

## 1. Načtení a příprava dat pro analýzu

---

- **Načtení souboru:**

*File → Open → Data → Look in (výběr cesty k souboru) → Files of type (výběr typu dat – .txt, .xlsx, .sav, ...) → File Name (výběr souboru) → Open*

1. **Soubor SPSS:**

GET

FILE='C:\...cesta...\nazev\_souboru.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

2. **Soubor z excelu:**

GET DATA /TYPE=XLSX

/FILE='C:\... cesta.....\nazev\_souboru..xlsx'

/SHEET=name 'nazev\_listu'

/CELLRANGE=full

/READNAMES=on

/ASSUMEDSTRWIDTH=32767.

EXECUTE.

DATASET NAME DataSet1 WINDOW=FRONT.

- **Uložení souboru:**

*File → Save as → ... (Ize vybrat typ souboru, ukládat jen některé proměnné, zda se mají např. v excelu ukládat hodnoty nebo popisky atd.)*

1. **Uložení jako .SAV:**

SAVE OUTFILE='C:\...cesta...\nazev\_souboru.sav'

/COMPRESSED.

2. **Uložení jako .XLSX:**

SAVE TRANSLATE OUTFILE='C:\...cesta...\nazev\_souboru.xlsx'

/TYPE=XLS

/VERSION=12

/MAP

/REPLACE

/FIELDNAMES

/CELLS=VALUES. /\* CELLS=LABELS v případě uložení popisků namísto hodnot.

- **Aktivace / zavření datového souboru:**

- Aktivace datového souboru

DATASET ACTIVATE DataSet1.

DATASET CLOSE DataSet1.

- **Popis („olejblování“) hodnot:**

Variable View → Values → ... (rozkliknout políčko se třemi tečkami) → zadání hodnoty (value) a popisku (label) → Add (při změně Change, při smazání popisku Delete)

**Pozn. 1:** Při zadávání textových popisků nedávat uvozovky, ty jsou tam doplněny automaticky.

**Pozn. 2:** Hodnoty a textové popisky pak lze přepínat v hlavním panelu pomocí tlačítka (zobrazení v datech):



**Pozn. 3:** Lze nastavit, zda chcete ve výstupech zobrazovat jen hodnoty, jen popis hodnot nebo obojí (Edit → Options → Output → Outline Labeling).

**Pozn. 4:** Obdobně se chová i popis proměnných (Variable View → Label → Zadej popis proměnné), lze editovat, vypisovat ve výstupech pouze popis místo názvu proměnné apod.

### 1. Příklad – popis hodnot:

VALUE LABELS stadium

1 'Stadium 1'

2 'Stadium 2'

3 'Stadium 3'

4 'Stadium 4'.

val lab pohlavi

1 'Muž'

2 'Žena'.

### 2. Příklad – popis proměnné:

VARIABLE LABELS stadium 'Stadium CRC při stanovení diagnózy'.

var lab stadium\_akt 'Stadium CRC při poslední kontrole'.

- **Seřazení souboru podle nějakého parametru:**

Data → Sort Cases → Sort by (přesun proměnné podle které chcete data seřadit, příp. i více proměnných) → Sort Order (zaškrtnout Ascending - A nebo Descending - D) → OK

SORT CASES by var1 (A).

SORT CASES by var 1 (A) var2 (D).

- **Identifikace duplikací:**

Data → Identify duplicate cases → Define matching cases by: (výběr proměnné/proměnných pro zjištění duplikací, např. ID) → Zaškrtnout Last case nebo First case in each group is primary (záleží na situaci)

**Pozn. 1:** Last case nebo First case in each group is primary – záleží na situaci, seřazení hodnot, pokud je cílem duplikované hodnoty smazat / filtrovat apod.

- **Vyfiltrování / smazání hodnot:**

Data → Select cases → If condition is satisfied (zadání podmínky, které hodnoty se mají vyfiltrovat nebo smazat) → Continue → Output (lze si vybrat druh výstupu: „Filter out unselected cases“, „Copy selected cases to a new dataset“ nebo „Delete unselected cases“) → OK

### 1. Příklad – vyfiltrování duplikovaných případů:

USE ALL.

COMPUTE filter\_\$(PrimaryFirst = 1).

VARIABLE LABELS filter\_\$(PrimaryFirst = 1 (FILTER)).

VALUE LABELS filter\_\$(0 'Not Selected' 1 'Selected'.

```
FORMATS filter_$. (f1.0).  
FILTER BY filter_$.  
EXECUTE.
```

... následné analýzy jsou provedeny pouze na vybraných případech, dokud se filtr nezruší...

```
USE ALL.
```

... dále se zase pracuje se všemi případy ...

nebo (je-li proměnná binární – 0/1):

```
USE ALL.  
FILTER BY PrimaryFirst.  
EXECUTE.
```

## 2. Příklad – smazání duplikovaných případů:

```
FILTER OFF.  
USE ALL.  
SELECT IF (PrimaryFirst = 1).  
EXECUTE.
```

## 3. Příklad – dočasná filtrace případů (POUZE pro následující proceduru):

```
TEMPORARY.  
SELECT IF vek < 60 AND pohlavi = 'M'.
```

```
temp.
```

```
sel if vek < 60 AND pohlavi = 'M'.
```

- **Spojování souborů:**

*Data → Merge Files → Add Variables (propojení dle proměnných) → výběr datového souboru (již otevřeného v SPSS nebo uloženého) → Continue → zaškrtnout Match cases on key variables → do kolonky Key Variables vybrat proměnné, podle kterých se proměnné spojují → výběr způsobu spojení → OK*

**Pozn. 1:** Oba datové soubory musí být seřazeny podle proměnných, podle kterých se spojují (!!!).

**Pozn. 2:** Způsob spojení datových souborů – výběr datového souboru, který se připojuje (vnitřní / vnější / křížové spojení).

**Pozn. 3:** Lze přidávat i případy – připojení nových řádků z jiné tabulky (*Data → Merge Files → Add Cases → ...*).

## 1. Příklad – připojení všech proměnných z DataSet2 k DataSet1 dle shodného identifikátoru pacienta (ID).

```
dataset activate DataSet1.  
MATCH FILES /FILE=*  
  /TABLE='DataSet2'  
  /BY ID.  
EXECUTE.
```

- **Agregace** (např. když pro každého pacienta máme víc řádků v tabulce (např. opakovaný rozbor krve) a chceme vytvořit novou tabulku, kde bude jeden pacient jen na jednom řádku (každá složka krve pak bude sumarizována mediánem či průměrem naměřených hodnot):

*Data → Aggregate → Break Variables (výběr proměnných, podle kterých se bude soubor agregovat – např. ID pacienta + datum vyšetření) → Summaries of Variables (výběr funkce, co a jak se bude agregovat) → lze vytvořit nový datový soubor nebo přidat agregovanou proměnnou jako nový sloupeček v původních datech → OK*

```
DATASET DECLARE agregovany_soubor.  
AGGREGATE  
  /OUTFILE=' agregovany_soubor '  
  /PRESORTED  
  /BREAK=ID datum_vysetreni  
  /N_BREAK = N  
  /glykemie_min = MIN(glykemie)  
  /glykemie_max = MAX(glykemie)  
  /glykemie_mean = MEAN(glykemie).
```

- **Transformace proměnných:**

**1. Logaritmická transformace:**

*Transform → Compute Variable... → v aritmetických funkcích (Function group: Arithmetic) vybrat Ln*

```
COMPUTE naklady.LN=LN(naklady).  
COMPUTE naklady.LOG10=LOG10(naklady).  
EXECUTE.
```

**2. Standardizace:**

*Analyze → Descriptive Statistics → Descriptives → zatrhnout Save standardized values as variables*

```
DESCRIPTIVES VARIABLES=naklady  
  /SAVE  
  /STATISTICS=MEAN STDDEV MIN MAX.
```

**3. Kategorizace:**

*Transform → Visual Binning...*

**4. Překódování textových proměnných na proměnné s číselnými kategoriemi:**

*Transform → Automatic Recode... (pokud jsou v datech chybějící hodnoty, je vhodné zatrhnout „Treat blank string values as user-missing“)*