -> do Long name napsat např. =27,5 (hodnota konstanty, se kterou chceme srovnávat) -> OK
Poté můžeme použít pro výpočet párový Wilcoxonův test: Statistics -> Nonparametrics -> Comparing two dependent samples (variables) -> OK -> zvolit *proměnné* -> OK -> Wilcoxon matched pairs test (Je možné spočítat i znaménkový test (Sign test), který je též nepar. alternativou jednovýb. t-testu.)

Wilcoxonův test – párový

Statistics -> Nonparametrics -> Comparing two dependent samples (variables) -> OK -> zvolit *proměnné* -> OK -> Wilcoxon matched pairs test (Je možné spočítat i znaménkový test (Sign test), který je též nepar. alternativou párového t-testu.)

Mannův-Whitneyův test

Statistics -> Nonparametrics -> Comparing two independent samples (groups) -> OK -> zvolit *proměnné (kvantitativní proměnnou do Dependent variable list a kategoriální proměnnou jako Indep. (grouping) variable)* -> OK -> M-W U test

Kruskalův-Wallisův test

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

7. Ověření předpokladů statistických testů

Vykreslení N-P grafů pro jednotlivé skupiny

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

Ověření normality pomocí histogramů pro jednotlivé skupiny

Graphs -> Histogram -> Variables -> zvolit *proměnnou* -> 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 *kategoriální proměnnou* -> OK -> zatrhnout Output to a single folder -> přepnout Sorting of Groups na Ascending -> OK -> OK

Ověření normality pomocí Basic Statistics pro jednotlivé skupiny

Statistics -> Basic Statistics -> Descriptive statistics -> Variables -> zvolit *proměnnou* -> OK -> na záložce Advanced případně zatrhnout Median -> By Group -> Grouping Variable(s) -> vybrat *kategoriální proměnnou* -> OK -> zatrhnout Output to a single folder a Accumulate tabular results in a

single spreadsheet -> *přepnout* Sorting of Groups *na* Ascending -> OK -> *na záložce* Normality *zatrhnout* Kolmogorov-Smirnov & Lilliefors test for normality *a* Shapiro-Wilk's W test -> Histograms -> *pomocí záložky* Prob. & Scatterplots *lze vykreslit i* Normal probability plot

8. Analýza kontingenčních tabulek

Kontingenční tabulka absolutních četností

Statistics -> Basic Statistics -> Tables and banners -> OK -> Specify tables (select variables) (*zvolit jednu kategoriální proměnnou do List 1 a druhou kategoriální proměnnou do List 2*)-> OK -> OK -> Summary

Kontingenční tabulka procent

Statistics -> Basic Statistics -> Tables and banners -> OK -> Specify tables (select variables) -> *zvolit proměnné do List 1 a 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 and banners -> OK -> Specify tables (select variables) -> *zvolit proměnné do List 1 a List 2* -> OK -> OK -> *na záložce* Options *zatrhnout* Expected frequencies -> Summary

Pearsonův chí-kvadrát test

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

9. Korelační analýza

Výpočet Pearsonova korelačního koeficientu

Statistics -> Basic Statistics -> Correlation matrices -> OK -> One variable list -> *zvolit proměnné* -> *na záložce* Options *zvolit* Display r, p-values, and N's -> Summary

Výpočet Spearmanova korelačního koeficientu (výpočet čtvercové korelační matice)

Statistics -> Nonparametrics -> Correlations (Spearman, Kendall tau, gamma) -> OK -> *zvolit proměnné* -> OK -> Spearman rank R

Výpočet Spearmanova korelačního koeficientu (výpočet detailní tabulky)

Statistics -> Nonparametrics -> Correlations (Spearman, Kendall tau, gamma) -> OK -> *zvolit* Detailed report (*místo* Square matrix) -> *zvolit proměnné do* First variable list *a do* Second variable list -> OK -> Spearman rank R