

# TRANSFORMACE A VYTVÁŘENÍ NOVÝCH PROMĚNNÝCH + VÝBĚR PŘÍPADŮ

Deskriptivní analýza kvantitativních dat

# Opáčko

MODUS:

Nejčastější hodnota

3 4 7 8 10 11



MEDIÁN:

Střední hodnota

3 4 7 10 11 11



SMĚRODATNÁ ODCHYLKA:

Vyjadřuje, jak moc jsou jednotlivé hodnoty rozptýleny kolem průměru

SD = 3.25

PRŮMĚR:

3 4 7 8 10 11 11

$M = 7.71$

# Jak reportovat základní charakteristiky jednotlivých proměnných v textu/úkolu

- N (počet validních hodnot) a navíc:
  - *Nominální/ krátké ordinální: (relativní) četnosti kategorií (tj. Procenta)*
  - *Kardinální (škálové): průměr, SD, min-max*

# Cvičení

- V datasetu *transformace\_lekce* zkuste vyjet četnosti proměnné, která měří míru důvěry k Televizi Barrandov + její modus
  - *Analyze → Descriptive Statistics → Frequencies*
    - Tam kliknout na Statistics a naklikat, co potřebujete (viz minulá hodina)

- Co když tuším, že medián důvěry k Televizi Barrandov by se mohl lišit s ohledem na věk respondenta?

A) transformovat proměnnou věk tak, abychom měli menší množství kategorií

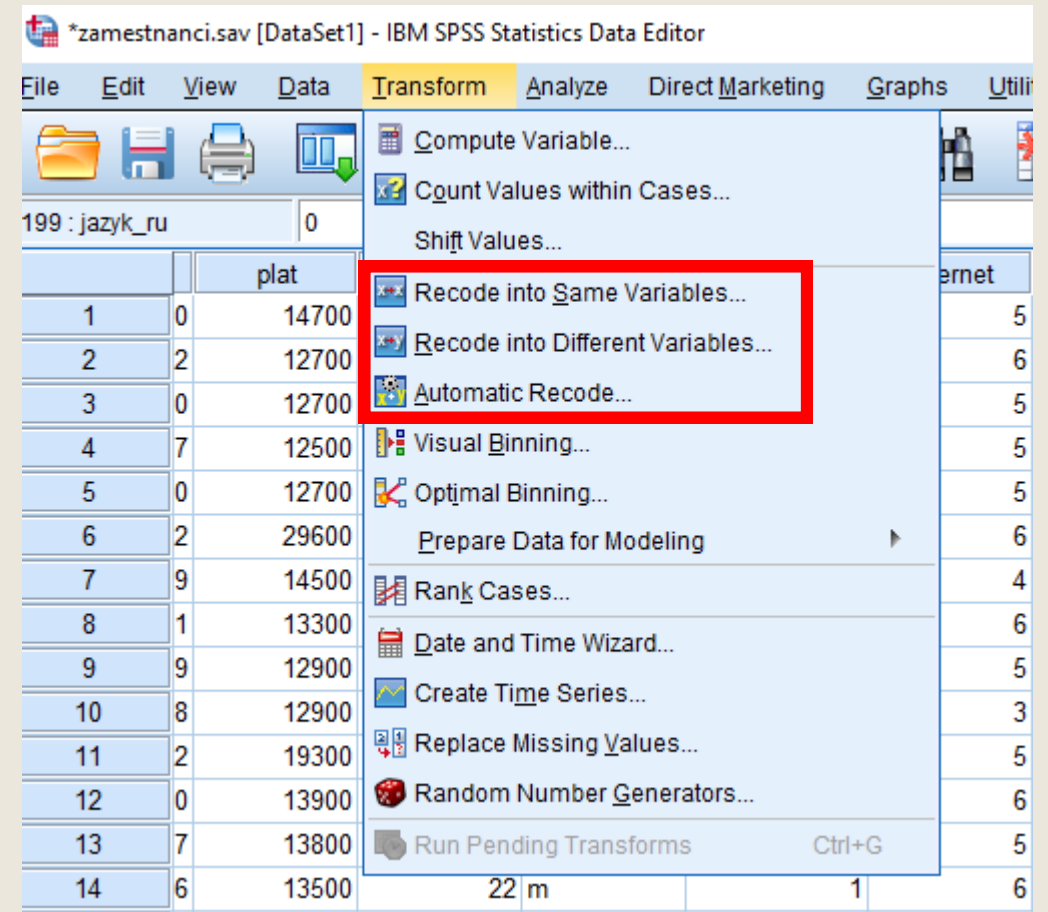
B) zobrazit tuto proměnnou vzhledem k jiné kategorii

# Rekódování proměnných (transformace)

1. *Máme jiný typ proměnné, než potřebujeme*
  - Nejčastěji *string* a potřebujeme *numeric*
2. *Máme proměnnou kardinální a chceme z ní udělat ordinální*
  - Typicky se tento typ transformace používá u věku – shlukujeme do kategorií
3. *Proměnná je nakódovaná jinak, než chceme*
  - Nejčastěji měníme orientaci škály – například, kdy chceme, aby se stoupající stupnicí stoupal i postoj k dané proměnné, ale orientaci škály je opačná
  - **Nepsaný zvyk: Čím vyšší skór mám, tím více se u mě projevuje měřená charakteristika** (náзор, postoj, vlastnost...)
4. *Chceme vytvořit průměrnou hodnotu z „příbuzných“ proměnných – tj máme několik položek, které nám měří dohromady jednotlivé dimenze nějakého jevu*

# Máme jiný typ proměnné než potřebujeme

- Procedura *Transform* → *recode into different variable (string na numeric)*





Recode into Different Variables: Old and New Values



Old Value

Value:

System-missing

System- or user-missing

Range:

through

Range, LOWEST through value:

Range, value through HIGHEST:

All other values

New Value

Value:

System-missing

Copy old value(s)

Old --> New:

f --> 1

'm' --> 2

Add

Change

Remove

Output variables are strings

Width:

8

Convert numeric strings to numbers ('5' -> 5)

Continue

Cancel

Help



# Když máme hodnoty v jiném měřítku než v jakém je mít chceme

ic	8	0	Kterými jazyky se dorozumíte? francouzština	{0, ne}...	None
ic	8	0	Kterými jazyky se dorozumíte? ruština	{0, ne}...	None

**1.** Numeric Variable -> Output Variable:  
věk --> věk\_kat

Output Variable  
Name: věk\_kat  
Label: věkové kategorie

**2.** Old and New Values...  
If... (optional case selection condition)

Old Value  
 Range:  
30 through 39  
 Range, LOWEST through value:  
 Range, value through HIGHEST:  
 All other values

New Value  
 Value: 2  
 System-missing  
 Copy old value(s)

Old --> New:  
20 thru 29 --> 1

Output variables are strings Width: 8  
Convert numeric strings to numbers ('5'->5)

Value Labels



Value Labels

Value:

Spelling...

Label:

- 1 = "18-24"
- 2 = "25-34"
- 3 = "35-44"
- 4 = "45-54"
- 5 = "55-64"
- 6 = "65 a vice"

- Add
- Change
- Remove

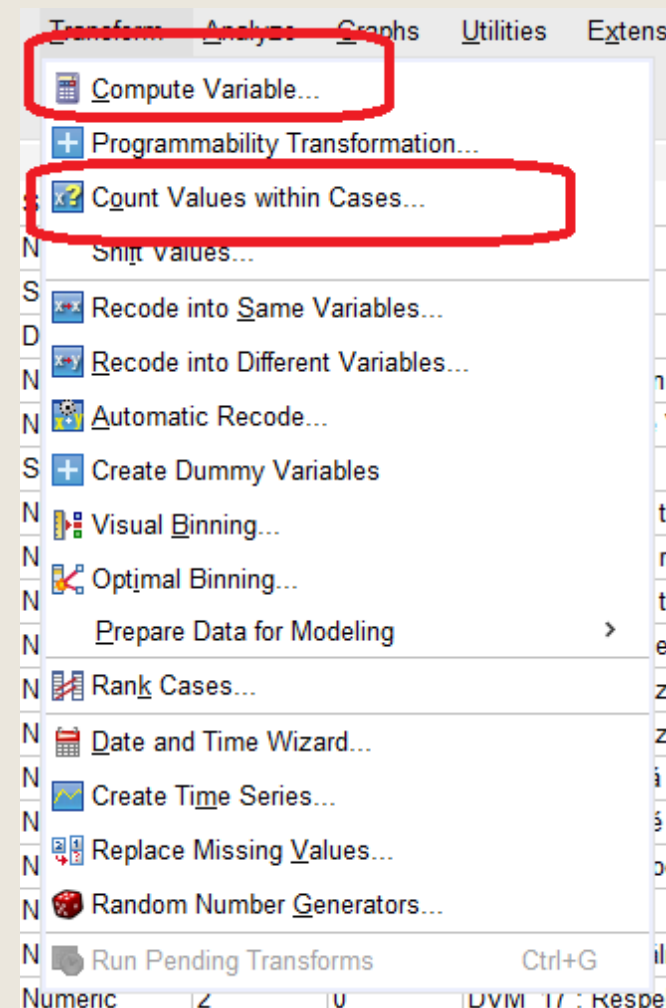
OK Cancel Help

# Když máme proměnnou nakódovanou jinak než jak ji chceme mít

- Standardně platí, že pravidlo, že u delších škál znamená **nejmenší hodnota nejmenší míru** a **největší hodnota největší míru**
- Často u škálových proměnných používáme několik položek za sebou, protože díky tomu můžeme docílit přesnějšího změření jevu:
  - *R12\_1 : Příjmy domácnosti, ve které žiju, pokrývají všechny její potřeby.*
  - *R12\_2 : Myslím, že se mám finančně lépe než většina lidí v této zemi.*
  - *R12\_3 : Obávám se, že na tom v budoucnu budu finančně hůř, než jak na tom jsem nyní.*
    - (na 5bodové stupnici, kde 1 znamená „rozhodně nesouhlasím“ a 5 znamená „rozhodně souhlasím“)

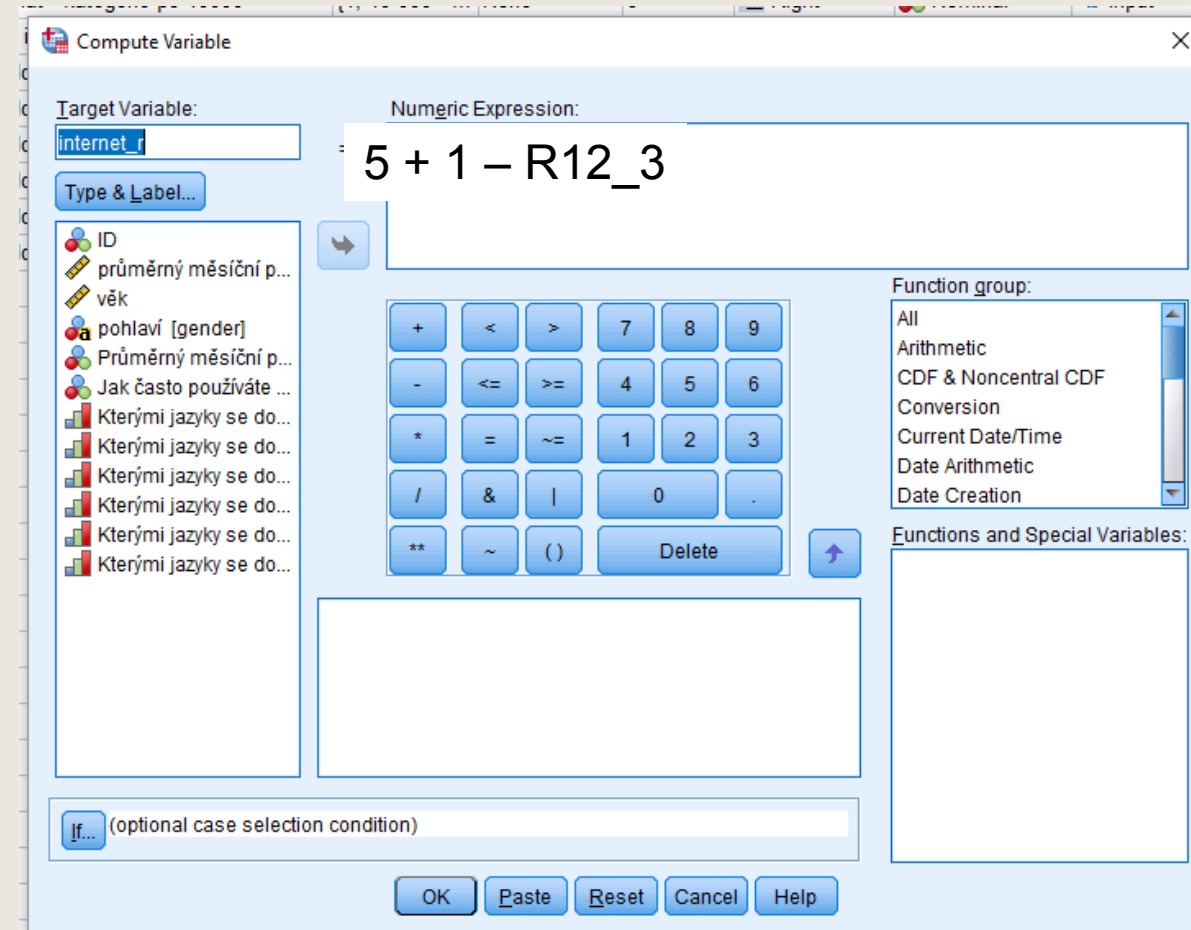
# Komputa položek

- Když potřebujeme být trochu „kreativní“
- Umožňuje nám
  - *Comupute variable*
    - Vytvářet/přetvářet položky podle různých matematických vzorců (sčítání, odčítání, násobení...)
  - *Count values*
    - Spočítat hodnoty napříč položkama



# Dvě možná řešení

- Rekódování pomocí procedury „recode“ – manuálně převrátíme škálu (1=10, 2=9, ...)
- Rekódování pomocí metody *compute* – spočítáme pomocí vzorečku
  - ***nová hodnota = nejvyšší hodnota +1 – stará hodnota***



# Když chceme z „příbuzných“ proměnných zjistit průměrnou hodnotu (vytvořit tzv. index)

- Kdy to potřebujeme: když chceme spočítat skór napříč položkami
- Například máme proměnné, které měří jednotlivé aspekty postoje k sexuální liberalizaci a chceme zjistit, jak průměrně je respondent liberální:
  - *Jak moc akceptovatelný je pro vás:*
    - VAL1\_1 : Potrat
    - VAL1\_2 : Homosexualita
    - VAL1\_3 : Rozvod
    - VAL1\_4 : Sex bez závazku

Numeric Expression:

```
(VAL1_1 + VAL1_2 + VAL1_3 + VAL1_4)/4
```

Jak si  
rozdělit výsledky  
podle vybrané  
proměnné

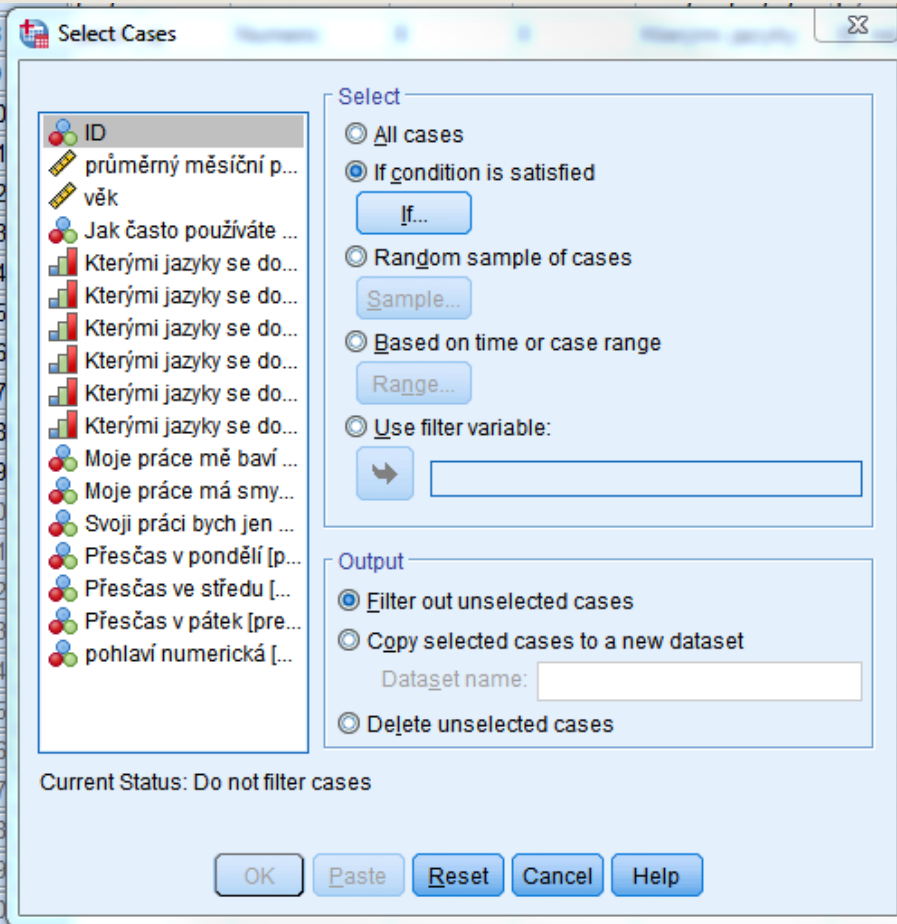
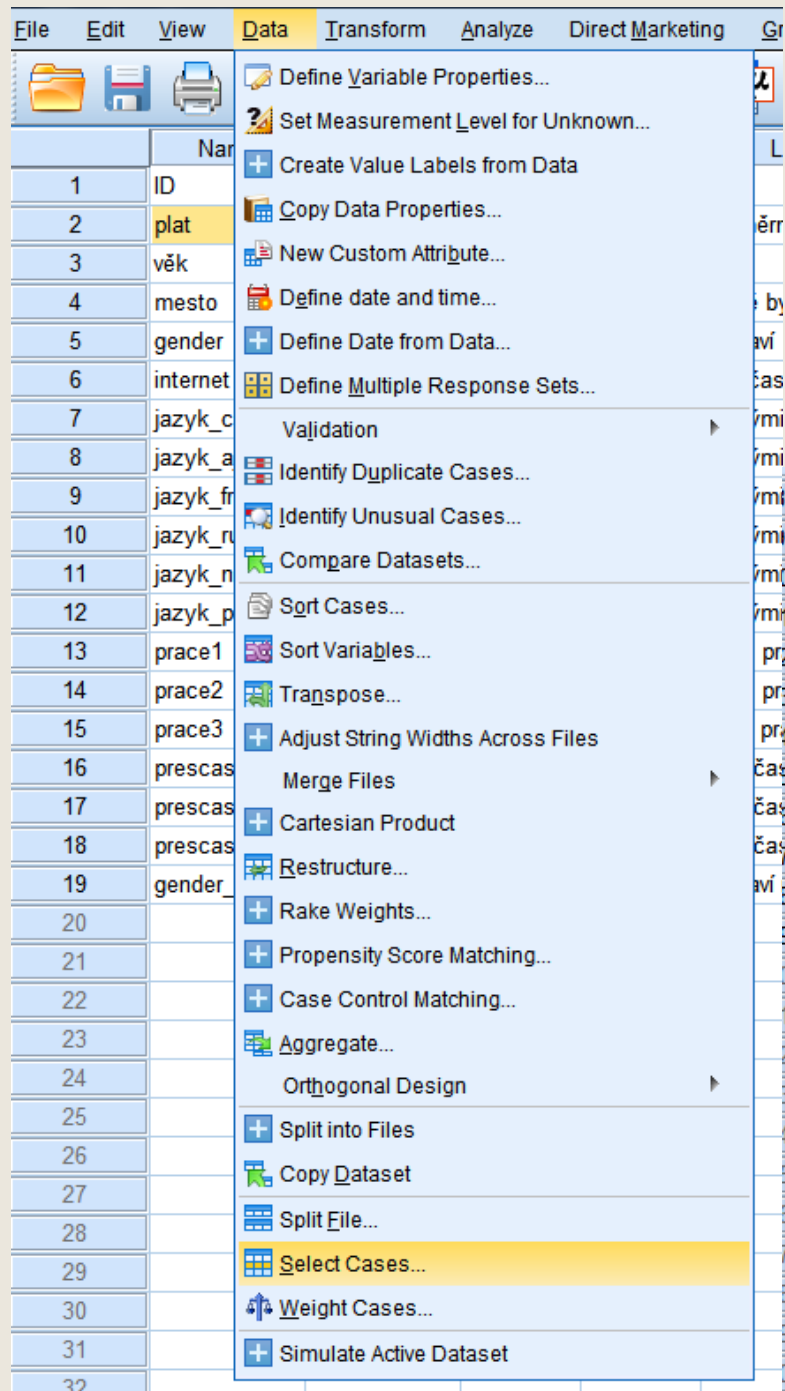


# Select cases

- 1. Data → Select cases
- 2. If condition is satisfied

Co to dělá:

- Vybere z našeho celé vzorku takové respondenty, které splňují naše nastavené podmínky
- Každé další příkazy (statistiky), které SPSS zadáme, pak provádí pouze na vybrané subskupině





# Select cases

Chceme zobrazit četnosti pro kategorii nejmladších respondentů

1. naklikneme proměnnou Vek\_kat pomocí šipky

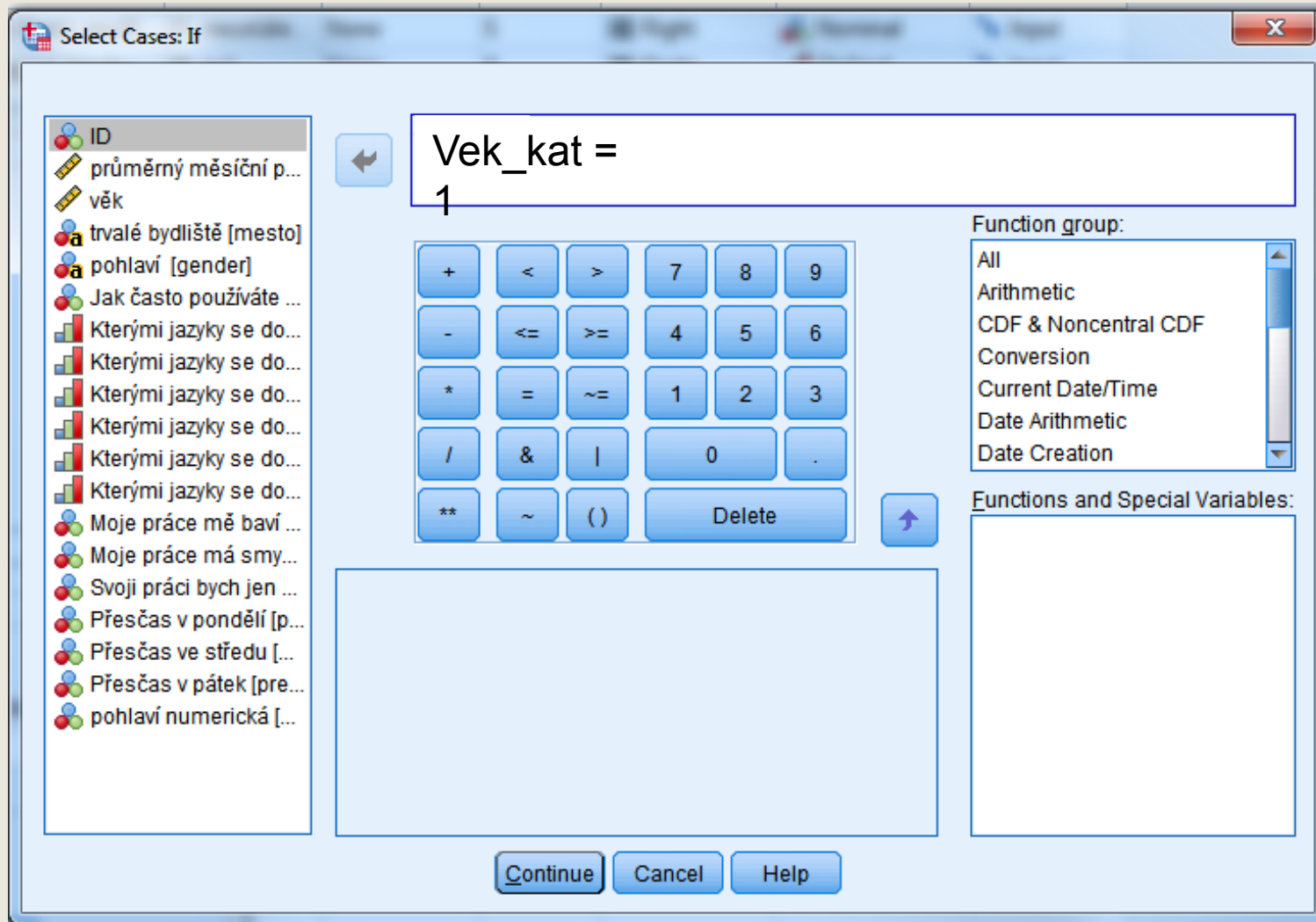
2. specifikujeme podmínku

- Vek\_kat je numerická
- **Vek\_kat = 1**

3. continue a ok

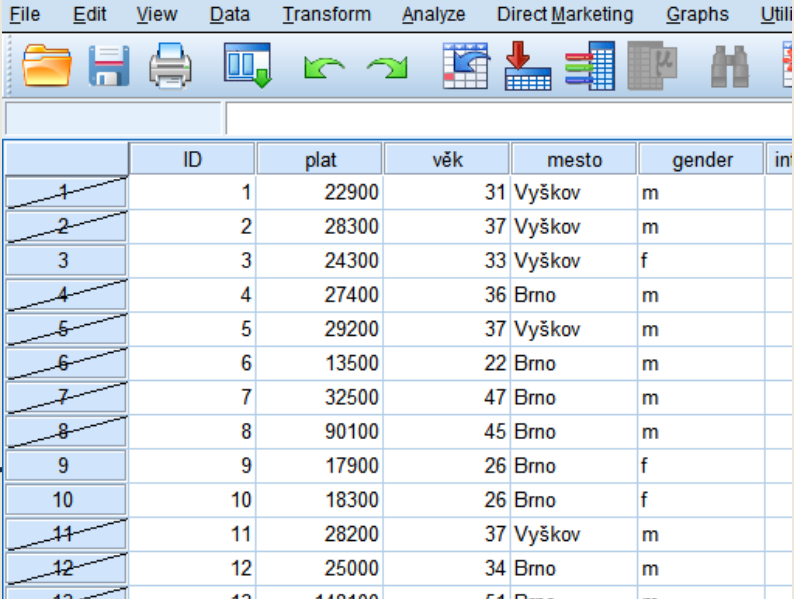
4. následně znovu vyjedeme četnosti (Analyze → Descriptive Statistics →

Frequencies), které se nyní zobrazí jen pro



# Select cases

- V datasetu v „Data view“ vidíme nepoužití respondenty (jsou vyškrtnutí)



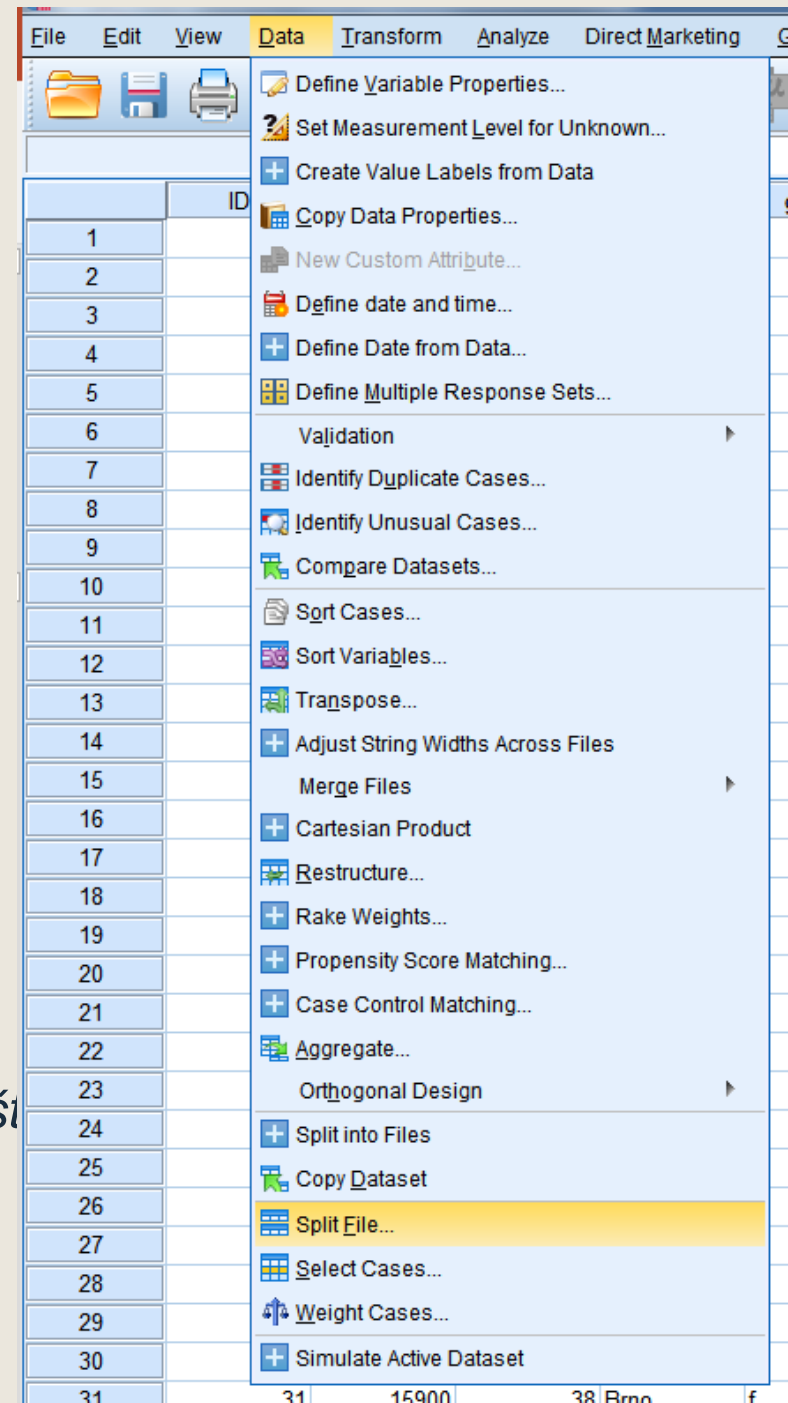
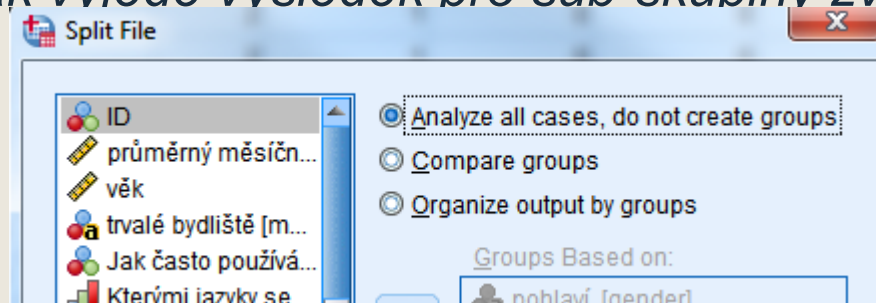
	ID	plat	věk	mesto	gender	in
<del>1</del>	1	22900	31	Vyškov	m	
<del>2</del>	2	28300	37	Vyškov	m	
3	3	24300	33	Vyškov	f	
<del>4</del>	4	27400	36	Brno	m	
<del>5</del>	5	29200	37	Vyškov	m	
<del>6</del>	6	13500	22	Brno	m	
<del>7</del>	7	32500	47	Brno	m	
<del>8</del>	8	90100	45	Brno	m	
9	9	17900	26	Brno	f	
10	10	18300	26	Brno	f	
<del>11</del>	11	28200	37	Vyškov	m	
<del>12</del>	12	25000	34	Brno	m	
13	13	148100	51	Brno	m	

- Pokud chceme opět pracovat s celým vzorkem, podmínku zrušíme přes příkaz
- Data → Select cases → All cases

# Split file

- 1. Data → Split file
- 2. Compare groups
- 3. Vložíme proměnnou, podle které se má output rozdělit
  - Tj. Pokud podle věkových kategorií, tak vek\_kat
- Co to dělá:
  - *Rozdělí vzorek podle zadané proměnné*
  - *V outputu pak uvíde výsledek pro sub-skupiny zvlášť*

**ZRUŠENÍ →**



TAKŽE JAKÉ JE ŘEŠENÍ NAŠÍ ÚVODNÍ  
OTÁZKY, ZDA SE PRŮMĚRNÁ DŮVĚRA  
K TELEVIZI BARRANDOV LIŠÍ S OHLEDEM NA  
VĚK RESPONDENTA?