

Metodologie pro ISK 2

Kontrola dat

Popis kategorizovaných dat

A close-up photograph of a document, likely a technical drawing or a data table. The document features a grid of lines. A prominent feature is a diagonal hatched area, possibly representing a specific material or a warning zone. Below this area, the words "ACTUAL NORMAL" are printed in a bold, sans-serif font, oriented diagonally. The background is slightly blurred, emphasizing the foreground elements.

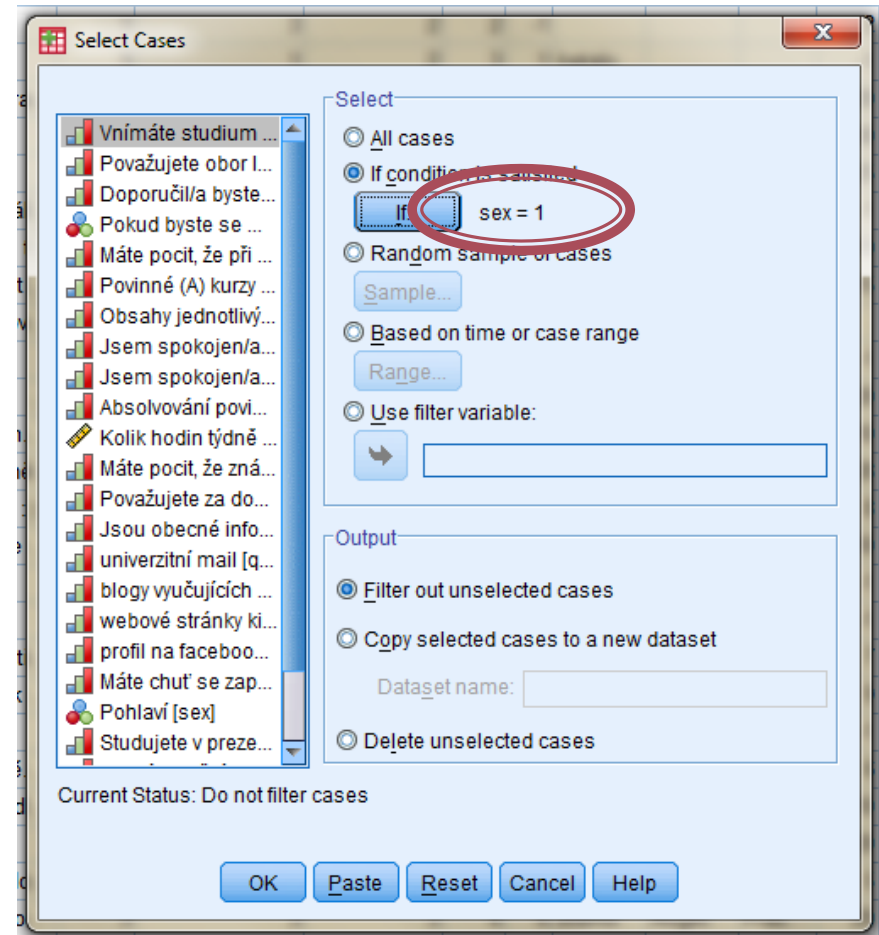
ACTUAL NORMAL

Popisná (deskriptivní) statistika

- Začínáme **zkoumáním jednotlivých proměnných**, teprve potom analyzujeme jejich vztahy
- Účelem analýzy je „**zpřístupnit data** graficky, tabulkově a výpočtem statistických charakteristik“ (Hendl 2009)

Výběr případů

- Pokud nechceme pracovat s celým souborem
- **Data → Select Cases**
- Lze vybírat náhodně nebo dle kritéria – pokud např. chceme pracovat jen s muži, pak musíme použít proceduru IF



Kontrola dat: GIGO

- „Garbage in, garbage out!“



Kontrola dat: GIGO

- **Chyby při zpracování**
 - Posuny desetinných čárek
 - Záměna znaků (nepoužívat „o“ pro missing values)
 - Chyby při přepisování
- **Kontrola dat**
 - Kontrola okem („vytisknout a řádek po řádku zkontrolovat“ *Hendl, 2009*)
 - Předběžné grafické zobrazení dat

Co s missing values?

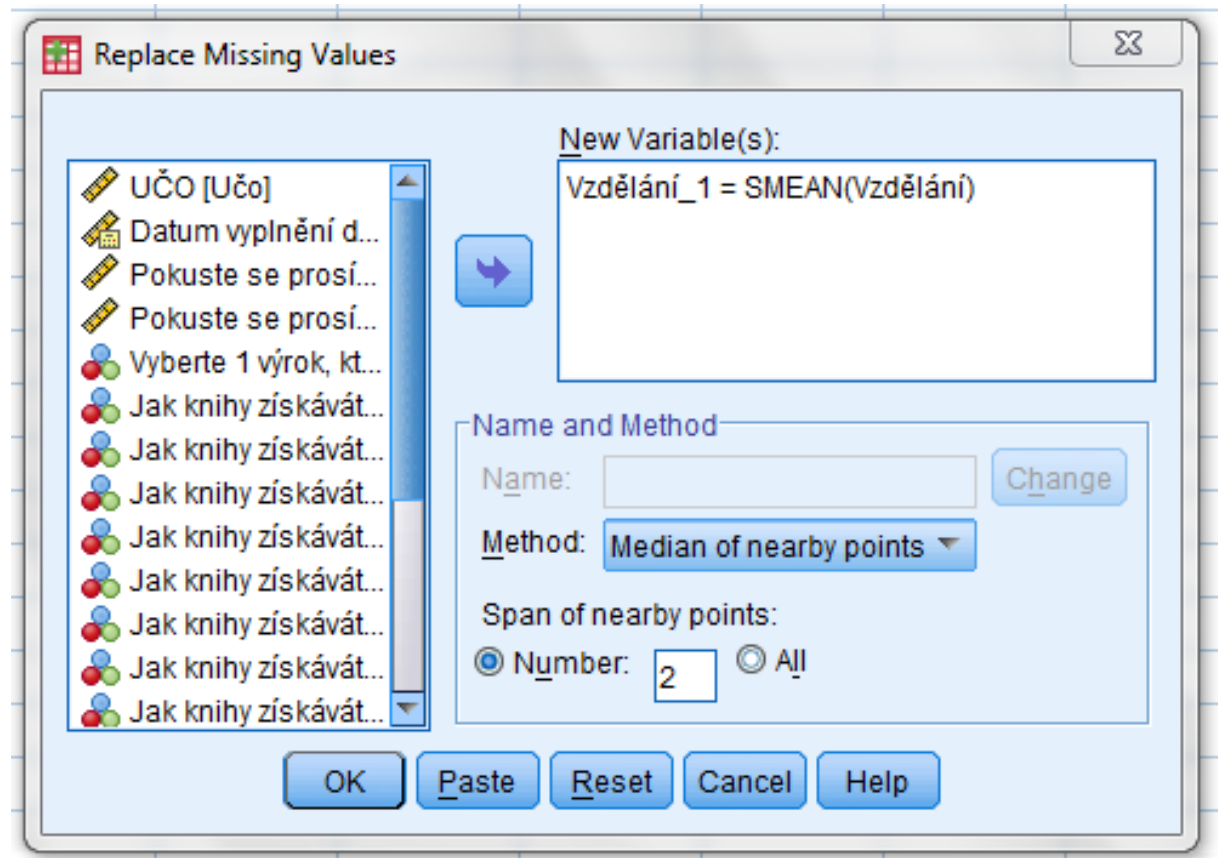
- **Hodnota „neodpověď/a“**
- Jsou výskyty rozděleny náhodně?
- Způsoby kontroly:
 - rozdělit soubor na skupiny záznamů s missing values a bez nich, porovnat charakteristiky obou souborů
 - Kontrola korelací vyplnění/nevyplnění s jinou proměnou

Co s missing values?

- Pokud potřebujeme pracovat jen s úplnými záznamy:
 - Vymazání jednotky
 - Vymazání proměnné
 - Imputace
 - Odhad scházející hodnoty (regresní analýza)
 - Nahrazení průměrnou hodnotou subpopulace

Co s missing values?

- Transform → Replace Missing Values



Kategorizovaná a nekategorizovaná data

- **Kategorizovaná data:** nominální, ordinální i kardinální proměnné s malým počtem variant (pohlaví, vzdělání, počet dětí)
- **Nekategorizovaná data:** kardinální proměnné s velkým počtem variant (plat)

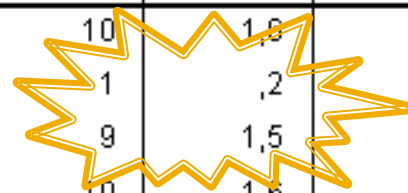

Kontrola kategorizovaných dat

- Analyze → Descriptive Statistics → Frequencies

Jméno výzkumníka

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------------------|-----------|---------|---------------|--------------------|
| Valid | Andrea Szászová | 10 | 1,6 | 1,6 | 1,6 |
| | Babora Ondrušová | 1 | ,2 | ,2 | 1,8 |
| | Barbora Ondrušová | 9 | 1,5 | 1,5 | 3,3 |
| | Barbora Pitašová | 10 | 1,6 | 1,6 | 4,9 |
| | Blanka Justová | 10 | 1,6 | 1,6 | 6,5 |
| | Blanka Svobodová | 10 | 1,6 | 1,6 | 8,1 |
| | Dagmar Chládková | 11 | 1,8 | 1,8 | 9,9 |
| | Dagmar Šiková | 10 | 1,6 | 1,6 | 11,6 |
| | Dalibor Bláha | 10 | 1,6 | 1,6 | 13,2 |
| | Daniela Králová | 10 | 1,6 | 1,6 | 14,8 |

Chybný
zápis jména



Kontrola kategorizovaných dat

- Analyze → Descriptive Statistics → Frequencies

| Statistics | | |
|------------------------|---------|-----|
| Jaké je Vaše vzdělání? | | |
| N | Valid | 608 |
| | Missing | 6 |

Jaké je Vaše vzdělání?

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|---------------------------------------|-----------|---------|---------------|--------------------|
| Valid | Základní | 46 | 7,5 | 7,6 | 7,6 |
| | Základní vyučen /střední bez maturity | 62 | 10,1 | 10,2 | 17,8 |
| | Střední s maturitou | 307 | 50,0 | 50,5 | 68,3 |
| | Pomaturitní nástavba, VOS | 40 | 6,5 | 6,6 | 74,8 |
| | Vysokoškolské | 153 | 24,9 | 25,2 | 100,0 |
| | Total | 608 | 99,0 | 100,0 | |
| Missing | System | 6 | 1,0 | | |
| Total | | 614 | 100,0 | | |

Co s missing values?

Kontrola kategorizovaných dat

- Analyze → Descriptive Statistics → Frequencies

Statistics

Pohlaví

| | | |
|---|---------|-----|
| N | Valid | 613 |
| | Missing | 1 |

Pohlaví

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid | Muž | 279 | 45,4 | 45,5 | 45,5 |
| | Žena | 333 | 54,2 | 54,3 | 99,8 |
| | 7 | 1 | ,2 | ,2 | 100,0 |
| | Total | 613 | 99,8 | 100,0 | |
| Missing | System | 1 | ,2 | | |
| Total | | 614 | 100,0 | | |

Chyba:
proměnná
„Pohlaví“ by
neměla
nabývat
hodnoty 7

Kontrola kategorizovaných dat

- Nalezení dat: CTRL+F
- Edit → Find

Statistics

| Pohlaví | | |
|---------|---------|-----|
| N | Valid | 613 |
| | Missing | 1 |

Pohlaví

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | Muž | 279 | 45,4 | 45,5 | 45,5 |
| | žena | 333 | 54,2 | 54,3 | 99,8 |
| | 7 | 1 | ,2 | ,2 | 100,0 |
| | Total | 613 | 99,8 | 100,0 | |
| | | 1 | ,2 | | |
| | | 614 | 100,0 | | |

Find and Replace - Data View

Column: Pohlaví

Find: 7

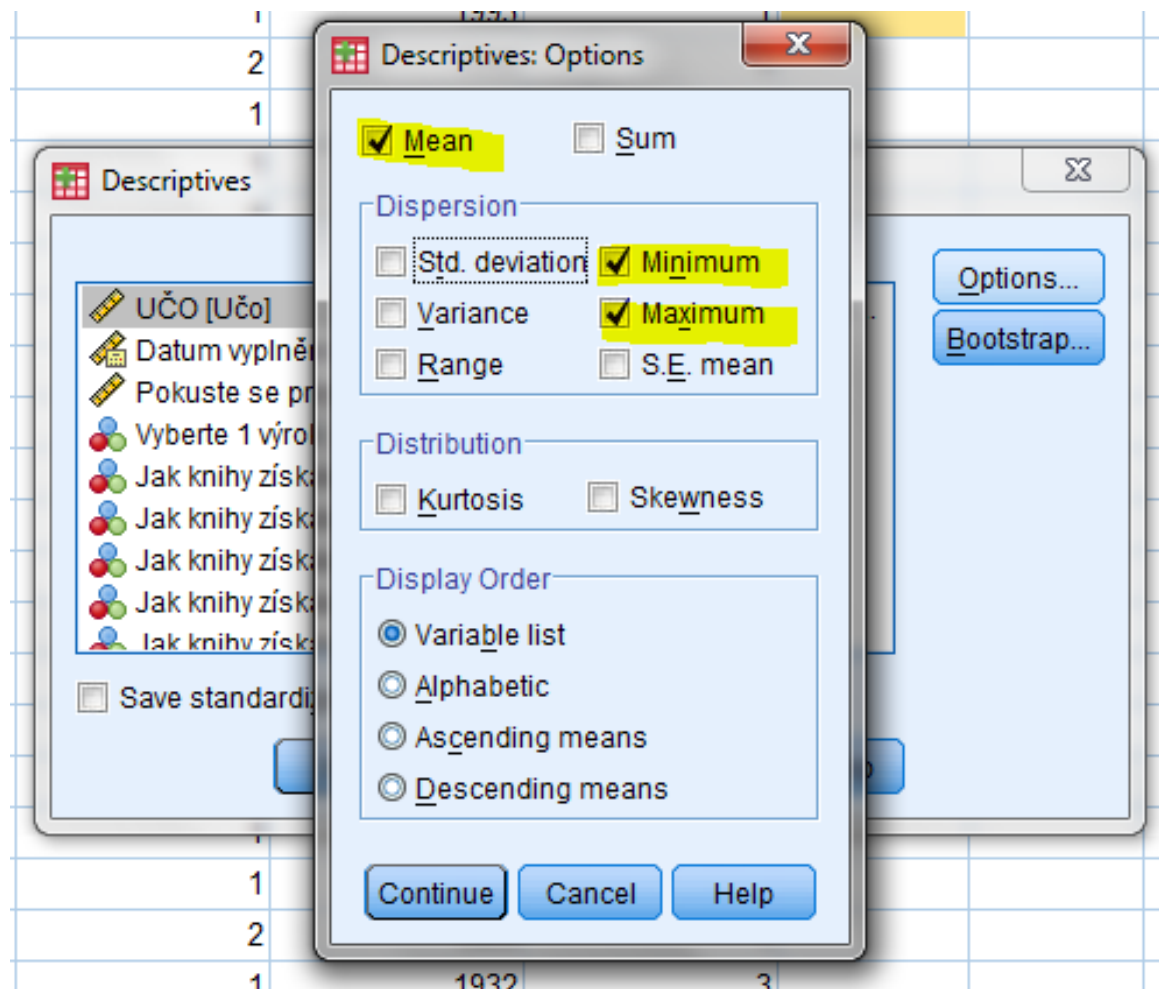
Replace

Replace with:

Show Options >>

Find Next Replace Replace All Close Help

Kontrola nekategorizovaných dat



Kontrola nekategorizovaných dat

- Analyze → Descriptive Statistics → Descriptives

Descriptive Statistics

| | N | Minimum | Maximum | Mean |
|---------------------------|-----|---------|---------|------|
| Jaký je Váš rok narození? | 613 | 1927 | 1997 | |
| Valid N (listwise) | 613 | | | |



Kontrola nekategorizovaných dat

- Analyze → Descriptive Statistics → Descriptives

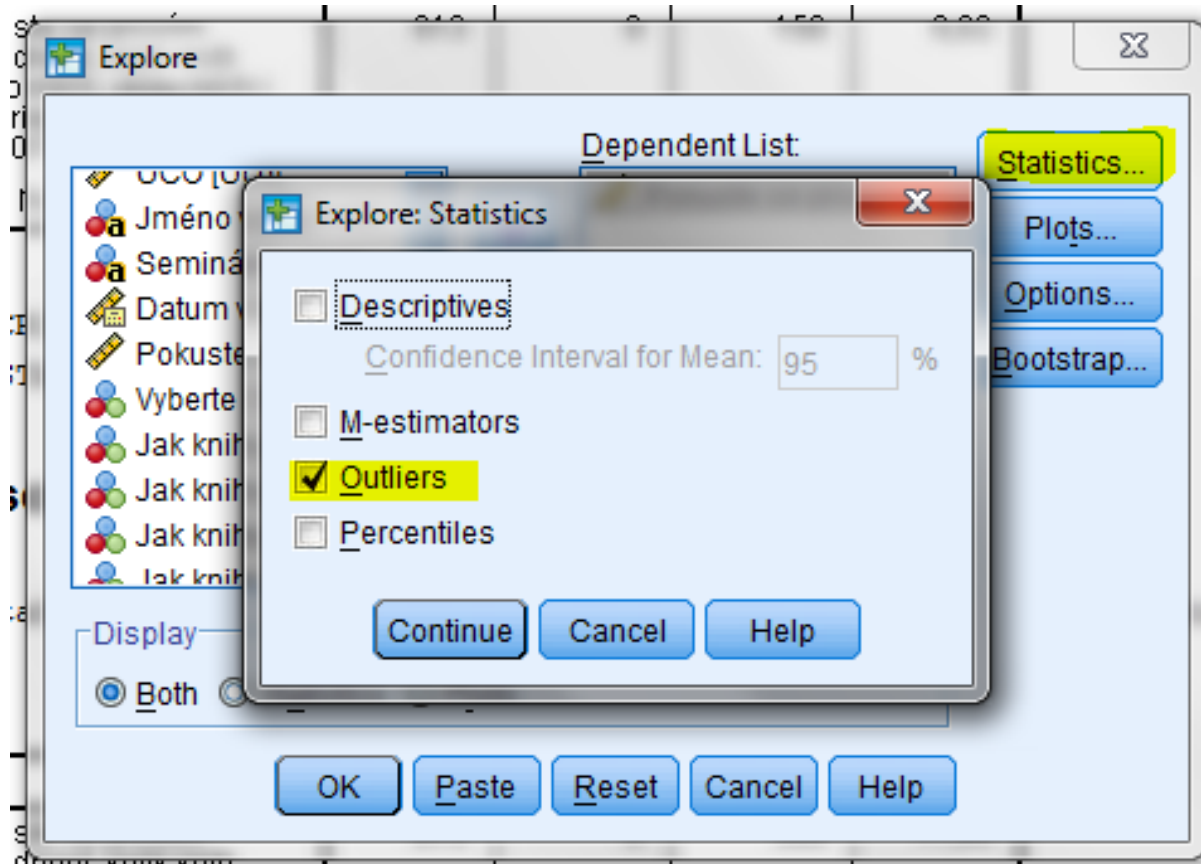
Descriptive Statistics

| | N | Minimum | Maximum | Mean |
|--|-----|---------|---------|-------|
| Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste přečetl za rok 2011. | 613 | 0 | 300 | 17,93 |
| Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste nakoupil za rok 2011. | 613 | 0 | 150 | 6,62 |
| Valid N (listwise) | 613 | | | |



Kontrola nekategorizovaných dat

- Analyze → Descriptive Statistics → Explore



Kontrola nekategorizovaných dat

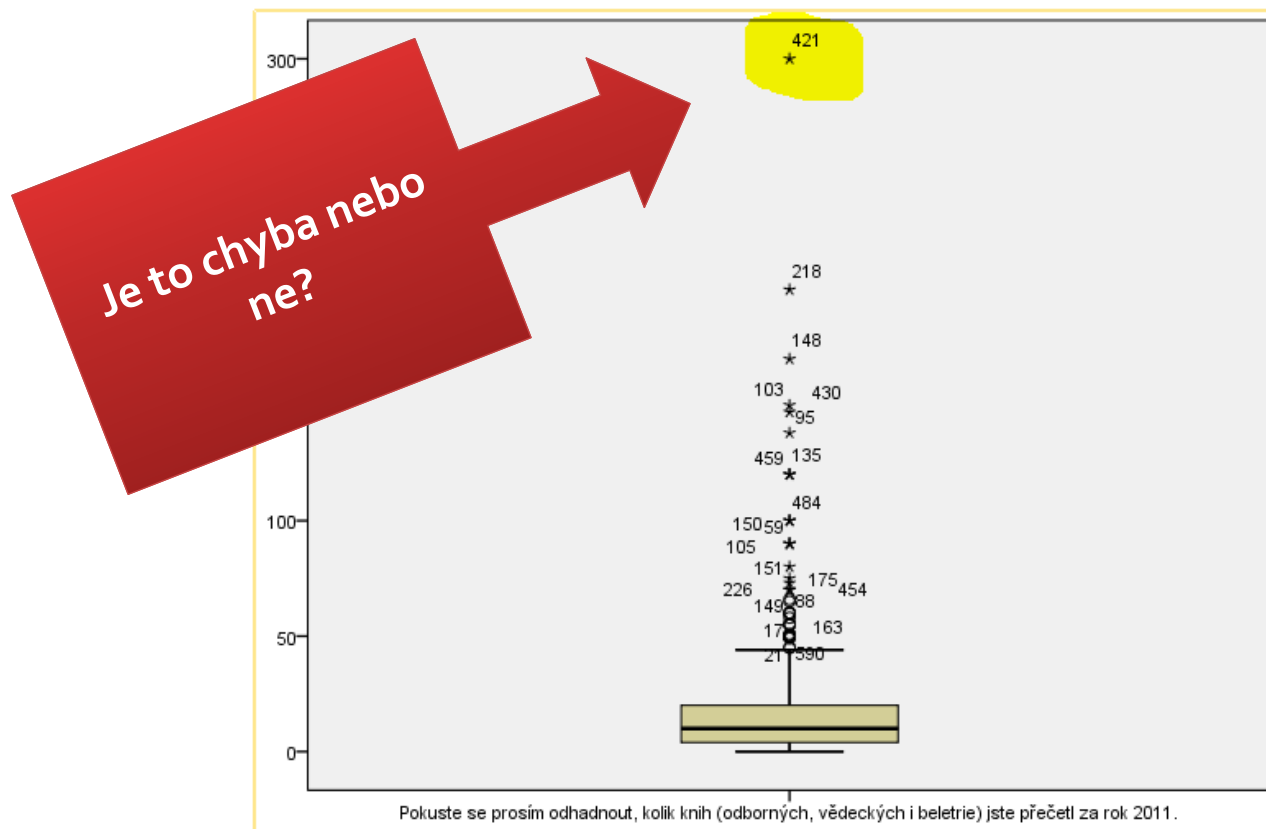
- Analyze → Descriptive Statistics → Explore

| | | | Case Number | Value |
|---|---------|---|-------------|----------------|
| Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste přečetl za rok 2011. | Highest | 1 | 421 | 300 |
| | | 2 | 218 | 200 |
| | | 3 | 148 | 170 |
| | | 4 | 103 | 150 |
| | | 5 | 430 | 147 |
| | Lowest | 1 | 599 | 0 |
| | | 2 | 582 | 0 |
| | | 3 | 562 | 0 |
| | | 4 | 548 | 0 |
| | | 5 | 505 | 0 ^a |

a. Only a partial list of cases with the value 0 are shown in the table of lower extremes.

Kontrola nekategorizovaných dat

- Analyze → Descriptive Statistics → Explore



Analýza kategorizovaných dat

- Třídění prvního stupně (univariační analýza)
- Analyze → Descriptive Statistics → Frequencies

Statistics

Jste spokojen/a s nabídkou knih a služeb ve veřejných knihovnách?

| | | |
|---|---------|-----|
| N | Valid | 459 |
| | Missing | 154 |

Jste spokojen/a s nabídkou knih a služeb ve veřejných knihovnách?

| | | Frequency | Percent | Valid Percent | Cumul Perc |
|---------|----------------------------------|-----------|---------|---------------|------------|
| Valid | Velmi spokojen/a | 83 | 13,5 | 18,1 | |
| | Spíše spokojen/a | 258 | 42,1 | 56,2 | |
| | Ani spokojen/a, ani nespokojen/a | 87 | 14,2 | 19,0 | |
| | Spíše nespokojen/a | 28 | 4,6 | 6,1 | |
| | Velmi nespokojen/a | 3 | ,5 | ,7 | |
| | Total | 459 | 74,9 | 100,0 | |
| Missing | Nevím / nemohu odpovědět | 154 | 25,1 | | |
| Total | | 613 | 100,0 | | |

„Podezřele“ moc missing values

Možný postup: rekódovat jako novou proměnnou (odpověděli/neodpověděli) a sledovat, zda výskyty hodnot souvisí s jinými proměnnými

Tabulka četností

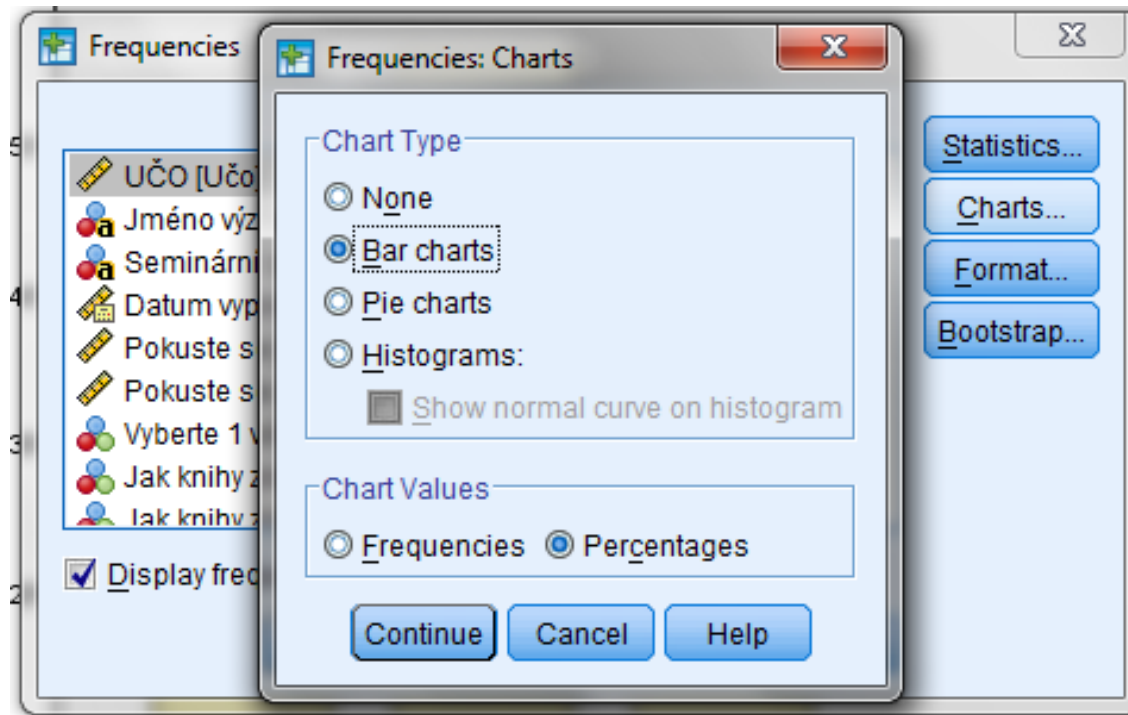
- Analyze → Descriptive Statistics → Frequencies

Jste spokojen/a s nabídkou knih a služeb ve veřejných knihovnách?

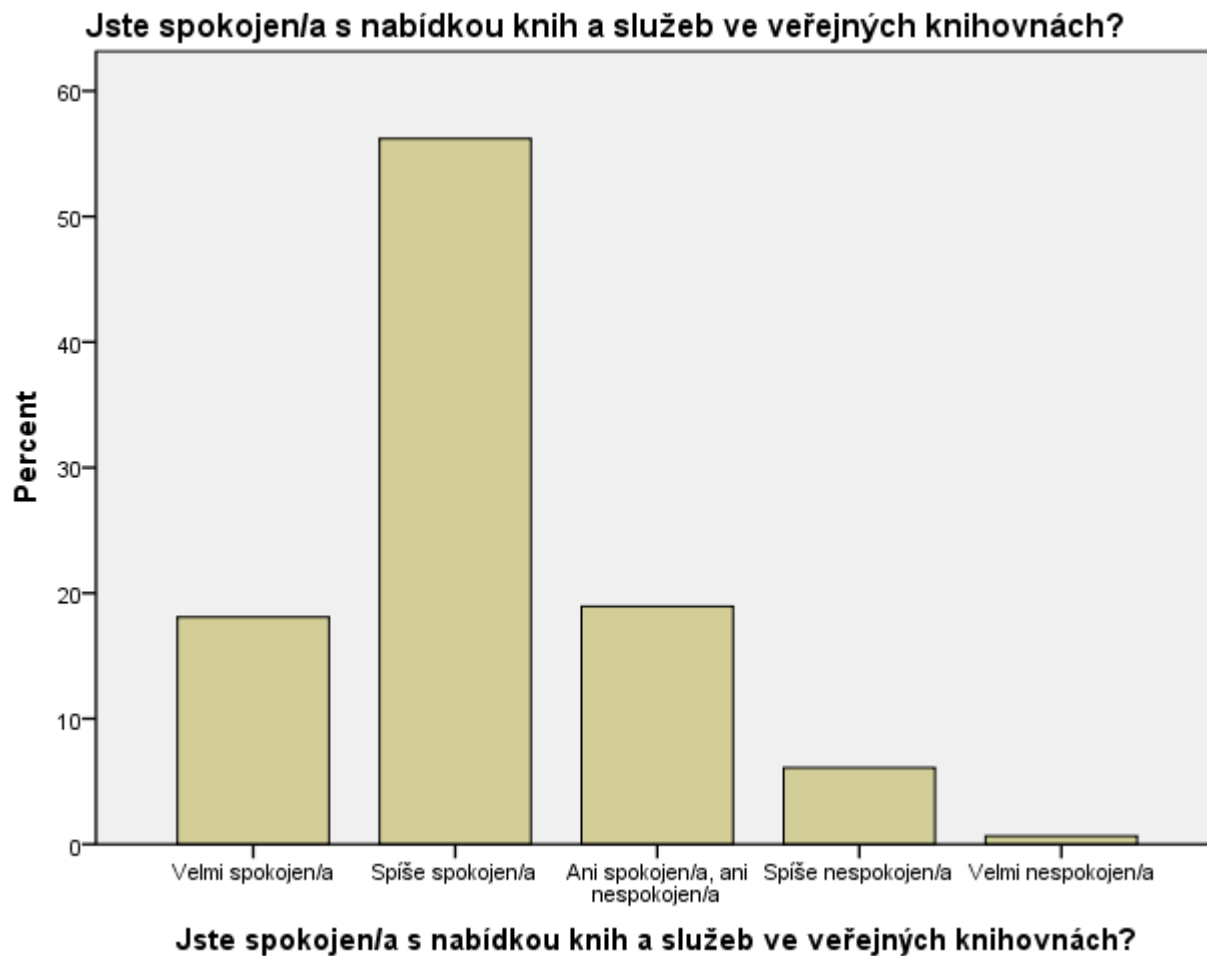
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|----------------------------------|-----------|---------|---------------|--------------------|
| Valid | Velmi spokojen/a | 83 | 13,5 | 18,1 | 18,1 |
| | Spíše spokojen/a | 258 | 42,1 | 56,2 | 74,3 |
| | Ani spokojen/a, ani nespokojen/a | 87 | 14,2 | 19,0 | 93,2 |
| | Spíše nespokojen/a | 28 | 4,6 | 6,1 | 99,3 |
| | Velmi nespokojen/a | 3 | ,5 | ,7 | 100,0 |
| | Total | | 459 | 74,9 | 100,0 |
| Missing | Nevím / nemohu odpovědět | 154 | 25,1 | | |
| Total | | 613 | 100,0 | | |

Analýza kategorizovaných dat

- Analyze → Descriptive Statistics → Frequencies → **Charts** → **Bar Charts**

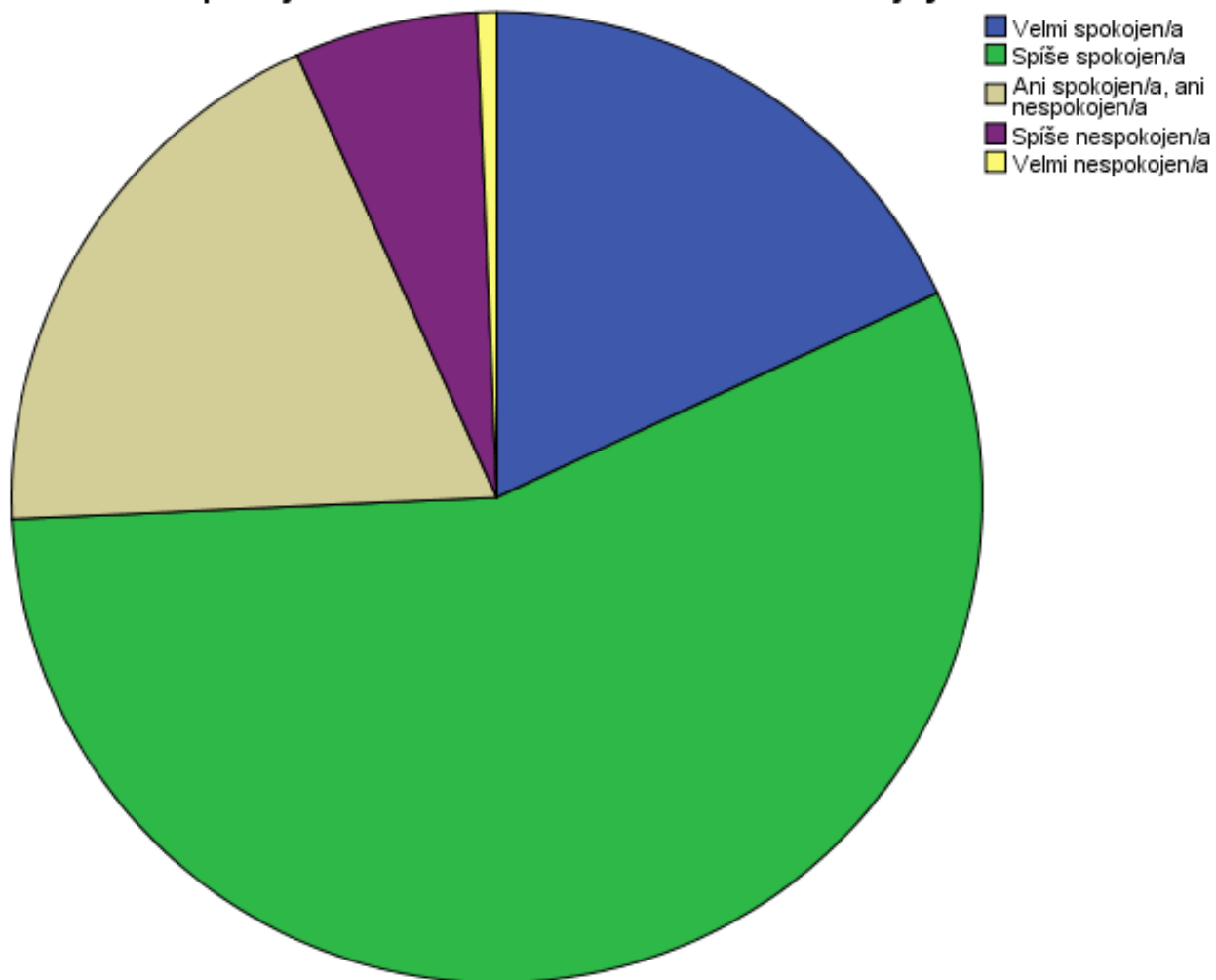


Sloupcové grafy



Koláčové grafy

Jste spokojen/a s nabídkou knih a služeb ve veřejných knihovnách?



Zobrazování dat

■ Tabulka četností

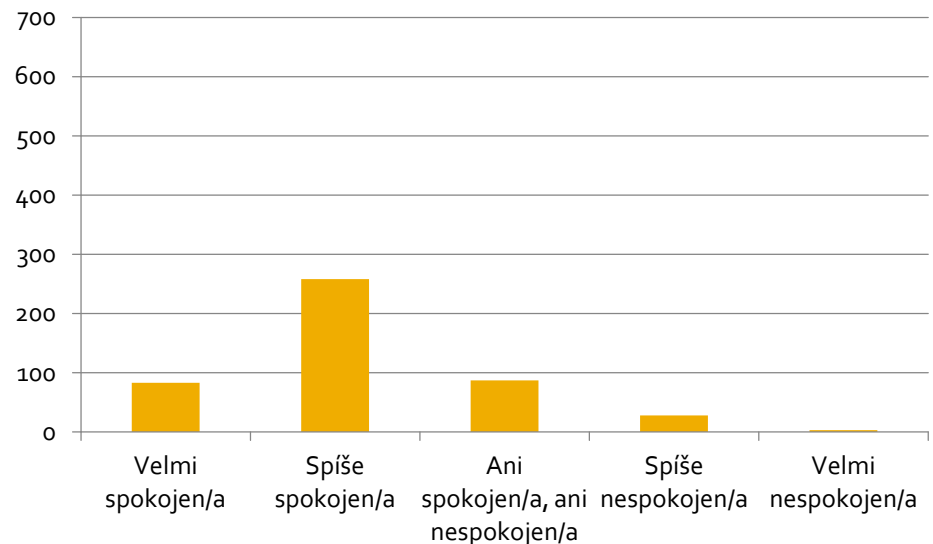
- První přehled výsledků měření
- Vhodné pro uvedení přesných čísel
- Absolutní / relativní / validní / kumulativní četnosti

■ Graf četností

- Názornější
- Používají se validní četnosti
- Osa X: hodnoty proměnné
- Osa Y: četnosti

Zobrazování dat

- **Sloupcový graf:**
 - Nezávislá proměnná na vodorovné ose
 - Závislá proměnná na svislé ose
 - Pozor na měřítko:



Míry centrální tendence

- Nominální proměnná: **MODUS**
- Ordinální proměnná: **MODUS, MEDIÁN**

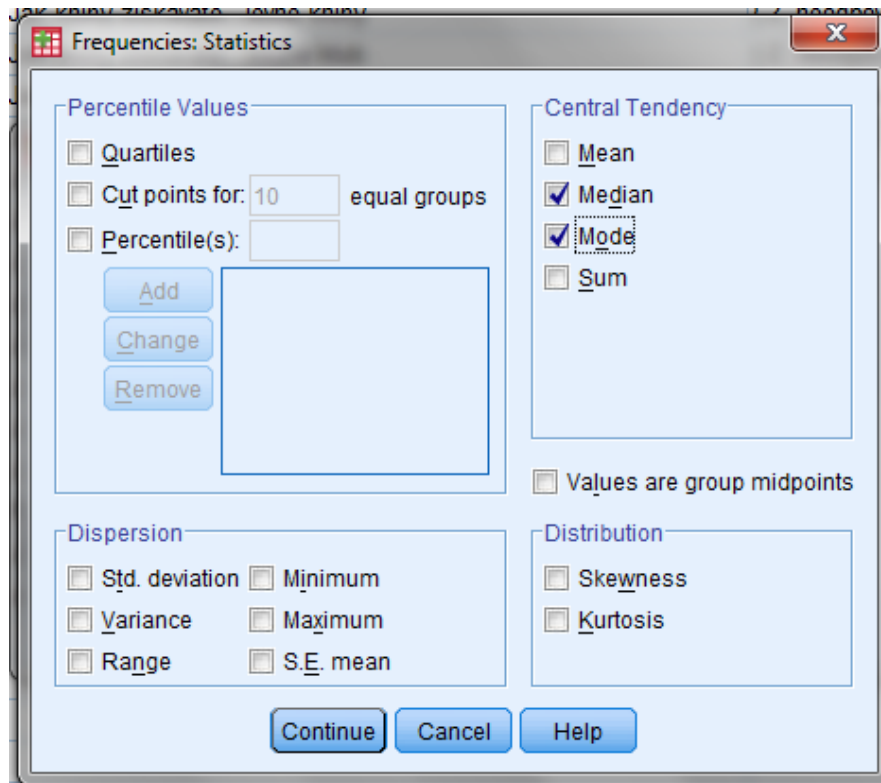
MODUS je hodnota, která se v datech vyskytuje nejčastěji.

MEDIÁN dělí řadu výsledků seřazených podle velikosti na dvě stejně početné poloviny

MEDIÁNOVÁ KATEGORIE je ta, ve které je dosaženo 50% všech údajů, postupujeme-li od první kategorie výše.

Míry centrální tendence

- Analyze → Descriptive Statistics → Frequencies → **Statistics → Median, Mode**



- Obrázky: [kevin dooley](#)