# Metodologie pro Informační studia a knihovnictví 2

## Modul III:

## Popis a kontrola dat

# Metodologie pro Informační studia a knihovnictví 2

## Modul 3: Popis, kontrola a čištění dat.

### Co se dozvíte v tomto modulu?

- Proč je potřeba dbát na kvalitu dat na vstupu
- Jak popsat výběrový soubor a na jaké hodnoty proměnných dávat pozor při kontrole?
- Jak vybrat jen určité případy (nový dataset)
- Jak postupovat v Excelu a v Google Spreadsheets?

V tomto modulu si připravíme dataset k samotné analýze. To, zda budete mít na konci analýzy smysluplné výsledky, do značné míry záleží právě na tom, jakou míru pozornosti budete věnovat počáteční kontrole dat.

# Popis a kontrola dat

Prvním úkolem výzkumníka je popis výběrového souboru. Charakteristikou vzorku by měla začít každá analýza i analytická kapitola v bakalářské či diplomové práci. Zajímá nás například:

- Kolik je ve výběrovém souboru jednotek?
- Kolik je v souboru mužů a žen?
- Kolik je v souboru lidí se ZŠ/SŠ/VŠ vzděláním?
- Jak je v souboru distribuován věk?

Toto rozložení může být vyjádřeno v **absolutních**, **relativních**, či **kumulativních relativních četnostech**.

- **Absolutní četnost** udává absolutní číslo – hodnotu četnosti varianty proměnné v souboru.
  *Například: V souboru je 1456 mužů a 1201 žen.*
- **Relativní četnost** udává **podíl** četnosti varianty proměnné v souboru.
  *Například: V souboru je 24 % osob se základním vzděláním.*
- **Kumulativní relativní četnost** udává kumulativní podíly variant proměnné v souboru (nejsou použitelné pro nominální proměnné).
  *Například: V souboru je 36 % respondentů, kteří mají alespoň maturitu (tedy nejen úspěšní středoškoláci s maturitou, ale také vysokoškoláci se všemi variantami diplomů).*

## *Popis a kontrola kategorizovaných dat*

### *Tabulky četností*

Pro zobrazení základních hodnot popisu rozložení hodnot kategorizovaných proměnných (tedy proměnných nominálních a ordinálních s menším počtem variant odpovědí) se používá tzv. **tabulka četností**. Ta obsahuje jak absolutní, tak relativní četnosti hodnot proměnných. Takto vypadá správná a kompletní tabulka četností:

| ***Jaké je Vaše vzdělání?*** | | | | |
|---|---|---|---|---|
| | | **Četnost odpovědí** | **Relativní četnost** | **Validní relativní četnost** |
| Validní hodnoty | Základní | 46 | 7,5 % | 7,6 % |
| | Základní vyučen /střední bez maturity | 62 | 10,1 % | 10,2 % |
| | Střední s maturitou | 307 | 50,1 % | 50,5 % |
| | Pomaturitní nástavba, VOŠ | 40 | 6,5 % | 6,6 % |
| | Vysokoškolské | 153 | 25,0 % | 25,2 % |
| | Celkem validní hodnoty | 608 | 99,2 % | 100,0 % |
| Chybějící hodnoty (neví, neodpověděl/a) | Chybějící hodnoty | 5 | 0,8 % | |
| **Celkem** | | **613** | **100,0 %** | |

V praxi se často používá jen zkrácená verze tabulky obsahující pouze validní četnosti:

| ***Jaké je Vaše vzdělání?*** | **Četnost odpovědí** | **Validní relativní četnost** |
|---|---|---|
| Základní | 46 | 7,6 % |
| Základní vyučen /střední bez maturity | 62 | 10,2 % |
| Střední s maturitou | 307 | 50,5 % |
| Pomaturitní nástavba, VOŠ | 40 | 6,6 % |
| Vysokoškolské | 153 | 25,2 % |
| **Celkem** | **608** | **100,0 %** |

Před počítáním četností je ale potřeba zkontrolovat data. Kontrolujeme, zda se nachází v platném intervalu (například proměnná pohlaví nabývá v našem souboru pouze hodnot 1 a 2, všechny ostatní varianty by měly být omyly).

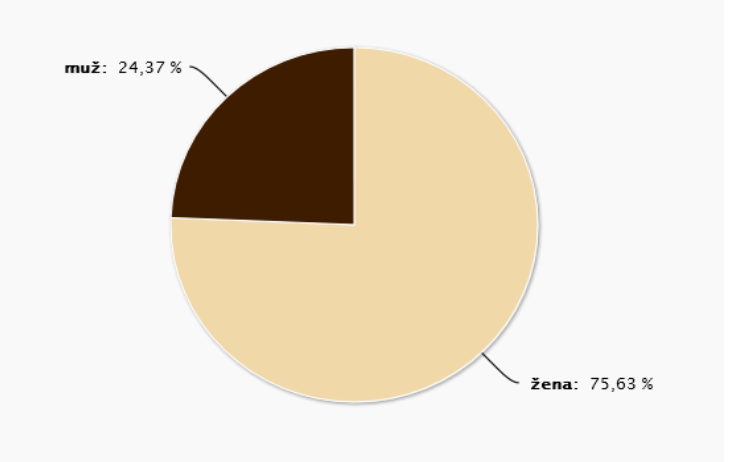
### *Grafy četností*

Pro znázornění rozložení četností se využívají i grafy znázorňující četnosti hodnot proměnných. Nejznámějšími variantami jsou koláčový a sloupcový graf.

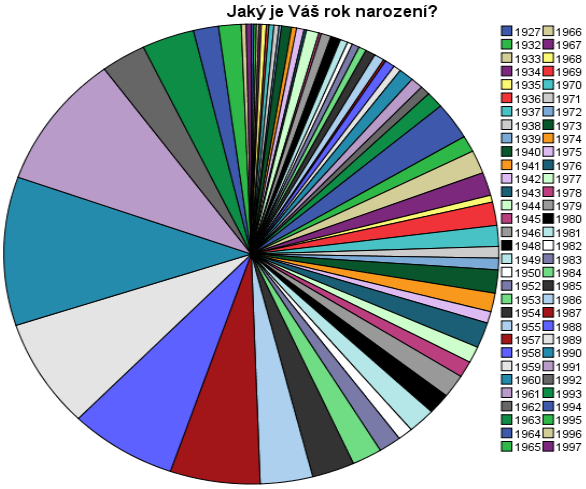
**Koláčový graf** je vhodný:

- pro třídění prvního stupně (jedna datová řada),
- pro porovnání četností u nominálních proměnných, které nemají příliš mnoho hodnot (méně než 7),
- pokud hodnoty, které chcete vykreslit, nejsou nulové,
- pokud hodnoty představují část celku.

Příklad proměnné, kde je vhodné využít koláčový graf:



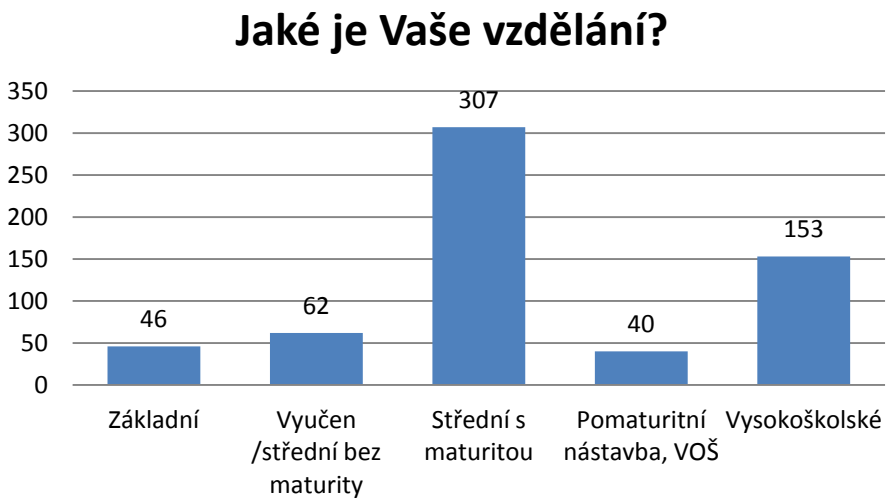
Příklad proměnné, kde NENÍ vhodné využít koláčový graf:




**Sloupcový graf** je vhodný pro:

- porovnání položek,
- ordinální proměnné a kardinální proměnné s menším počtem kategorií,
- znázornění změn za časové období (třídění druhého stupně).

Příklad sloupcového grafu:

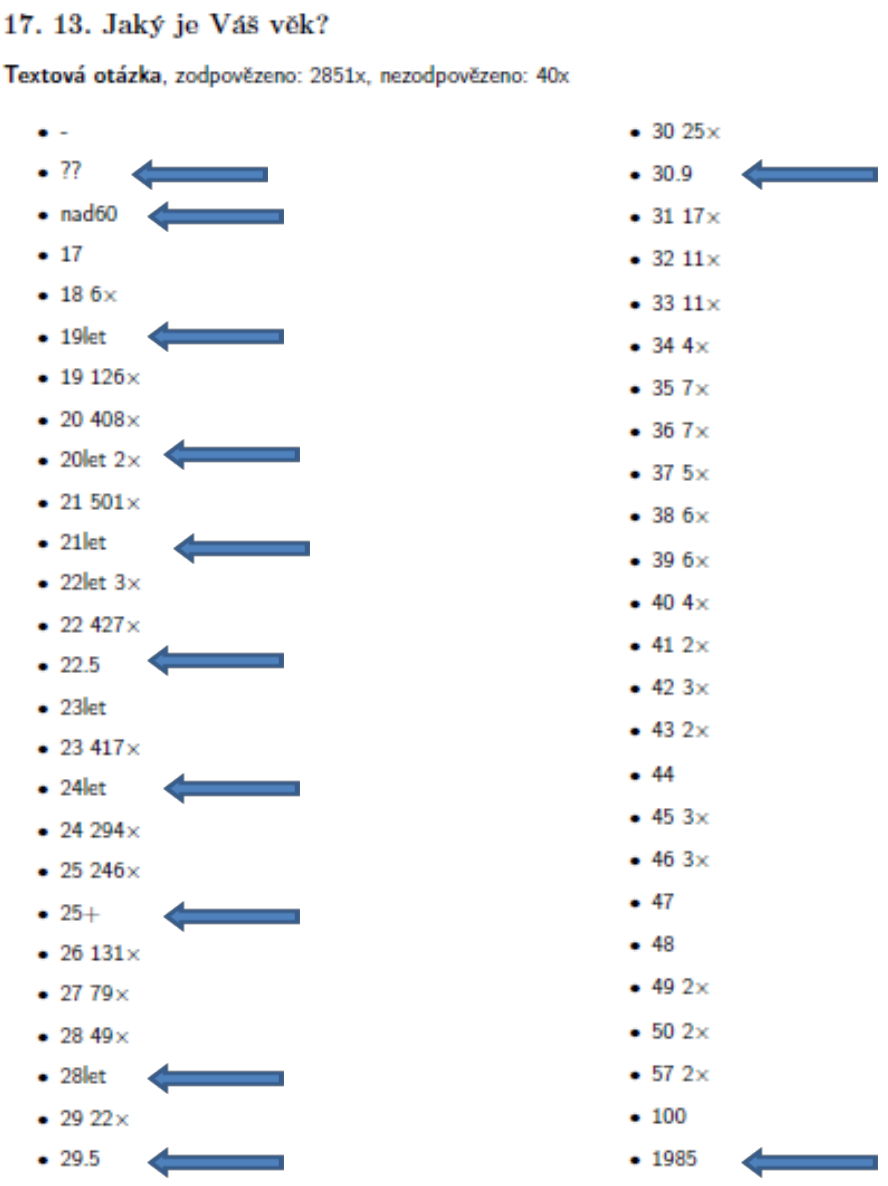



Grafy se v Excelu vkládají pomocí funkce „**Grafy**“ na listu „**Vložení**“.

## *Popis a kontrola nekategorizovaných dat*

Pro první kontrolu nekategorizovaných dat nám bude stačit podívat se na **minimální** a **maximální** hodnoty dat. Například u proměnné „rok narození“ by naši respondenti neměli být narozeni později než v roce 1995 (máme rok 2013 a respondenti měli být starší 18 let). Dřívější datum narození není jasné, ale nejstarší občance ČR je momentálně 109 let, držme se tedy limitu 1904 jako nejmenšího možného roku narození. U hodnot 1904–1995 tedy máme důvod domnívat se, že jsou v pořádku. Často se však mohou vyskytnout chyby vzniklé při zápisu (např. rok 11982 či naopak vynechání číslice – rok 198). Tato data je potřeba opravit.

Někdy se může stát, že respondenti nevědí, jak odpovědět. Potom můžete na jednoduchou otázku (*„Kolik je vám let“*) získat velmi různé formáty odpovědí:



## *Co s chybnými daty?*

Narazíme-li na chybnou hodnotu, máme v zásadě několik možností:

- **Zjistit chybu a nahradit chybný zápis správnou hodnotou.** Například pokud chyba vznikla při přepisu papírového dotazníku do elektronické tabulky, je možné dotazník dohledat a chybu opravit. Stejně postupujeme i v případě, že respondenti nevyplnili pole tak, jak jsme chtěli (např. hodnotu „23let" si překódujeme jen na „23").
- Pokud není možné zjistit chybu, můžeme **prohlásit odpověď za chybějící** a nakládat s ní, jako by nebyla otázka vůbec zodpovězena. Variantně můžeme respondenta úplně vyřadit ze souboru.

## *Co s chybějícími daty?*

Kromě chybných dat je potřeba zkoumat i **chybějící hodnoty**. Vyplatí se před samotnou analýzou zkontrolovat, kolikrát se vyskytly v odpovědích varianty „nevím / nemohu odpovědět".

Jsou odpovědi rozděleny náhodně? Nemá výskyt nevím souvislost s nějakou jinou proměnnou?

Pro kontrolu můžeme rozdělit soubor na skupiny záznamů s chybějícími hodnotami a bez nich, porovnat charakteristiky obou souborů, nebo nechat korelovat vyplnění/nevyplnění s jinou proměnnou (o korelacích bude řeč v dalších modulech).

# 3 Práce s datovým souborem

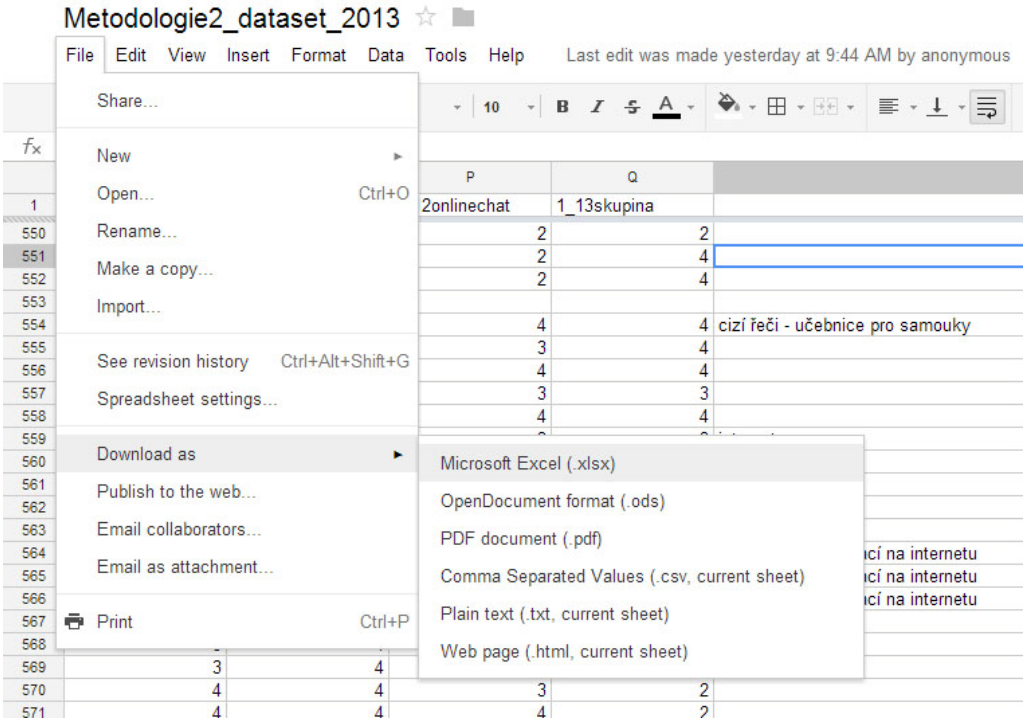
Dřív než začneme pracovat s datovým souborem, je potřeba zmínit několik zásad.

1. Ať už pracujeme v jakémkoliv programu, je vždy důležité pravidelně **zálohovat data**. Ponechte si zálohovaný původní datový soubor, ať se k němu v případě nejistot můžete vrátit. Zálohujte si také průběžnou práci – při analýze často vytváříte nové proměnné, o které byste mohli bez zálohování přijít. Při nepozornosti si také můžete přemazat některá data, proto je vhodné mít zazálohovaných několik posledních verzí souborů s daty.
2. Pokud pracujete ve **sdíleném souboru**, dbejte na to, aby byly kroky jednotlivých výzkumníků odlišitelné a zpětně dohledatelné. Pokud to prostředí neumožňuje, zvažte jinou variantu způsobu práce s daty.
3. Než začnete analyzovat, data **zkontrolujte** a **pečlivě popište**.

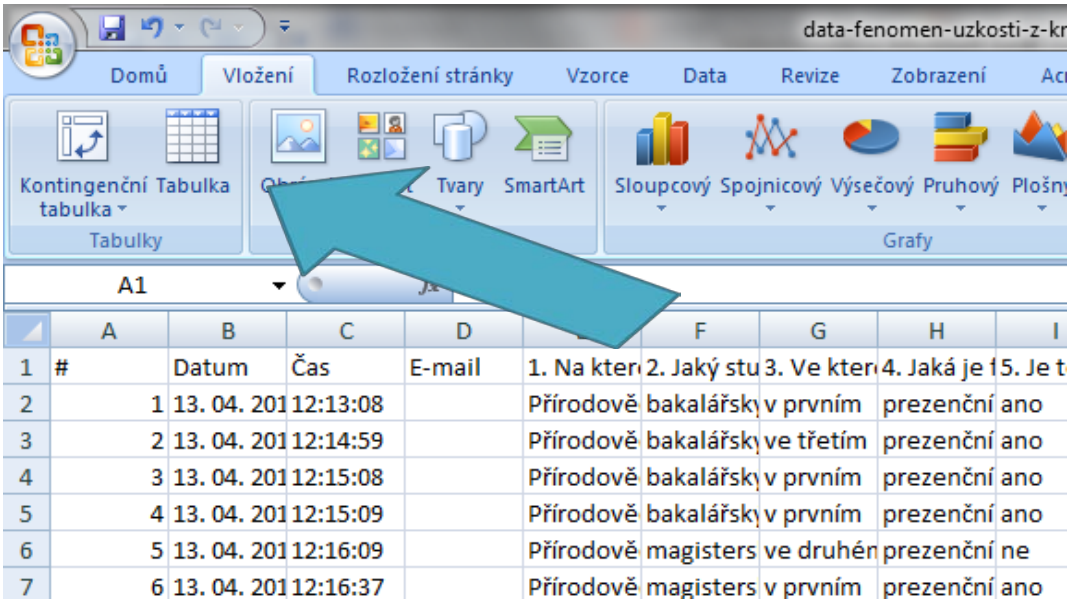
## *Stažení tabulky*

V tomto semestru budeme pracovat se souborem, který jsme si společně vytvořili v Google dokumentech. Většinu operací, které budeme používat, lze provádět přímo v Google Spreadsheets. Pro práci v Excelu je možné si stáhnout tabulku z Google dokumentů pomocí funkce „**Download as**“.

*Stažení souboru ve formátu .xls:*



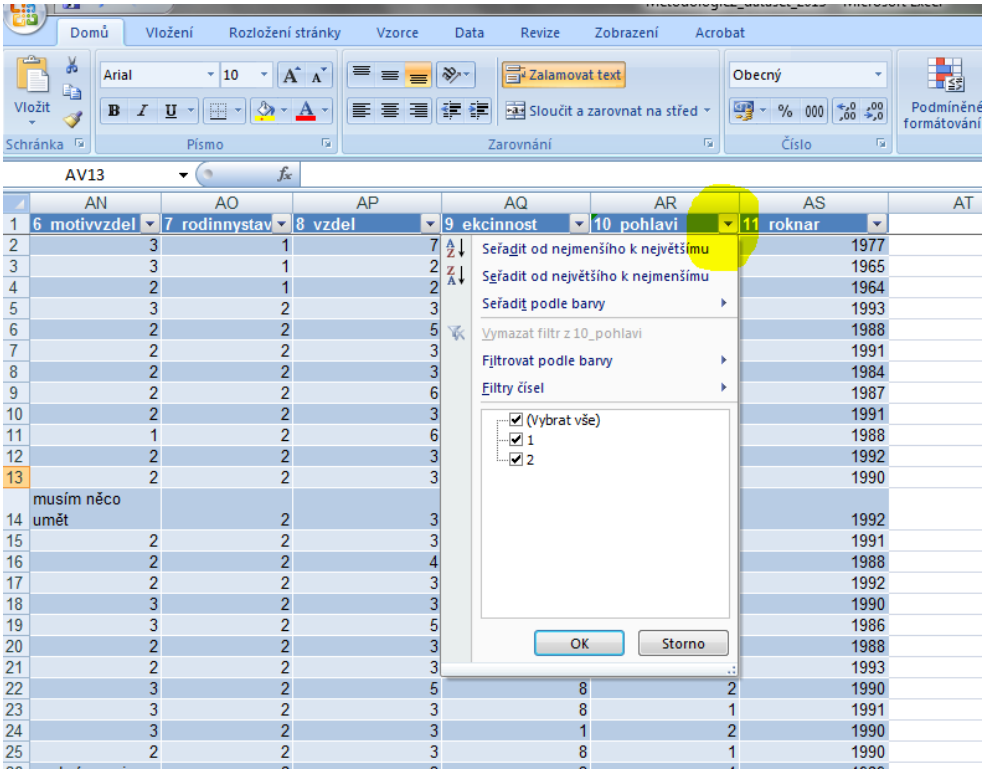

V Excelu je poté pro práci s daty vhodné data převést na inteligentní tabulku pomocí funkce „**Tabulka**“ v listu “**Vložení**”:




| | A | B | C | D | | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | # | Datum | Čas | E-mail | 1. Na kter | 2. Jaký stu | 3. Ve kter | 4. Jaká je t | 5. Je t |
| 2 | 1 | 13. 04. 201 | 12:13:08 | | Přírodově | bakalářsk | v prvním | prezenční | ano |
| 3 | 2 | 13. 04. 201 | 12:14:59 | | Přírodově | bakalářsk | ve třetím | prezenční | ano |
| 4 | 3 | 13. 04. 201 | 12:15:08 | | Přírodově | bakalářsk | v prvním | prezenční | ano |
| 5 | 4 | 13. 04. 201 | 12:15:09 | | Přírodově | bakalářsk | v prvním | prezenční | ano |
| 6 | 5 | 13. 04. 201 | 12:16:09 | | Přírodově | magisters | ve druhém | prezenční | ne |
| 7 | 6 | 13. 04. 201 | 12:16:37 | | Přírodově | magisters | v prvním | prezenční | ano |

Excel rozpozná záhlaví a převede data na přehlednější tabulku.

Někdy nechceme pracovat s celým datovým souborem, ale zajímají nás například pouze ženy. V Excelu si můžeme jednoduše vyfiltrovat rozkliknutím položky v záhlaví:




| | AN | AO | AP | AQ | AR | AS | AT |
|---|---|---|---|---|---|---|---|
| 1 | 6_motivvzdel | 7_rodinnystav | 8_vzdel | 9_ekcinnost | 10_pohlavi | 11_roknar | |
| 2 | 3 | 1 | 7 | | | 1977 | |
| 3 | 3 | 1 | 2 | | | 1965 | |
| 4 | 2 | 1 | 2 | | | 1964 | |
| 5 | 3 | 2 | 3 | | | 1993 | |
| 6 | 2 | 2 | 5 | | | 1988 | |
| 7 | 2 | 2 | 3 | | | 1991 | |
| 8 | 2 | 2 | 3 | | | 1984 | |
| 9 | 2 | 2 | 6 | | | 1987 | |
| 10 | 2 | 2 | 3 | | | 1991 | |
| 11 | 1 | 2 | 6 | | | 1988 | |
| 12 | 2 | 2 | 3 | | | 1992 | |
| 13 | 2 | 2 | 3 | | | 1990 | |
| 14 | musím něco umět | 2 | 3 | | | 1992 | |
| 15 | 2 | 2 | 3 | | | 1991 | |
| 16 | 2 | 2 | 4 | | | 1988 | |
| 17 | 2 | 2 | 3 | | | 1992 | |
| 18 | 3 | 2 | 3 | | | 1990 | |
| 19 | 3 | 2 | 5 | | | 1986 | |
| 20 | 2 | 2 | 3 | | | 1988 | |
| 21 | 2 | 2 | 3 | | | 1993 | |
| 22 | 3 | 2 | 5 | 8 | 2 | 1990 | |
| 23 | 3 | 2 | 3 | 8 | 1 | 1991 | |
| 24 | 3 | 2 | 3 | 1 | 2 | 1990 | |
| 25 | 2 | 2 | 3 | 8 | 1 | 1990 | |

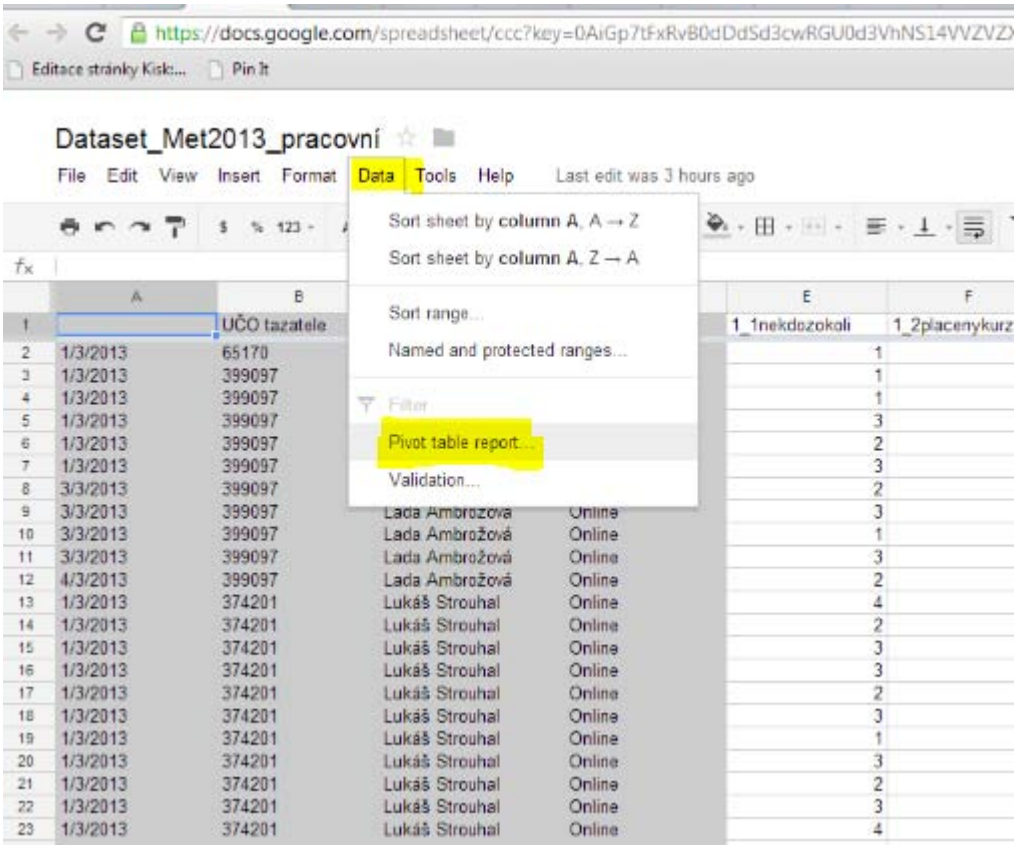
## *Popis rozložení hodnot proměnných*

Pro počítání absolutních četností v Excelu slouží příkaz **COUNTIF**.

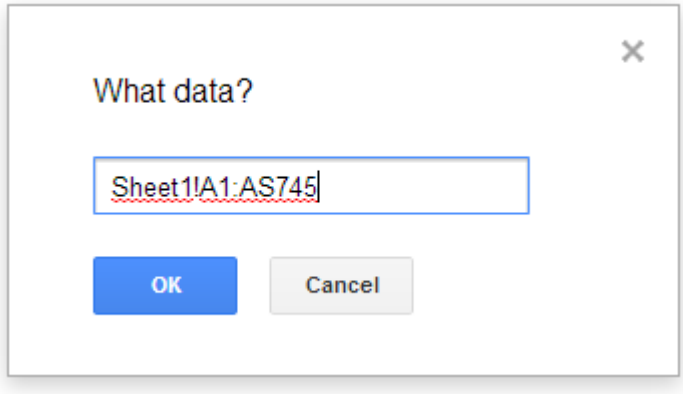
| A | B |
|---|---|
| **Prodejce** | **Faktura** |
| Novák | 15 000 |
| Novák | 9 000 |
| Horák | 8 000 |
| Horák | 20 000 |
| Novák | 5 000 |
| Veselý | 22 500 |
| **Vzorec** | **Popis (výsledek)** |
| =COUNTIF(A2:A7;"Novák") | Počet faktur od Nováka (3) |
| =COUNTIF(A2:A7;A4) | Počet faktur od Horáka (2) |
| =COUNTIF(B2:B7,"< 20000") | Počet faktur s hodnotou nižší než 20 000 (4) |
| =COUNTIF(B2:B7,">="&B5) | Počet faktur s hodnotou vyšší nebo rovnou 20 000 (2) |

*Zdroj: http://office.microsoft.com*

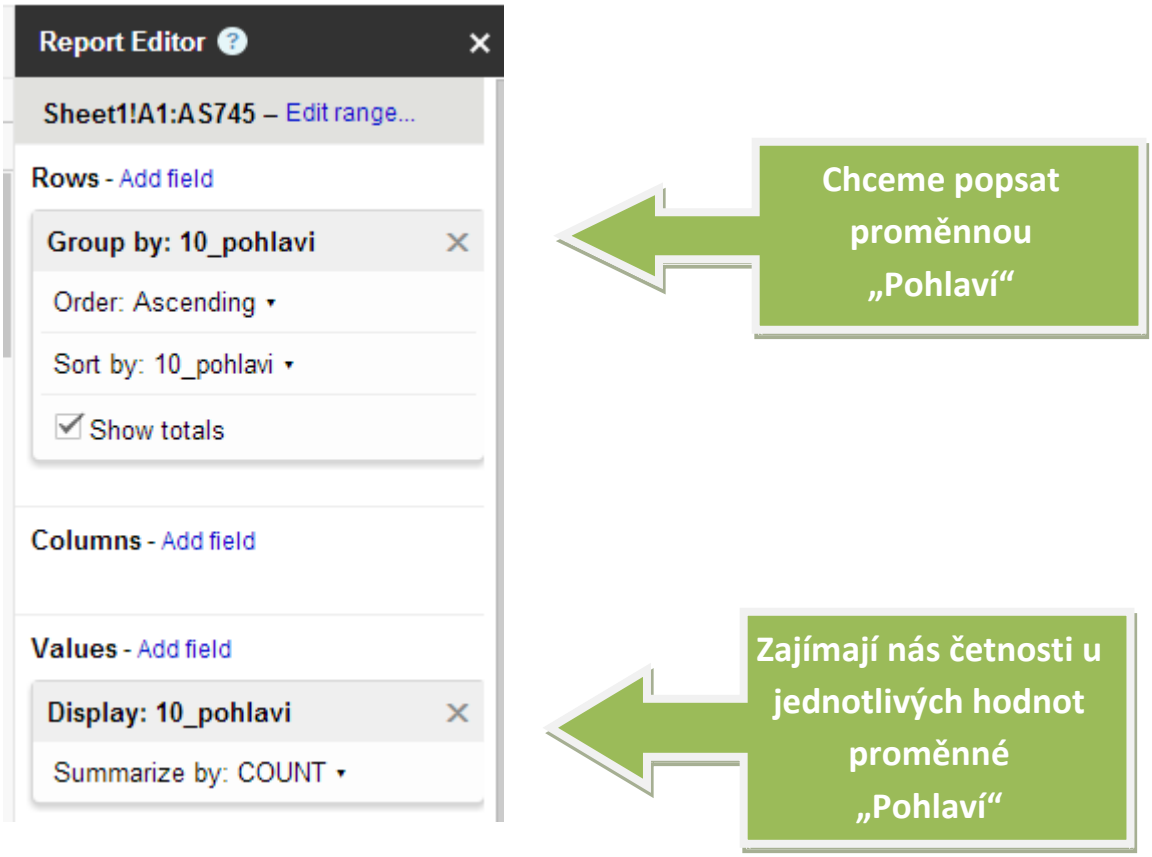
Příkaz COUNTIF nám spočítá výskyt konkrétní varianty hodnoty proměnné. Pro vytvoření tabulky četností je však užitečnější funkce „pivot tables“.  Najdete ji v sekci „Data“.




Aplikace se vás nejprve zeptá na rozsah dat. Dávejte si pozor, abyste zahrnuli celou tabulku.



Nová tabulka se vám objeví na novém listu. Tabulku četností vytvoříte tak, že v položce „Řádky“ / „Rows“ specifikujete proměnnou, kterou chcete popsat a proces výpočtu hodnot. Pro tabulku četností budeme nejčastěji používat příkaz „COUNT“.

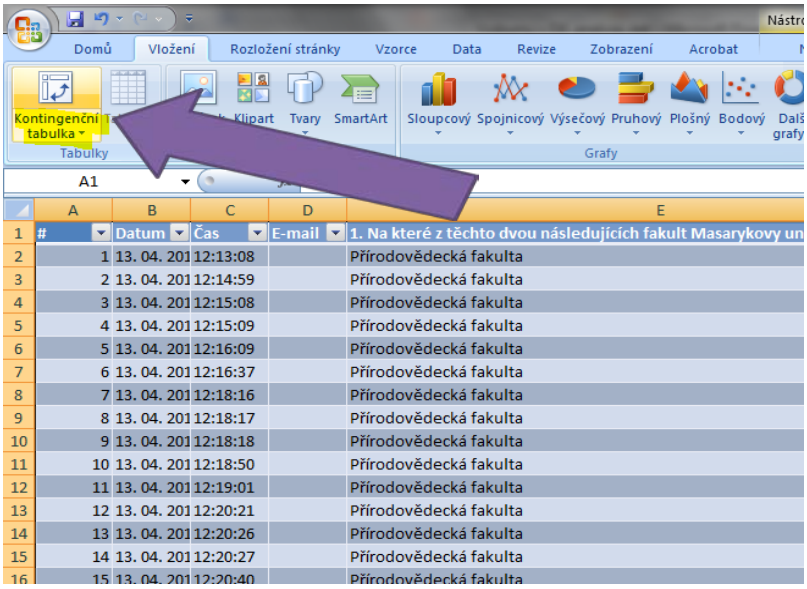



Zpracování v Google Spreadsheets může chvilku trvat, proto buďte trpěliví, pokud tabulka nebude hned reagovat na zadané změny.

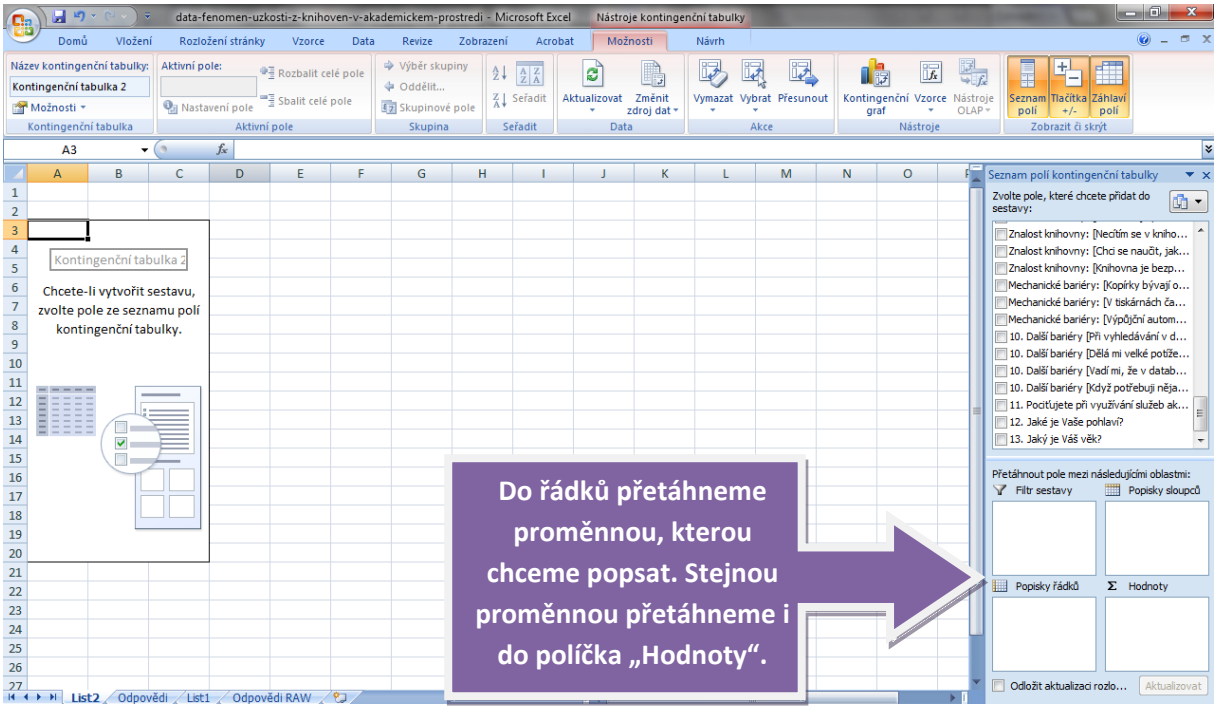
Pokud jste si nepřekódovali odpovědi předem, výsledná tabulka bude obsahovat naše kódy, před publikováním je tedy třeba ji ještě upravit – místo kódů (např. „1“) by výsledná tabulka měla obsahovat reálné hodnoty proměnných (např. „muž“).

| ***Jste:*** | **Četnost odpovědí** | **Validní relativní četnost** |
|---|---|---|
| Muž | 80 | 40 % |
| Žena | 120 | 60 % |
| **Celkem** | **200** | **100 %** |

Pokud jste se rozhodli pracovat v Excelu, je postup velmi podobný. Tabulku vytvoříte tak, že označíte data, se kterými chcete pracovat, a zvolíte možnost „**Kontingenční tabulka**“ na kartě „**Vložení**“.




Na novém listu se objeví prostředí pro tvorbu kontingenčních tabulek. Pro tvorbu tabulek četností budeme využívat zatím jen možnosti popisů řádků:

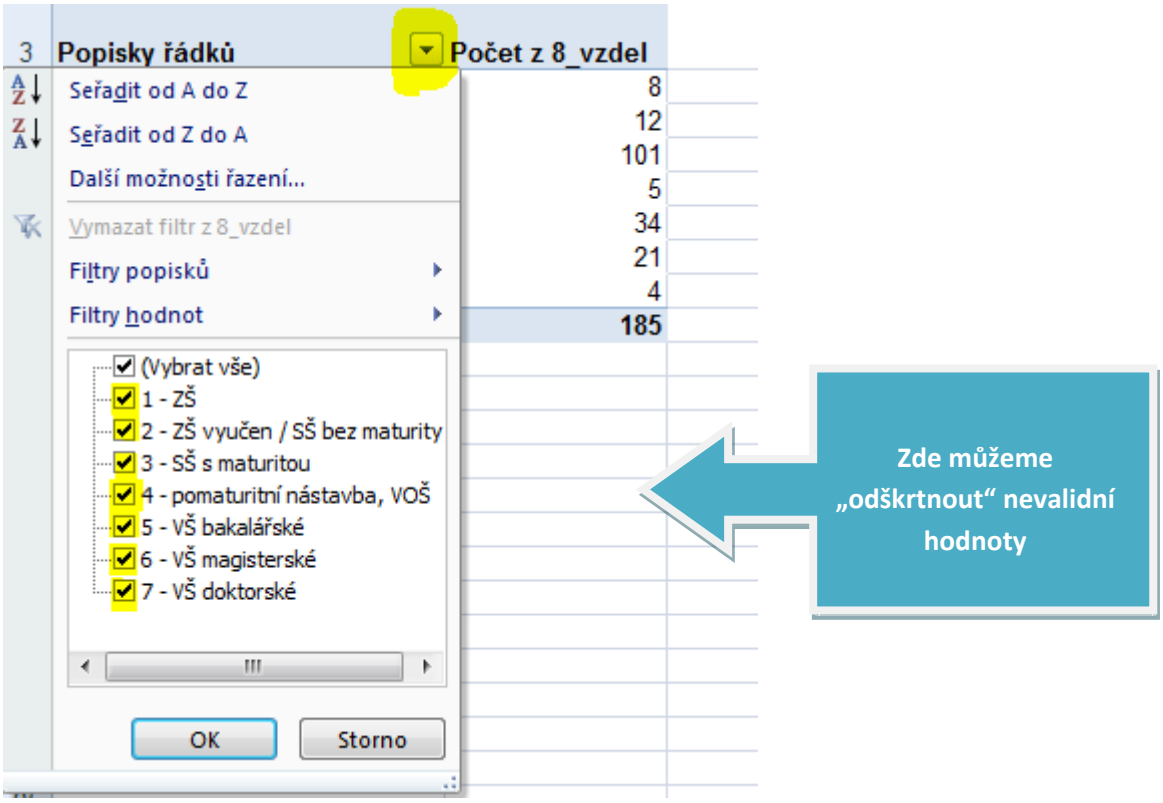



Pro ukázku si vytvořme tabulku se vzděláním:

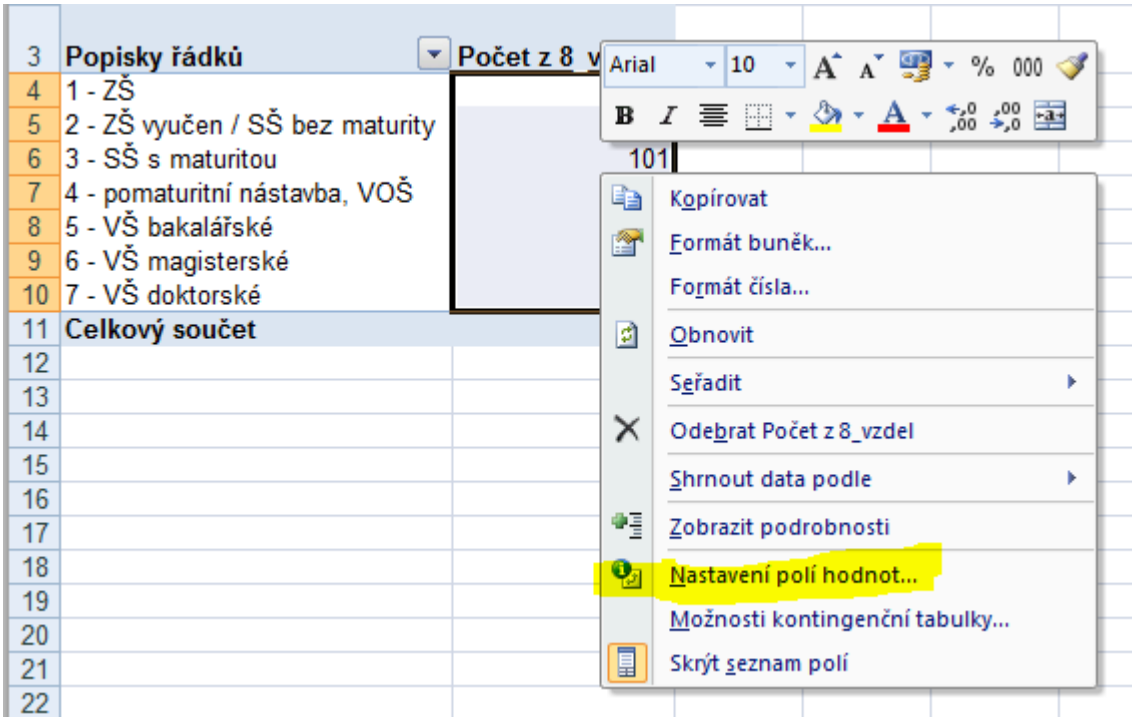
| Popisky řádků | Počet z 8_vzdel |
|---|---|
| 1 - ZŠ | 8 |
| 2 - ZŠ vyučen / SŠ bez maturity | 12 |
| 3 - SŠ s maturitou | 101 |
| 4 - pomaturitní nástavba, VOŠ | 5 |
| 5 - VŠ bakalářské | 34 |
| 6 - VŠ magisterské | 21 |
| 7 - VŠ doktorské | 4 |
| **Celkový součet** | **185** |

**Tato tabulka ukazuje hodnoty výskytu jednotlivých variant odpovědí u proměnné „Dokončené vzdělání“**

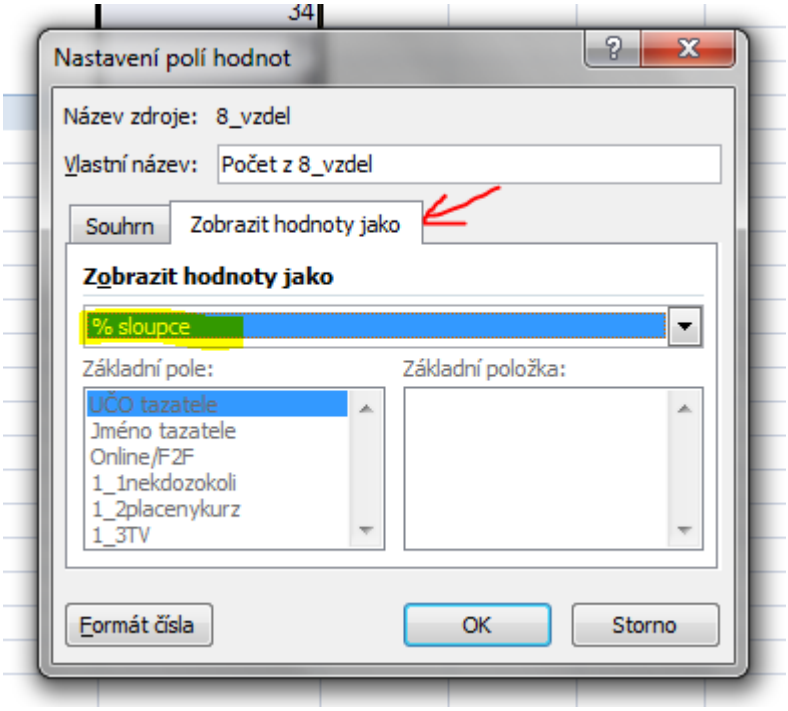
Pokud máme v otázce varianty odpovědí, které nechceme zahrnovat do analýzy (tzv. nevalidní odpovědi – tedy odpovědi typu „nevím“, „neodpověděl“), můžeme je odškrtnout v rozbalovacím menu:




Chceme-li přepočítat absolutní četnosti na relativní četnosti, klikneme na datovou oblast pravým tlačítkem myši a zvolíme možnost „**Nastavení polí hodnot**“:




Vybereme záložku „**Zobrazit hodnoty jako**“ a zvolíme „% sloupce“. Absolutní hodnoty se přepočítají na procenta:




Získáme tak **relativní četnosti**:

| Popisky řádků | Počet z 8_vzdel |
|---|---|
| 1 - ZŠ | 4,32% |
| 2 - ZŠ vyučen / SŠ bez maturity | 6,49% |
| 3 - SŠ s maturitou | 54,59% |
| 4 - pomaturitní nástavba, VOŠ | 2,70% |
| 5 - VŠ bakalářské | 18,38% |
| 6 - VŠ magisterské | 11,35% |
| 7 - VŠ doktorské | 2,16% |
| **Celkový součet** | **100,00%** |

## *Minimální a maximální hodnoty*

Minimální a maximální hodnoty lze rozpoznat už z popisu rozložení proměnných. U spojitých nekategorizovaných dat ale popis rozložení četností nepoužíváme, proto je výhodnější znát příkaz na rychlé zjištění minimálních a maximálních hodnot. V Excelu i v Google Spreadsheet se tyto hodnoty zjišťují pomocí funkce MIN a MAX. Zapisují se do políčka jako příkaz ve tvaru

**„=MIN(datová oblast)“** či **„=MAX(datová oblast)“**

| | A |
|---|---|
| 1 | Data |
| 2 | 10 |
| 3 | 7 |
| 4 | 9 |
| 5 | 27 |
| 6 | 2 |

| Vzorec | Popis (výsledek) |
|---|---|
| =MIN(A2:A6) | Nejmenší z výše uvedených čísel (2) |
| =MIN(A2:A6;0) | Nejmenší z výše uvedených čísel a čísla 0 (0) |

| | A |
|---|---|
| 1 | Data |
| 2 | 10 |
| 3 | 7 |
| 4 | 9 |
| 5 | 27 |
| 6 | 2 |

| Vzorec | Popis (výsledek) |
|---|---|
| =MAX(A2:A6) | Největší z výše uvedených čísel (27) |
| =MAX(A2:A6;30) | Největší z výše uvedených čísel a čísla 30 (30) |

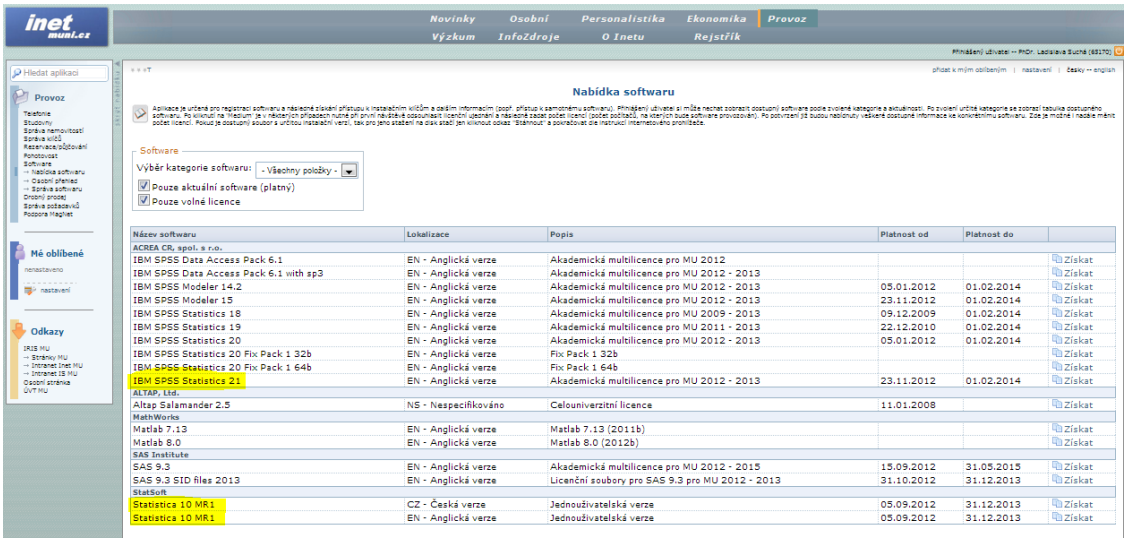
## *Využívejte podpory a nápovědy!*

Pokud si nejste jistí provedením příkazu, využívejte podpory Microsoft Office i Google Spreadsheets. Na internetu lze najít také spoustu videotutorialů a návodů. V nejhorším případě pište na sucha@phil.muni.cz ☺.

# Návod pro práci s SPSS

## Instalace programu

SPSS najdete v INETu. Po přihlášení se se svým UČO a sekundárním heslem najdete programy v sekci Provozní služby – Software – Nabídka softwaru.




Program si můžete stáhnout ve formátu ISO. Pro spuštění je tedy nutné jej vypálit na DVD nebo vytvořit virtuální disk. Při registraci nezapomeňte uvést registrační kód dostupný v INETu.

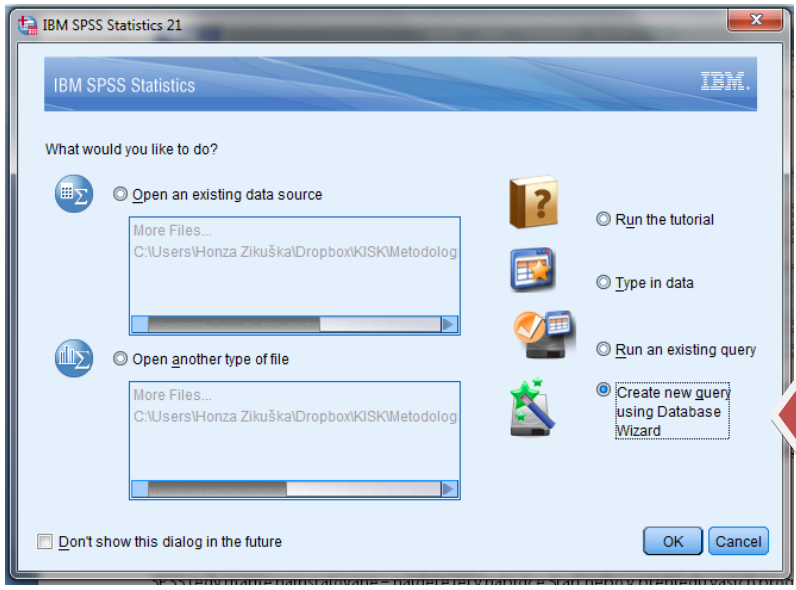
Dostupných je hned několik druhů licencí – doporučuji vybrat licenci **IBM SPSS Statistics 21** (nejnovější verze programu).

## Otevření souborů s daty

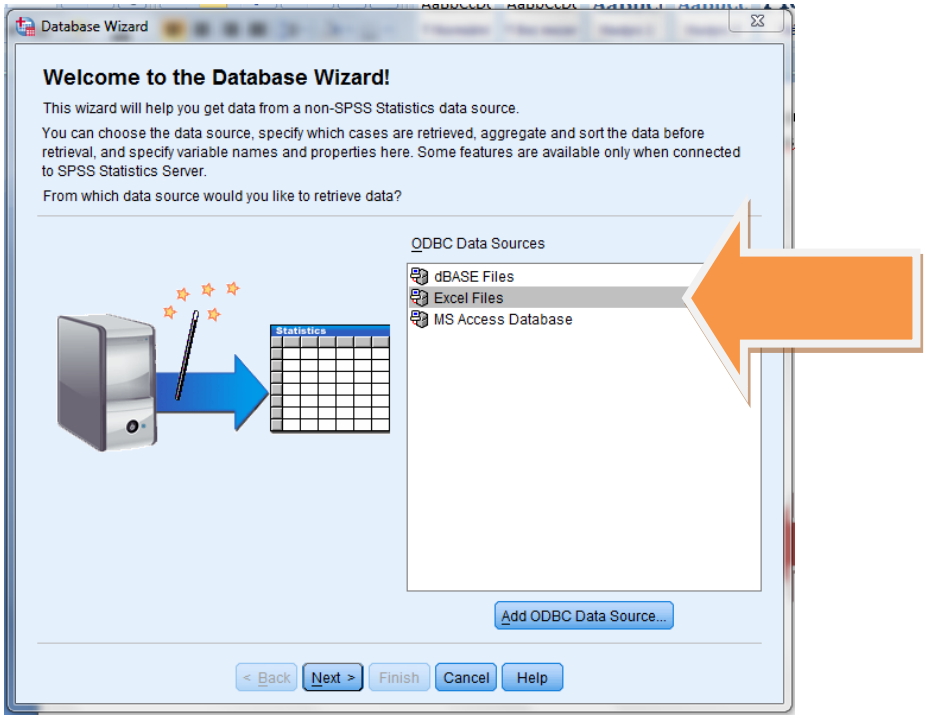
SPSS tedy máme nainstalované – najdete jej v nabídce Start nebo v přehledu vašich programů. Do SPSS můžete data dostat několika způsoby – ten nejzákladnější je přímé tvoření datasetu v SPSS. My ale budeme potřebovat pracovat s daty, která již máme ve formátu .xls.

Postupovat budeme následovně:

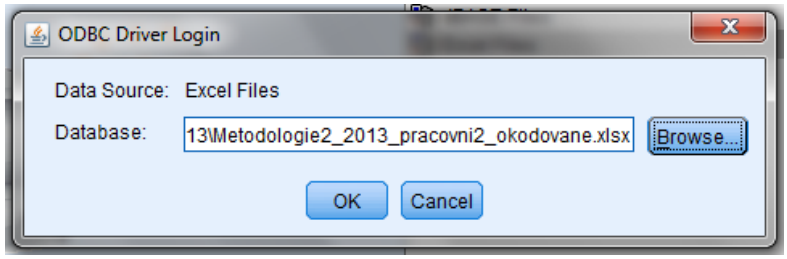
1. Uložíme si na své PC datový soubor ve formátu pro Excel (najdeme jej v ISu).
2. Pro převedení excelového souboru do souboru typu .sav spustíme „Database Wizzard“:




3. Z nabízených možností v dalším okně si vyberte „Excel files“:




4. Vyberte soubor ze svého PC:



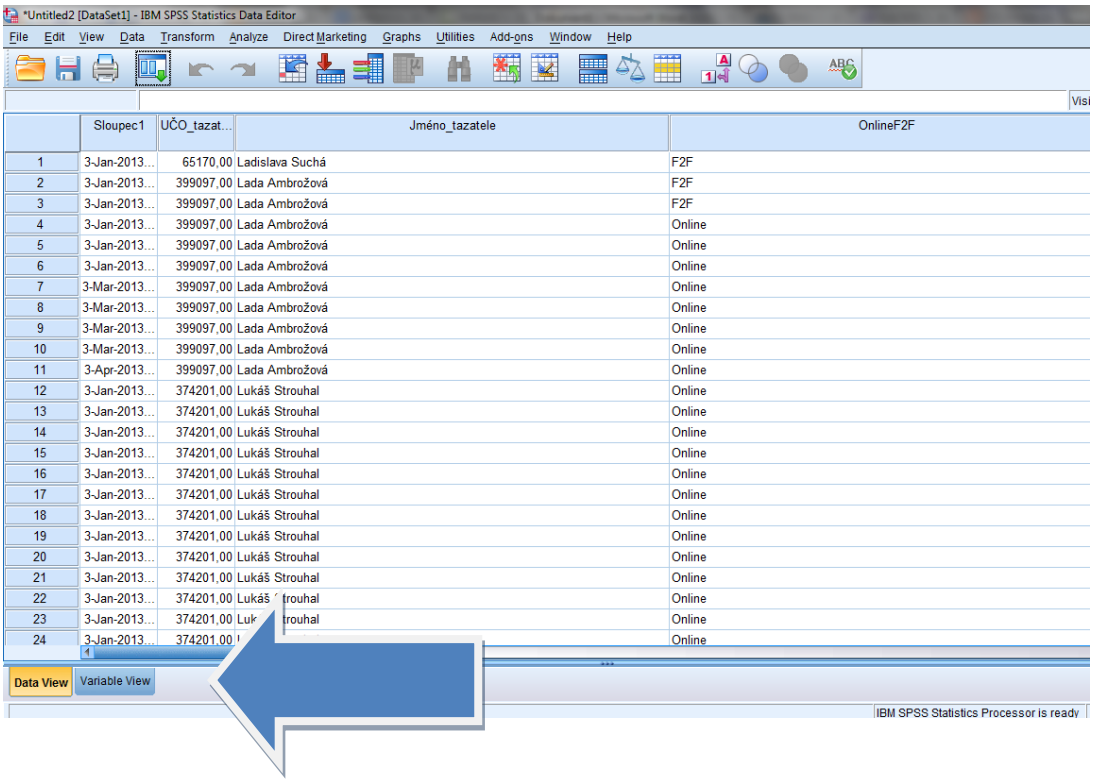

5. Vyberte si oblast, kterou chcete převést a poté potvrďte stistem „**Finish**“


6. V počítači se vám otevřou dvě nová okna. Jedno přímo s datasetem a druhé je tzv. „Output“ – okno, kam se zapisují procesy a výsledky operací SPSS.

## Práce s datasetem

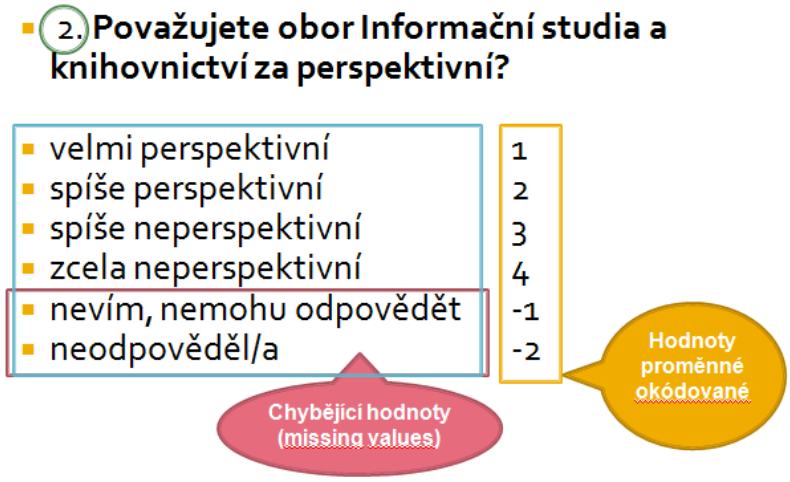
Dataset je neprve potřeba upravit a popsat. Všimněte si, že v SPSS lze přepínat mezi dvěma druhy zobrazení:

- pohled na data,
- pohled na proměnné.

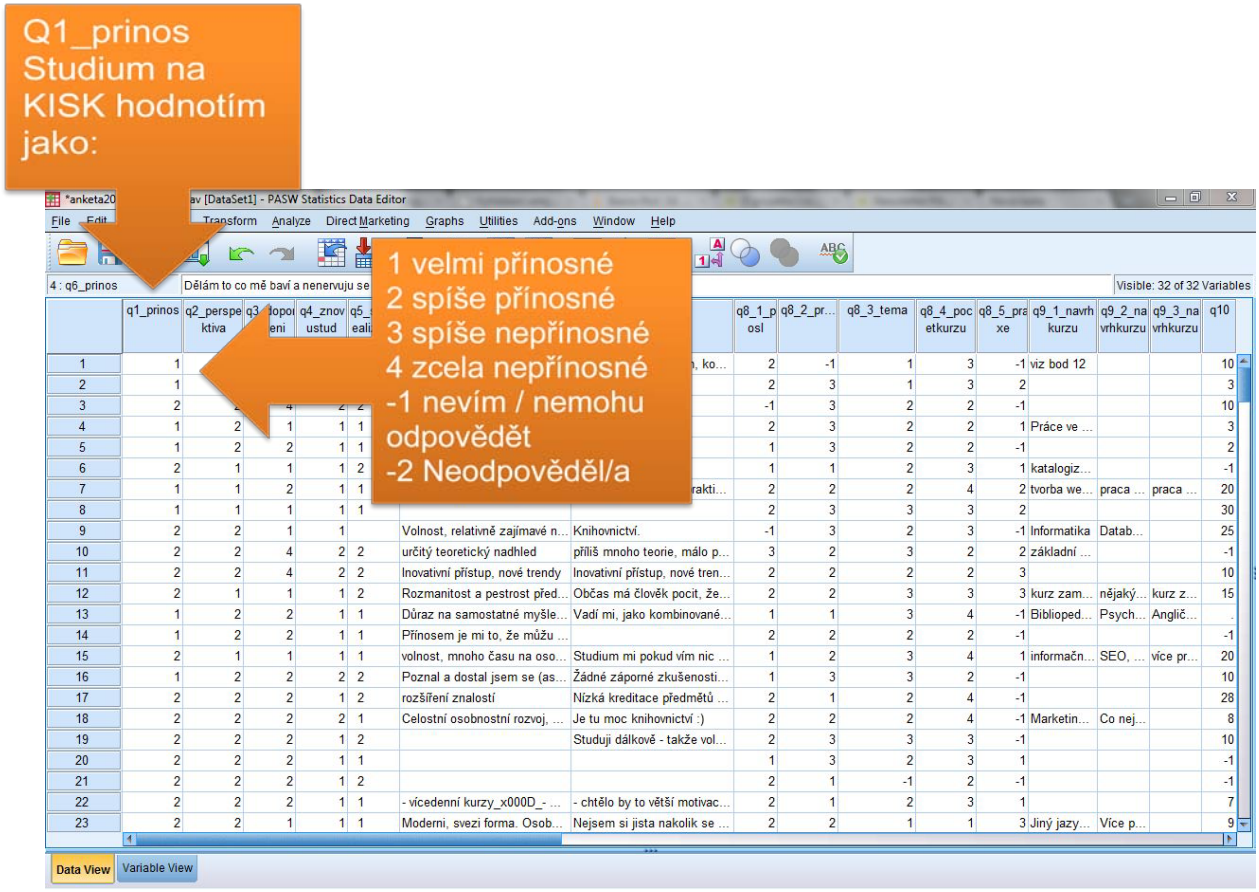



**Pohled na data** je velmi podobný tomu, co znáte z Excelu – co řádek, to respondent, co sloupec, to proměnná. **Pohled na proměnné** upřesňuje parametry jednotlivých proměnných.

Ukažme si to na příkladu této otázky:



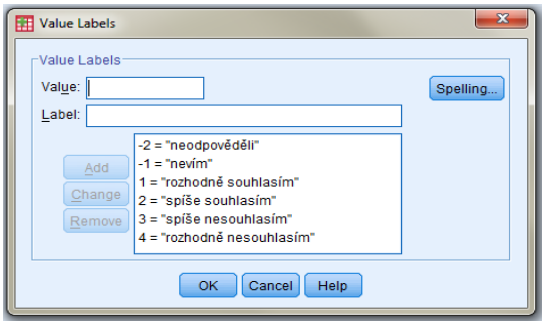

Takto bude vypadat matice dat:

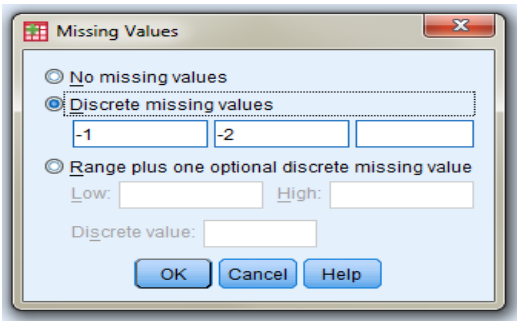



Zároveň je potřeba popsat jednotlivé proměnné na kartě **Variable view**:

- *Name: zkrácené označení proměnné.*
- *Typ: číselné/slovní (SPSS potřebuje vědět, jaké operace může provádět s jednotlivými proměnnými)*

- *Decimal: desetinná místa (pouze kardinální proměnné) – automaticky jsou nastavena dvě desetinná místa, snižte si jejich počet na 0.*
- *Label: většinou kopírujeme znění otázky.*
- *Value labels: hodnoty proměnné – popíšeme všechny hodnoty proměnné včetně „missing values“*
- *Missing values: které hodnoty nezahrnujeme do dané analýzy – SPSS s nimi v konkrétních operacích nebude počítat.*
- *Measure: typ proměnné (nominální/ordinální/kardinální)*



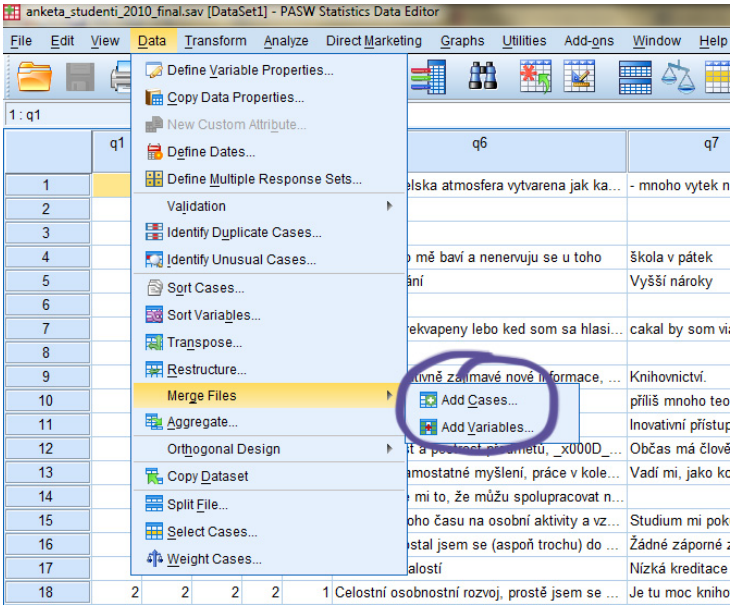




Ve studijních materiálech v ISu máte již datasety s popsanými proměnnými.

## *Slučování datových souborů*

Někdy potřebujeme sloučit více datových souborů. Máme na výběr dvě varianty:

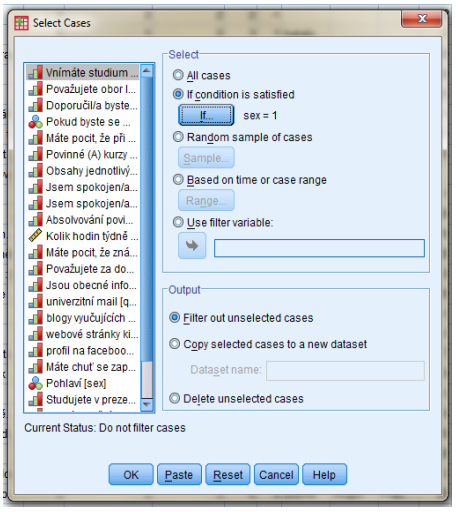
- Chceme sloučit více dat o stejných případech: Merge Files → Add variables
- Chceme sloučit soubory s různými jednotkami a stejnými proměnnými Merge Files → Add Cases



## *Výběr případů*

Někdy naopak potřebujeme pracovat jen s některými případy (například se ženami):

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



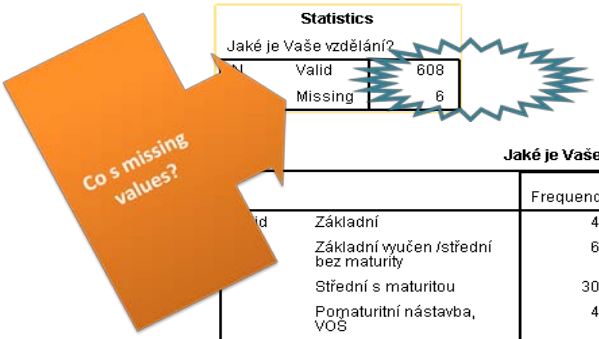

## *Kontrola dat*

V SPSS probíhá kontrola dat se stejnou logikou jako v jakénkoliv jiném programu. Její provedení je jen jednodušší, protože SPSS je přizpůsobeno na provádění statistických operací. SPSS má také tu výhodu, že nám v Outputu dává tabulky již v té podobě, v jaké by se měly objevit v odborné práci – tedy kompletní tabulky četností s nevalidními validními absolutními i relativními hodnotami.

Pro použití v odborné práci je pouze třeba **přeložit popisky tabulek**.

## *Kontrola kategorizovaných dat*

SPSS nám prostřednictvím jednoduchého příkazu **Analyze → Descriptive Statistics → Frequencies** (zde si vyberete konkrétní proměnnou) vrátí počet validních a nevalidních hodnot proměnných. Výsledky najdeme v okně Output:




**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, VOŠ | 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 | | |

Stejně jako v případě SPSS nás bude zajímat výpis četností jednotlivých výskytů hodnot proměnné. Zde máme příklad chybného zápisu jména studentky či chybného zápisu v proměnné „pohlaví“:

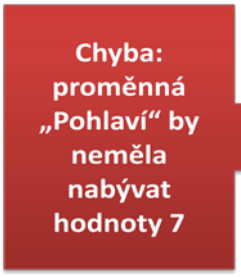

**Jméno výzkumníka**

| | | 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 |

**Statistics**

Pohlaví

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



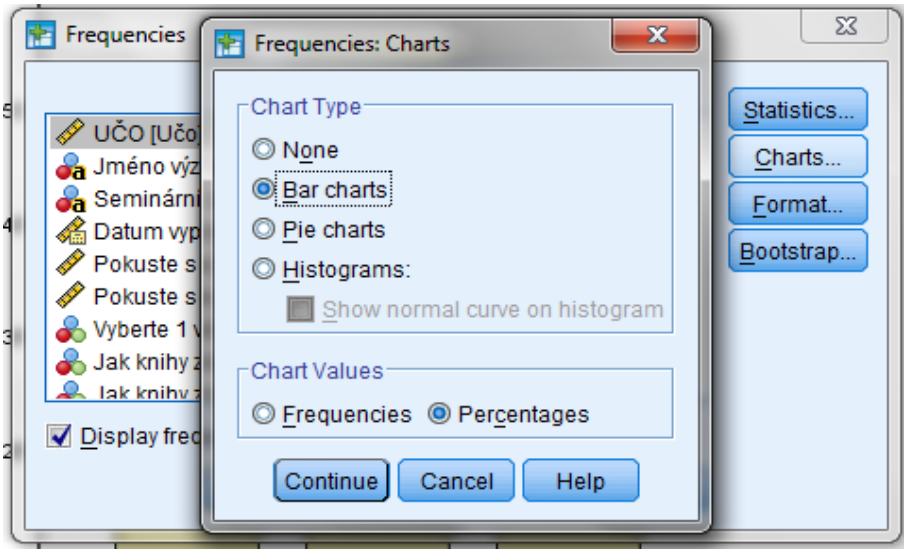

**Pohlaví**

| | | 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 | | |

Poté co naleznete chybná data, můžete je v datasetu vyhledat pomocí příkazu CTRL+F stejně jako v Excelu.

## *Tabulky četností a grafy v SPSS*

Tabulky četností v SPSS získáme příkazem Analyze → Descriptive Statistics → Frequencies . Grafy vytvoříme cestou Analyze → Descriptive Statistics → Frequencies → **Charts**.




## Literatura


Disman, M. (2002) Jak se vyrábí sociologická znalost. Praha: Karolinum.

Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124.

Wheelan, Ch. (2013) Naked Statistics. New York: W. W. Norton & Company Ltd.