

**Studijní opora předmětu**  
**Elementární zpracování dat z kvantitativního výzkumu**  
*Úvod do analýzy kvantitativních dat v prostředí statistického software PSPP/ SPSS*

**Jiří Šafr**

FHS UK

Email: <jiri.safr(AT)seznam.cz>

Poslední aktualizace: **14. prosince 2017, verze 1.0**

Vytvořeno 7. března. 2017, verze 0.1

**Obsah**

|  |    |
|--|----|
| Úvod: kvantitativní data a jejich záznam do datové matice .....                      | 3  |
| PSPP a SPSS – program pro statistickou analýzu dat .....                             | 4  |
| Prostředí programu PSPP/SPSS .....   | 5  |
| Druhy „oken“ v PSPP/ SPSS .....  | 5  |
| Data editor a datový soubor .....  | 6  |
| Základní data management v PSPP/SPSS .....   | 7  |
| Vytvoření nové matice dat .....  | 7  |
| Označování popisků proměnných .....  | 10 |
| Vkládání dat .....   | 11 |
| Otevření již existujícího souboru dat .....  | 12 |
| Analýza dat (1): popisné statistiky, třídění dat 1.stupně .....                      | 17 |
| Popisné statistiky pomocí FREQUENCIES .....  | 17 |
| Grafy (1): třídění dat 1. stupně .....   | 19 |
| Zadání grafů v příkazu Frequencies (sloupcový, koláčový a histogram) .....           | 19 |
| Sloupcový graf a histogram .....   | 20 |
| Technické rady pro práci v PSPP .....  | 25 |
| Možnosti úprav výstupu (Output) z PSPP .....   | 25 |
| CSV – univerzální formát pro data .....  | 28 |
| Import CSV dat do PSPP .....   | 31 |
| Syntax – ovládání pomocí příkazů (úvod) .....  | 33 |
| Analýza dat (2): Logika vytváření kontingenčních tabulek a jejich interpretace ..... | 35 |
| Prezentace a interpretace kontingenční tabulky (výzkumná otázka a hypotéza) .....    | 35 |
| Uspořádání v kontingenční tabulce .....  | 38 |
| Postup vytváření kontingenčních tabulek v PSPP (SPSS) .....                          | 40 |
| Test dobré shody Chí kvadrát – (ne)závislost proměnných v kontingenční tabulce ..... | 46 |
| Znaménkové schéma – v kterých polích tabulky je jak silná souvislost? .....          | 47 |
| Zadání kontingenční tabulky pomocí syntaxu PSPP/SPSS .....                           | 49 |

Tato studijní opora částečně vychází z dřívějšího textu pro FHS Návod na software PSPP/SPSS (část 1. – Úvod a část 2. – Kontingenční tabulky). Publikace vznikla jako součást řešení projektu *Další profilování a inovování studijního programu na Katedře řízení a supervize v sociálních a zdravotnických organizacích* v rámci institucionálního plánu FHS UK pro roky 2016-2018.

© Jiří Šafr 2017

Upozornění: Jedná se o studijní materiál pro kurz **Praktikum - elementární zpracování dat z kvantitativního výzkumu (YMR13VOLI)** vyučovaný na Katedře řízení a supervize FHS UK, který **bude dále průběžně aktualizován**.

Cílem kurzu a tedy i této studijní opory je pomoci studentům (druhého ročníku) zpracovávat data pro výzkumné zprávy a své diplomové práce pomocí kvantitativní metodologie sociologického výzkumu. V jeho průběhu by se měli naučit na té nejjednodušší úrovni vytvářet a zpracovávat kvantitativní data (nejen) z vlastních výzkumných projektů. Výklad a praktické procvičování je zaměřeno na základní popisné statistiky a postupy k odhalení vztahů mezi proměnnými, tj. možnosti ověření platnosti hypotéz jednoduchým – deskriptivním způsobem (třídění dat 1. a 2. stupně). Předmětem výuky není inferenční statistika, tj. statistické testování hypotéz.

Text je členěn na dvě části. V první najdete úvod do prostředí programů PSPP a IBM-SPSS, základy data managementu (názvy proměnných, labelování, označování chybějících hodnot), nejzákladnější přístupy k třídění dat prvního stupně (popisné statistiky a grafy) a také technické rady specifické pro práci s PSPP. Na ní navazuje druhá část věnovaná vytváření tabulek třídění druhého stupně a interpretaci výsledků.

Před čtením tohoto textu, a tedy před prvním zapnutím počítače a spuštěním programu PSPP anebo IBM-SPSS, laskavému čtenáři důrazně doporučuji přečíst si nejprve *Studijní oporu předmětu Kvantitativní metody výzkumu v praxi*. Ta vám objasní, co to vůbec obnáší „dělat sociálně-vědní výzkum“, tedy především co vše předchází tomu, než budete mít k dispozici data k analýze. Dozvíte se zde nejzákladnější principy, jak a hlavně proč ve výzkumu postupovat a také čeho se vyvarovat, zejména s ohledem na designování jednoduchých studií především v oblasti řízení a sociální práce/zdravotnictví.

Zatím poslední verzi 07 najdete na [http://kmvp.wz.cz/KMVP\\_opora07.pdf](http://kmvp.wz.cz/KMVP_opora07.pdf)

Také doporučuji přečíst si nejprve prezentace k přednáškám KMVP, zejména tyto dvě:

*Příklad postupu ověření platnosti hypotézy. Základní postupy explorační analýzy pomocí třídění dat 2. stupně.* [http://kmvp.wz.cz/KMVP\\_hypoteza\\_overeni\\_priklad.pptx](http://kmvp.wz.cz/KMVP_hypoteza_overeni_priklad.pptx)

*Výběry vzorků z populace, příprava dat, popisné statistiky a základní postup při ověřování hypotéz.* [http://kmvp.wz.cz/KMVP\\_4a\\_statistika.pptx](http://kmvp.wz.cz/KMVP_4a_statistika.pptx)

Aktuální verze obou studijních textů spolu s powepointovými prezentacemi kurzů KMVP a Praktikum – elementární zpracování dat z kvantitativního výzkumu najdete na stránce <http://kmvp.wz.cz>.

Data, která používám v následujících příkladech jsou dostupná na stránce <http://metodykv.wz.cz> kde také najdete podrobnější prezentace k analýze kvantitativních dat v prostředí programu IBM-SPSS.

## Úvod: kvantitativní data a jejich záznam do datové matice

Pokud jako metodu sběru dat používáme dotazník, a řekneme, že jsme zvolili přístup face-to-face rozhovor (tazatel se ptá respondenta) se záznamem do papírového archu – dotazníku, pak po provedení všech rozhovorů máme před sebou „hromadu dotazníků“. V zásadě máme dvě možnosti, jak s tímto datovým materiálem naložit a jak jej dále analyzovat: datovou matici (v elektronické podobě) si nevytvoříme a tak k analýze můžeme použít pouze metodu „čárkování“ (případně neustálého přeskupování na hromádky podle třídících znaků), k čemuž nám prakticky stačí tužka a papír případně kalkulačka (lepší je Excel či jiný tabulkový procesor). Touto možností se dále zabývat nebudeme, jednak protože je nespolehlivá a může být při vícenásobném třídění pekelně úmorná a také „data“ vlastně neexistují, není je možné sdílet, zálohovat a tedy ani archivovat.

Druhou cestou, kterou se určitě vydáme, je pořídit ze zápisů v dotaznících elektronický záznam dat do *datové matice*.<sup>1</sup> Ta má nejčastěji podobu, kterou ukazuje následující obrázek, kde ve sloupcích jsou proměnné a v řádcích případy, v našem případě respondenti. Ale mohly by to být také například různé organizace, proměnné ve sloupcích by pak představovaly kupříkladu jejich příjmy a výdaje za různé oblasti, počty klientů atd. Pokud máte data z jednoduchého výzkumu s využitím jejich sběru pomocí dotazníku, pak jeden datový řádek představuje jednoduše jeden dotazník, tedy jednoho respondenta.

Datová struktura ale může být složitější, a také ve výzkumu organizací často bývá. Někdy totiž v datech máme individuální případy (úroveň 1) sloučené do určitých přirozených klastrů (úroveň 2), které jsou v matici upořádány vždy pod sebou. Například provádíme výzkum zaměstnanců (1), kteří jsou z různých poboček firmy (2), analyzovat pak můžeme jedince – zaměstnance i organizační jednotky jako celky mezi sebou ale také obojí navzájem (o tomto uspořádání dat podrobněji v kapitole Úrovně měření, analytická jednotka a volba designu výzkumu, kde máme jednoduchý příklad takovéto víceúrovňové datové matice).

Výhoda záznamu dat do matice v elektronickém formátu spočívá v tom, že je lze jednoduše dále upravovat (transformovat proměnné, například slučovat hodnoty nebo vytvářet znaky nové na základě kombinací jiných původních znaků) a také úsporně archivovat a zálohovat. Jak datovou matici vytvořit, data do ní vložit a dále upravovat si povíme v další části, nejprve se ale seznámíme se dvěma statistickými programy, které nám v tom pomohou a ve kterých budeme pracovat: PSPP resp. IBM-SPSS, což je pro naše základní analýzy a úpravy dat vlastně jeden a ten samý program.

---

<sup>1</sup> Tím úmyslně pomímám mnohem přívětivější situaci, kdy záznam dotazníku probíhá pomocí nějakého elektronického způsobu zápisu, např. při vyplňování on-line dotazníku samotným respondentem, díky němuž datová matice vzniká víceméně automaticky (popravdě někdy méně než více..., tím mám na mysli různé problémy se způsobem záznamu a konverzí formátu).

**Obrázek 1. Datová matice (2-way data: případy x proměnné)**

**Datová matice**

|    | A  | B       | C   | D   | E   | F    | G    | H    | I    | J       |
|----|----|---------|-----|-----|-----|------|------|------|------|---------|
|    | ID | pohlaví | ot1 | ot2 | ot3 | ot4a | ot4b | ot4c | ot4d | ot5_vek |
| 2  | 1  | muž     | 1   | 0   | 1   | 1    | 2    | 2    |      | 45      |
| 3  | 2  | muž     | 3   | 1   | 2   | 3    | 1    | 2    |      | 38      |
| 4  | 3  | muž     | 1   | 0   | 2   | 2    | 2    | 1    |      | 54      |
| 5  | 4  | žena    | 1   | 0   | 3   | 3    | 3    | 3    |      | 18      |
| 6  | 5  | žena    | 1   | 0   | 2   | 2    | 1    | 2    |      | 23      |
| 7  | 6  | žena    | 1   | 0   | 1   | 1    | 1    | 1    |      | 21      |
| 8  | 7  | žena    | 2   | 1   | 4   | 3    | 2    | 3    |      | 25      |
| 9  | 8  | žena    | 3   | 1   | 1   | 2    | 2    | 2    |      | 30      |
| 10 | 9  | žena    | 3   | 1   | 4   | 2    | 3    | 2    |      | 43      |
| 11 | 10 | žena    | 3   | 1   | 3   | 2    | 3    | 2    |      | 21      |
| 12 | 11 | žena    | 3   | 1   | 2   | 1    | 2    | 2    |      | 24      |
| 13 | 12 | žena    | 3   | 1   | 1   | 1    | 3    | 1    |      | 29      |
| 14 |    |         |     |     |     |      |      |      |      |         |
| 15 |    |         |     |     |     |      |      |      |      |         |

## PSSP a SPSS – program pro statistickou analýzu dat

PSSP je program určený pro statistickou analýzu kvantitativních dat, který částečně nahrazuje profesionální programový balík IBM SPSS Statistics.<sup>2</sup> Vytvořen byl pod GNU General Public License, tudíž je distribuovaný zdarma. Datové formáty, provedení prostředí programu a hlavně příkazový řádek (syntax) PSPP je kompatibilní s SPSS.

Hlavní stránku projektu s podrobnými návody a popisem naleznete na

**<http://www.gnu.org/software/pspp>** . Zde jsou k dispozici ale pouze jen verze instalací pro Linux a nebo starší verze. Download neustále aktualizovaných verzí pro Windows naleznete na **<http://pspp.awardspace.com>** . Doporučuji vždy se čas od času se podívat po nové verzi (k 3/3/2017 byla poslední verze 10.4).

Nové verze včetně opravných balíčků najdete také na <https://sourceforge.net/projects/pspp4windows/files/2010-04-14/pspp-master-single-user-20100414-Setup.exe/download> nebo <http://ufpr.dl.sourceforge.net/project/pspp4windows/2010-04-14/pspp-master-20100414-Setup.exe>

**Ve verzi 10.4** (z 2016/09/27) byly k dispozici:

- transformace a management dat (rekodování, agregování, spojování, převrácení, vážení atd.)
- popisné statistiky včetně grafů (histogram, vousaté krabíčky, Q-Q plot),
- kontingenční tabulky s míry asociace (korelace Pearson, Spearman, TauB, CC),
- testování průměrů (t-testy a jednoduchá analýza rozptylu OneWayAnova),
- Pearsonova korelace (korelační matice),
- Klastrová analýza pomocí metody K-means cluster,
- analýza položkové reliability (Cronbachovo Alpha, Half split),
- faktorová analýza resp. analýza hlavních komponent (PCA, PAF),
- lineární regrese (OLS), binární logistická regrese,
- obecný lineární model (GLM) resp. vícerozměrná analýza rozptylu,
- neparametrické testy,
- prediktivní ROC křivky.

<sup>2</sup> SPSS znamená Statistical Package for the Social Sciences. Tento statistický softwarový „balík“ má dlouhou historii spadající na počátek 70. let minulého století. Dále v textu pro tento program, prodáváný dnes pod obchodním názvem „IBM SPSS Statistics“, používám pro jednoduchost pouze tuto původní zkratku SPSS.

Dodejme, že profesionální program IBM-SPSS Statistics toho umí mnohem více. Pokud jde o základní postupy, tak určitě v něm najdete mnohem lepší (vizuálně i flexibilitou úprav) a pokročilejší grafy. Pro základní data management a jednoduchou explorační analýzu však dnes již PSPP dostačuje (příkazy FREQUENCIES, DESCRIPTIVES, MEANS, EXPLORE, CROSSTABS, korelace), otázkou je ale uživatelská přívětivost (pokud jde o výstupy) a především spolehlivost (program může nečekaně „spadnout“, ale neustále aktualizované nové verze chyb postupně odstraňují). Proto si PSPP průběžně aktualizujte.

## Prostředí programu PSPP/SPSS

### Druhy „oken“ v PSPP/ SPSS

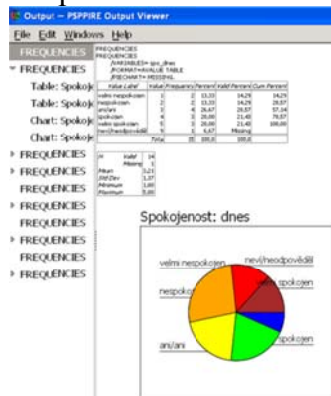
V programu PSPP (i SPSS) pracujeme se třemi „okny“:

1. **Data editor** – samotná data a jejich management, jde o hlavní okno programu (datové soubory mají koncovku „\*.sav“)
2. **Output – výstupy analýz** (zde je zatím největší nevýhoda PSPP, které nemá vlastní formát pro outputy a tak output lze zatím uložit jen jako export do nějakého formátu, přičemž bohužel ztratíte některé formátové vlastnosti. Koncovka pak závisí na volbě exportu do formátu, k dispozici je „pdf, txt, htm, odt“ a další; MS word ani Excel (zde je alespoň CSV) mezi nimi ale není ...)
3. **Syntax– editor přímých příkazů** (koncovka souborů je „\*.sps“). Ten nemusíte vůbec používat, pokud program budete ovládat pomocí menu a klikání myši. My se ho ale používat naučíme, protože je to šikovný pomocník: přehledně zaznamenává vaši práci, lze ji kdykoliv kontrolovat a zopakovat mj. na jiných datech a také tvoří most mezi PSPP a SPSS (vše ze syntaxu v PSPP by mělo fungovat v SPSS, naopak to ale pochopitelně neplatí).

#### Data editor

|    | id | sex | age | married | age_2009 | age_2010 | jobless | job_2009 | job_2010 | job_label |
|----|----|-----|-----|---------|----------|----------|---------|----------|----------|-----------|
| 1  | 1  | 1   | 3   | 0       | 5        | 4        | 1       | 1        | 1        | 1         |
| 2  | 2  | 2   | 2   | 0       | 3        | 4        | 1       | 1        | 1        | 0         |
| 3  | 3  | 3   | 5   | 0       | 3        | 3        | 1       | 0        | 0        | 1         |
| 4  | 4  | 4   | 5   | 0       | 3        | 0        | 2       | 0        | 0        | 1         |
| 5  | 5  | 5   | 3   | 0       | 3        | 3        | 1       | 1        | 1        | 1         |
| 6  | 6  | 4   | 4   | 0       | 3        | 3        | 1       | 1        | 1        | 0         |
| 7  | 7  | 7   | 3   | 0       | 5        | 4        | 2       | 1        | 0        | 0         |
| 8  | 8  | 4   | 4   | 0       | 3        | 3        | 1       | 0        | 1        | 1         |
| 9  | 9  | 10  | 4   | 0       | 4        | 4        | 1       | 0        | 0        | 0         |
| 10 | 12 | 5   | 5   | 0       | 4        | 4        | 1       | 1        | 0        | 0         |
| 11 | 13 | 3   | 3   | 0       | 4        | 4        | 1       | 0        | 0        | 0         |
| 12 | 14 | 3   | 3   | 0       | 4        | 3        | 2       | 1        | 0        | 0         |
| 13 | 15 | 2   | 1   | 0       | 5        | 5        | 1       | 0        | 0        | 0         |
| 14 | 16 | 4   | 4   | 0       | 4        | 2        | 1       | 0        | 0        | 0         |
| 15 | 19 | 1   | 1   | 0       | 3        | 3        | 0       | 0        | 1        | 1         |

#### Output



#### Syntax editor

```
*Syntax - PSPP Syntax Editor
*Akter 2009, ot.12 Do jaké míry Vás zajímají následující oblasti dějiny?
missing value q12a q12b q12c q12d q12e q12f q12g q12h q12i q12j q12k q12l
desc q12a to q12q.

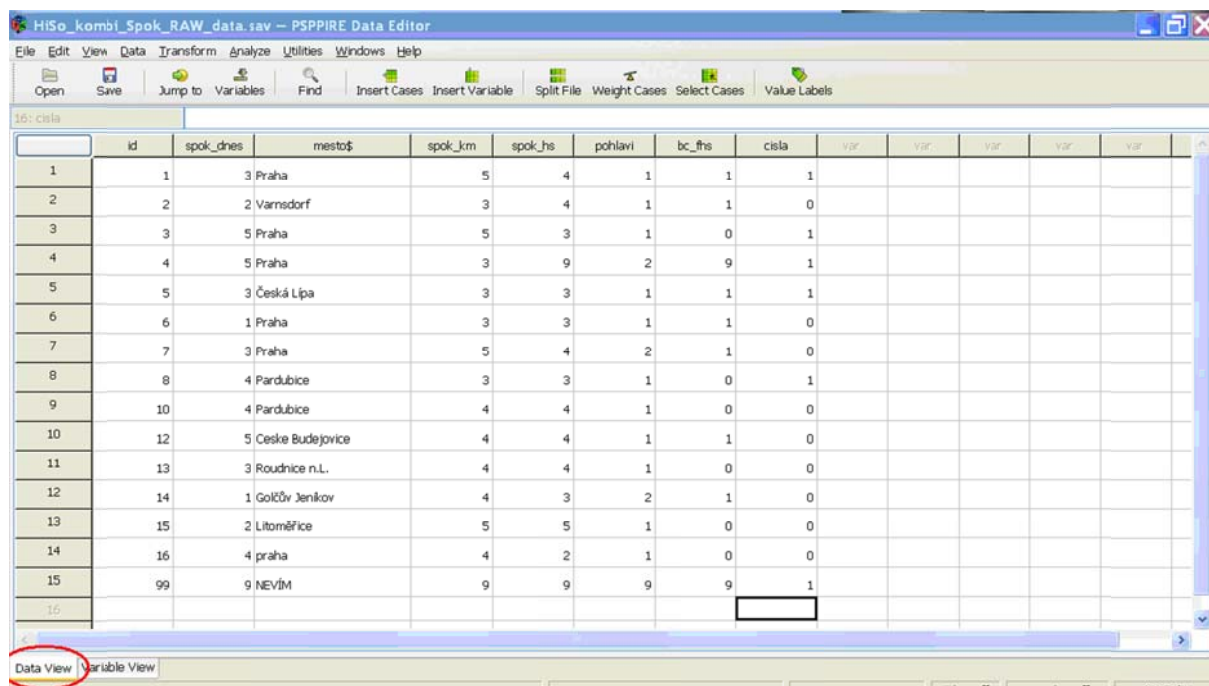
TITLE 'Akter 2009 q12: Do jaké míry Vás zajímají následující oblasti dějiny?'
FACTOR VARIABLES = q12a q12b q12c q12d q12e q12f q12g q12h q12i q12j q12k
/PRINT UNIVARIATE INITIAL CORRELATION EXTRACTION ROTATION
/EXTRACTION=APC
/CRITERIA = MINERGEN (1)
/PLOT = EIGEN
/METHOD = CORRELATION.
```

Na rozdíl od SPSS, kde můžete mít u každého typu okna otevřeno více oken (např. více outputů a mezi nimi přepínat, dokonce lze mít otevřeno i několik datových souborů současně), tak v PSPP to platí jen u okna syntaxu pro zápis příkazového řádku (můžete tak mít otevřený například jeden syntax pro úpravu dat a druhý pro samotné analýzy). Přepínání mezi okny je možné buď pomocí standardní klávesové zkratky levý Alt+Tab nebo přes menu v rubrice „Windows“ se zobrazí všechna aktuálně otevřená okna (to platí i pro SPSS, kde se to jmenuje „Window“).

## Data editor a datový soubor

Data editor má dvě záložky:

1. **Data view** – zde jsou v řádcích případy, ve sloupcích proměnné. Jde o datovou matici, zde vkládáme data



|    | id | spok_dnes | mesto\$          | spok_km | spok_hs | pohlavi | bc_fhs | cista |  |  |  |  |  |
|----|----|-----------|------------------|---------|---------|---------|--------|-------|--|--|--|--|--|
| 1  | 1  | 3         | Praha            | 5       | 4       | 1       | 1      | 1     |  |  |  |  |  |
| 2  | 2  | 2         | Varnsdorf        | 3       | 4       | 1       | 1      | 0     |  |  |  |  |  |
| 3  | 3  | 5         | Praha            | 5       | 3       | 1       | 0      | 1     |  |  |  |  |  |
| 4  | 4  | 5         | Praha            | 3       | 9       | 2       | 9      | 1     |  |  |  |  |  |
| 5  | 5  | 3         | Česká Lípa       | 3       | 3       | 1       | 1      | 1     |  |  |  |  |  |
| 6  | 6  | 1         | Praha            | 3       | 3       | 1       | 1      | 0     |  |  |  |  |  |
| 7  | 7  | 3         | Praha            | 5       | 4       | 2       | 1      | 0     |  |  |  |  |  |
| 8  | 8  | 4         | Pardubice        | 3       | 3       | 1       | 0      | 1     |  |  |  |  |  |
| 9  | 10 | 4         | Pardubice        | 4       | 4       | 1       | 0      | 0     |  |  |  |  |  |
| 10 | 12 | 5         | Ceske Budejovice | 4       | 4       | 1       | 1      | 0     |  |  |  |  |  |
| 11 | 13 | 3         | Roudnice n.L.    | 4       | 4       | 1       | 0      | 0     |  |  |  |  |  |
| 12 | 14 | 1         | Golčův Jeníkov   | 4       | 3       | 2       | 1      | 0     |  |  |  |  |  |
| 13 | 15 | 2         | Litoměřice       | 5       | 5       | 1       | 0      | 0     |  |  |  |  |  |
| 14 | 16 | 4         | praha            | 4       | 2       | 1       | 0      | 0     |  |  |  |  |  |
| 15 | 99 | 9         | NEVIM            | 9       | 9       | 9       | 9      | 1     |  |  |  |  |  |
| 16 |    |           |                  |         |         |         |        |       |  |  |  |  |  |

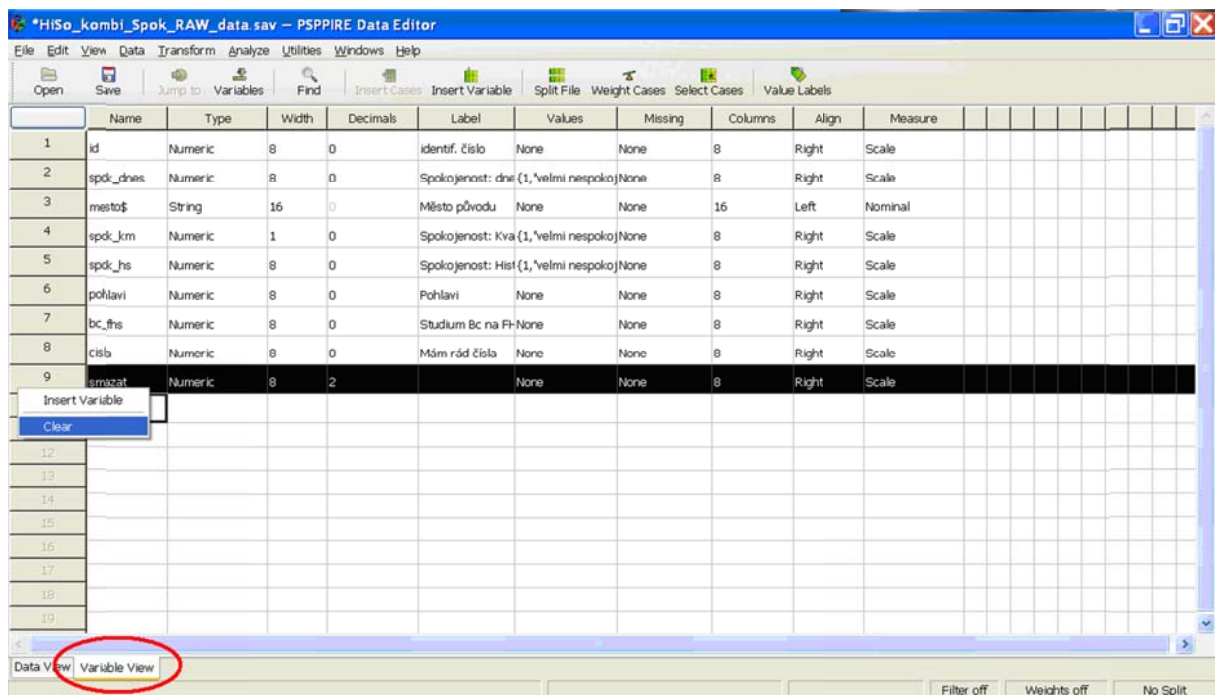
## 2. Variable view

Zde můžeme upravovat parametry proměnných: název – Name, typ, zobrazená šířka-počet znaků, label-pojmenování, labels kategorií hodnot, chybějící hodnoty – missing, zarovnání a úroveň měření – Measure.

Rovněž lze při označení proměnné (řádku) pomocí pravého tlačítka myši proměnnou smazat nebo vložit novou (prázdnou).

V tomto okně si většinou nejprve připravíme datovou matici pro nová data. Minimálně musíme vepsat do sloupce Name názvy proměnných (podle např. otázek v dotazníku), můžeme navíc i olabelovat jak proměnné tak jejich hodnoty (viz dále). V okně Data view se pak ve sloupcích objeví jména proměnných a již můžeme vkládat data – přepisovat například odpovědi z dotazníků.



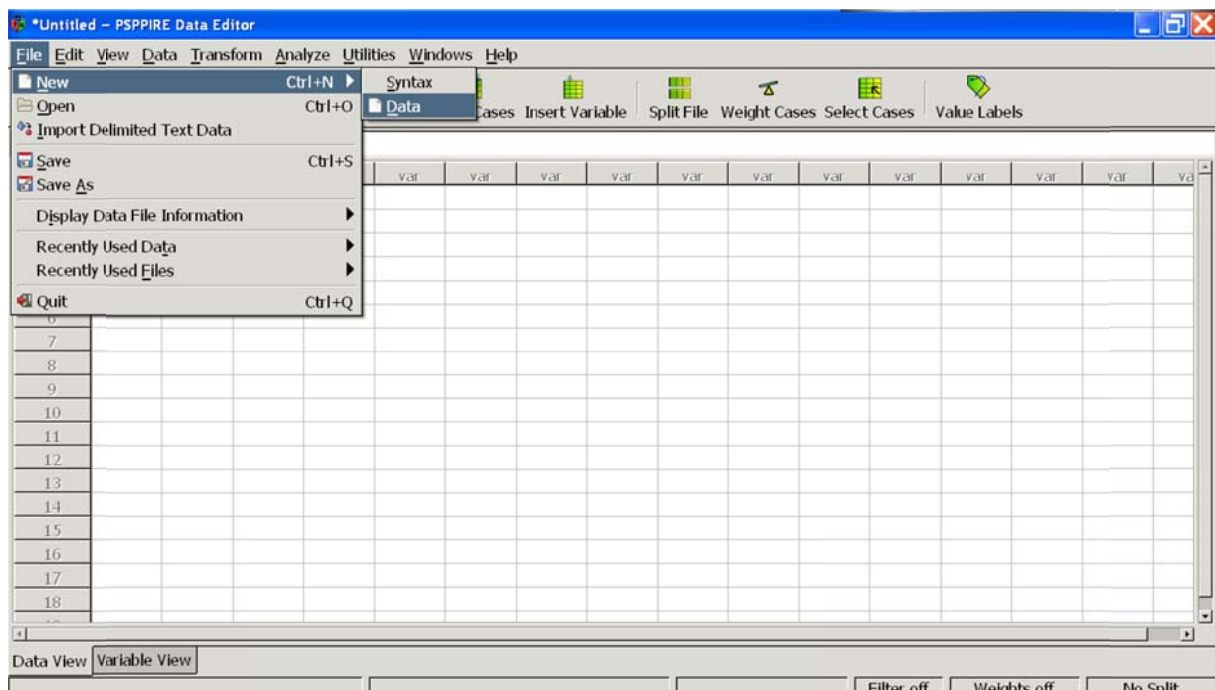


## Základní data management v PSPP/SPSS

### Vytvoření nové matice dat

Otevřete program PSPP (nebo SPSS). V menu klikněte na File → New → Data.

Pokud máte již otevřena jiná data, tak se ujistěte, že jsou uložena, protože ty jinak zmizí.



V záložce Variable View klikněte do prvního pole **Name** a zadejte název první proměnné. Zde je to **ID** – znak označující číslo dotazníku. Následuje proměnná q1 (např. první otázka v dotazníku), q2, vek, pohlavi atd.

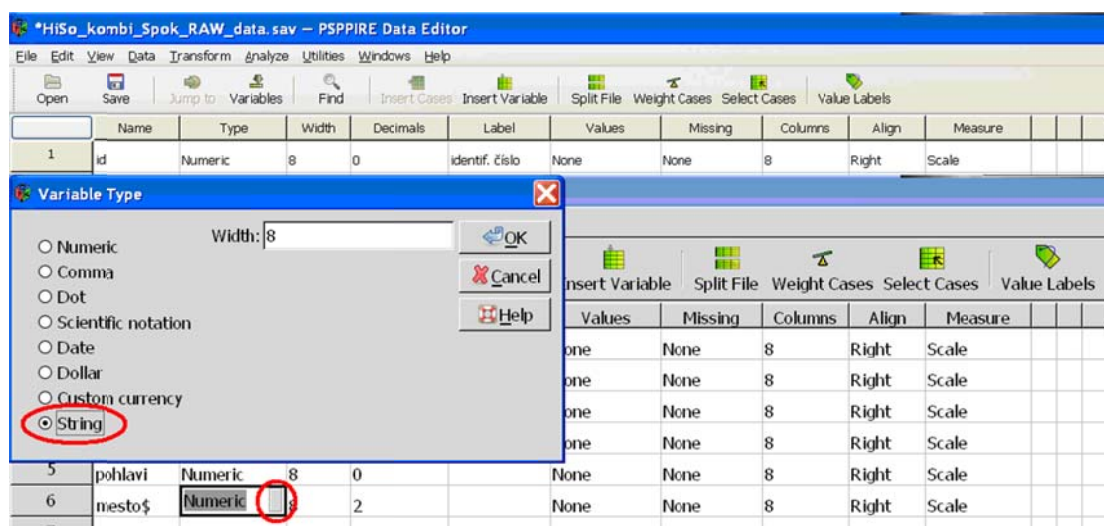
|   | Name    | Type    | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|---|---------|---------|-------|----------|-------|--------|---------|---------|-------|---------|
| 1 | ID      | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 2 | q1      | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 3 | q2      | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 4 | vek     | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 5 | pohlavi | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |

Poté co vždy napíšete nový název proměnné, tak se automaticky doplní její vlastnosti na přednastavené parametry. Ty je někdy třeba upravit, zejména pokud půjde o nominální – tzv. **stringový znak**. Jeho příkladem je zde proměnná mesto\$, která bude vkládána jako řetězec písmen (slova), nikoliv jako číslo.

Pozor: uvědomte si, že v datech **za čísly v kategoriích mohou být i nominální znaky**, kdy číslo pouze označuje nějakou kategorii (například pohlaví 1=muž, 2 =žena), tzn. nejde o numerický (kardinální či ordinální) znak v pravém slova smyslu!

Stringové proměnné používáme nejčastěji pro záznam odpovědí na otevřené otázky, protože často dopředu neznáme přesný kódový klíč, ve kterém bychom přiřadili kód=číslo, ke kterému bychom pak mohli přiřadit Value label. Zvykem bývá označovat je pomocí znaku \$ (ale to v některých jiných programech pak zlobí).

Volbu záznamu dat pomocí typu stringové proměnné (jsou zde i jiné typy záznamu) označíme kliknutím v políčku **Type** u příslušné proměnné na šedivý čtvereček. Objeví se menu, kde místo (přednastaveného) typu Numeric vybereme **String**.

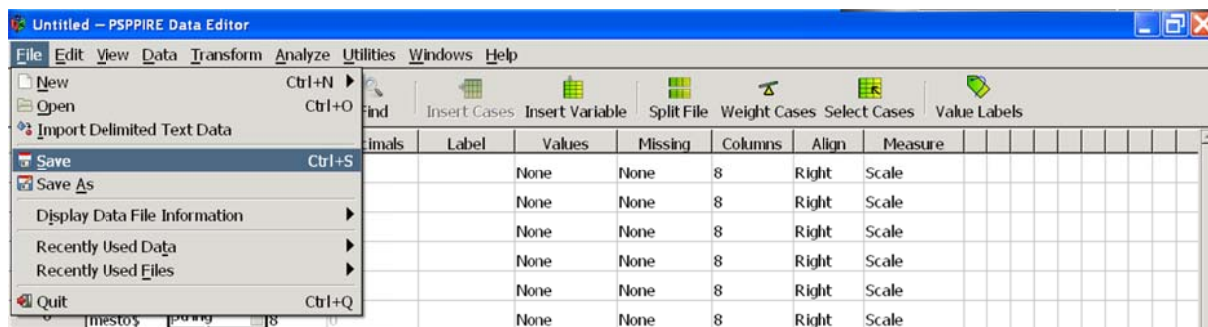


**Názvy proměnných** (tzv. VAR NAME) dělejte **pouze na 8 znaků** (i když PSSPP umí i více, ale to pak neotevřete v SPSS ve starších verzích než 11), **bez diakritiky, bez mezer** (to vám stejně ani nedovolí). Název vždy musí **začínat písmenem**, nikoliv číslem (odtud q1 pro otázku 1).

Poté co vytvoříte datovou matici, tak si ji uložte. Přípona pro datové soubory je „sav“. Pokud nezadáte název, tak se uloží pod automatickým názvem Untitled.sav. Při vkládání dat pak,



**ukládejte průběžně**, zamezíte ztrátě času, pokud vám program spadne (nebo vypnou proud). Průběžné ukládání dat platí v PSPP dvojnásobně, protože program je někdy nestabilní a může spadnout.



## Označování popisků proměnných

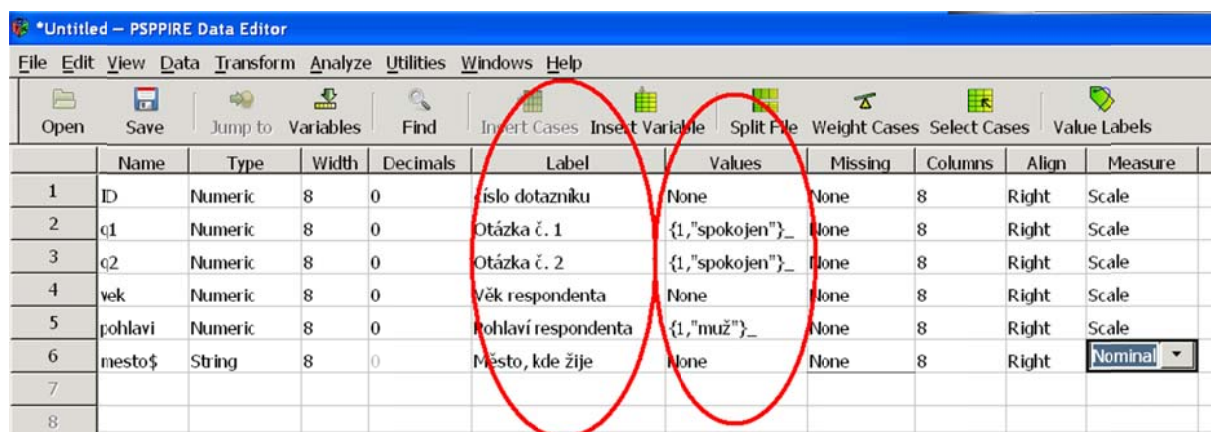
Krátká rekapitulace: v PSSP / SPSS rozlišujeme tři informace o proměnných:

|                 |           |   |
|-----------------|-----------|---|
| jméno proměnné  | VAR NAME  | max. 8 znaků bez diakritiky (např. pohlavi)   |
| popis proměnné  | VAR LABEL | Delší popis znaku, lze s diakritikou, např. (zkrácené) znění otázky v dotazníku, který se bude objevovat ve výstupech (např. Pohlaví respondenta) |
| popis kategorií | VAL LABEL | Popis kategorií znaku, které se budou objevovat ve výstupech (např. 1 = muž, 2 = žena)  |

Jméno proměnné je povinné, pokud ho ne zadáte nemůžete dál (v SPSS se sám vygeneruje univerzální název VAR0001 atd.).

Popis proměnné a jejích kategorií je už pak nepovinný, záleží na vás, zda si je zadáte nebo ne. Lze je zadat i později po porřízení dat. Umožní vám lepší orientaci v datech a bude se automaticky objevovat ve výstupech.

Zadání labelů se provádí v příslušném sloupci nebo lépe pomocí syntaxu (o tom později). U popisu proměnné píšeme přímo do kolonky **Label**.



|   | Name    | Type    | Width | Decimals | Label               | Values          | Missing | Columns | Align | Measure |
|---|---------|---------|-------|----------|---------------------|-----------------|---------|---------|-------|---------|
| 1 | ID      | Numeric | 8     | 0        | číslo dotazníku     | None            | None    | 8       | Right | Scale   |
| 2 | q1      | Numeric | 8     | 0        | Otázka č. 1         | {1,"spokojen"}_ | None    | 8       | Right | Scale   |
| 3 | q2      | Numeric | 8     | 0        | Otázka č. 2         | {1,"spokojen"}_ | None    | 8       | Right | Scale   |
| 4 | vek     | Numeric | 8     | 0        | Věk respondenta     | None            | None    | 8       | Right | Scale   |
| 5 | phlavi  | Numeric | 8     | 0        | Pohlaví respondenta | {1,"muž"}_      | None    | 8       | Right | Scale   |
| 6 | mesto\$ | String  | 8     | 0        | Město, kde žije     | None            | None    | 8       | Right | Nominal |
| 7 |         |         |       |          |                     |                 |         |         |       |         |
| 8 |         |         |       |          |                     |                 |         |         |       |         |

V případě popis kategorií (**Values**) se klikne na šedivý čtvereček v příslušné buňce a objeví se menu, kde napíšeme hodnotu - Value, např. 1 a vyplníme Value label, např. muž. Pak je třeba kliknout na tlačítko **Add**. Změnit popis lze kliknutím na příslušnou kategorii, pak je třeba kliknout na **Apply**. Smazání popisu kategorie provedeme pomocí **Remove**.

Poznámka: Olabelované nemusí být všechny kategorie hodnot znaku (u ordinálních znaků často stačí jen ty krajní, u pětibodové škály např. 1=spokojen 5=nespokojen).

Value Labels

Value: 2

Value Label: žena

+

Add

✓

Apply

−

Remove

1 = "muž"

OK

Cancel

Help

Windows

Help

Insert Cases

Insert Variable

Split File

Weight Cases

Select Cases

Value Labels

| Label               | Values          | Missing | Columns | Align | Measure |
|---------------------|-----------------|---------|---------|-------|---------|
| číslo dotazníku     | None            | None    | 8       | Right | Scale   |
| Otázka č. 1         | None            | None    | 8       | Right | Scale   |
| Otázka č. 2         | {1,"spokojen"}_ | None    | 8       | Right | Scale   |
| Věk respondenta     | None            | None    | 8       | Right | Scale   |
| Pohlaví respondenta | {1,"muž"}_      | None    | 8       | Right | Scale   |
| Město, kde žije     | None            | None    | 8       | Right | Nominal |

Pokud budete data přenášet do jiného programu, který není specializovaný na statistickou analýzu dat, např. Excel, tak labels (proměnných i jejich kategorií) se ztratí, ale jméno proměnné vždy je (musí zůstat).

Například v Excelu by to vypadalo takto (nejsou tam labels ani pro název proměnné ani pro jejich kategorie).

|   | A  | B  | C  | D   | E  | F     | G       |
|---|----|----|----|-----|----|-------|---------|
| 1 | ID | q1 | q2 | vek | po | hlavi | mesto\$ |
| 2 | 1  | 1  | 4  | 63  | 2  | Praha |         |
| 3 | 2  | 1  | 2  | 18  | 1  | Brno  |         |
| 4 | 3  | 4  | 3  | 45  | 2  | Plzeň |         |

## Vkládání dat

Poté co jste si zadali datovou matici (nebo alespoň její část), můžete přistoupit k ukládání samotných dat (tedy nejčastěji dotazníků) a to v záložce **Data View**. Postupně do buněk v jednom řádku vkládáme hodnoty proměnných (jejich jména vidíme v prvním šedivém řádku) pro jeden případ – nejčastěji jde o jeden dotazník od jednoho respondenta. U stringových proměnných vepisujeme znaky (slova).

|   | ID | q1 | q2 | vek | po | hlavi | mesto\$ | var | var | var | var | var |
|---|----|----|----|-----|----|-------|---------|-----|-----|-----|-----|-----|
| 1 | 1  | 2  | 3  | 63  | 2  | Praha |         |     |     |     |     |     |
| 2 |    |    |    |     |    |       |         |     |     |     |     |     |
| 3 |    |    |    |     |    |       |         |     |     |     |     |     |

Podobně jako ve správě proměnných, řádek – jeden dotazník můžeme vymazat označením do bloku a menu myši pravým tlačítkem a Clear.

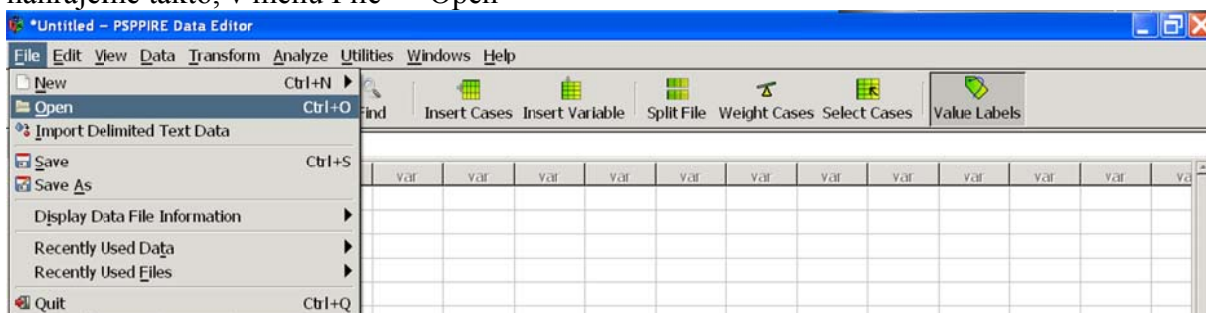
|   | ID | q1 | q2 | vek | po | hlavi | mesto\$ | var | var | var | var | var |
|---|----|----|----|-----|----|-------|---------|-----|-----|-----|-----|-----|
| 1 | 1  | 1  | 2  | 63  | .  |       |         |     |     |     |     |     |
| 2 | 2  | .  | .  | 22  | .  |       |         |     |     |     |     |     |
| 3 |    |    |    | 555 | 7  | nikde |         |     |     |     |     |     |

## Soubor s daty si průběžně ukládáme!

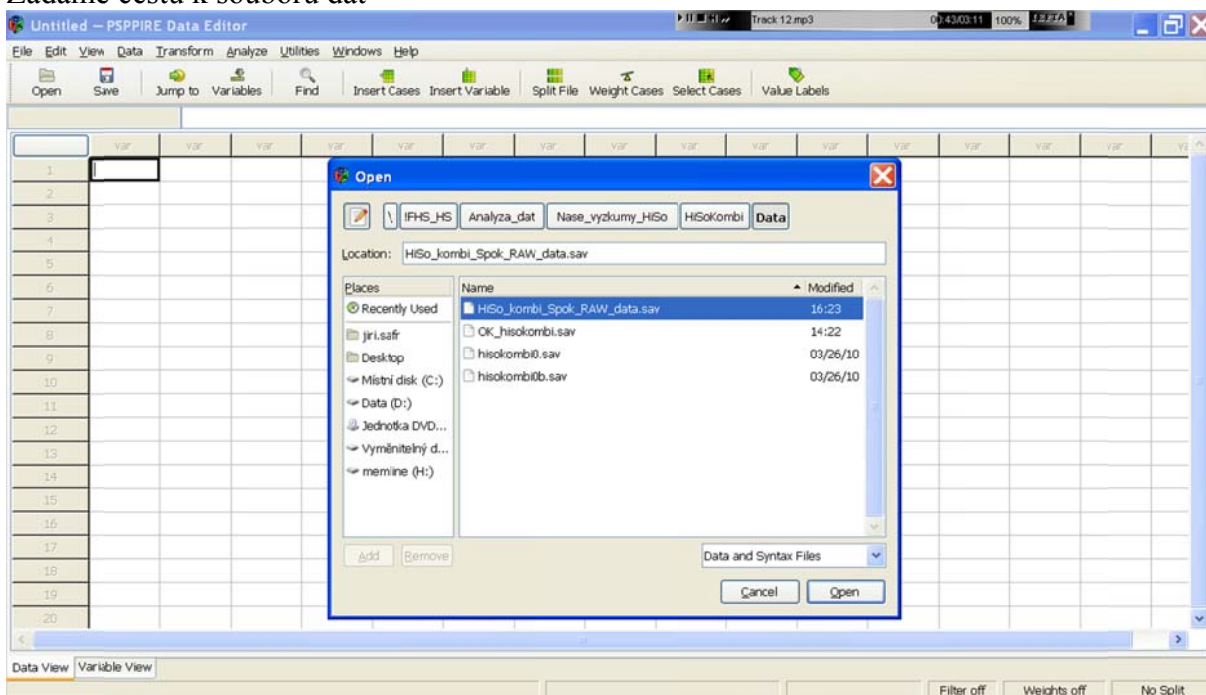
Poznámka: data lze také vložit ze syntax, ale o tom později.

## Otevření již existujícího souboru dat

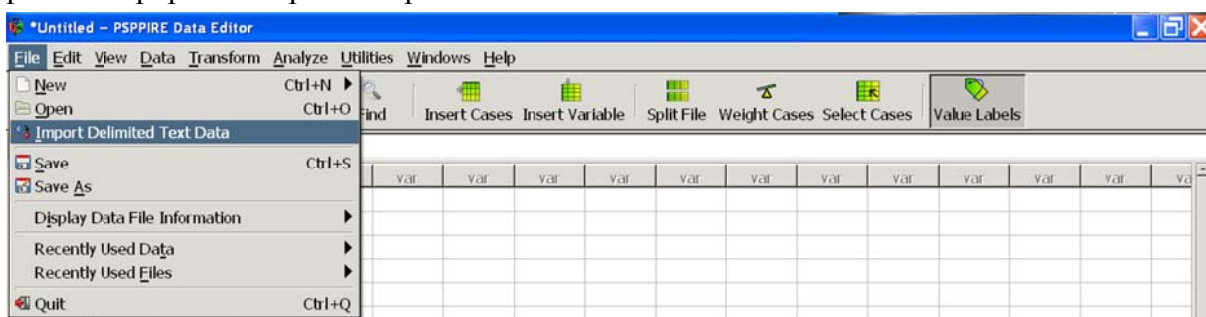
Nejčastěji máme data ve formátu SPSS (přípona „sav“). Již dříve vytvořený soubor dat nahrajeme takto, v menu File → Open



Zadáme cestu k souboru dat



Alternativně můžeme importovat též soubory v univerzálním formátu CSV (viz dále), kdy jsou data vytvořena v nějakém jiném programu (např. Excelu). Tento formát v sobě většinou nenese informace o labelech. Použijeme v menu **Import Delimited Text Data**. Postup je podrobně popsán v kapitole Import CSV dat do PSPP.



Nyní máte data ze svého vlastního výzkumu uložena nebo případně z nějakého jiného zdroje otevřena. Také je máte víceméně již „olabelovaná“, tj. máte přiřazeny popisky u názvů

proměnných (VARIABLE LABEL), to bývá nejčastěji zjednodušené znění otázek z dotazníku a také máte popisky u hodnot znaků (například názvy kategorií odpovědí). Pokud labely ještě nemáte, nevadí, dají se doplnit průběžně později, hlavní je v datech se vyznat (co která proměnná a její hodnota znamená), k čemuž poslouží i separátně vytvořený codebook – dokumentace k datům (ten se ale dá jednoduše během vteřiny vytvořit z již olabelovaných dat).

Uvědomte si, že jde zatím jen o „hrubá – nevyčištěná“ data. Než přikročíte k jejich věcné analýze, tak nejprve musíte data zkontrolovat, což většinou znamená, že je v prvním kroku bude třeba takzvaně „vyčistit“ a případně i nějak ještě upravit, například vytvořit umělé – syntetické proměnné. (u převzatých dat to pochopitelně tak být nemusí, ale zkontrolovat je musíme stejně vždycky).

Prvním krokem při čištění dat je zjištění jaké se v datech objevuje rozmezí hodnot u všech znaků (například odpovídají možným hodnotám v dotazníku?). Pokud se jedná o nějakou systematickou chybu či výpadek při samotném zadávání dat, bude třeba data manuálně v data editoru nejprve opravit. I když jsou hodnoty „v pořádku“, přesto nemusí zcela odpovídat hodnotám, které budeme chtít analyzovat, tj. platné hodnoty. Příkladem jsou odpovědi „Nevím“ nebo „Neodpověděl“. Takovéto hodnoty znaku, které nebudeme dále chtít věcně analyzovat, ale přesto je chceme v datech zachovat, můžeme v datech označit jako chybějící, tj. „missing“. (A také bychom kvalitu dat v tomto ohledu měli nejprve zkontrolovat; o tom „kdo nám neodpovídá“ a zda to má nějaké důsledky pro reprezentativitu výsledků si ale povíme na jiném místě).



## Označení chybějících hodnot – missing values

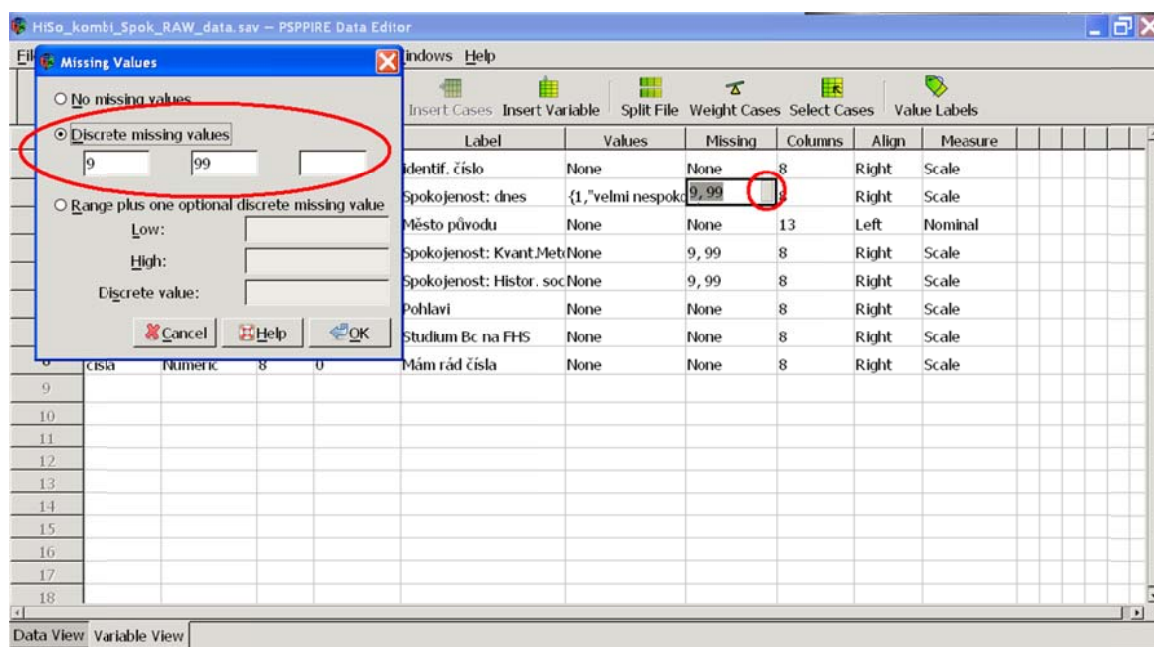
Chybějící hodnota (missing value) je taková hodnota znaku, u které nechceme, aby byla součástí běžné analýzy. Rozlišujeme **systémově definovanou chybějící hodnotu (SYMIS)**, která se vytvoří automaticky, pokud k proměnné nevložíme u daného případu-respondenta žádnou hodnotu (políčko datech je prázdné, resp. je v něm tečka) a **uživatelsky definovanou chybějící hodnotu (MISSING VALUE)**, kdy označíme určité hodnoty za neplatné.

Programy PSPP a SPSS umožňují vedle systémově definovaných chybějících hodnot (SYMIS) definovat i **tři hodnoty uživatelské**, což je výhodné, protože se objevují ve výstupech samostatně (i původním rozlišením a labelem), což nám umožňuje kontrolovat jejich závislost na jiných znacích (např. zda na určitou otázku odmítají či neumí odpovědět určité kategorie respondentů).

U uživatelských missing values se nejčastěji jedná o kategorie odpovědí jako „neodpověděl“ nebo „neví“, pro něž je zvykem označovat je pomocí hodnot, které jsou již na první pohled mimo rozpětí platných kategorií (odpovědí). Například u čtyřbodové škály souhlasu může uživatelsky definovaný missing představovat hodnota 8 = neví a 9 = neodpověděl/odmítl odpovědět (pro znaky s více číslicemi by to bylo, např. u věku 999). Tyto hodnoty nebudou pak figurovat v analýzách, resp. budou (např. u frekvencí) zobrazeny odděleně. Jde-li o kardinální znak (či ordinální škálu), nebudou tyto hodnoty zahrnuty například do výpočtu průměru či směrodatné odchylky.

Poznámka: správně bychom měli rozlišovat mezi „prázdným políčkem“ v datech (= systémově definovaná chybějící hodnota, SYMIS, která mohla vzniknout i chybou při ukládání dat) a určitou hodnotou označující, že respondent „neodpověděl“ (např. 9). V praxi se však toto rozlišení většinou neprovádí.

V data editoru na záložce Variable View provedeme označení určité hodnoty za missing ve sloupci **Missing**. Objeví se menu, kde můžeme vyjmenovat až 3 hodnoty, které chceme z analýz vynechat, alternativně lze zadat dva rozsahy (do/od nejnižší a/nebo nejvyšší hodnoty a jednu diskretní hodnotu).

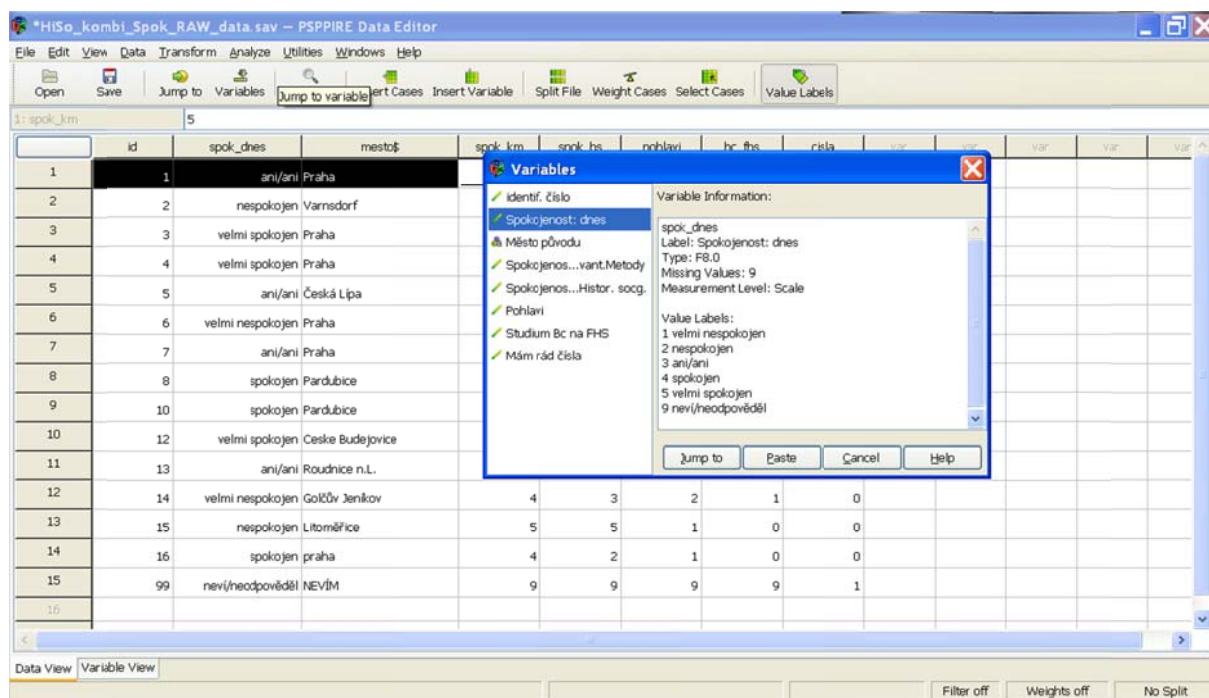


Optimální je zadávat chybějící hodnoty pomocí Syntaxu (bude popsáno v další části).

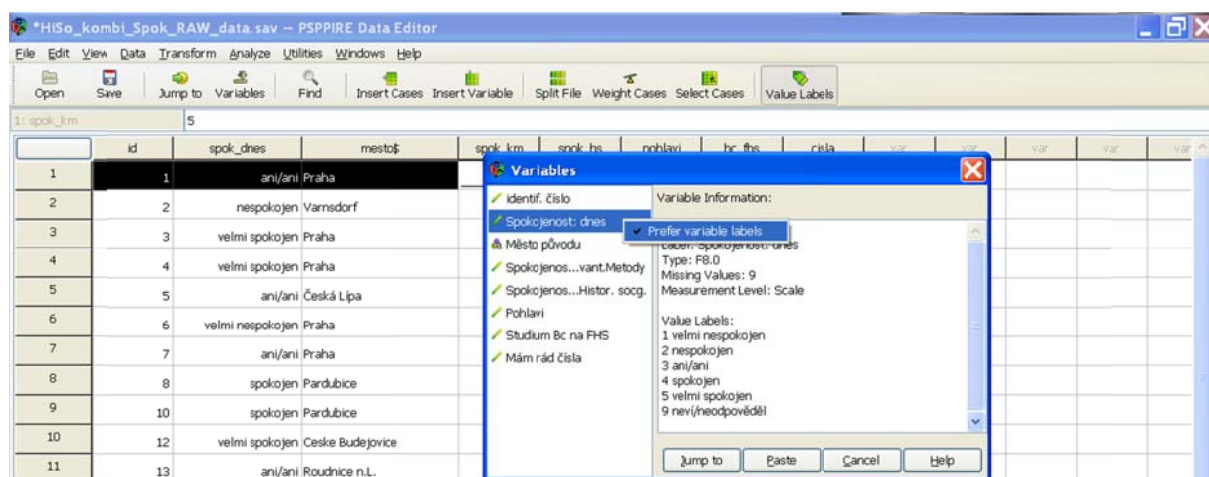
Hodnoty Missing Values si můžeme také průběžně pro určité analýzy „zapínat/vypínat“. Lze tak například efektivně selektovat určitý rozsah kategorií znaku. Např. označením hodnot nižších než 21 u proměnné věk, vyfiltrujeme respondenty mladší než 21 let, aniž bychom je vymazaly, či jinak nenávratně transformovaly původní proměnnou.

## Ovládání programu a seznam poměných

Krom standardní cesty skrze menu (nápis v šedivém pruhu: File / Edit / View / Data atd.) lze využívat zkratk pomocí ikoněk. Například velmi užitečná je funkce Variables, která umožňuje: přesun na danou proměnnou v datech (**Jump**), ukazuje její parametry: úroveň měření, hřebějící hodnoty (missingy) a labels. Také když kliknete na **Paste**, tak vám její název vloží do syntaxu (bohužel na rozdíl od SPSS, do pokaždé nového syntaxu, takže pokud ji chcete použít v rozepsaném syntaxu, musíte si ji tam ještě zkopírovat a tento nový syntax zavřít).



Standardně je v PSPP bohužel nastaveno zobrazení Variable labels, nikoliv **Variable names**, což bývá přehlednější (protože nejčastěji proměnné označujeme podle čísla otázky v dotazníku např., q1, q2, ...). Sice PSPP na rozdíl od SPSS nemá Options, kde by se to dalo stabilně nastavit, ale na jména proměnných lze přepnout pomocí pravého tlačítka myši, když najedete na jakoukoliv proměnnou (bohužel při otevírání okna to musíte udělat pokaždé znovu). To platí i pro okýnka při analýzách v menu Analyze.



|    | id | spok_dnes                       | mesto\$ | spok_km | spok_hs | pohlavi | bc_fhs | cisla | var | var | var |
|----|----|---------------------------------|---------|---------|---------|---------|--------|-------|-----|-----|-----|
| 1  | 1  | ani/ani Praha                   |         |         |         |         |        |       |     |     |     |
| 2  | 2  | nespokojen Varnsdorf            |         |         |         |         |        |       |     |     |     |
| 3  | 3  | velmi spokojen Praha            |         |         |         |         |        |       |     |     |     |
| 4  | 4  | velmi spokojen Praha            |         |         |         |         |        |       |     |     |     |
| 5  | 5  | ani/ani Česká Lipa              |         |         |         |         |        |       |     |     |     |
| 6  | 6  | velmi nespokojen Praha          |         |         |         |         |        |       |     |     |     |
| 7  | 7  | ani/ani Praha                   |         |         |         |         |        |       |     |     |     |
| 8  | 8  | spokojen Pardubice              |         |         |         |         |        |       |     |     |     |
| 9  | 10 | spokojen Pardubice              |         |         |         |         |        |       |     |     |     |
| 10 | 12 | velmi spokojen Ceske Budejovice |         |         |         |         |        |       |     |     |     |
| 11 | 13 | ani/ani Roudnice n.L.           |         |         |         |         |        |       |     |     |     |

| Variables |                          |
|-----------|--------------------------|
| id        | Variable Information:    |
| spok_dnes | spok_dnes                |
| mesto\$   | Label: Spokojenost: dnes |
| spok_km   | Type: F8.0               |
| spok_hs   | Missing Values: 9        |
| pohlavi   | Measurement Level: Scale |
| bc_fhs    | Value Labels:            |
| cisla     | 1 velmi nespokojen       |
|           | 2 nespokojen             |
|           | 3 ani/ani                |
|           | 4 spokojen               |
|           | 5 velmi spokojen         |
|           | 9 nev/neodpověděl        |

V datovém editoru – v záložce Data View si můžete přepnout mezi zobrazením hodnot čísel kategorií a nebo popisků → ikonka Value Labels.

| *HiSo_kombi_Spok_RAW_data.sav – PSPPIRE Data Editor          |      |                    |           |        |              |                 |            |              |              |                                     |
|--|------|--------------------|-----------|--------|--------------|-----------------|------------|--------------|--------------|-------------------------------------|
| File Edit View Data Transform Analyze Utilities Windows Help |      |                    |           |        |              |                 |            |              |              |                                     |
| Open   | Save | Jump to            | Variables | Find   | Insert Cases | Insert Variable | Split File | Weight Cases | Select Cases | Value Labels Show/hide value labels |
| 2: id  |      |                    |           |        |              |                 |            |              |              |                                     |
|  | id   | spo_dnes           | mesto\$   | spo_KM | spo_HiSo     | pohlavi         | bc_fhs     | cisla        | var          | var                                 |
| 1  | 1    | 3 Praha            | 5         | 4      | 1            | 1               | 1          |              |              |                                     |
| 2  | 2    | 2 Varnsdorf        | 3         | 4      | 1            | 1               | 0          |              |              |                                     |
| 3  | 3    | 5 Praha            | 5         | 3      | 1            | 0               | 1          |              |              |                                     |
| 4  | 4    | 5 Praha            | 3         | 9      | 2            | 9               | 1          |              |              |                                     |
| 5  | 5    | 3 Česká Lipa       | 3         | 3      | 1            | 1               | 1          |              |              |                                     |
| 6  | 6    | 1 Praha            | 3         | 3      | 1            | 1               | 0          |              |              |                                     |
| 7  | 7    | 3 Praha            | 5         | 4      | 2            | 1               | 0          |              |              |                                     |
| 8  | 8    | 4 Pardubice        | 3         | 3      | 1            | 0               | 1          |              |              |                                     |
| 9  | 10   | 4 Pardubice        | 4         | 4      | 1            | 0               | 0          |              |              |                                     |
| 10   | 12   | 5 Ceske Budejovice | 4         | 4      | 1            | 1               | 0          |              |              |                                     |
| 11   | 13   | 3 Roudnice n.L.    | 4         | 4      | 1            | 0               | 0          |              |              |                                     |
| 12   | 14   | 1 Golčův Jeníkov   | 4         | 3      | 2            | 1               | 0          |              |              |                                     |
| 13   | 15   | 2 Litoměřice       | 5         | 5      | 1            | 0               | 0          |              |              |                                     |
| 14   | 16   | 4 praha            | 4         | 2      | 1            | 0               | 0          |              |              |                                     |
| 15   | 99   | 9 NEVÍM            | 9         | 9      | 9            | 9               | 1          |              |              |                                     |

| *HiSo_kombi_Spok_RAW_data.sav – PSPPIRE Data Editor          |      |                                 |                |                |              |                 |            |              |              |              |
|--|------|---------------------------------|----------------|----------------|--------------|-----------------|------------|--------------|--------------|--------------|
| File Edit View Data Transform Analyze Utilities Windows Help |      |                                 |                |                |              |                 |            |              |              |              |
| Open   | Save | Jump to                         | Variables      | Find           | Insert Cases | Insert Variable | Split File | Weight Cases | Select Cases | Value Labels |
| 1: id  |      |                                 |                |                |              |                 |            |              |              |              |
|  | id   | spo_dnes                        | mesto\$        | spo_KM         | spo_HiSo     | pohlavi         | bc_fhs     | cisla        | var          | var          |
| 1  | 1    | ani/ani Praha                   | relmi spokojen | spokojen       | muž          | ano             | ano        |              |              |              |
| 2  | 2    | nespokojen Varnsdorf            | ani/ani        | spokojen       | muž          | ano             | ne         |              |              |              |
| 3  | 3    | velmi spokojen Praha            | relmi spokojen | ani/ani        | muž          | ne              | ano        |              |              |              |
| 4  | 4    | velmi spokojen Praha            | ani/ani        | neodpověděl    | žena         | 9               | ano        |              |              |              |
| 5  | 5    | ani/ani Česká Lipa              | ani/ani        | ani/ani        | muž          | ano             | ano        |              |              |              |
| 6  | 6    | velmi nespokojen Praha          | ani/ani        | ani/ani        | muž          | ano             | ne         |              |              |              |
| 7  | 7    | ani/ani Praha                   | relmi spokojen | spokojen       | žena         | ano             | ne         |              |              |              |
| 8  | 8    | spokojen Pardubice              | ani/ani        | ani/ani        | muž          | ne              | ano        |              |              |              |
| 9  | 10   | spokojen Pardubice              | spokojen       | spokojen       | muž          | ne              | ne         |              |              |              |
| 10   | 12   | velmi spokojen Ceske Budejovice | spokojen       | spokojen       | muž          | ano             | ne         |              |              |              |
| 11   | 13   | ani/ani Roudnice n.L.           | spokojen       | spokojen       | muž          | ne              | ne         |              |              |              |
| 12   | 14   | velmi nespokojen Golčův Jeníkov | spokojen       | ani/ani        | žena         | ano             | ne         |              |              |              |
| 13   | 15   | nespokojen Litoměřice           | relmi spokojen | relmi spokojen | muž          | ne              | ne         |              |              |              |
| 14   | 16   | spokojen praha                  | spokojen       | nespokojen     | muž          | ne              | ne         |              |              |              |
| 15   | 99   | nevi/neodpověděl NEVÍM          | í/neodpověděl  | í/neodpověděl  | 9            | 9               | ano        |              |              |              |

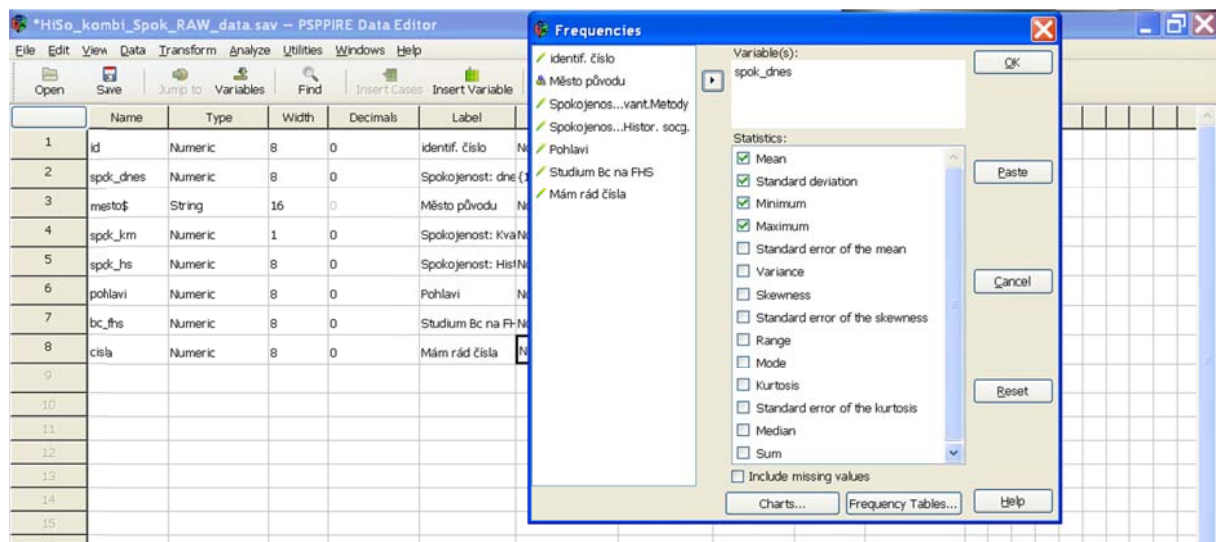
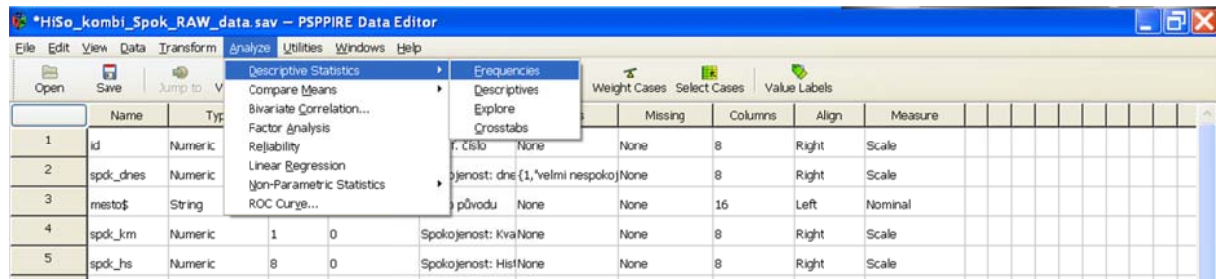
Poznámka: Proměnná mesto\$ je stringová proto se zobrazují jména města jako hodnoty.

## Analýza dat (1): popisné statistiky, třídění dat 1.stupně

### Popisné statistiky pomocí FREQUENCIES

#### 1. FREQUENCIES Třídění prvního stupně: jednoduchá tabulka frekvencí

V menu: Analyze → Descriptive statistics → Frequencies



Ze seznamu proměnných (okýnko vlevo) přenést pomocí šipky proměnnou (či více proměnných) do okýnka **Variable(s)** napravo. Ve **Statistics** lze nastavit, co chceme v analýze. Prednastaveno máme průměr, směrodatnou odchylku, minimum, maximum. Vytvořit můžeme také grafy (viz dále).

The screenshot shows the PSPPIRE Output Viewer. The output for the 'Frequencies' analysis is displayed. It includes a table of frequencies and a summary of statistics.

| Value Label      | Value | Frequency | Percent      | Valid Percent | Cum Percent |
|------------------|-------|-----------|--------------|---------------|-------------|
| velmi nespokojen | 1     | 2         | 13,33        | 14,29         | 14,29       |
| nespokojen       | 2     | 2         | 13,33        | 14,29         | 28,57       |
| ani/ani          | 3     | 4         | 26,67        | 28,57         | 57,14       |
| spokojen         | 4     | 3         | 20,00        | 21,43         | 78,57       |
| velmi spokojen   | 5     | 3         | 20,00        | 21,43         | 100,00      |
| nevi/neodpověděl | 9     | 1         | 6,67         | Missing       |             |
| <b>Total</b>     |       | <b>15</b> | <b>100,0</b> | <b>100,0</b>  |             |

| N       | Valid | Missing |
|---------|-------|---------|
| 14      |       | 1       |
| Mean    | 3,21  |         |
| Std Dev | 1,37  |         |
| Minimum | 1,00  |         |
| Maximum | 5,00  |         |

Vidíme rozložení kategorie odpovědí pro proměnnou spok\_dnes (Jak jste dnes spokojen/a?):

Value = číslo kategorie

Frequency = absolutní četnost

Percent = procenta ze všech případů (tj. včetně případných missingů)

Valid percent = procenta platných odpovědí, kdy jsou missingy vyřazeny (je třeba si je ale nejprve označit!)

Cum percent = kumulativní procento (součet % kategorií nad danou hodnotou)

Total = celkem

Druhá tabulka uvádí souhrnné statistiky: počet platných a chybějících hodnot (*Valid*, *Missing*) průměrnou hodnotu *Mean* (což dává smysl pouze u kardinálních znaků; případně i ordinálních, jako je tomu zde), směrodatnou odchylku průměru *Std Dev* („rozptýlení“ odpovědí od průměru) a v souboru se vyskytující minimální a maximální hodnotu (*Minimum*, *Maximum*).

Jak s výsledky v okně výstupů - Outputu dále pracovat najdete v kapitole Možnosti úprav výstupu (Output) z PSPP.

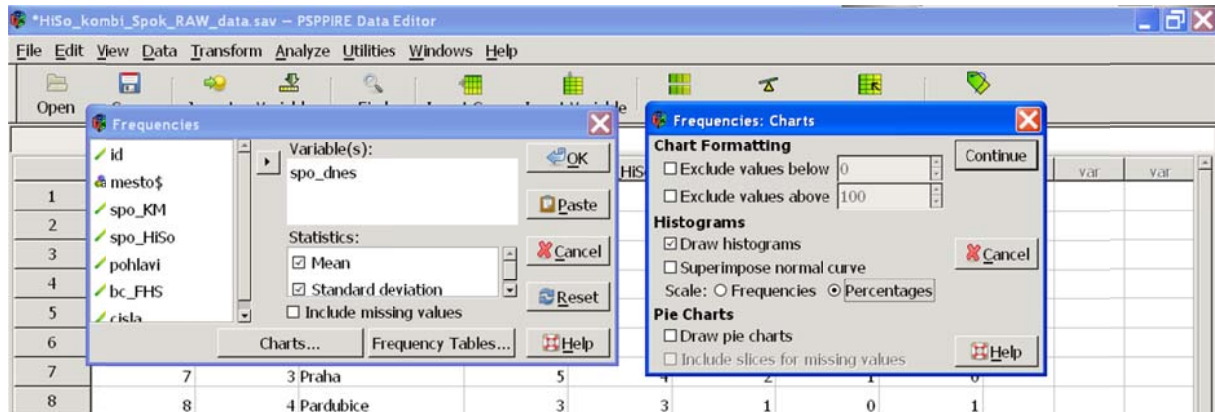
Třídění dat druhého stupně, tj. vzájemné souvislosti dvou (kategoriálních) znaků se věnuje kapitola Logika vytváření kontingenčních tabulek a jejich interpretace.



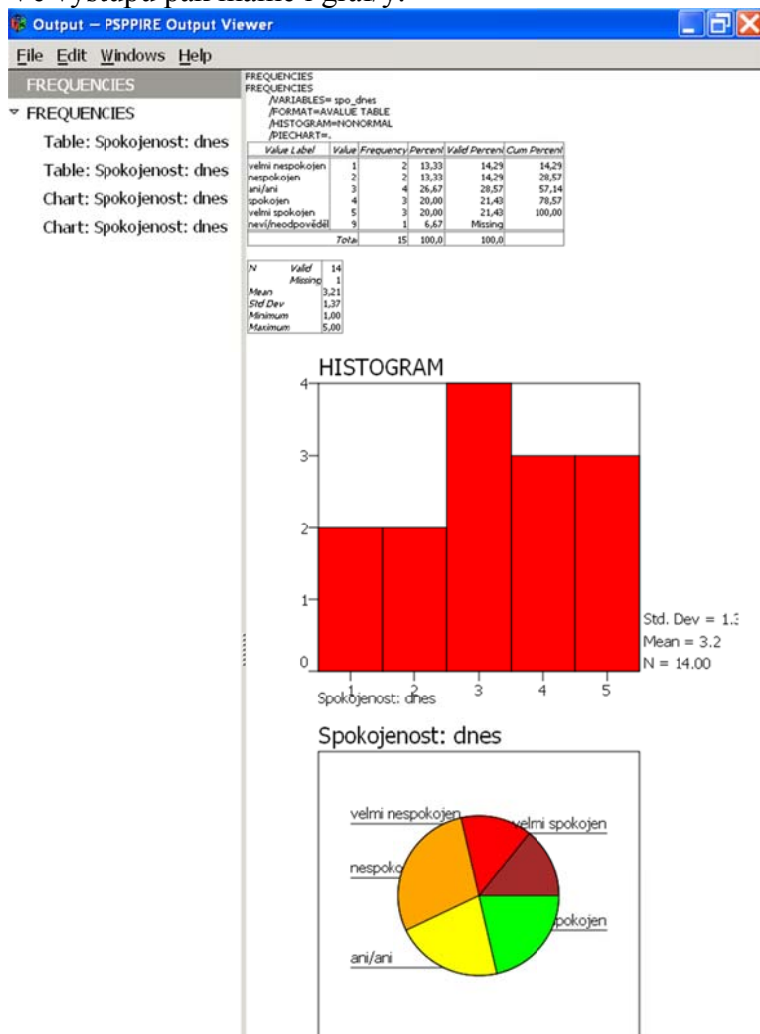
## Grafy (1): třídění dat 1. stupně

### Zadání grafů v příkazu Frequencies (sloupcový, koláčový a histogram)

V předcházejícím příkladu lze zadat také grafy v sekci **Charts**, např. Histogram s procenty.

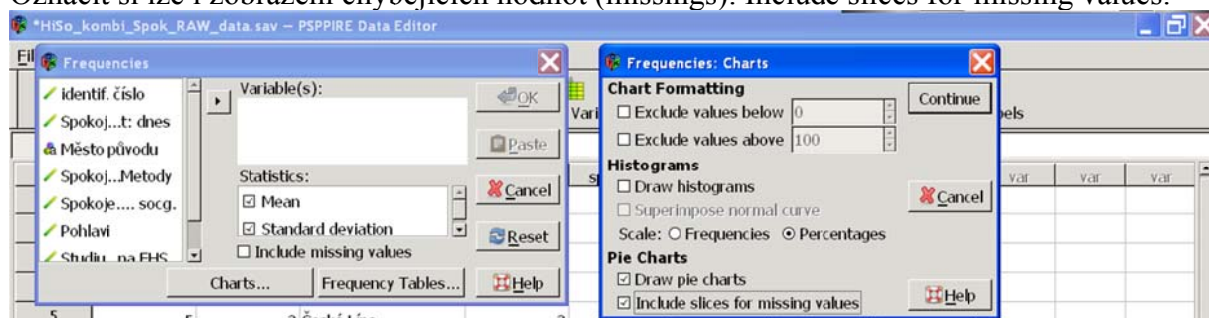


Ve výstupu pak máme i graf/y:

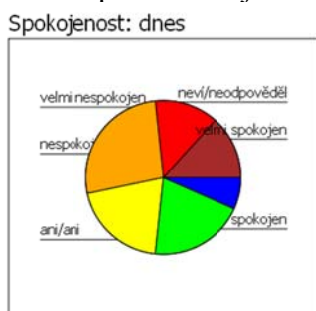


Pozor Histogram ukazuje frekvence, zatímco Pie chart (koláčový graf) a Barchart (sloupcový graf) může uvádět jak počet případů tak i procenta (zadání Percentages). (viz dále rozdíl mezi Histogramem a Sloupcovým grafem v %)

Označit si lze i zobrazení chybějících hodnot (missings): Include slices for missing values.



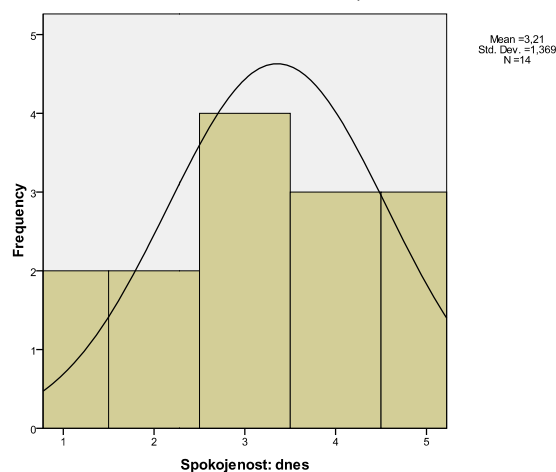
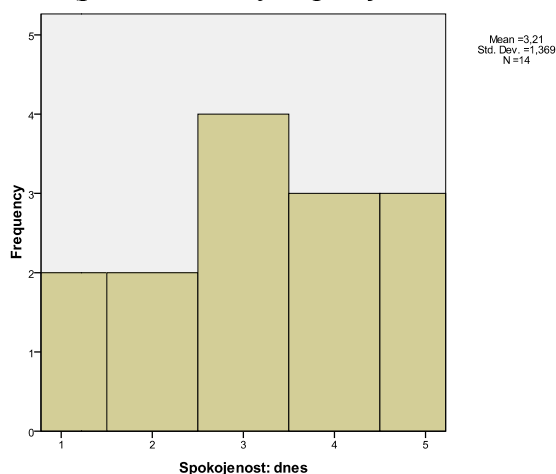
výsledek pak obsahuje i kategorii „neví/neodpověděl“ definovanou jako missing:



## Sloupcový graf a histogram

**Histogram** (graf znázorňující četnost)

Histogram je sloupcový graf, v němž každé kategorii přiřadíme její četnost (nejde tedy o procenta!). Zobrazuje distribuci kategorií numerického- kardinálního (nebo ordinálního) znaku (proto ho někdy doplňujeme o distribuční křivku normálního rozdělení).

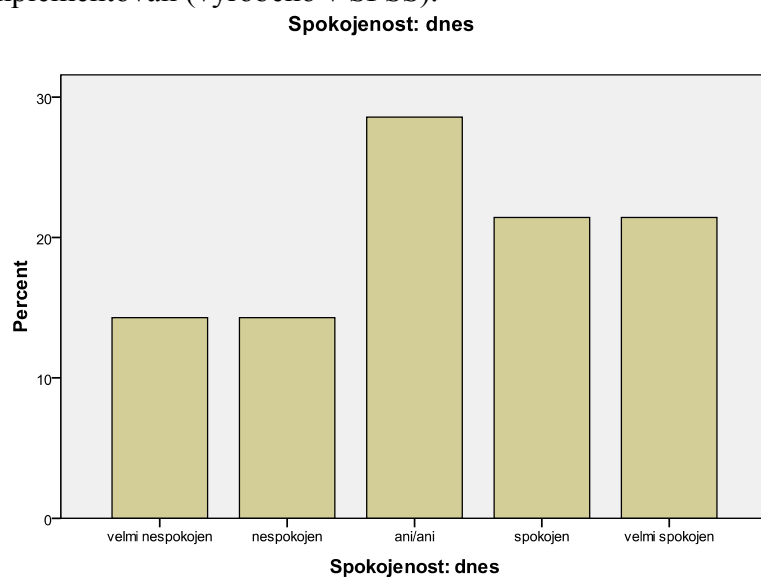


Zdroj: Studenti HiSo-Kombi LS 2010

Frequency na ose Y znázorňuje počet případů v dané kategorii (absolutní četnost tedy nikoliv procenta), např. v kategorii 1-velmi nespokojen máme dva respondenty, zatímco v kategorii 3-ani/ani jsou čtyři.

### Sloupcový graf v % (Bar Chart)

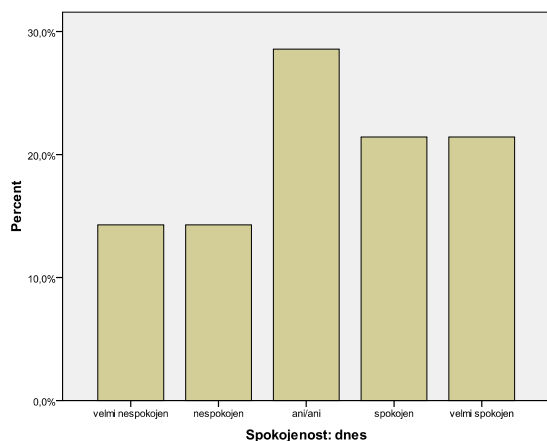
Sloupcový graf pokud je zadán v % znázorňuje relativní podíly kategorií (ty lze dále je porovnávat pro určité podskupiny). Pokud bychom ho zadali v absolutních četnostech, dostaneme vlastně histogram. Sloupcový graf v procentech (Bar chart) není v PSPP dosud implementován (vyrobena v SPSS).



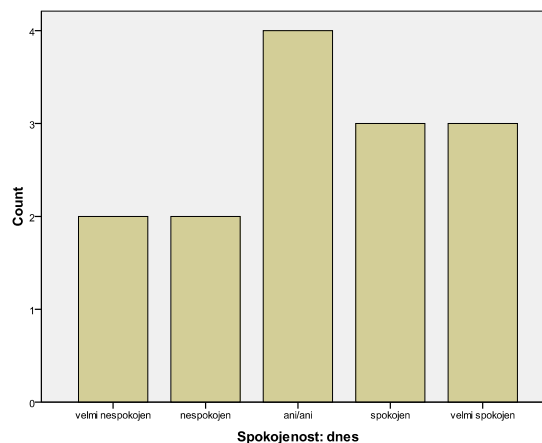
Zdroj: Studenti HiSo-Kombi LS 2010

## Porovnání různých druhů grafů.

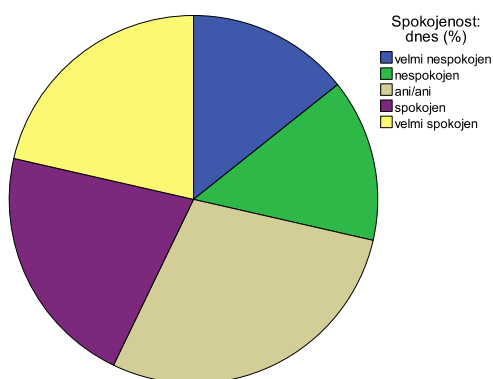
Bar chart v %



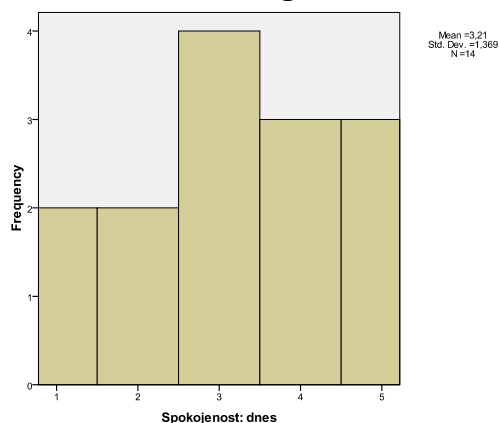
Bar chart v absolutních četnostech



Koláčový graf v % (Pie Chart)



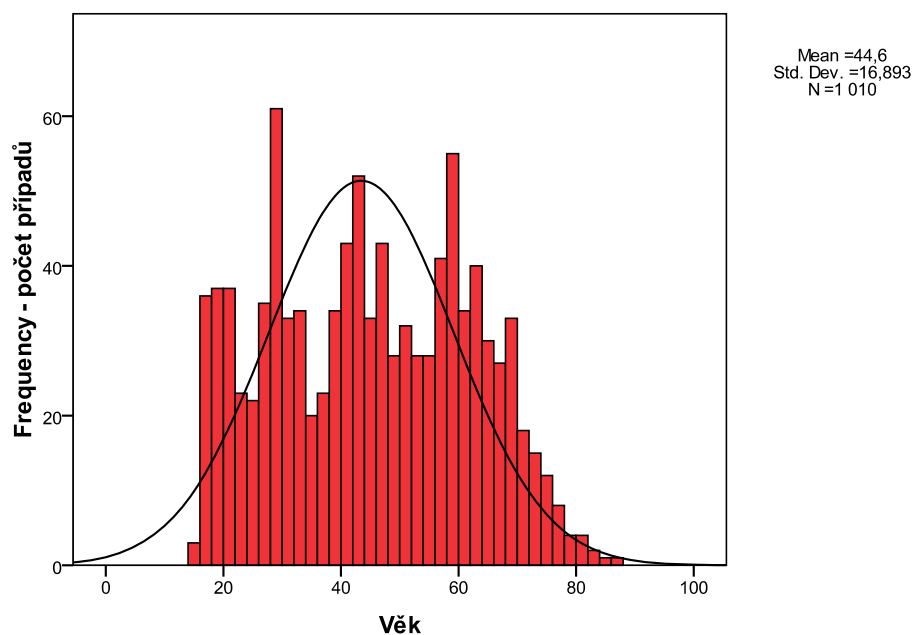
Histogram



Zdroj: Studenti HiSo-Kombi LS 2010

Obecně platí, že histogram používáme pro zobrazování distribuce kategorií numerického (kardinálního) znaku, který může mít větší počet hodnot, jakým je kupříkladu věk (viz ukázka z výzkumu CVVM 2007/04), zatímco Sloupcový graf v % (nebo koláčový) používáme pro zobrazení menšího počtu kategorií, které nemusí být hierarchicky (ordinálně) uspořádány (nominální znaky). Viz sloupcový graf pro sloučený věk do 4 kategorií.

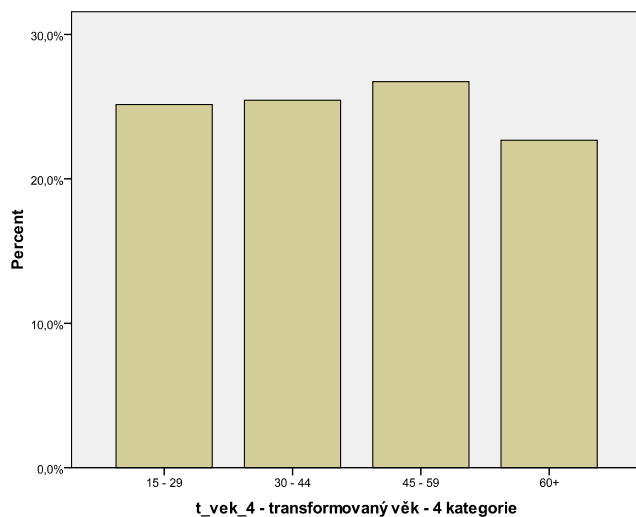
**Graf. Histogram pro věk respondenta.**  
Histogram



Zdroj: CVVM 2007/04

Všimněte si, že v histogramu jsou sloupce na sebe těsně navázány – zobrazujeme numerické hodnoty znaku, které plynule vzrůstají a navazují na sebe. Naproti tomu ve sloupcovém grafu v % ukazujeme nominální (či ordinální) znak, kde bychom také mohli kategorie mezi sebou případně i proházet.

**Graf. Sloupcový graf pro věkové kategorie (%)**

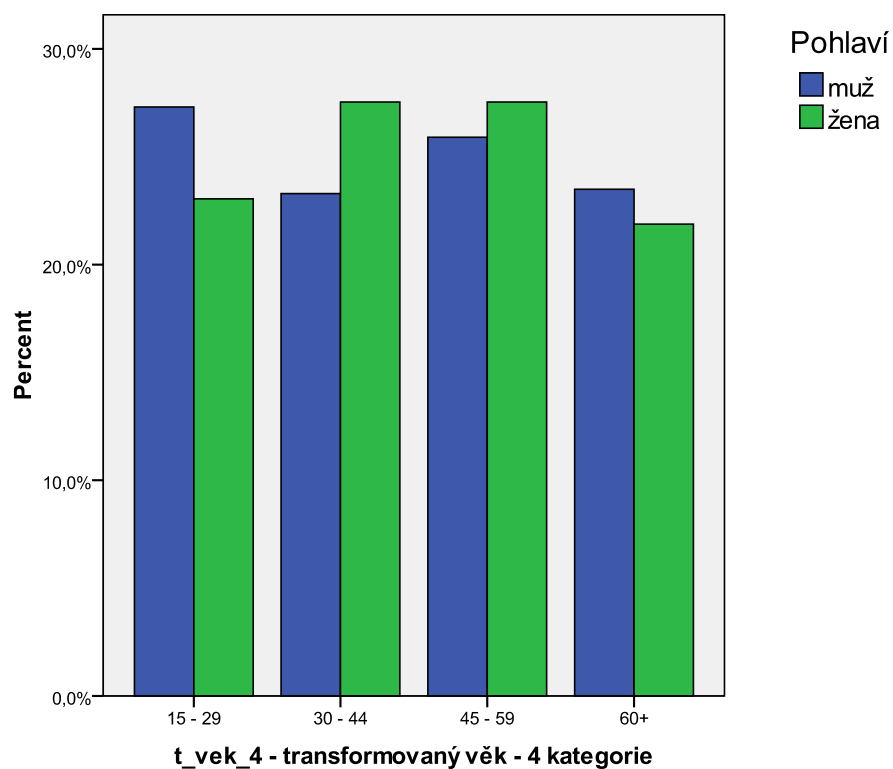


Zdroj: CVVM 2007/04

Dále platí, že sloupcový graf v % je vhodný pro porovnání podílů mezi různými skupinami, což histogram neumožňuje, protože záleží na tom, kolik případů je v dané kategorii, pro níž frekvence třídíme. Příkladem budiž věkové kategorie a pohlaví respondenta.



**Graf. Sloupcový graf pro věkové kategorie podle pohlaví respondenta (%)**



Zdroj: CVVM 2007/04

O třídění dat druhého stupně si povíme více v druhé části Opory.

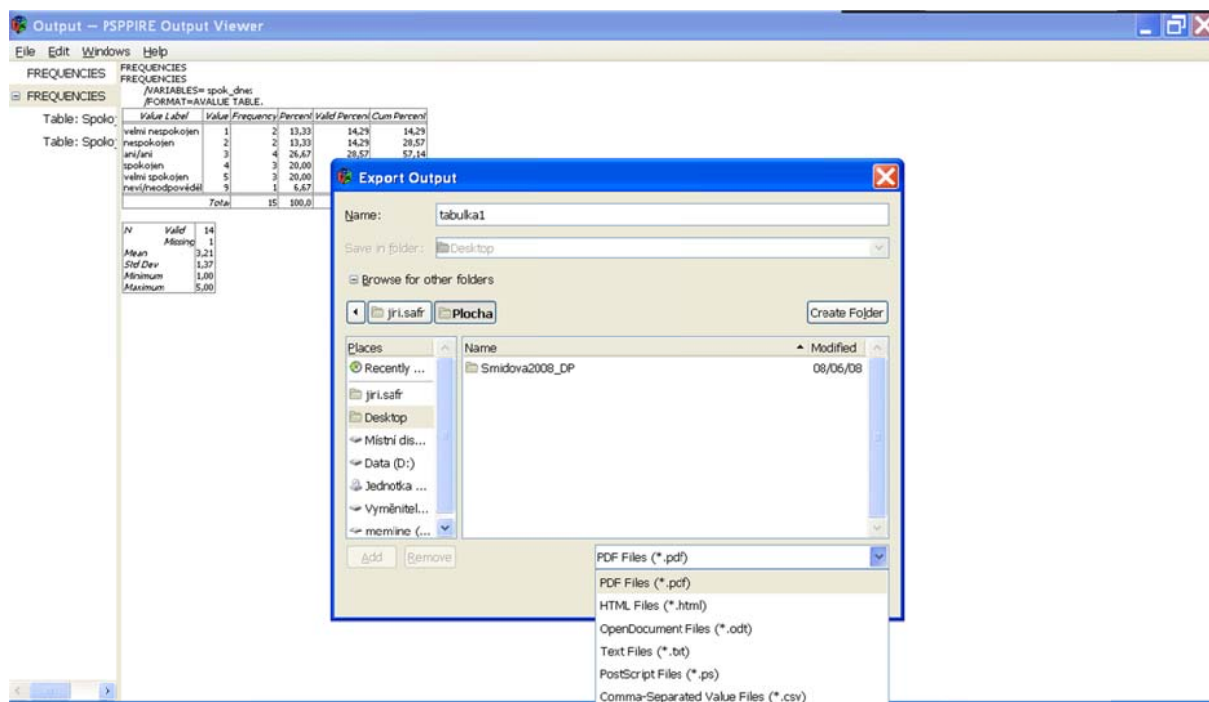
## Technické rady pro práci v PSPP

### Možnosti úprav výstupu (Output) z PSPP

Nevýhodou PSPP je na rozdíl od SPSS, že nemůžete tabulku či jiné výstupy v okně přímo upravovat. Je třeba nejprve celý výstup uložit a teprve ten pak lze upravit. Ukládat můžeme do pdf (přesný obraz z PSPP), txt, htm. Vhodný je formát htm, protože zachovává tabulky v tabulkovém formátu a lze ho poté otevřít i v MS Wordu. Také již lze výstup i kopírovat a vkládat přímo do MS Wordu (od verze PSPP 10). Bohužel to lze jen jako celý výstup (označení je krkolonné, přes menu takto: Edit / Select All a pak Copy)<sup>3</sup> a navíc grafy se nám uloží jen jako obrázky (ve formátu png) do velmi divného a skoro nedohledatelného místa v počítači (nejčastěji někde jako: users / jméno uživatele / temp, naštěstí vám to při vložení do Wordu vypíše cestu, kde obrázky najdete). Lze používat i klávesové zkratky Ctrl+A (výběr všeho) a pak Ctrl+C (kopírovat), ale je třeba být v Outputu v levém podokně se seznamem výsledků (nejprve na něj kliknout).

V SPSS je to vše mnohem jednodušší a komfortnější. Tabulky a grafy jsou v Outputu samostatné objekty, které můžete dále upravovat (dvakrát na něj kliknete a otevře se příslušný editor), zkopírovat si můžete jen vybrané výstupy a celé to lze uložit ve vlastním formátu (\*.spv) a také exportovat mj. do Excelu při zachování formátování. Navíc grafy lze kopírovat a vkládat jako vektorovou grafiku, která zachovává maximální ostrost obrazu a obrázky jsou extrémně malé (klikněte na objekt 1x a pravým tlačítkem vyvolejte menu, zvolte: Copy Special a zaklikněte „Metafile (WMF, EMF)“, poté při vkládání např. do MS Wordu dejte Vložit jako a vyberte jeden z těchto vektorových formátů). Vyhněte se, pokud to lze, formátům obrázků určených pro fotografie jako jsou jpg, png stejně jako bitmapového bmp, jsou neostré a zabírají moc místa.

V PSPP výstup uložíme takto: File → Export, pak zadat název souboru a jeho formát.



<sup>3</sup> Pozor zvolit nejprve Select All je nezbytné, jinak se nic nezkopíruje. Takže nelze jako v SPSS selektivně vybírat jen některé tabulky či grafy.

Pozor: je možné, že po uložení v názvu souboru bude chybět koncovka označující typ souboru. Tu je pak třeba ručně zadat-změnit, např. ve WinComanderu nebo v Průzkumníkovi Windows.

ukázka outputu v PDF

25. Apr. 2010 - Page 1  
G:\U\pspp\0.7.4-g8c314e-1686-pc-mingw32

FREQUENCIES

FREQUENCIES

/VARIABLES= spok\_dnes

/FORMAT=AVALUE TABLE.

Spokojenost: dnes

| Value Label      | Value | Frequency | Percent      | Valid Percent | Cum Percent |
|------------------|-------|-----------|--------------|---------------|-------------|
| velmi nespokojen | 1     | 2         | 13,33        | 14,29         | 14,29       |
| nespokojen       | 2     | 2         | 13,33        | 14,29         | 28,57       |
| ani/ani          | 3     | 4         | 26,67        | 28,57         | 57,14       |
| spokojen         | 4     | 3         | 20,00        | 21,43         | 78,57       |
| velmi spokojen   | 5     | 3         | 20,00        | 21,43         | 100,00      |
| nevi/neodpovědi  | 9     | 1         | 6,67         | Missing       |             |
| <b>Total</b>     |       | <b>15</b> | <b>100,0</b> | <b>100,0</b>  |             |

Spokojenost: dnes

|         |         |      |
|---------|---------|------|
| N       | Valid   | 14   |
|         | Missing | 1    |
| Mean    |         | 3,21 |
| Std Dev |         | 1,37 |
| Minimum |         | 1,00 |
| Maximum |         | 5,00 |

ukázka outputu v TXT

output\_txt - Poznámkový blok

Soubor Úpravy Formát Zobrazení nápověda

FREQUENCIES

FREQUENCIES

/VARIABLES= spok\_dnes

/FORMAT=AVALUE TABLE.

Spokojenost: dnes

| Value Label      | Value | Frequency | Percent      | Valid Percent | Cum Percent |
|------------------|-------|-----------|--------------|---------------|-------------|
| velmi nespokojen | 1     | 2         | 13,33        | 14,29         | 14,29       |
| nespokojen       | 2     | 2         | 13,33        | 14,29         | 28,57       |
| ani/ani          | 3     | 4         | 26,67        | 28,57         | 57,14       |
| spokojen         | 4     | 3         | 20,00        | 21,43         | 78,57       |
| velmi spokojen   | 5     | 3         | 20,00        | 21,43         | 100,00      |
| nevi/neodpovědi  | 9     | 1         | 6,67         | Missing       |             |
| <b>Total</b>     |       | <b>15</b> | <b>100,0</b> | <b>100,0</b>  |             |

Spokojenost: dnes

|         |         |      |
|---------|---------|------|
| N       | Valid   | 14   |
|         | Missing | 1    |
| Mean    |         | 3,21 |
| Std Dev |         | 1,37 |
| Minimum |         | 1,00 |
| Maximum |         | 5,00 |

A při uložení do formátu htm to vypadá takto:

FREQUENCIES

FREQUENCIES

/VARIABLES= spok\_dnes

/FORMAT=AVALUE TABLE.

Spokojenost: dnes

| Value Label      | Value | Frequency | Percent      | Valid Percent | Cum Percent |
|------------------|-------|-----------|--------------|---------------|-------------|
| velmi nespokojen | 1     | 2         | 13,33        | 14,29         | 14,29       |
| nespokojen       | 2     | 2         | 13,33        | 14,29         | 28,57       |
| ani/ani          | 3     | 4         | 26,67        | 28,57         | 57,14       |
| spokojen         | 4     | 3         | 20,00        | 21,43         | 78,57       |
| velmi spokojen   | 5     | 3         | 20,00        | 21,43         | 100,00      |
| nevi/neodpovědi  | 9     | 1         | 6,67         | Missing       |             |
| <b>Total</b>     |       | <b>15</b> | <b>100,0</b> | <b>100,0</b>  |             |

Spokojenost: dnes

|         |         |      |
|---------|---------|------|
| N       | Valid   | 14   |
|         | Missing | 1    |
| Mean    |         | 3,21 |
| Std Dev |         | 1,37 |
| Minimum |         | 1,00 |
| Maximum |         | 5,00 |

Poznámka: jak vidno, PSPP v htm nezvládá zcela českou diakritiku. Záleží ale na verzi češtiny ve Windows.

Tabulku pak lze zkopírovat do Wordu a tam jí upravit, například takto (zkopírováno z htm formátu otevřeného v internetovém prohlížeči, ale můžete také celý soubor htm otevřít rovnou ve Wordu):

**Tabulka 1. Odpovědi na otázku „Jak jste dnes spokojen?“, procenta.**

| <b>Odpovědi</b>     | <b>Hodnota</b> | <b>Počet</b> | <b>%</b>       | <b>Platná %</b> | <b>Kumulativní %</b> |
|---------------------|----------------|--------------|----------------|-----------------|----------------------|
| Velmi nespokojen    | 1              | 2            | 13,3           | <b>14,3</b>     | 14,3                 |
| Nespokojen          | 2              | 2            | 13,3           | <b>14,3</b>     | 28,6                 |
| Ani spk./ani nespk. | 3              | 4            | 26,7           | <b>28,6</b>     | 57,1                 |
| Spokojen            | 4              | 3            | 20,0           | <b>21,4</b>     | 78,6                 |
| Velmi spokojen      | 5              | 3            | 20,0           | <b>21,4</b>     | 100,0                |
| Neví/neodpověl/a    | 9              | 1            | 6,7            | -               |                      |
| <b>Celkem</b>       |                | <b>15</b>    | <b>100,0 %</b> | <b>100,0 %</b>  |                      |

### Spokojenost: dnes

|                    |                |     |
|--------------------|----------------|-----|
| <i>N</i>           | <i>Platné</i>  | 14  |
|                    | <i>Missing</i> | 1   |
| <i>Průměr</i>      |                | 3,2 |
| <i>Směr. odch.</i> |                | 1,4 |
| <i>Minimum</i>     |                | 1,0 |
| <i>Maximum</i>     |                | 5,0 |

Výsledky ve formátu htm a csv lze otevírat a následně upravovat v Excelu.

|    | A                   | B       | C         | D       | E             | F           |
|----|---------------------|---------|-----------|---------|---------------|-------------|
| 1  | Spokojenost: dnes   |         |           |         |               |             |
| 2  | Value Label         | Value   | Frequency | Percent | Valid Percent | Cum Percent |
| 3  | velmi nespokojen    | 1       | 2         | 13,33   | 14,29         | 14,29       |
| 4  | nespokojen          | 2       | 2         | 13,33   | 14,29         | 28,57       |
| 5  | an/ani              | 3       | 4         | 26,67   | 28,57         | 57,14       |
| 6  | spokojen            | 4       | 3         | 20      | 21,43         | 78,57       |
| 7  | velmi spokojen      | 5       | 3         | 20      | 21,43         | 100         |
| 8  | neví/-neodpověděl/a | 9       | 1         | 6,67    | Missing       |             |
| 9  | Total               |         | 15        | 100     | 100           |             |
| 10 |                     |         |           |         |               |             |
| 11 | Spokojenost: dnes   |         |           |         |               |             |
| 12 | N                   | Valid   | 14        |         |               |             |
| 13 |                     | Missing | 1         |         |               |             |
| 14 | Mean                |         | 3,21      |         |               |             |
| 15 | Std Dev             |         | 1,37      |         |               |             |
| 16 | Minimum             |         | 1         |         |               |             |
| 17 | Maximum             |         | 5         |         |               |             |

|    | A                        | B       | C         | D       | E          | F           | G | H | I | J | K | L |
|----|--------------------------|---------|-----------|---------|------------|-------------|---|---|---|---|---|---|
|    | Value Label              | Value   | Frequency | Percent | Valid Perc | Cum Percent |   |   |   |   |   |   |
| 2  | velmi nespokojen         | 1       | 2         | 13,33   | 14,29      | 14,29       |   |   |   |   |   |   |
| 3  | nespokojen               | 2       | 2         | 13,33   | 14,29      | 28,57       |   |   |   |   |   |   |
| 4  | ani/ani                  | 3       | 4         | 26,67   | 28,57      | 57,14       |   |   |   |   |   |   |
| 5  | spokojen                 | 4       | 3         | 20      | 21,43      | 78,57       |   |   |   |   |   |   |
| 6  | velmi spokojen           | 5       | 3         | 20      | 21,43      | 100         |   |   |   |   |   |   |
| 7  | nevi/neodpověděl         | 9       | 1         | 6,67    | Missing    |             |   |   |   |   |   |   |
| 8  | Total                    |         | 15        | 100     | 100        |             |   |   |   |   |   |   |
| 10 | Table: Spokojenost: dnes |         |           |         |            |             |   |   |   |   |   |   |
| 11 | N                        | Valid   | 14        |         |            |             |   |   |   |   |   |   |
| 12 |                          | Missing | 1         |         |            |             |   |   |   |   |   |   |
| 13 | Mean                     |         | 3,21      |         |            |             |   |   |   |   |   |   |
| 14 | Std Dev                  |         | 1,37      |         |            |             |   |   |   |   |   |   |
| 15 | Minimum                  |         | 1         |         |            |             |   |   |   |   |   |   |
| 16 | Maximum                  |         | 5         |         |            |             |   |   |   |   |   |   |

Poznámka: viz následující osvětlení, jak funguje formát CSV.

## CSV – univerzální formát pro data

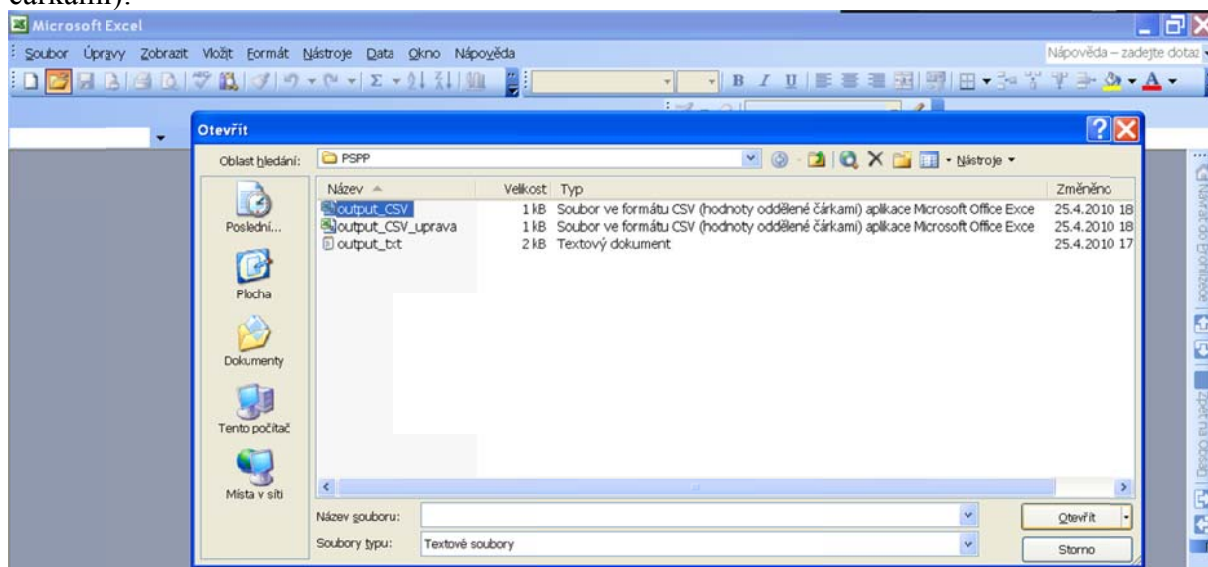
**Formát CSV** (Comma-separated values) je jednoduchý univerzální souborový formát určený pro výměnu tabulkových dat, kdy hodnoty jsou oddělené čárkami.

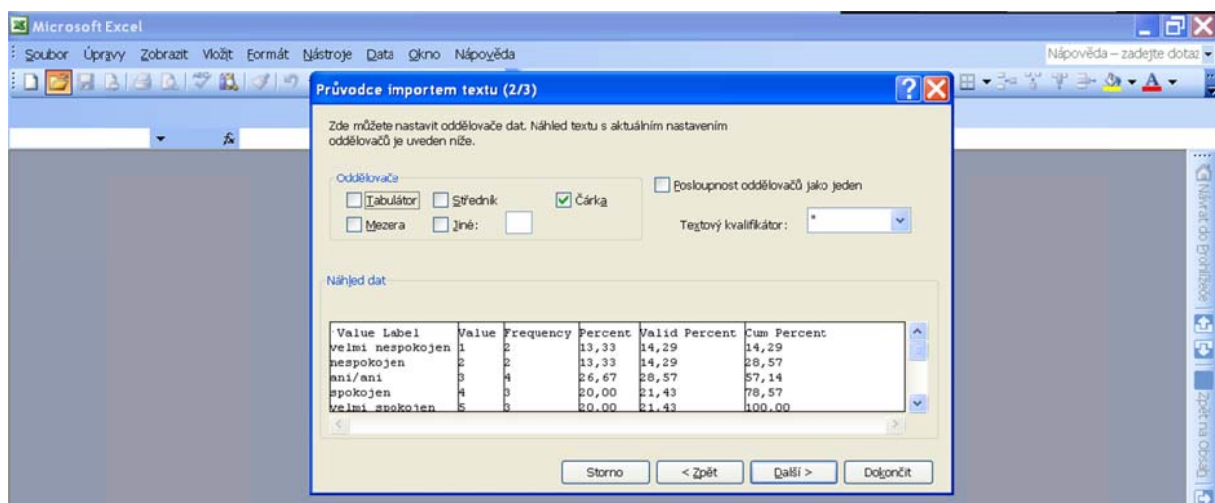
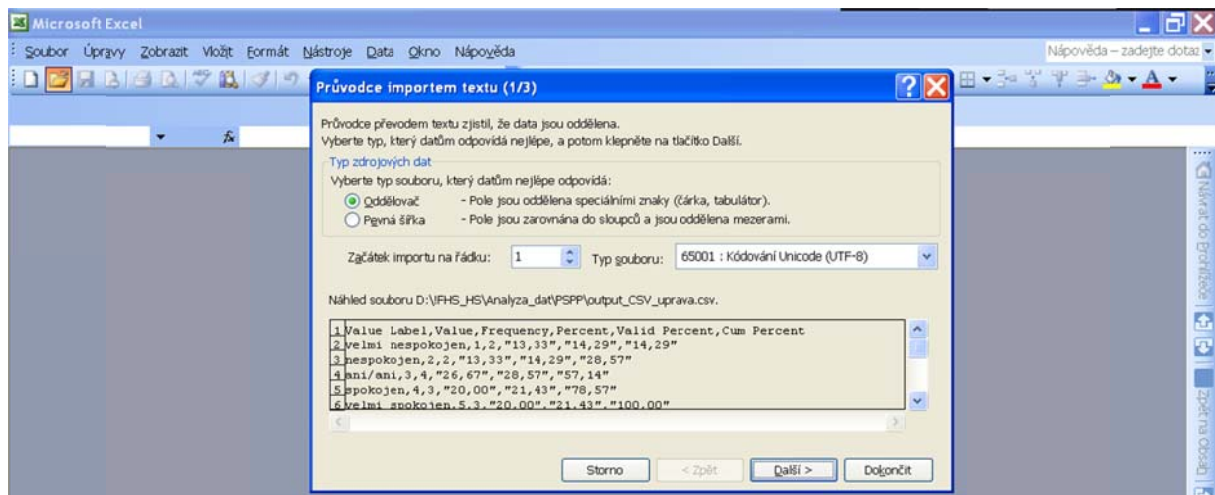
Použit ho můžete také pro přenos dat mezi různými programy: např. PSPP / SPSS ↔ Excel. Zejména pak pro import dat z online dotazníků (např. [www.vyplnto.cz](http://www.vyplnto.cz))

Problém je, že v české verzi Excelu se standardně používají místo čárek středníky, to protože čárkou v češtině oddělujeme desetinné místo. PSPP ale generuje CSV s čárkami (SPSS umí se středníky). Aby se odlišila desetinná hodnota od sloupce používají se pro označení čísel/údajů v buňce ještě navíc uvozovky (více na <http://cs.wikipedia.org/wiki/CSV>).

Existují dvě řešení jak CSV z PSPP otevřít v Excelu.

1. V Excelu při otvírání zadat jako Oddělovač čárku, tj. označit, že sloupce budou odděleny čárkami):



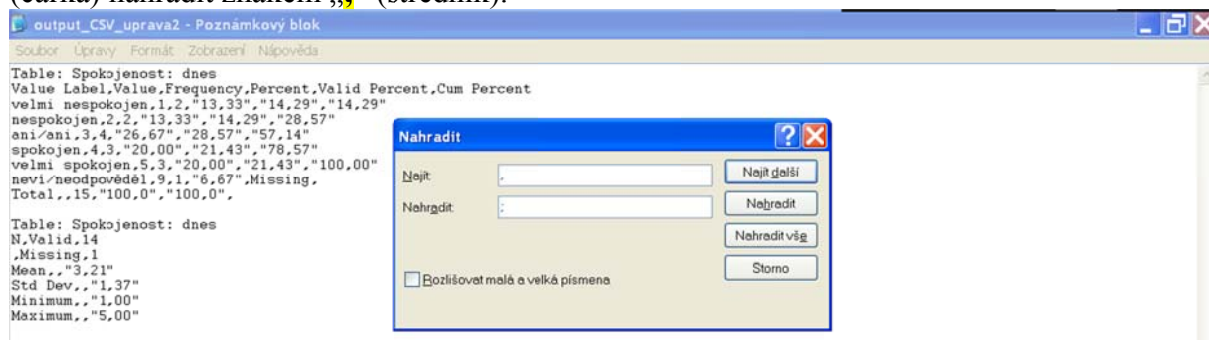


výsledkem je pak přiřazení hodnot do sloupců

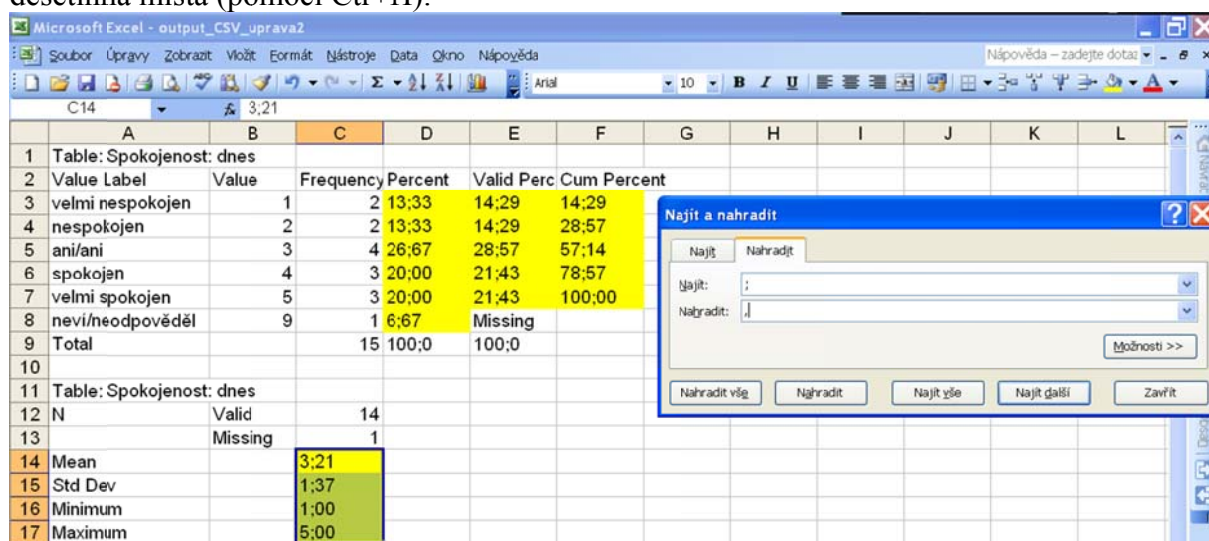
| Microsoft Excel - output_CSV_uprava                              |                          |         |           |         |            |             |   |   |   |   |   |   |   |   |  |
|--|--------------------------|---------|-----------|---------|------------|-------------|---|---|---|---|---|---|---|---|--|
| Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda |                          |         |           |         |            |             |   |   |   |   |   |   |   |   |  |
| A1 Value Label   |                          |         |           |         |            |             |   |   |   |   |   |   |   |   |  |
| A  | B                        | C       | D         | E       | F          | G           | H | I | J | K | L | M | N | O |  |
| 1  | Value Label              | Value   | Frequency | Percent | Valid Perc | Cum Percent |   |   |   |   |   |   |   |   |  |
| 2  | velmi nesp               | 1       | 2         | 13,33   | 14,29      | 14,29       |   |   |   |   |   |   |   |   |  |
| 3  | nespokojen               | 2       | 2         | 13,33   | 14,29      | 28,57       |   |   |   |   |   |   |   |   |  |
| 4  | ani/ani                  | 3       | 4         | 26,67   | 28,57      | 57,14       |   |   |   |   |   |   |   |   |  |
| 5  | spokojen                 | 4       | 3         | 20      | 21,43      | 78,57       |   |   |   |   |   |   |   |   |  |
| 6  | velmi spok               | 5       | 3         | 20      | 21,43      | 100         |   |   |   |   |   |   |   |   |  |
| 7  | nevi/neodp               | 9       | 1         | 6,67    | Missing    |             |   |   |   |   |   |   |   |   |  |
| 8  | Total                    |         | 15        | 100     | 100        |             |   |   |   |   |   |   |   |   |  |
| 9  |                          |         |           |         |            |             |   |   |   |   |   |   |   |   |  |
| 10   | Table: Spokojenost: dnes |         |           |         |            |             |   |   |   |   |   |   |   |   |  |
| 11   | N                        | Valid   | 14        |         |            |             |   |   |   |   |   |   |   |   |  |
| 12   |                          | Missing | 1         |         |            |             |   |   |   |   |   |   |   |   |  |
| 13   | Mean                     |         | 3,21      |         |            |             |   |   |   |   |   |   |   |   |  |
| 14   | Std Dev                  |         | 1,37      |         |            |             |   |   |   |   |   |   |   |   |  |
| 15   | Minimum                  |         | 1         |         |            |             |   |   |   |   |   |   |   |   |  |
| 16   | Maximum                  |         | 5         |         |            |             |   |   |   |   |   |   |   |   |  |



2. Nebo otevřít CSV soubor nejprve v např. Poznámkovém bloku nebo Wordu a znak „;“ (čárka) nahradit znakem „;“ (středník).



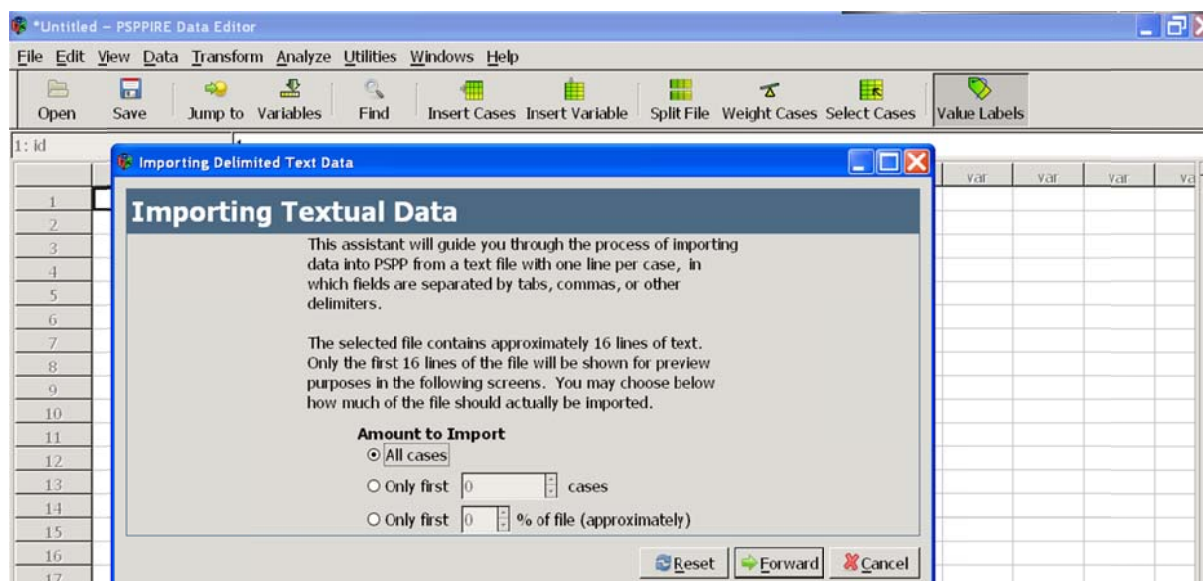
Jenomže pak je třeba ještě v Excelu naradit všechny středníky zase zpátky na čárky oddělující desetinná místa (pomocí Ctr+H).



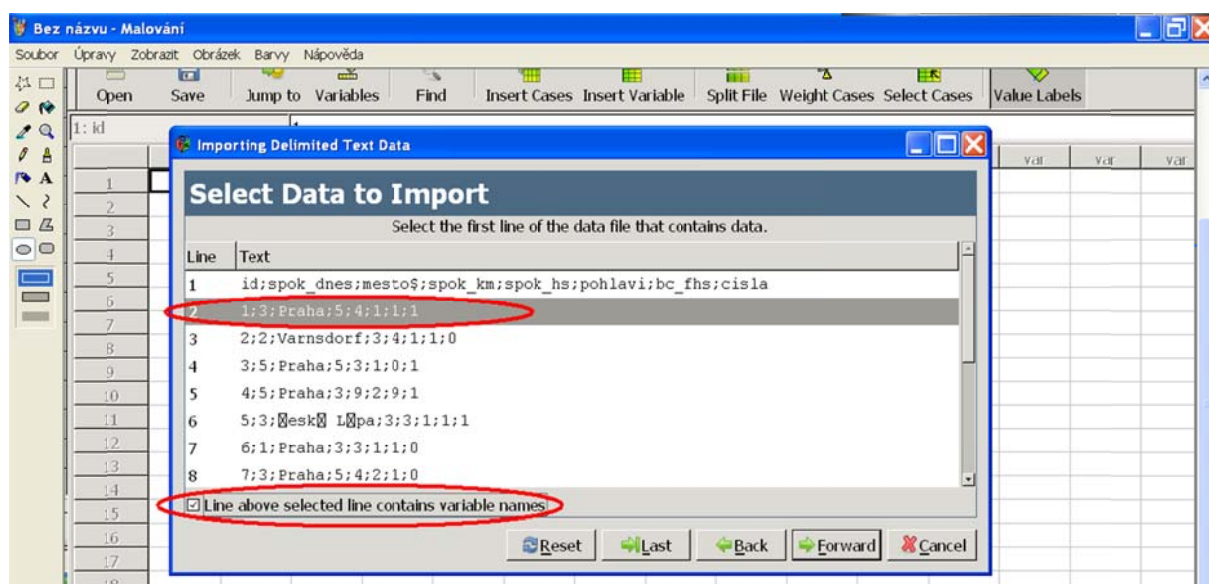
Při nahrazování je třeba hlídat, zda se někde znaky nezměnily nechtěným způsobem (např. pokud by v labelech byly někde použity čárky).

## Import CSV dat do PSPP

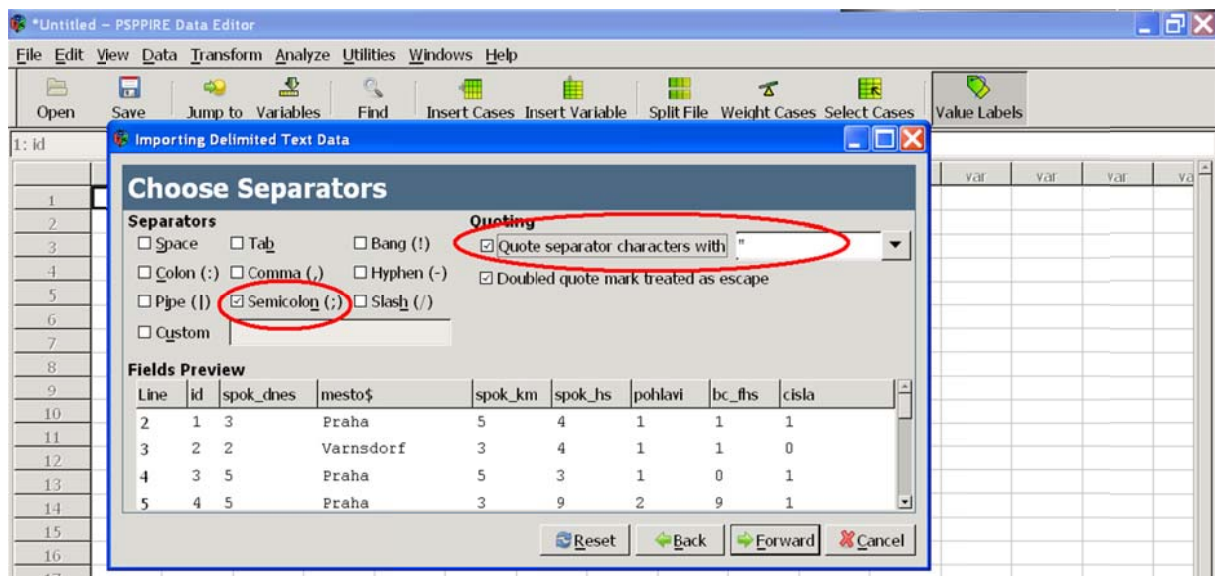
Formát CSV se používá především pro import dat z jiných programů (neobsahuje ovšem labely jen názvy proměnných).



Musíme zadat, zda první řádek obsahuje názvy proměnných → **Line above selected line contains variable names** a současně klikneme na druhý řádek (musí být šedivý).



Následuje zadání oddělovače sloupců = proměnných. Zde volíme středník (Semicolon), použít lze i jiné znaky (např. čárku), záleží na tom, jak jsou sloupce odděleny, což vidíme v ukázce datové matice **Fields Preview** dole. Zadat lze také, zda znaky ve sloupcích jsou navíc označeny uvozovkami (což umožňuje rozeznání čárek či středníků, které neznamenají oddělovače sloupců; viz výše).



## Syntax – ovládání pomocí příkazů (úvod)

Program PSPP/ SPSS lze ovládat klikáním na menu v grafickém rozhraní (případně na ikonky) a nebo také pomocí příkazového řádku, tzv. syntaxu. Syntaxe slouží k efektivnímu provádění transformací, úprav dat a analýz. Výhodou je především rychlost a opakovatelnost – pro zadání nové analýzy stačí změnit názvy proměnných. Pokud si syntax uložíme (koncovka .SPS), pak můžeme operace opakovat a kontrolovat (vidíme i po čase, co jsme s daty prováděli).

V podstatě se jedná o jednoduchý textový editor, v němž zadáváme příkazové řádky. Příkazový řádek vždy obsahuje příkaz a jméno proměnné(ých), doplněn může být o další specifikace analýzy.

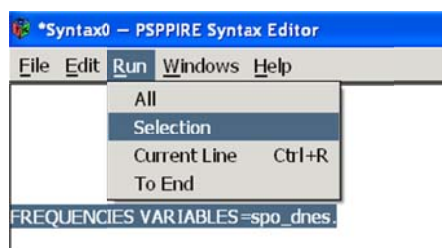
### Vždy končí tečkou !

Uveďme si zadání příkazu FREQUENCIES, který jsme používali v předchozí části pro vytvoření jednoduché tabulky. Do okna editoru Syntaxu vepíšeme (nebo zkopírováním vložíme) řádek:

```
FREQUENCIES VARIABLES=spo_dnes.
```

Máme zde příkaz k operaci (zde tabulku frekvencí) a jméno proměnné (zde spo\_dnes, která je odpovědí na otázku „Jak jste dnes spokojen?“)

Příkazový řádek (nebo i více řádků najednou) pak označíme do bloku a v menu klikneme na **Run** a zvolíme **Selection**. Použít lze také klávesovou zkratku Ctrl+R. Poté se podíváme do Outputu na výsledek (na rozdíl od SPSS se nám toto okno samo automaticky nepřepne).



Příkazy lze zadávat ve zkrácené podobě (a v tom je právě kouzlo rychlého zadávání příkazů) u Frequencies stačí takto:

```
FRE spo_dnes.
```

Nicméně příkazy lze i dále specifikovat – specifikace v rámci jednoho řádku odděluje / (lomítko). U Frequencies můžeme například přidat zadání tisku koláčového grafu (**PIECHART**) a nebo také seřazení kategorií podle dosažené četnosti (**DFREQ**).

```
FREQUENCIES  
  /VARIABLES= spo_dnes  
  /FORMAT=DFREQ TABLE  
  /PIECHART= MISSING.
```

Použijeme-li jako příkaz \*. pak se nám ve výstupu objeví náš komentář, např.:

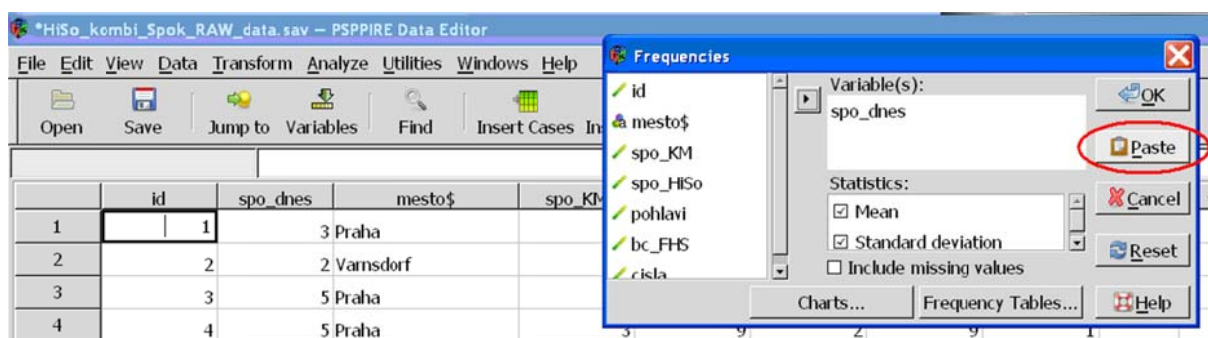
```
*Popisné statistiky příkazem FREQUENCIES pro znak spo_dnes.
```

```
FREQUENCIES spo_dnes.
```

První řádek s hvězdičkou \* je zde pouze váš (nepovinný) komentář, který se objeví ve výstupu, nic neprovádí, jen informuje (objeví se text co je mezi \* a tečkou). V druhém řádku je pak samotný příkaz Frequencies.

Je dobrým zvykem samotné příkazy psát velkými písmeny, ulehčí to orientaci, ale není to nutné.

Příkaz pro syntax si můžete jednoduše nechat vytvořit při zadávání pomocí klikání myši v menu. Před (nebo místo) kliknutí na tlačítko OK, kliknete na **Paste**. Vygeneruje se vám celý příkaz do nového okna syntax editoru. Bohužel vždy do nového, nikoliv aktivního okna (jako je tomu v SPSS), takže pokud chcete mít všechny příkazy v jednom souboru, musíte text příkazu vždy vybrat a vložit do vašeho syntax souboru, pomocí menu myši – pravé tlačítko (pozor nefungují klávesové zkratky ctrl+c /+v).



Všimněte si také, že syntax se nám automaticky objevuje v okně výstupů, před samotnými výsledky. Přehledně tak vidíme, o co v analýze jde a můžeme jí také později i zopakovat.

| Output – PSPPIRE Output Viewer |       |           |         |               |             |
|--------------------------------|-------|-----------|---------|---------------|-------------|
| FREQUENCIES                    |       |           |         |               |             |
| FREQUENCIES                    |       |           |         |               |             |
| /VARIABLES= spo_dnes           |       |           |         |               |             |
| /FORMAT=DFREQ TABLE.           |       |           |         |               |             |
| Value Label                    | Value | Frequency | Percent | Valid Percent | Cum Percent |
| ani/ani                        | 3     | 4         | 26,67   | 28,57         | 28,57       |
| spokojen                       | 4     | 3         | 20,00   | 21,43         | 50,00       |
| velmi spokojen                 | 5     | 3         | 20,00   | 21,43         | 71,43       |
| velmi nespokojen               | 1     | 2         | 13,33   | 14,29         | 85,71       |
| nespokojen                     | 2     | 2         | 13,33   | 14,29         | 100,00      |
| neví/neodpověděl               | 9     | 1         | 6,67    | Missing       |             |
| Total                          |       | 15        | 100,0   | 100,0         |             |
| N                              |       |           |         |               |             |
| Valid                          | 14    |           |         |               |             |
| Missing                        | 1     |           |         |               |             |
| Mean                           | 3,21  |           |         |               |             |
| Std Dev                        | 1,37  |           |         |               |             |
| Minimum                        | 3,00  |           |         |               |             |
| Maximum                        | 2,00  |           |         |               |             |

Důležité je, že některé funkce či analýzy jsou v PSPP (to ale platí i pro SPSS) dostupné pouze pomocí zadání v syntaxu, tzn. že je nenajdete v menu.

Skvělý návod na syntax PSPP naleznete na

<http://www.gnu.org/software/pspp/manual/pspp.html>

Také prezentace pro předměty Analýza kvantitativních dat I., II. III. uvádí příklady příkazů pro SPSS (PSPP), Najdete je na stránce <http://metodykv.wz.cz/>, kde je také archiv různých syntaxů, co jsme dělali na seminářích AKD. Jednoduché syntaxy s tím „co jsme dělali“ na kurzu Praktikum elementární analýzy kvantitativních dat najdete také na stránce <http://kmvp.wz.cz/>.



## Analýza dat (2): Logika vytváření kontingenčních tabulek a jejich interpretace

### Prezentace a interpretace kontingenční tabulky (výzkumná otázka a hypotéza)

V zásadě k tabulkám můžeme přistoupit dvojím způsobem. Buď postupujeme jen víceméně deskriptivně – explorativně, kdy nemáme žádnou předem danou otázku/ hypotézu a jen systematicky popisujeme rozdíly v závislé (vysvětlované) proměnné z hlediska třídění jinými znaky. Anebo máme předem definované výzkumné otázky (obecné i ty konkrétní, nejčastěji formulované již ve fázi přípravy výzkumu) a z nich odvozené hypotézy (tj. vlastně odpovědi na tyto otázky na základě teorie) a postupujeme tak, že tabulku či graf konstruujeme podle těchto otázek/hypotéz, tak abychom přímo ověřili předpokládané vztahy. Viz následující příklad ze cvičného výzkumu studentů na FHS s názvem „TV a knihy“.

**Výzkumná otázka (RQ):** *Souvisí počet přečtených knih s dobou sledování TV?*

**Hypotéza „sociologická“, tj. věcně formulovaná:** *Počet přečtených knih roste s dobou strávenou sledováním TV.* → předpokládáme *pozitivní souvislost*

Nulová „statistická“ hypotéza ( $H_0$ ): Počet přečtených knih se neliší v závislosti na době strávené u TV. → Statistickým postupem je očekávat a ověřovat „*žádný vztah*“.

K ní lze formulovat tzv. alternativní hypotézu ( $H_A$ ) – kdy  $H_0$  neplatí, tj. vztah existuje.

Dodejme ještě, že vztahy nemusí být takto jednoduché, vyjádřené jako lineární souvislosti (u kategoriálních znaků spíše hovoříme o monotónním vztahu). Ostatně uvedený příklad pracuje se dvěma ordinálními znaky (koneckonců mohli bychom místo tabulky spočítat jen ordinální–pořadový korelační koeficient, ale tím bychom ztratili mnoho podstatných informací), ale kontingenční tabulky používáme také pro hlavně čistě nominální znaky (zde by se hodilo například zkoumat souvislost mezi tím, jaký literární žánr studenti čtou a jaké typy televizních pořadů sledují). Navíc vztahy se mohou projevovat třeba jen v určitých podskupinách (tj. jen za určitých podmínek).

Z toho vyplývá, že **zamítnutím jednoduše formulované hypotézy bychom se neměli spokojit**. To pochopitelně platí i pro hypotézy „potvrzené“, resp. správně řečeno – nezamítnuté. Navíc mnoho vztahů, kde chceme usuzovat na kauzální působení, je zprostředkovaných a tudíž bivariátní vztah dvou znaků může být způsoben (podmíněn) působením třetí proměnné. Práci výzkumníka je právě v dalším kroku odhalovat i tyto komplikovanější vztahy (o podobě vztahu hodně napoví kontingenční tabulka a nebo pro spojitý znaky X-Y bodový – scatterplot graf a následné třídění třetího stupně).



**Tabulka 2. Počet knížek přečtený za rok (tercily) podle sledování TV v hodinách za den (tercilu), sloupcová procenta, studenti FHS předmětů J. Šafra v letech 2011 až 2012.**

|   |                 | Sledování TV (tercily) |                |                 | Celkem |
|---|-----------------|------------------------|----------------|-----------------|--------|
|   |                 | I. tercil (-)          | II. tercil (0) | III. tercil (+) |        |
| <b>Knihy přečtené celkem za rok (tercily)</b> | I. tercil (-)   | 15,8                   | 23,1           | 38,9            | 25,4%  |
|   | II. tercil (0)  | 42,1                   | 46,2           | 44,4            | 44,4%  |
|   | III. tercil (+) | 42,1                   | 30,8           | 16,7            | 30,2%  |
| Celkem  |                 | 100 %                  | 100 %          | 100 %           | 100 %  |
| N   |                 | (19)                   | (26)           | (18)            | (63)   |

Zdroj: FHS TV a knihy 2011-2012

N (validní) = 63 (chybějící hodnoty 12,5 %).

Tabulka 2 ukazuje relativní podíl přečtených knih za rok (soubor jsme rozdělili na tři stejně velké skupiny čtenářů podle tercilů) ve skupinách definovaných podle doby sledování televize v běžném dni (rovněž kategorizováno na tercily). V tabulce ověřujeme **hypotézu, podle níž počet přečtených knih roste s dobou strávenou s TV.**

Zatímco v kategorii podprůměrného množství přečtených knih (I. tercil) je pouze 16 %, těch kdo televizi sledují málo (I. tercil), tak studentů, kteří se na televizi dívají nadprůměrně často (III. tercil) je v této kategorii zhruba 2,5 krát více (39 %). Obdobně v kategorii nejvíce přečtených knih (III. tercil) je 42 % těch, kdo se na televizi v podstatě nedívají a zároveň jen 17% těch, kteří se na ní dívají velmi často (III. tercil). Vidíme tak, že mezi čtením knih a sledováním televize existuje negativní souvislost (vyjádřeno pomocí kontingenčního koeficientu souvislost je tato souvislost středně silná,  $CC = 0,25$ ). Naši hypotézu o pozitivní souvislosti mezi počtem přečtených knih a sledováním TV nám tedy nezbyvá než „zamítnout“ (resp. **ji na základě našich dat nemůžeme potvrdit**), **vztah je totiž přesně obrácený: čím více studenti čtou, tím méně se dívají na televizi.** ← To byl ale pouze popis vztahů v tabulce, následovat musí věcné vysvětlení, většinou ho odvozujeme z určité teorie. Zde alespoň krátký náznak:

Jedním z možných vysvětlení by mohl být vzájemný konkurenční vztah konzumace médií, den má jen 24 hodin a člověk se musí rozhodnout, jaké médium zvolí... Můžeme také uvažovat o tom, že tato média reprezentují poněkud odlišný kulturní svět (knihy má blízko k intelektuální – vysoké kultuře, zatímco televize spíše k masové – zábavné kultuře) a tak v pozadí výběru jednoho nebo druhého média může působit odlišný kulturní vkus. Nicméně k detailnějšímu posouzení těchto vysvětlení bude třeba přistoupit k dalším analýzám, které zhodnotí souvislosti čtení s jinými aktivitami volného času a zejména ověří, zda tento vztah je platný v rámci různých sociálních skupin (např. dle pohlaví, věku apod.).

**Při interpretaci dat nebud'te pouze popisní, mechanicky nekopírujte, co čtenář vidí v tabulce.**

**Vyberte pouze podstatné vztahy z hlediska vaší hypotézy – ty které ji potvrzují a zejména ty, které ji vyvrací (!).**

**Interpretujete jen skutečně věcně významné rozdíly – zamyslete se, jak velké jsou rozdíly v závislé–vysvětlované proměnné mezi kategoriemi nezávislé proměnné, pokud vezmete v úvahu metriku (počty či vzdálenosti kategorií; u numerických znaků směrodatnou odchylku) závislé proměnné.**

**Nabízejte věcná vysvětlení zasazená do kontextu toho, co zkoumáte.**

**Ale pozor na interpretaci vztahů jako kauzálních: „X ovlivňuje Y“. Takový to závěr, resp. ověření hypotézy o kauzálním – příčinném vlivu, vždy vyžaduje mnohem více než jednu tabulku souvislosti dvou proměnných (viz Opora KMVP s. 24-33).**

**Pište čtivý text, který nebude čtenáře nudit.**

Prezentace tabulek v textu (např. diplomové práce nebo výzkumné zprávy) má svá pravidla. Povšimněte si zvýrazněných částí v předchozí tabulce 2. Vždy je nutné uvádět absolutní počet, z něž počítáte procenta a také by mělo být jasné, jaký je podíl chybějících hodnot (tj. kolik respondentů neodpovědělo). Více si o základních pravidlech pro úpravu tabulek i to, jak z nich psát text, povíme na semináři.

Pro zajímavost, předchozí tabulku bylo nejprve třeba upravit, původně vypadala ve výstupu SPSS takto:

| Case Processing Summary                    |  |       |         |         |         |        |
|--|--|-------|---------|---------|---------|--------|
| Cases                                      |  |       |         |         |         |        |
|  |  | Valid |         | Missing |         | Total  |
|  |  | N     | Percent | N       | Percent | N      |
| knihy_celk3t * TV3t Sledování TV (tercily) |  | 63    | 87,5%   | 9       | 12,5%   | 72     |
|  |  |       |         |         |         | 100,0% |

| knihy_celk3t * TV3t Sledování TV (tercily) Crosstabulation |               |                                      |             |              |               |        |
|--|---------------|--------------------------------------|-------------|--------------|---------------|--------|
|  |               | TV3t Sledování TV (tercily)          |             |              |               |        |
|  |               |                                      | 1 I. tercil | 2 II. tercil | 3 III. tercil | Total  |
| knihy_celk3t   | 1 I. tercil   | Count                                | 3           | 6            | 7             | 16     |
|  |               | % within TV3t Sledování TV (tercily) | 15,8%       | 23,1%        | 38,9%         | 25,4%  |
|  | 2 II. tercil  | Count                                | 8           | 12           | 8             | 28     |
|  |               | % within TV3t Sledování TV (tercily) | 42,1%       | 46,2%        | 44,4%         | 44,4%  |
|  | 3 III. tercil | Count                                | 8           | 8            | 3             | 19     |
|  |               | % within TV3t Sledování TV (tercily) | 42,1%       | 30,8%        | 16,7%         | 30,2%  |
| Total  |               | Count                                | 19          | 26           | 18            | 63     |
|  |               | % within TV3t Sledování TV (tercily) | 100,0%      | 100,0%       | 100,0%        | 100,0% |

## Uspořádání v kontingenční tabulce

Než si ukážeme, jak kontingenční tabulku vytvořit v prostředí programu PSPP/SPSS připomeňme si obecný postup a logiku vytváření **kontingenčních tabulek** (viz obrázek Uspořádání tabulky). Dodejme, že dále se budeme zabývat pouze tabulkami pro kategoriální znaky, tj. situací, kdy počítáme s absolutními nebo relativními četnostmi (N; procenta nebo pravděpodobnost).<sup>4</sup>

Základní princip sestavení kontingenční tabulky:

**V kategoriích nezávislé proměnné ukazujeme kompletní (100 %) distribuci závislé proměnné.**

*závislá proměnná* = je v hypotéze ovlivňována, způsobována (nejčastěji je v řádcích)  
*nezávislá(é) proměnná* = vysvětluje, ovlivňuje závislou (ne nutně kauzálně–příčinně)

Ale pozor! **Směr kauzality je vždy věcí teorie, nelze ji určit z dat samotných.**

| Uspořádání tabulky  |                          |            |           |                |
|---|--------------------------|------------|-----------|----------------|
| sloupcová procenta:   |                          |            |           |                |
| V kategoriích nezávislé proměnné ukazujeme kompletní (100 %) distribuci závislé proměnné. |                          |            |           |                |
| ZÁVISLÁ -<br>vysvětlovaná   | NEZÁVISLÁ - vysvětlující |            |           |                |
|   | Pohlaví                  |            |           | Celkový součet |
|   | Spokojenost              | Muž        | Žena      |                |
|   | 1 (nespokojen)           | 41 % (5)   | 22 % (2)  | 7              |
|   | 2                        | 41 % (5)   | 11 % (1)  | 6              |
|   | 3 (spokojen)             | 16 % (2)   | 66 % (6)  | 8              |
| Celkový součet  |                          | 100 % (12) | 100 % (9) | 21             |

Nejčastěji bývá **závislá** proměnná nalevo v řádcích a **nezávislá** (vysvětlující) ve sloupcích.      Praktikum KMVP část 2      18

<sup>4</sup> Tabulky pochopitelně mohou zobrazovat i jiné míry, např. střední hodnoty pro kardinální znaky (průměr, medián). Logika je podobná, dokonce na pochopení je to mnohem jednodušší. Každý jistě rozumí principu porovnání průměrů v podskupinách a také ví, jak ho spočítat. Třeba pokud jde o rozdíl v průměrném příjmu čtyř skupin zaměstnanců podle vzdělání, což je v podstatě stejná úloha jako našem následujícím příkladu, kde ovšem máme příjem kategorizován do čtyř skupin na základě kvintilů a tudíž musíme vztah vzdělání – příjem analyzovat pomocí kontingenční tabulky.

Uved'me si to na dalším příkladu, kde máme opět dva ordinální znaky, což většinou znamená, že očekáváme jejich prakticky vzato lineární souvislost (nebo alespoň monotóní vztah).

**tabulku ukazující závislost příjmu na vzdělání:** příjem (kvartily) a vzdělání (4 kategorie), data reprezentují dospělou populaci ČR v roce 2007 a pochází z mezinárodního výzkumného projektu ISSP 2007 (proveden byl náhodný výběr v populaci starší 18-let).

prijem4 Příjem-osobní – kvartily (s24) \* vzd4 Vzdělání Crosstabulation

| % within vzd4 Vzdělání                 |                   | vzd4 Vzdělání |        |        |        | Total  |
|--|-------------------|---------------|--------|--------|--------|--------|
|  |                   | 1 ZŠ          | 2 VYUČ | 3 SŠ   | 4 VŠ   |        |
| prijem4 Příjem-osobní - kvartily (s24) | 1 I. do 7 tis.    | 62,6%         | 26,4%  | 28,9%  | 19,6%  | 31,6%  |
|  | 2 II. 7-9 tis.    | 26,2%         | 31,9%  | 15,7%  | 17,6%  | 24,1%  |
|  | 3 III. 9-15 tis.  | 7,5%          | 29,7%  | 29,2%  | 21,6%  | 26,2%  |
|  | 4 IV. nad 15 tis. | 3,7%          | 11,9%  | 6,2%   | 41,2%  | 18,1%  |
| Total                                  |                   | 100,0%        | 100,0% | 100,0% | 100,0% | 100,0% |

\*V kategoriích nezávislé proměnné (zde Vzdělání) ukazujeme kompletní (100 %) distribuci závislé proměnné (prijem4).

\*Informace z tabulky interpretujeme tak, že porovnáme navzájem podskupiny nezávislé proměnné (zde stupně vzdělání) podle vlastností závislé proměnné (zde kvartily příjmu)

→ **tabulku čteme „po řádcích“** (pokud máme nezávislý znak ve sloupcích, závislý v řádcích a sloupcová procenta jako zde, což je nejobvyklejší, ale lze to cele otočit o 90st).

A nezapomeňte, před sestavením kontingenční tabulky vždy nejprve formulujte výzkumnou otázku (případně i odpověď na ní ve formě hypotézy). Ta vám definuje závislou a nezávislou proměnnou (a případně i další – kontrolní).

Nyní, jsa vybaveni „teoretickými“ znalostmi, jak se sestavuje a interpretuje kontingenční tabulka, můžeme konečně přistoupit k analýze dat v prostředí programů PSPP nebo SPSS. Příklad ukazují v PSPP. V SPSS je to obdobné, logika grafického rozhraní je podobná, jen zde najdete mnohem větší nabídku, což je vidět již na první pohled, kdy v SPSS máte mnohem více různých ikon ke klikání (což jsou vlastně takové zkratky do prostředí menu). Zadávaní pomocí příkazového řádku je u těchto základních analýz a procedur stejný (obecně k tomuto principu viz kapitolu Syntax – ovládání pomocí příkazů (úvod) a na konci máte i přehled příkazů k zadávání kontingenčních tabulek). Nanejdříve se může stát, že v PSPP nebude fungovat něco, na co jsme zvyklí z SPSS, ale PSPP vám v takovém případě většinou oznámí že „tento příkaz zatím ještě není podporován“. Takže připomínám, pravidelně si aktualizujte verzi PSPP (je to free-projekt v neustálém vývoji).

## Postup vytváření kontingenčních tabulek v PSPP (SPSS)

Nejprve si otevřeme datový soubor, zde je to ISSP2007\_v2\_1.sav.

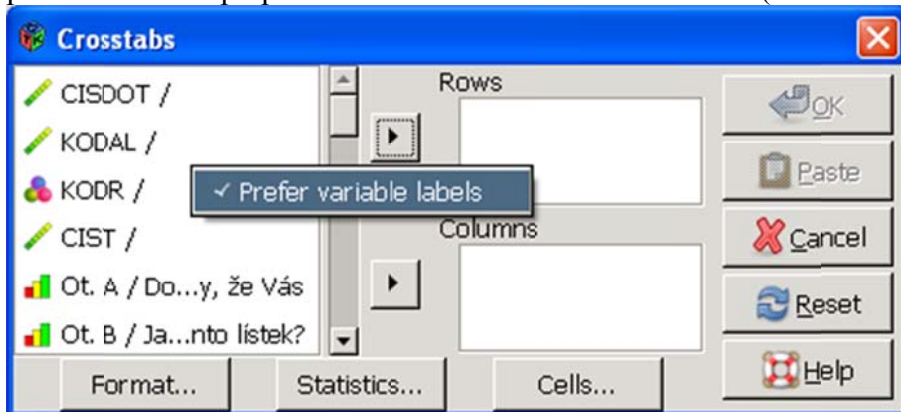
V PSPP přes menu zadáme:

Analyze → Descriptive Statistics → Crosstabs.

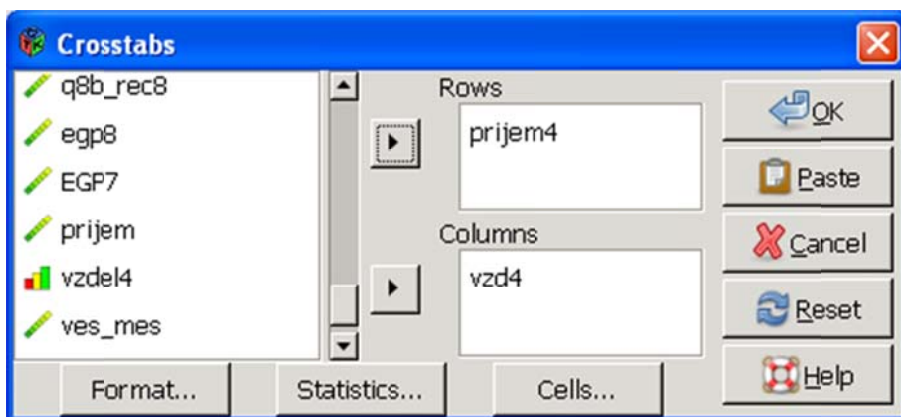
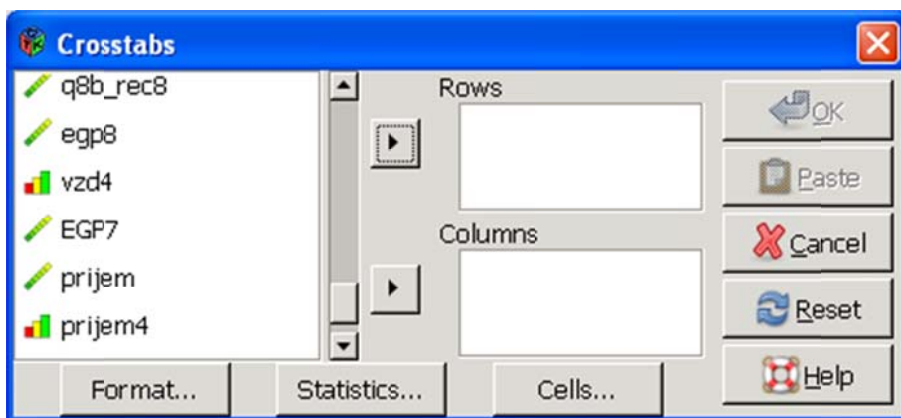
The screenshot shows the PSPP Data Editor window titled "ISSP2007\_v2\_1.sav – PSPPIRE Data Editor". The menu bar includes File, Edit, View, Data, Transform, **Analyze**, Utilities, Windows, and Help. The 'Analyze' menu is open, showing options: Descriptive Statistics, Compare Means, Bivariate Correlation..., Factor Analysis, Reliability, Linear Regression, Non-Parametric Statistics, and ROC Curve... The 'Descriptive Statistics' submenu is also open, showing Frequencies, Descriptives, Explore, and **Crosstabs**. The data table below has 17 rows and 7 columns. The first column is 'Name', the second is 'Type', and the next five columns are 'C1', 'C2', 'C3', 'C4', and 'C5'. The last column is 'Label'. The 'Name' column contains values from 1 to 17. The 'Type' column contains 'Numeric' for all rows. The 'Label' column contains labels for each row, such as 'C1 / Dostal(a)', 'C2 / Jakou bar', etc.

| Name | Type    | C1 | C2 | C3 | C4 | C5 | Label            |
|------|---------|----|----|----|----|----|------------------|
| 1    | Numeric |    |    |    |    |    | C1 / Dostal(a)   |
| 2    | Numeric |    |    |    |    |    | C2 / Jakou bar   |
| 3    | Numeric |    |    |    |    |    | C3 / Kolik členů |
| 4    | Numeric | 9  | 0  |    |    |    | C4 / Věk členů   |
| 5    | Numeric | 1  | 0  |    |    |    | C5 / Věk členů   |
| 6    | Numeric | 1  | 0  |    |    |    | C6 / Věk členů   |
| 7    | Numeric | 1  | 0  |    |    |    | C7 / Věk členů   |
| 8    | Numeric | 1  | 0  |    |    |    | C8 / Věk členů   |
| 9    | Numeric | 1  | 0  |    |    |    | C9 / Věk členů   |
| 10   | Numeric | 1  | 0  |    |    |    | C10 / Věk členů  |
| 11   | Numeric | 1  | 0  |    |    |    | C11 / Věk členů  |
| 12   | Numeric | 2  | 0  |    |    |    | C12 / Věk členů  |
| 13   | Numeric | 2  | 0  |    |    |    | C13 / Věk členů  |
| 14   | Numeric | 2  | 0  |    |    |    | C14 / Věk členů  |
| 15   | Numeric | 2  | 0  |    |    |    | C15 / Věk členů  |
| 16   | Numeric | 2  | 0  |    |    |    | C16 / Věk členů  |
| 17   | Numeric | 2  | 0  |    |    |    | C17 / Věk členů  |

Pro jednodušší orientaci si můžeme pomoci myši – najedeme na seznam proměnných – a jejího pravého tlačítka přepnout na zobrazování Variable Names (místo labels).

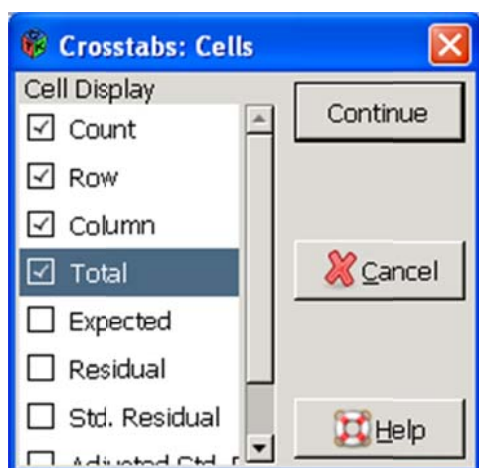


Najdeme si proměnné, které chceme do tabulky a do řádků = Rows dáme **závislou (zde příjem4)** a do sloupců = **Rows (vzdel4)**

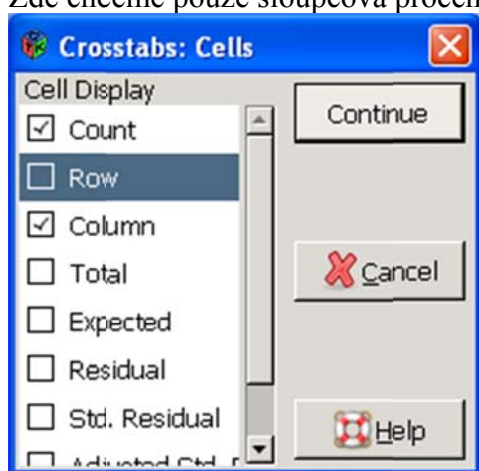


Dále musíme nastavit, co bude v políčkách tabulky v sekci **Cells**:  
 PSPP má přednastaveny všechny možnosti (četnosti, řádková, sloupcová i celková %), vhodné je vybrat pouze to, co budeme potřebovat, aby tabulka byla přehledná.



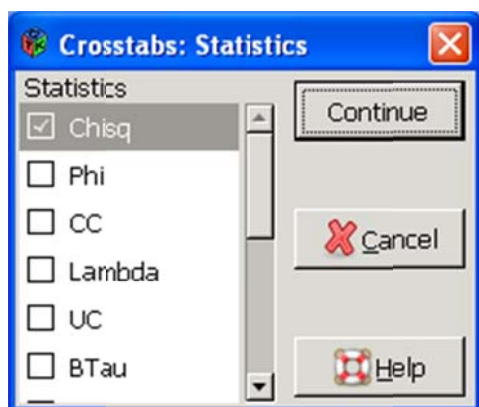


Zde chceme pouze sloupcová procenta (**Column**) a ponecháme též absolutní četnosti (**Count**).



### Statistics: Statistické testy a míry asociace

Pokud ponecháme nastavení – zakliknuto je **Chisq**, tak se nám ve výstupu rovněž objeví základní test asociací v tabulce tzv. Test dobré shody Chisquare (Chíkvadrát), který testuje nulovou hypotézu, že kategorie proměnných jsou na sobě nezávislé. Můžeme ho vypnout, ale doporučuji ponechat, je to základní test, interpretace je jednoduchá (navíc je to test neparametrický, takže je vhodný i na malé výběrové soubory). Hlídat ale musíme dostatečné četnosti v jednotlivých polích tabulky (viz dále).



**Výstup (output)** a jak ho čteme:

Nejprve dostaneme tabulku Summary informující nás o počtu platných případů (průnik za obě proměnné v kontingenční tabulce).

Summary.

|   | Cases |         |         |         |       |         |
|---|-------|---------|---------|---------|-------|---------|
|   | Valid |         | Missing |         | Total |         |
|   | N     | Percent | N       | Percent | N     | Percent |
| Příjem-osobní – kvartily (s24) * Vzdělání | 843   | 69.0%   | 379     | 31.0%   | 1222  | 100.0%  |

Zde vidíme, že z celkového počtu případů (respondentů ve výzkumu) 1222 odpovědělo na obě otázky pouze 843, takže máme 31% missingů – chybějících hodnot, což je hodně. Nepísané pravidlo pro velké výběrové soubory (cca nad 100 případů) říká, že je-li chybějících hodnot do 5%, pak to můžeme ignorovat, jinak bychom měli zjistit, „kdo jsou ti, kdo neodpovídají“. Zde je to způsobeno odpovědí na otázku po příjmu (to by nám řekla tabulka třídění prvního stupně pomocí FREQUENCIES – viz návod na PSPP část 1).

Hlavní tabulka, která nás zajímá a kterou později upravíme do zprávy/ diplomky je samotná **kontingenční tabulka**:

prijem4 \* vzd4 [count, column %].

| prijem4         | vzd4            |                 |                 |                | Total           |
|-----------------|-----------------|-----------------|-----------------|----------------|-----------------|
|                 | ZŠ              | VYUČ            | SŠ              | VŠ             |                 |
| I. do 7 tis.    | 67,0<br>62,6%   | 95,0<br>26,4%   | 94,0<br>28,9%   | 10,0<br>19,6%  | 266,0<br>31,6%  |
| II. 7–9 tis.    | 28,0<br>26,2%   | 115,0<br>31,9%  | 51,0<br>15,7%   | 9,0<br>17,6%   | 203,0<br>24,1%  |
| III. 9–15 tis.  | 8,0<br>7,5%     | 107,0<br>29,7%  | 95,0<br>29,2%   | 11,0<br>21,6%  | 221,0<br>26,2%  |
| IV. nad 15 tis. | 4,0<br>3,7%     | 43,0<br>11,9%   | 85,0<br>26,2%   | 21,0<br>41,2%  | 153,0<br>18,1%  |
| Total           | 107,0<br>100,0% | 360,0<br>100,0% | 325,0<br>100,0% | 51,0<br>100,0% | 843,0<br>100,0% |

Tabulka v sobě obsahuje jak relativní podíly (%) tak absolutní četnosti. Můžeme je s výhodou rozdělit při zadávání na dvě – provedeme analýzu 2x za sebou jednou s Count, podruhé s Column.

Samotné absolutní četnosti vypadají takto (zde pozor na nesmyslné desetinné místo):

prijem4 \* vzd4 [count].

| prijem4         | vzd4  |       |       |      | Total |
|-----------------|-------|-------|-------|------|-------|
|                 | ZŠ    | VYUČ  | SŠ    | VŠ   |       |
| I. do 7 tis.    | 67,0  | 95,0  | 94,0  | 10,0 | 266,0 |
| II. 7–9 tis.    | 28,0  | 115,0 | 51,0  | 9,0  | 203,0 |
| III. 9–15 tis.  | 8,0   | 107,0 | 95,0  | 11,0 | 221,0 |
| IV. nad 15 tis. | 4,0   | 43,0  | 85,0  | 21,0 | 153,0 |
| Total           | 107,0 | 360,0 | 325,0 | 51,0 | 843,0 |

Tuto tabulku absolutních četností nemůžeme interpretovat, protože četnosti ve spoluvýskytu kategorií jsou závislé na tzv. marginálních četnostech, tj. rozložení kategorií obou znaků.

**Ale nezapomeňte zde kontrolovat, že v každém políčku tabulky máte alespoň cca 5 případů (v absolutní četnosti).** Pokud ne, tak můžete některé kategorie sloučit (minimum je 2 x 2 tabulka), například pomocí příkazu RECODE (o tom v 3 dílu návodu na PSPP).

Samotné sloupcové četnosti vypadají takto:

prijem4 \* vzd4 [column %].

| prijem4         | vzd4   |        |        |        | Total  |
|-----------------|--------|--------|--------|--------|--------|
|                 | ZŠ     | VYUČ   | SŠ     | VŠ     |        |
| I. do 7 tis.    | 62,6%  | 26,4%  | 28,9%  | 19,6%  | 31,6%  |
| II. 7–9 tis.    | 26,2%  | 31,9%  | 15,7%  | 17,6%  | 24,1%  |
| III. 9–15 tis.  | 7,5%   | 29,7%  | 29,2%  | 21,6%  | 26,2%  |
| IV. nad 15 tis. | 3,7%   | 11,9%  | 26,2%  | 41,2%  | 18,1%  |
| Total           | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

**Interpretace** provádíme takto: nejprve zkontrolujeme absolutní četnosti (v Count), zda máme dostatečné množství případů v políčkách tabulky. Pak čteme sloupcová procenta a to pořádkem porovnáváme, zda se liší % zastoupení v příjmové kategorii podle vzdělání. Opakuji grafický příklad:

% within vzd4 Vzdělání

|  |                   | vzd4 Vzdělání |        |        |        | Total  |
|--|-------------------|---------------|--------|--------|--------|--------|
|  |                   | 1 ZŠ          | 2 VYUČ | 3 SŠ   | 4 VŠ   |        |
| prijem4 Příjem-osobní - kvartily (s24) | 1 I. do 7 tis.    | 62,6%         | 26,4%  | 28,9%  | 19,6%  | 31,6%  |
|  | 2 II. 7–9 tis.    | 26,2%         | 31,9%  | 15,7%  | 17,6%  | 24,1%  |
|  | 3 III. 9–15 tis.  | 7,5%          | 29,7%  | 29,2%  | 21,6%  | 26,2%  |
|  | 4 IV. nad 15 tis. | 3,7%          | 11,9%  | 26,2%  | 41,2%  | 18,1%  |
| Total                                  |                   | 100,0%        | 100,0% | 100,0% | 100,0% | 100,0% |

Také můžeme sledovat kupení četností na diagonále, které indikuje lineární závislost.

|  |                   | vzd4 Vzdělání |        |        |        | Total  |
|--|-------------------|---------------|--------|--------|--------|--------|
|  |                   | 1 ZŠ          | 2 VYUČ | 3 SŠ   | 4 VŠ   |        |
| prijem4 Příjem-osobní – kvartily (s24) | 1 I. do 7 tis.    | 62,6%         | 26,4%  | 28,9%  | 19,6%  | 31,6%  |
|  | 2 II. 7–9 tis.    | 26,2%         | 31,9%  | 15,7%  | 17,6%  | 24,1%  |
|  | 3 III. 9–15 tis.  | 7,5%          | 29,7%  | 29,2%  | 21,6%  | 26,2%  |
|  | 4 IV. nad 15 tis. | 3,7%          | 11,9%  | 26,2%  | 41,2%  | 18,1%  |
| Total                                  |                   | 100,0%        | 100,0% | 100,0% | 100,0% | 100,0% |

Zde se zdá, že vyšší příjmové kategorie nacházíme u vzdělanějších respondentů, tedy že příjem roste se vzděláním. Ale podrobněji viz dále znaménkové schéma.

Tabulku lze pochopitelně otočit o 90st., aniž by se změnil její význam, ale při zadání musíme **prohodit proměnné** (vzd4 a prijem4) **z řádků na sloupce** a zadat, že chceme **řádková procenta (Row)**. Pak čteme tabulku „po sloupcích“.

vzd4 \* prijem4 [row %].

| vzd4  | prijem4      |              |                |                 | Total  |
|-------|--------------|--------------|----------------|-----------------|--------|
|       | I. do 7 tis. | II. 7–9 tis. | III. 9–15 tis. | IV. nad 15 tis. |        |
| ZŠ    | 62,6%        | 26,2%        | 7,5%           | 3,7%            | 100,0% |
| VYUČ  | 26,4%        | 31,9%        | 29,7%          | 11,9%           | 100,0% |
| SŠ    | 28,9%        | 15,7%        | 29,2%          | 26,2%           | 100,0% |
| VŠ    | 19,6%        | 17,6%        | 21,6%          | 41,2%           | 100,0% |
| Total | 31,6%        | 24,1%        | 26,2%          | 18,1%           | 100,0% |

V principu to tedy vypadá takto:

| % within vzd4 Vzdělání |        | prijem4 Příjem-osobní – kvartily (s24) |                |                  |                   | Total  |
|------------------------|--------|--|----------------|------------------|-------------------|--------|
|                        |        | 1 I. do 7 tis.                         | 2 II. 7–9 tis. | 3 III. 9–15 tis. | 4 IV. nad 15 tis. |        |
| vzd4 Vzdělání          | 1 ZŠ   | 62,6%                                  | 26,2%          | 7,5%             | 3,7%              | 100,0% |
|                        | 2 VYUČ | 26,4%                                  | 31,9%          | 29,7%            | 11,9%             | 100,0% |
|                        | 3 SŠ   | 28,9%                                  | 15,7%          | 29,2%            | 26,2%             | 100,0% |
|                        | 4 VŠ   | 19,6%                                  | 17,6%          | 21,6%            | 41,2%             | 100,0% |
| Total                  |        | 31,6%                                  | 24,1%          | 26,2%            | 18,1%             | 100,0% |

Udělat správně zorientovanou tabulku nemusí stačit. Plně dostačuje pouze tam, kde máme data o úplné populaci (tj. census). Se samotnou tabulkou si také vystačíme v situaci, kdy nám jde o pouhou exploraci vztahů bez nároku na přesnější závěry, ale pokud máme data pocházející pouze z nějakého **reprezentativního výběru z populace** (zde např. pomocí dotazníkového šetření) a chceme zobecnit výsledky našeho výzkumu, tedy chceme-li tvrdit, že rozdíly mezi kategoriemi naměřené v našem výběrovém vzorku platí i pro celou zkoumanou populaci, pak musíme provést ještě statistický test této závislosti. O principech statistické inference (tj. o zobecňování ze vzorku na populaci) se dočtete v učebnicích statistiky, ale mějte vždy namysli, že pořád platí:

**Robustnost našich závěrů by měla být vždy patrná již ze samotných tabulek bez nějakých rádob „kouzelných“ statistických testů. Vždy proto sledujte věcnou významnost výsledků** (tj. „jak jsou velké“ a co znamenají ve vztahu ke zkoumanému fenoménu).

## Test dobré shody Chí kvadrát – (ne)závislost proměnných v kontingenční tabulce

Postup ověřování statistické hypotézy si ukážeme na našem příkladu vztahu dvou kategoriálních proměnných v kontingenční tabulce (dodejme, že existuje mnoho jiných postupů a statistických testů pro různé typy proměnných a forem jejich vzájemných vztahů).

Ale zpět do outputu. Ve výstupu máme ještě **tabulku s výsledkem statistického testu nezávislosti**.

Statistická tzv. nulová hypotéza zde říká, že příjem je na vzdělání nezávislý.

Hodnota testového kritéria Chíkvadrát je 122,42 při 9 stupních volnosti (df) a *p*-hodnota (Asympt.signif.) je 0,00.

Nulovou hypotézu zde nemůžeme přijmout neboť **dosažená hladina významnosti *p* je menší než 0,05** (což odpovídá zvolené chybě 5%). Znamená to, že i pro celou populaci ČR platí, že alespoň jedna kategorie příjmu je ovlivněna alespoň jednou kategorií vzdělání. Tento test nám ale neříká, které kategorie to jsou, ani kolik jich je (zda platí zde kvazi-lineární závislost, nebo zda je závislost jen mezi dvěma kategoriemi).

Intuitivně na to můžeme usoudit z rozdílů mezi sloupcovými procenty, čteme-li tabulku po řádcích. Můžeme použít statistický test, který by nám dále ukázal kde je souvislost a jak je silná, když si do políček tabulky přidáme tzv. Adjustované residua (viz dále).

Zdůrazňuji, že **smyslem testování hypotéz je usuzovat na to, zda vztahy ve vašem výběrovém souboru (vzorku, zde 1222 respondentů ve výzkumu ISSP) s přijatelnou mírou chyby (nejčastěji volíme 5 %) platí i pro celou populaci**. Nic jiného! **Neříká nic o tom, jak moc je vztah mezi proměnnými těsný**. K tomu bychom museli ještě spočítat nějakou míru souvislosti – těsnosti vzájemného vztahu (u ordinálních znaků i směru), například koeficient kontingence nebo pořadovou korelaci (v obou případech existují různé varianty těchto koeficientů), anebo se detailněji podívat „dovnitř“ tabulky, což uděláme v další části pomocí tzv. znaménkového schématu.

Navíc, k tomu abychom mohli tento postup používat, musí být splněno mnoho podmínek, mj.: výběr vzorku musí být **náhodný**, dostatečně velký (minimálně 30-50 případů, ale záleží na velikosti populace), z dostatečně velké populace, na druhé straně nesmí však být příliš velký (protože pak jsou všechny, byť triviální, souvislosti statisticky významné) a v polích tabulky musí být dostatečný počet případů (nesmí být prázdná).

K principu testování hypotéz viz více článek P. Soukupa Soukup, P. 2007. "Statisticky významný neznamená důležitý." na <http://www.socioweb.cz/index.php?disp=teorie&shw=298&lst=108>

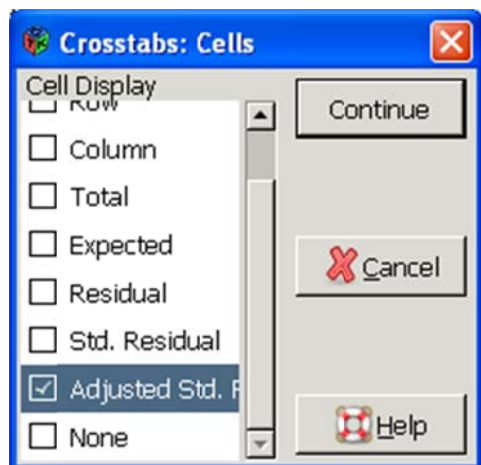
### Chi-square tests.

| Statistic                    | Value  | df | Asymp. Sig. (2-sided) |
|------------------------------|--------|----|-----------------------|
| Pearson Chi-Square           | 122,42 | 9  | ,00                   |
| Likelihood Ratio             | 123,44 | 9  | ,00                   |
| Linear-by-Linear Association | 70,15  | 1  | ,00                   |
| N of Valid Cases             | 843    |    |                       |

## Znaménkové schéma – v kterých polích tabulky je jak silná souvislost?

Při interpretaci souvislosti v tabulce používáme nejčastěji sloupcová procenta (viz výše), s výhodou lze využít tzv. Adjustovaných residuí, na jejichž základě se utvoří znaménkové schéma odchylek v polích. Jde o jednoduchý způsob k zčitelnění % odchylek od tzv. teoretického rozdělení četností v málo přehledné kontingenční tabulce tak, aby v ní byla rychlá optická orientace. Procentní odchylky převádíme na znaménka „+“ a „–“, podle vypočtené veličiny „z“ [Šafář 1969: 566; podrobně viz Řehák, Řeháková 1986].

V CROSSTABS si zadáme Adjustovaná standardizovaná residua (**Adjusted Std. Residuals**)



Výsledkem může být buď samostatná tabulka jako zde nebo další řádek v kontingenční tabulce:

prijem4 \* vzd4 [adj. resid.].

| prijem4         | vzd4 |      |      |      | Total |
|-----------------|------|------|------|------|-------|
|                 | ZŠ   | VYUČ | SŠ   | VŠ   |       |
| I. do 7 tis.    | 7,4  | -2,8 | -1,3 | -1,9 | ,0    |
| II. 7–9 tis.    | ,5   | 4,6  | -4,5 | -1,1 | ,0    |
| III. 9–15 tis.  | -4,7 | 2,0  | 1,6  | -,8  | ,0    |
| IV. nad 15 tis. | -4,1 | -4,0 | 4,8  | 4,4  | ,0    |
| Total           |      |      |      |      |       |

Platí, že čím vyšší hodnoty (uvažujeme ale **pouze vyšší než 1,96**), tím větší souvislost mezi kategoriemi nezávislé a závislé proměnné. Buď jsou kategorie v tom kterém políčku nadreprezentovány (kladné hodnoty) nebo podreprezentovány (záporné hodnoty) oproti teoretickému očekávání (to zohledňuje odlišně zastoupení kategorií uvažovaných znaků).



Graficky to lze vyjádřit pomocí tzv. **znaménkového schématu**, které slouží ke zčitelnější odchylnosti od tzv. teoretického rozdělení četností v málo přehledné kontingenční tabulce tak, aby ní byla rychlá optická orientace.

% odchylky převádíme na znaménka „+“ a „-“

- kde  $\text{abs}(z) \geq 3.29$  nahradíme +++ resp. ---,
- kde  $\text{abs}(z) \geq 2.58$  nahradíme ++ resp. --,
- kde  $\text{abs}(z) \geq 1.96$  nahradíme + resp. -.
- pro  $\text{abs}(z) < 1.96$  dáme 0

což odpovídá klíči:

0 = statisticky nevýznamné

+, - = významná odchylka na 5 % hladině statistické významnosti

++, -- = významnost od 0,1 % do 1 %,

+++, ---, = pravděpodobnost náhodného výskytu odchylky menší než 0,1 %

V našem případě to pak vypadá takto (vytvořeno v SPSS, ale znaménka lze dodatečně udělat v Excelu či Wordu – stačí přepsat Adjustované residua podle výše uvedeného pravidla):

|                                |                   | vzd4 Vzdělání |        |      |      |
|--------------------------------|-------------------|---------------|--------|------|------|
|                                |                   | 1 ZŠ          | 2 VYUČ | 3 SŠ | 4 VŠ |
| příjem4                        | 1 I. do 7 tis.    | +++           | --     | o    | o    |
| Příjem-osobní – kvartily (s24) | 2 II. 7–9 tis.    | o             | +++    | ---  | o    |
|                                | 3 III. 9–15 tis.  | ---           | +      | o    | o    |
|                                | 4 IV. nad 15 tis. | ---           | ---    | +++  | +++  |

Interpretace znamének je ve smyslu „nacházíme výrazně více lidí s určitou vlastností (zde kategorií vzdělání) v daných polích tabulky oproti tzv. očekávané četnosti“. Nejde tedy o žádný průměr.

Nezapomeňte, než uděláte znaménkové schéma, nejprve byste měli provést Chikvadrát test pro celou tabulku (což jsme udělali v předchozí části). Zjistit, zda můžete zamítnout nulovou hypotézu o tom, že žádná kategorie v tabulce se „statisticky významně“ neodlišuje od ostatních (tedy můžeme-li přijmout tzv. alternativní hypotézu, že alespoň v jednom poli se zjištěná četnost odlišuje). Připomínám, že tento test říká, že tento rozdíl platí nejen pro data v našem vzorku, ale že ho nalezneme i v celé populaci.

## Zadání kontingenční tabulky pomocí syntaxu PSPP/SPSS

Následující text lze zkopírovat do Syntax editoru v PSPP/SPSS (v menu: File / New / Syntax ) a spustit (pomocí Run nebo Ctrl+R). Nejprve je třeba **otevřít datový soubor** ISSP2007\_v2\_1.sav.

Tyto příkazy také lze použít při analýze jakýchkoliv jiných dat, stačí jen změnit názvy proměnných (*prijem4* a *vzd4*).

\*zadání kontingenční tabulky v SPSS / PSPP:  
pro **COL = sloupcová procenta.**

```
CROSSTAB prijem4 BY vzd4 /CELL=COL COUNT.
```

\*pro test nezávislosti Chí kvadrát přidáme.

```
CROSSTAB prijem4 BY vzd4 /CELL=COL COUNT /STATISTICS=CHISQ
```

\*pro přidání adjustovaných residuí ještě.

```
CROSSTAB prijem4 BY vzd4 /CELL=COL COUNT ASRESID /STATISTICS=CHISQ.
```

\*tabulky % a N samostatně dostaneme takto.

```
CROSSTAB prijem4 BY vzd4 /CELL= COUNT.
```

```
CROSSTAB prijem4 BY vzd4 /CELL= COL.
```

```
CROSSTAB prijem4 BY vzd4 /CELL= ASRESID.
```

\*nebo otočeno o 90st, tedy pro **ROW = řádková procenta** a **prohodíme proměnné!**.

```
CROSSTAB vzd4 BY prijem4 /CELL=ROW COUNT.
```

```
CROSSTAB vzd4 BY prijem4 /CELL=ROW COUNT ASRESID.
```