

Univariace

5. týden

ZURn4108 Deskriptivní analýza kvantitativních dat

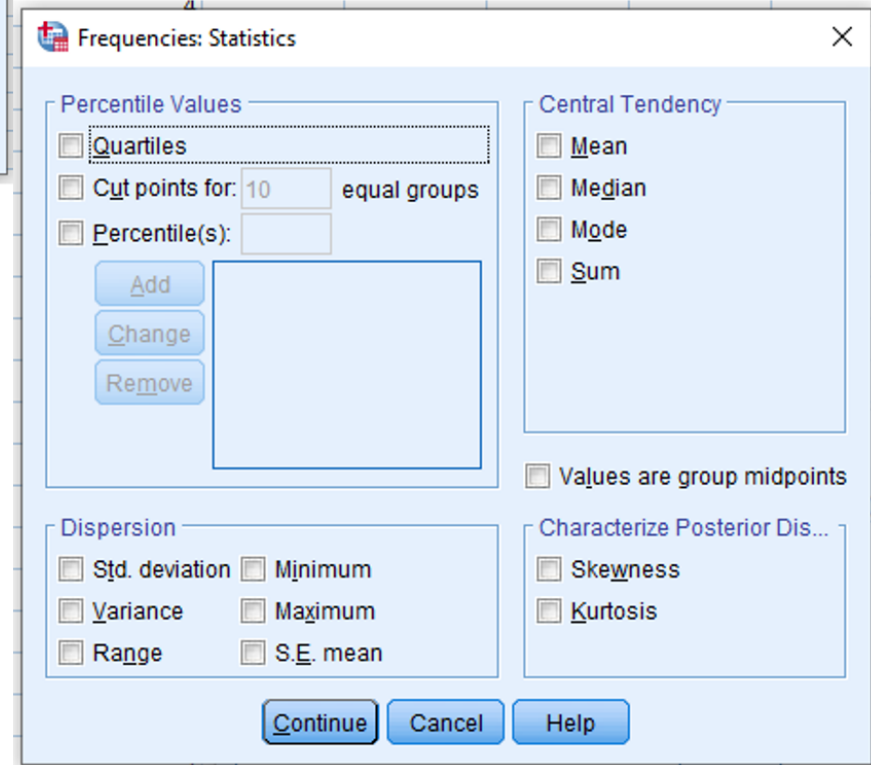
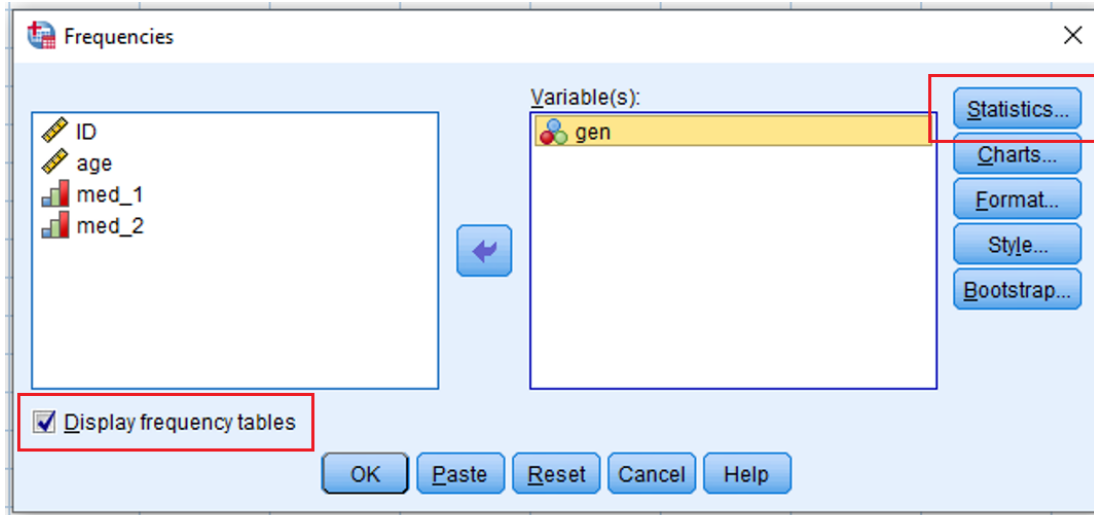
Čištění

- Ideálně ještě před analýzou (nebo deskripcí) – využívá ale stejné nástroje
- Proč?
 - Data mohou mít celou řadu chyb – překlepy, špatné kódování, invalid data ze strany respondenta...
- V základu nejčastěji
 - Minimum a Maximum (kontrola „outlierů“)
 - Zjištění základních četností hodnot proměnné (správnost hodnot, vyplněnost)
 - CTRL+F, případně rekódování (v dalších hodinách)

Co nám může číselná univariace ukázat?

- Četnosti (absolutní, relativní, kumulativní)
- Minimální a maximální hodnoty
- Střední hodnoty
- Míry variability
- Percentily

- Jak na základní univariaci?
 - Analyze – Descriptive Statistics – Frequencies - Statistics



Střední hodnoty

- V širším hledisku obsahuje větší množství ukazatelů – my si budeme povídat o 3:
 - Modus – nejčastější hodnota
 - Medián – hodnota, která výběr rozděluje na 2 početně stejné jednotky
 - Průměr – aritmetický průměr

Míra variability

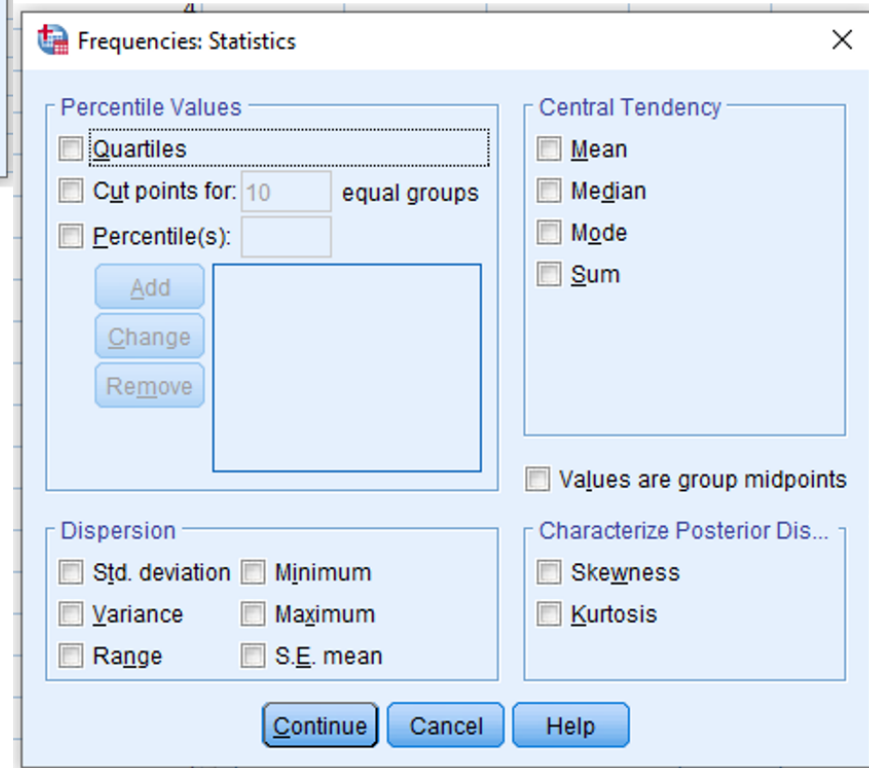
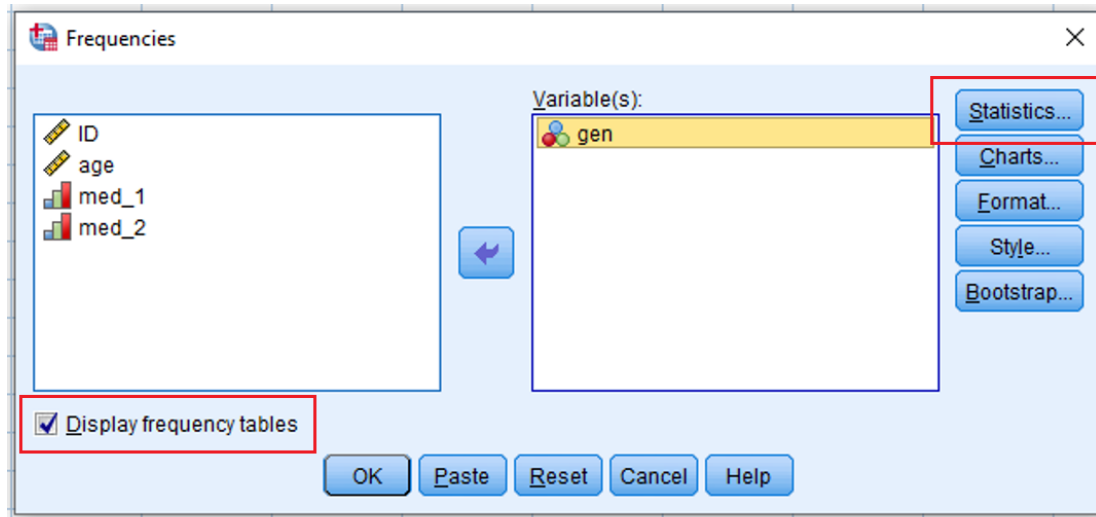
- Nominální – koncentrace, variační poměr
- Ordinální – variační rozpětí, ordinální rozptyl, variační poměr
- Obvykle nás zajímají pro průměr (tj. kardinální proměnné)
 - Rozptyl
 - Směrodatná odchylka

- **Variační poměr (nominální a ordinální)**
 - Od 1 odečteme procentuální zastoupení modusu vyjádřené v číslech
 - Modus známek studentů je „B“ – 30% studentů (nebo také podíl četnosti modusu a celkového vzorku)
 - Variační poměr vypočítáme jako $1 - 3/10 = 7/10$ („0,7“)
 - V SPSS není
- **Rozptyl (kardinální)**
 - Celkový součet kvadratických odchylek všech hodnot od průměru, který pak vydělíme celkovým $N - 1$
 - V SPSS jako VARIANCE
- **Směrodatná odchylka (kardinální)**
 - Odmocnina rozptylu
 - V SPSS jako STD. DEVIATION

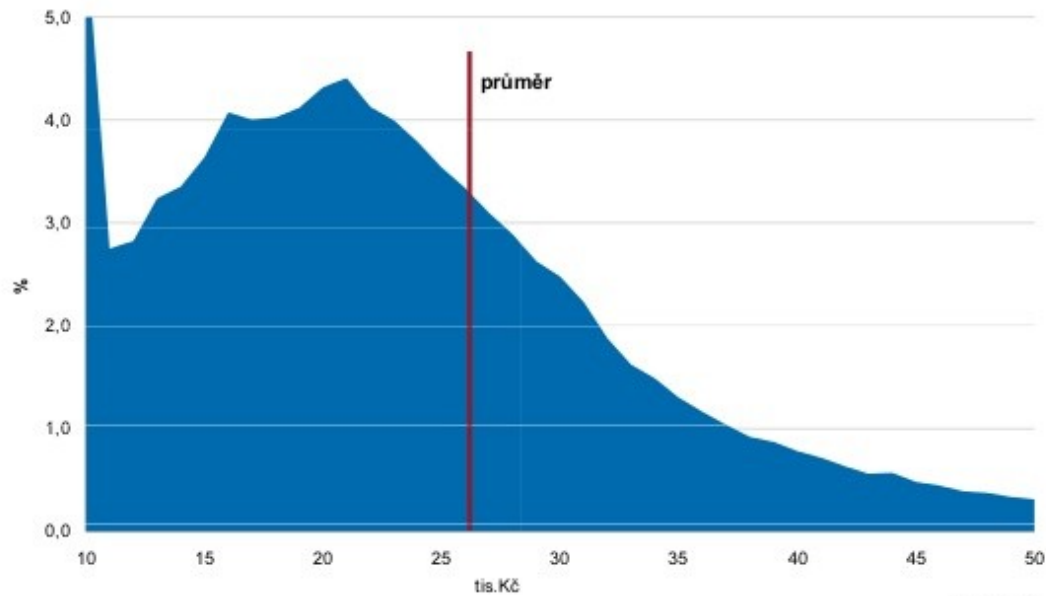
Percentily

- Někdy chceme vědět přesné hodnoty, v kterých se nám data dělí na konkrétně velké soubory
- Může se hodit při transformaci dat (téma na příští hodinu)
 - Soubor chceme rozdělit do pětín podle věku (nejmladší pětina až nejstarší pětina). Které hodnoty věku jsou hraniční těchto skupin?
- Percentily někdy můžeme nazývat podle konkrétních hodnot:
 - Percentily na 10 dílků: decily
 - Percentily na 4 dílky: kvartily (asi nejběžnější druh)

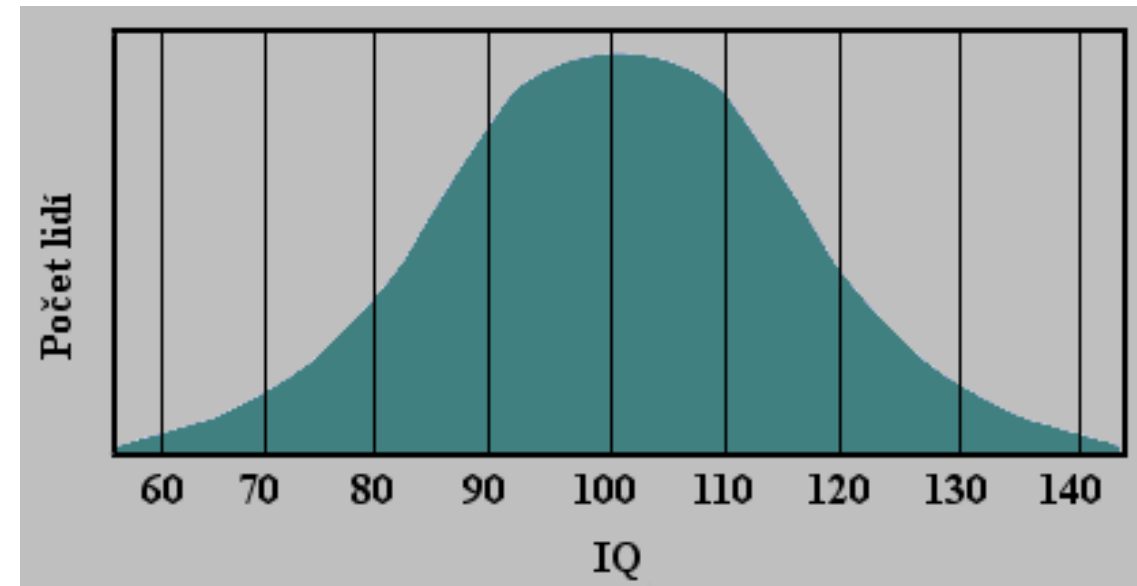
- základní univariaci znova:
 - Analyze – Descriptive Statistics – Frequencies - Statistics



Střední hodnoty – stejné X rozdílné

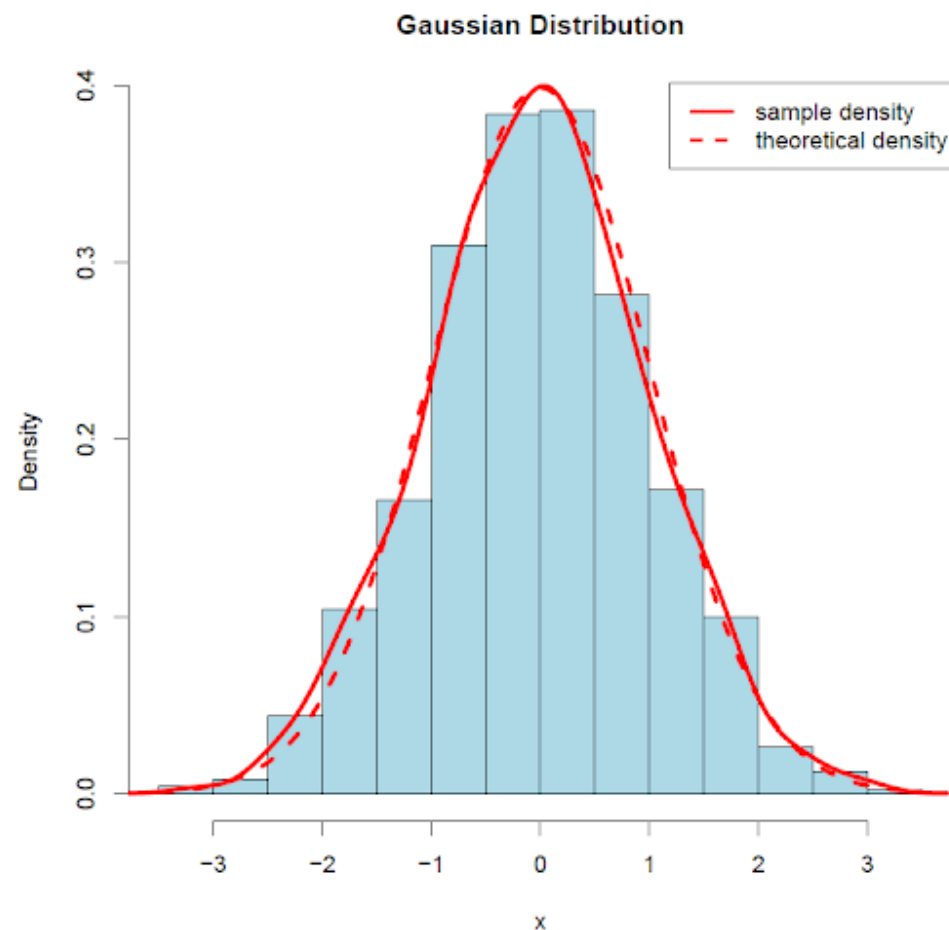


Dvě třetiny zaměstnanců jsou pod průměrnou mzdou



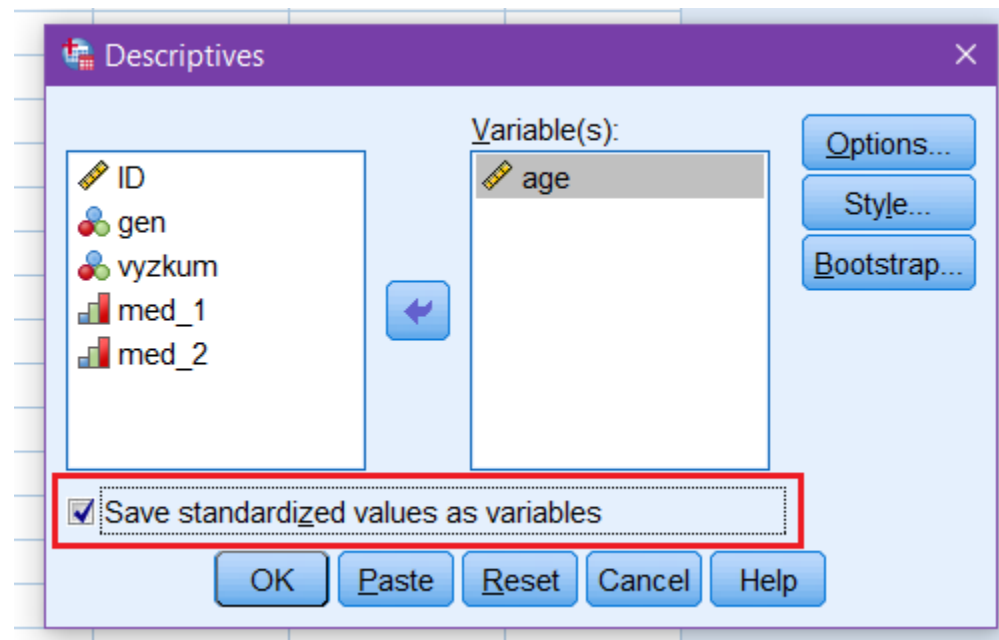
Normální rozdělení

- Nebo také „Gaussovo rozdělení“
- Modelové rozdělení náhodné veličiny
- V populaci poměrně běžné
- Rychlý test – šikmost a špičatost
 - Šikmost (skewness) – symetrie rozložení; blízká 0, pokud normální rozdělení
 - Špičatost (kurtosis) – míra soustředění hodnot kolem středu; blízká 0, pokud normální rozdělení
 - Analyze – Descriptive Statistics – Descriptives – Options



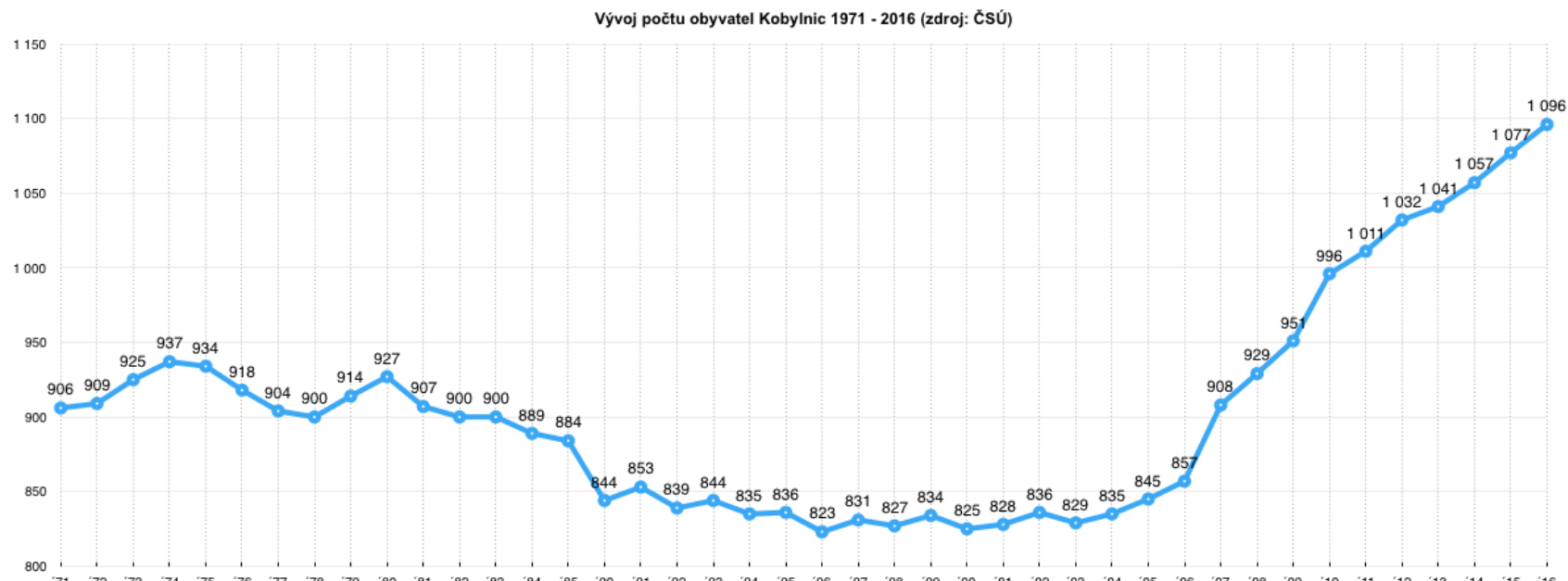
Z-skóre

- Analyse – Des. Statistics – Descriptives
- „standardizuje“ nám hodnoty položek

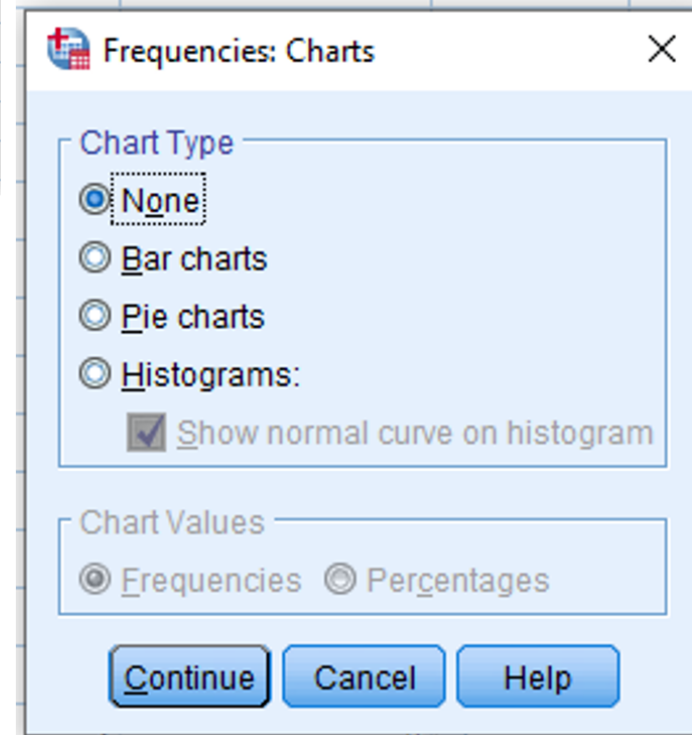
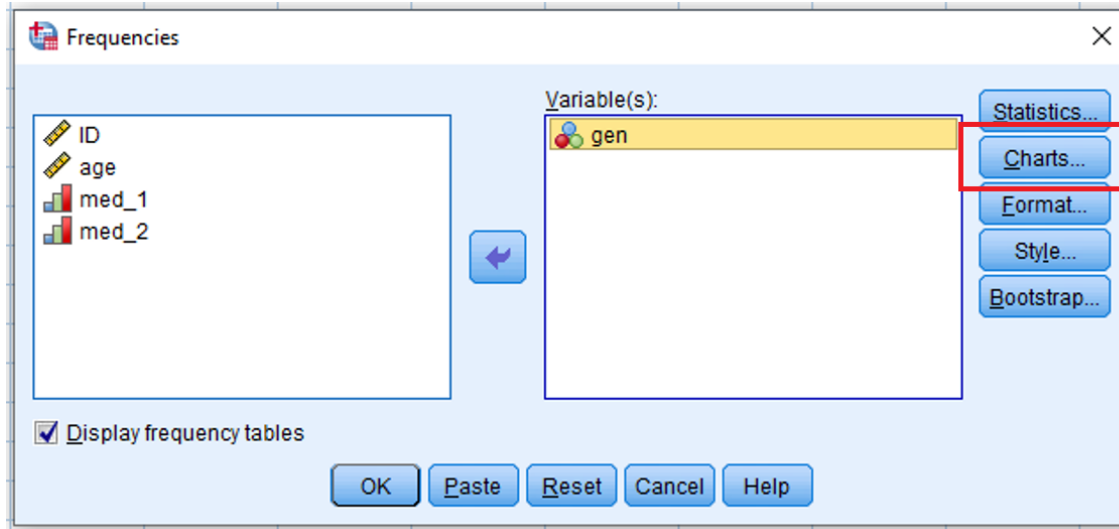


Grafy

- Několik základních druhů
 - Sloupcový
 - Výsečový (koláčový)
 - Histogram
 - Boxplot (krabicový)
 - Spojnicový



- Sloupnicový graf, výsečový graf a histogram
 - Analyze – Descriptive Statistics – Frequencies - Charts



- Box plot

- Graphs – Chart Builder (i alternativa k ostatním druhům grafu)
- Analyze – Descriptive Statistics – Explore – Plots – Boxplots

The screenshot shows the SPSS Chart Builder dialog box. The main window displays a 1-D Boxplot for the variable 'Jaký je Váš věk?'. The plot is a blue box with a black horizontal line representing the median, and whiskers extending to the minimum and maximum values. The plot is titled '1-D Boxplot of Jaký je Váš věk?'. The Y-axis is labeled 'Jaký je Váš věk?' and the X-axis is labeled 'Filter?'. The dialog box includes a 'Variables:' list on the left with 'Jaký je Váš věk? [age]' selected. Below the plot, there are tabs for 'Gallery', 'Basic Elements', 'Groups/Point ID', and 'Titles/Footnotes'. The 'Gallery' tab is active, showing a list of chart types: Favorites, Bar, Line, Area, Pie/Polar, Scatter/Dot, Histogram, High-Low, Boxplot (selected), and Dual Axes. The 'Basic Elements' tab is also visible, showing 'Edit Properties of:' with 'Box1' selected. The 'Statistics' section is expanded, showing 'Variable:' as 'Jaký je Váš věk?' and 'Statistic:' as 'Boxplot'. The 'Display error bars' section is checked, and the 'Error Bars Represent' section is expanded, showing 'Confidence intervals' selected with a 'Level (%)' of 95, and 'Standard error' and 'Standard deviation' also available with a 'Multiplier' of 2. The 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' buttons are at the bottom.

Konec prezentace

- Dotazy, opáčko