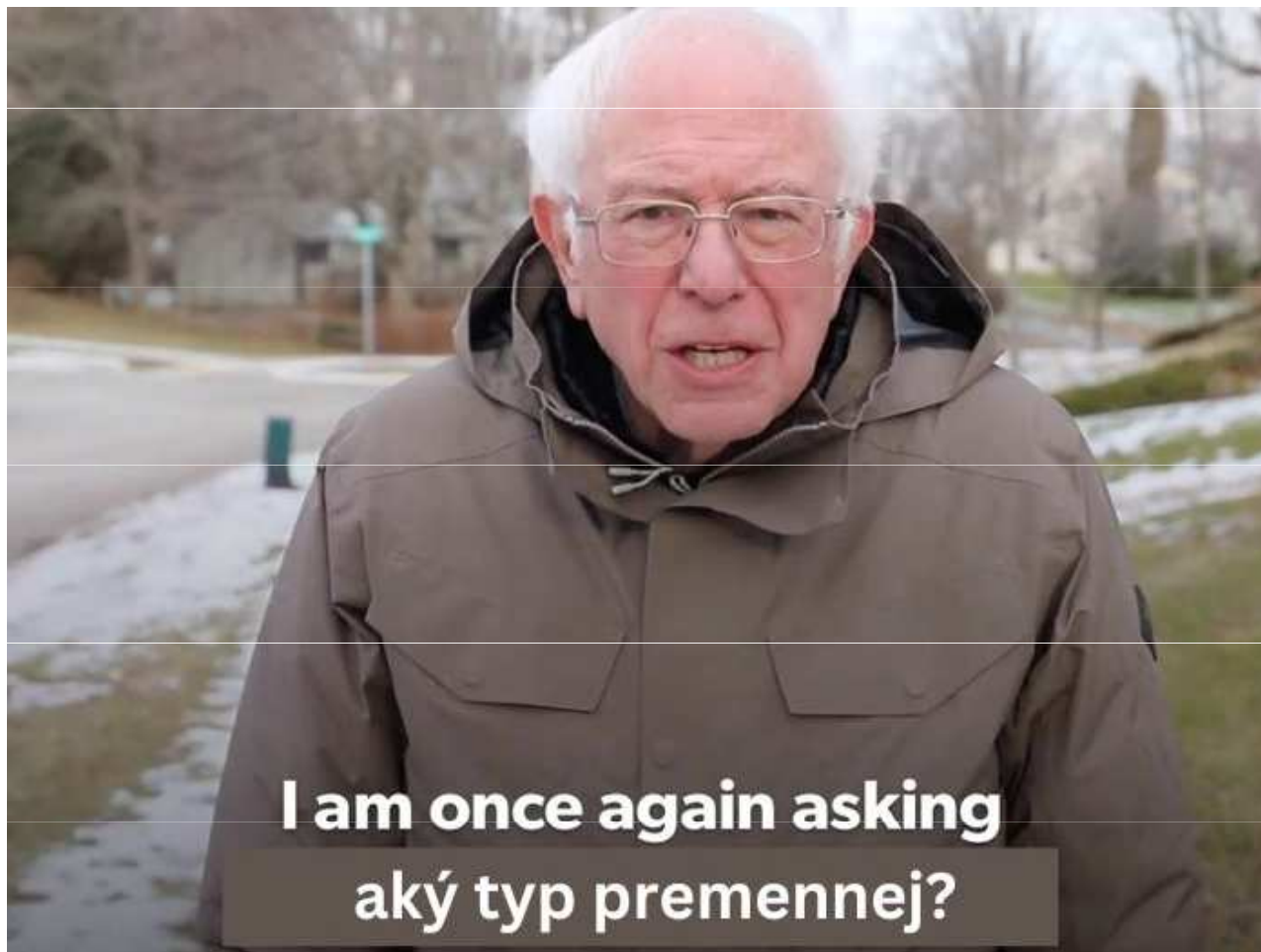


MUNI
FSS

Transformace a vytváření proměnných, výběr případů

... ale nejprve opakování



... ale nejprve opakování

Modus (nejčastější hodnota)

3 4 7 8 10 11 11

Medián (střední hodnota)

3 4 7 8 10 11 11

Směrodatná odchylka
(rozptýlení hodnot kolem průměru)
SD = 3.25

Průměr
M = 7.71

Ještě jedno opakování

Nominální proměnná

modus

Ordinální proměnná

modus, medián

Kardinální proměnná

modus, medián, průměr

Základní charakteristiky proměnných

Při reportování **vždy** uvádíme:

- N (počet validních hodnot)
- pro **nominální, krátké ordinální** proměnné: (relativní) četnost kategorií (procenta)
 - např. 51 % žen, 362 respondentů se středoškolským vzděláním s maturitou
- pro **kardinální proměnné**: průměr, SD, minimum, maximum
 - např. $M = 7.74$, $SD = 3.25$, $\min = 3$, $\max = 11$
- případně lze doplnit modus, medián (v případě univariace vhodné)

Cvičení

- stáhněte si dataset [transformace_prednáška.sav](#)
- zjistěte četnosti proměnné, která měří míru důvěry k Televizi Nova (DVM_9) a její modus
- Analyze → Descriptive Statistics → Frequencies
 - v možnosti „Statistics“ vybrat, co je třeba

Liší se medián důvěry k TV Nova napříč věkovými kategoriemi?

1. upravit (**transformovat**) proměnnou pro věk → vytvořit kategorie
2. zobrazit výsledky proměnné dle věkových kategorií

Rekódování proměnných (transformace)

Využijeme, když:

- a) máme jiný typ proměnné, než potřebujeme (např. string → numeric)
- b) máme proměnnou kardinální, ale chceme ji zjednodušit na ordinální (konkrétní věk → věková kategorie)
- c) máme proměnnou, kterou chceme nakódovat jiným způsobem (změna orientace škály)

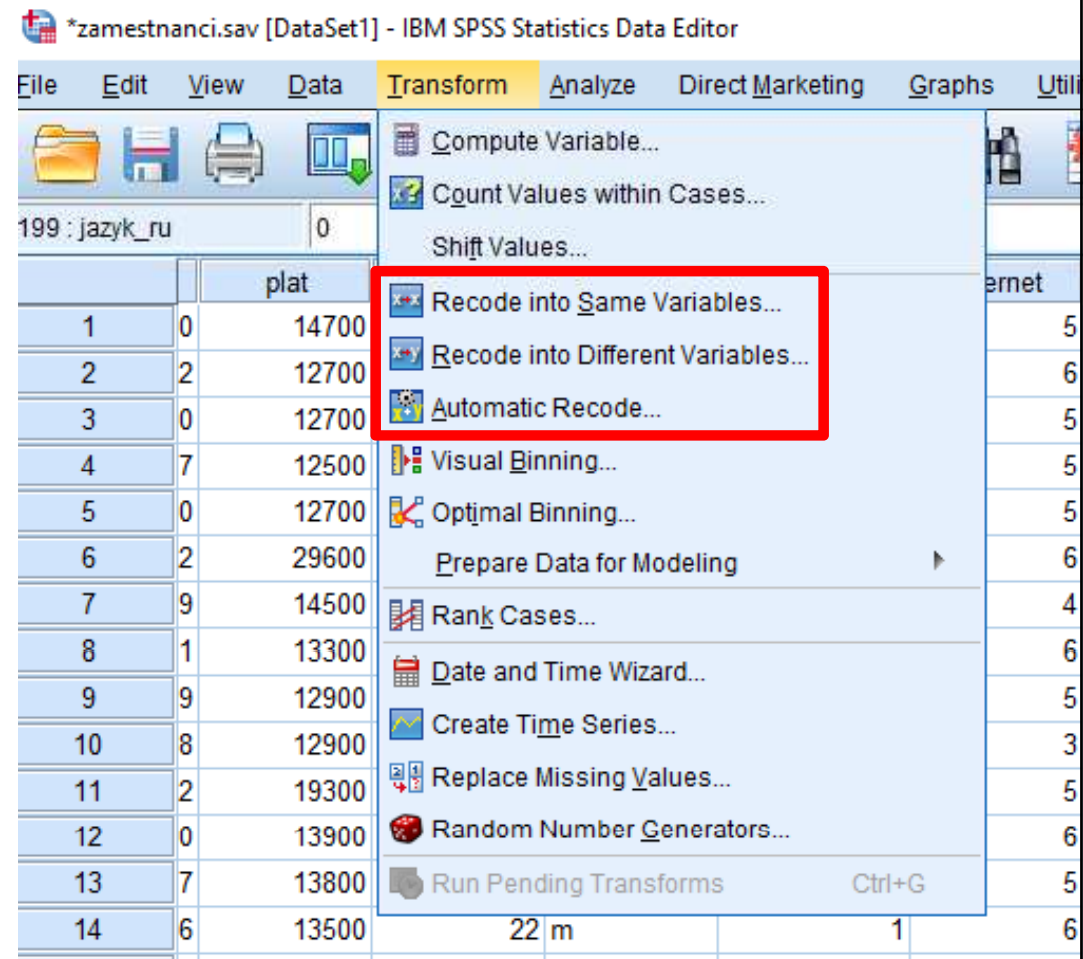
Nepsaný zvyk: čím vyšší skór (hodnotu pro daný případ), tím více se projevuje měřená charakteristika (názor, postoj, vlastnost).

- d) máme vícero „příbuzných“ proměnných a chceme z nich vytvořit index, průměrnou hodnotu apod. (různé dimenze jevu → jev)

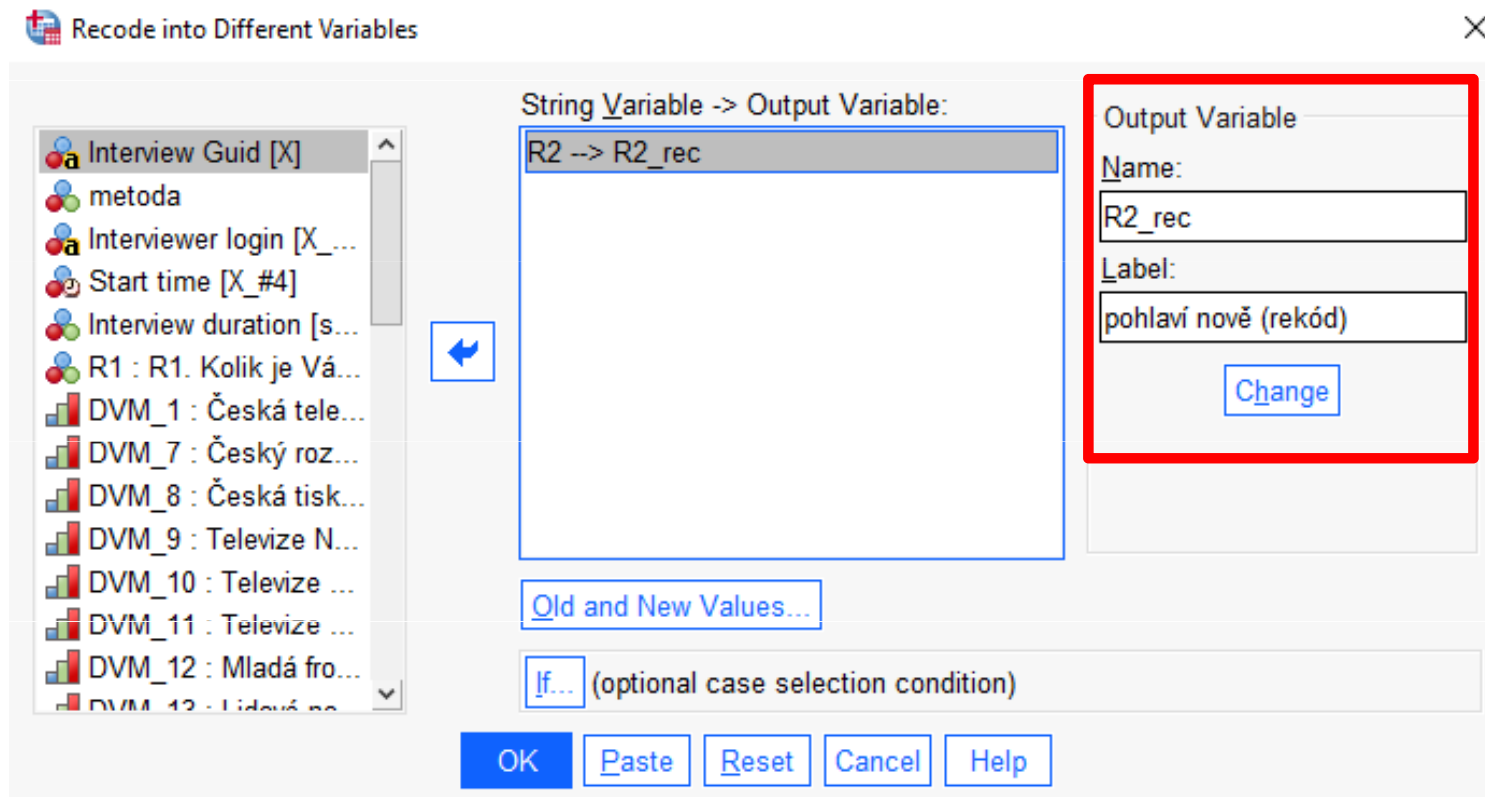
a) Jiný typ proměnné

– Transform → Recode

- **same variable** = stejná proměnná
- **different variable** = nová proměnná
- **automatic recode** = automaticky, ale pozor, vyžaduje velmi čistá data



a) Jiný typ proměnné (R2)



a) Jiný typ proměnné (R2)

Recode into Different Variables: Old and New Values

Old Value

New Value

Value:

System-missing

System- or user-missing

Range:

Range, LOWEST through value:

Range, value through HIGHEST:

All other values

Value: 2

System-missing

Copy old value(s)

Old -> New:

m -> 1

Add

Change

Remove

Output variables are strings Width: 8

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

Continue Cancel Help

b) Zjednodušení, kategorizace (R1)

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

Recode into Different Variables

Numeric Variable -> Output Variable:
věk --> věk_kat

Output Variable
Name: věk_kat
Label: věkové kategorie

Old and New Values...

if... (optional case selection condition)

Recode into Different Variables: Old and New Values

Old Value
 Value:
 System-missing
 System- or user-missing
 Range:
30 through 39
 Range, LOWEST through value:
 Range, value through HIGHEST:
 All other values

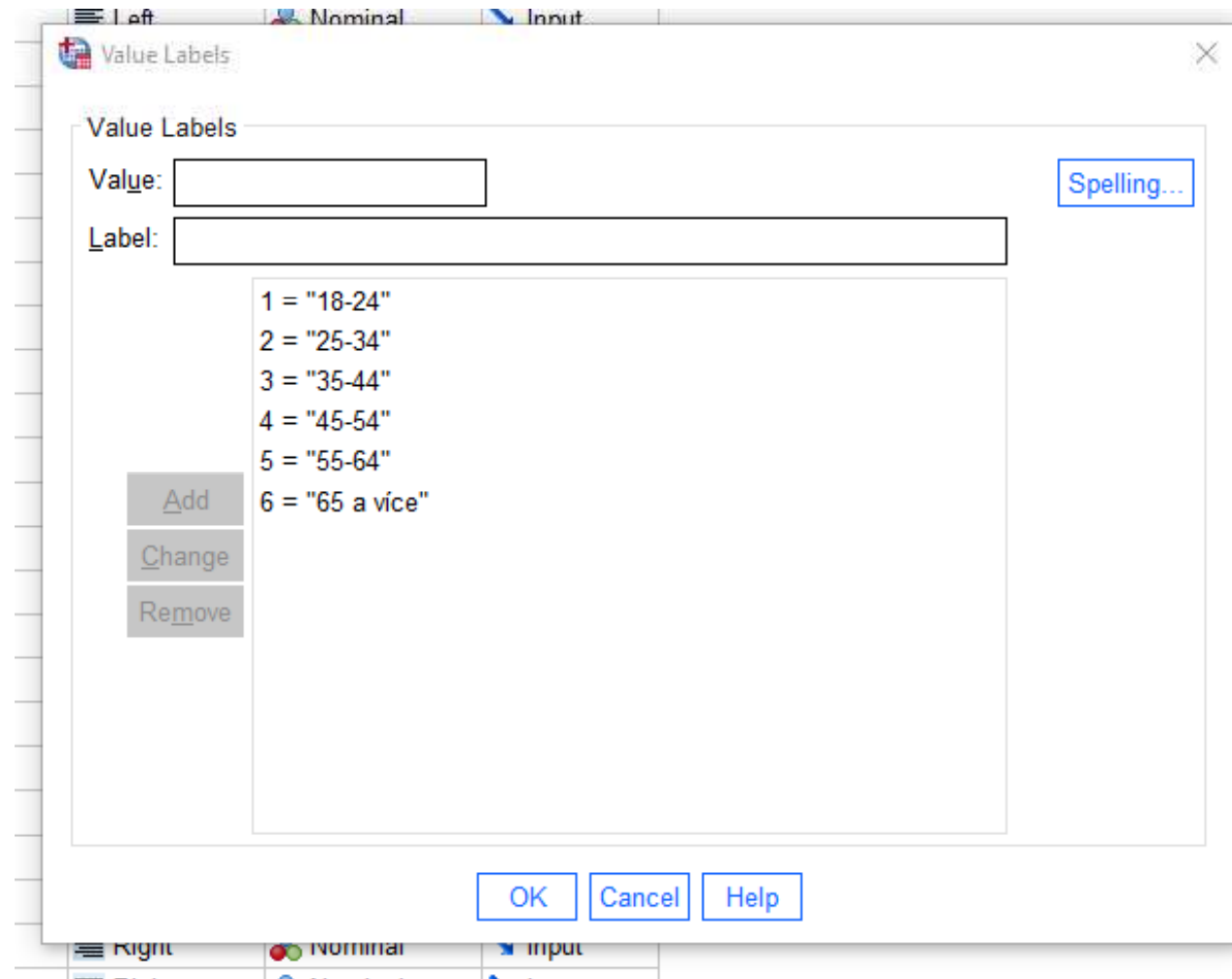
New Value
 Value: 2
 System-missing
 Copy old value(s)

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

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

b) Zjednodušení, kategorizace (R1)

Nezapomenout hned po dokončení správně zaznačit Value Labels.



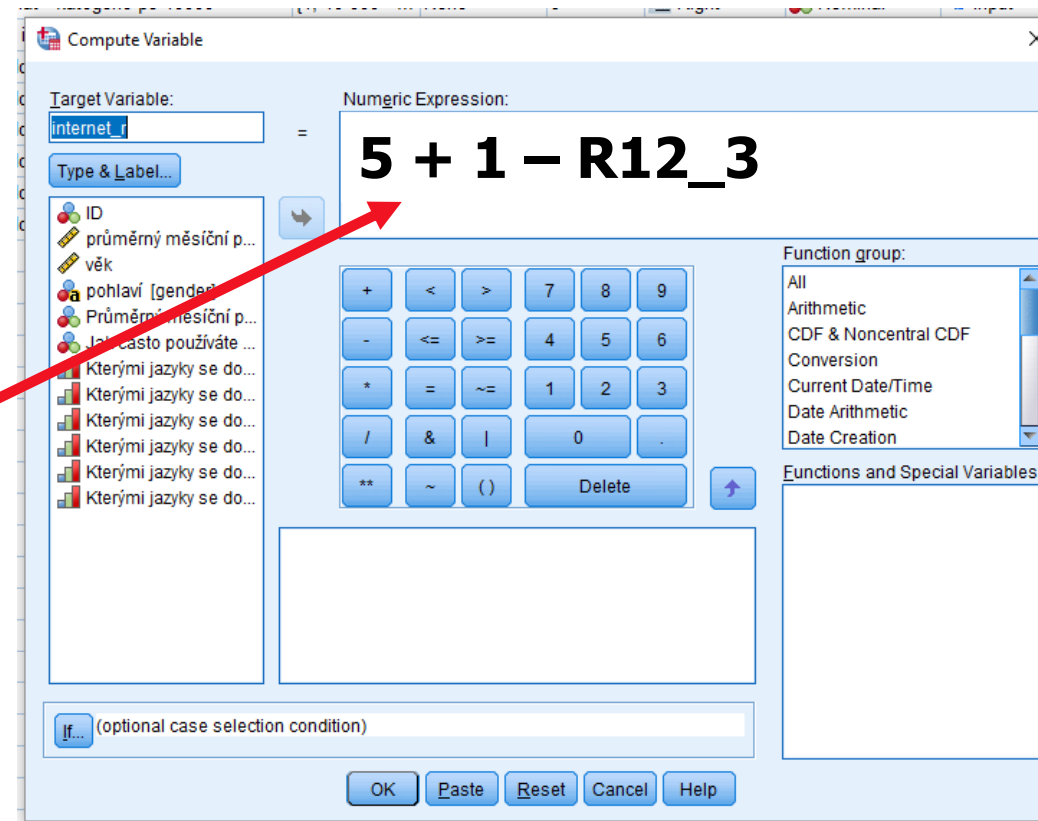
c) Změna způsobu kódování

- často u škálových proměnných používáme několik položek za sebou s **odlišným významem nejvyšší hodnoty** pro celkové vyhodnocení jevu, 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.

Stupnice o 5 bodech, kde 1 znamená „rozhodně nesouhlasím“ a 5 znamená „rozhodně souhlasím“.

c) Změna způsobu kódování

- využití Recode into...
- využití Compute Variable
 - spočítáme pomocí vzorečku
 - **nová hodnota = nejvyšší hodnota + 1 – stará hodnota**



Bonus: Komputa položek

- když potřebujeme být trochu „kreativní“ a provádět pokročilé přepočty hodnot
- **Comupute Variable**
 - vytvářet/přetvářet položky podle různých matematických vzorců (sčítání, odčítání, násobení...)
- **Count values Within Cases**
 - spočítat hodnoty napříč položkami

d) Práce s „příbuznými“ proměnnými

- nejčastěji pro spočítání skóru (**indexu**) napříč jednotlivými položkami (jednotlivými proměnnými)
- např. proměnné měřící jednotlivé aspekty postoje k politické participaci → zjišťujeme, jakou **celkovou** pozici v rámci postojů k jednotlivým aktivitám politické participace má daný respondent
- **Compute Variable**

d) Práce s „příbuznými“ proměnnými

Postoj k sexuální liberalizaci:

Jak moc akceptovatelný je pro Vás:

- VAL1_1 : Potrat
- VAL1_2 : Homosexualita
- VAL1_3 : Rozvod
- VAL1_4 : Sex bez závazku

Výpočet indexu
pomocí průměru

Numeric Expression:

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

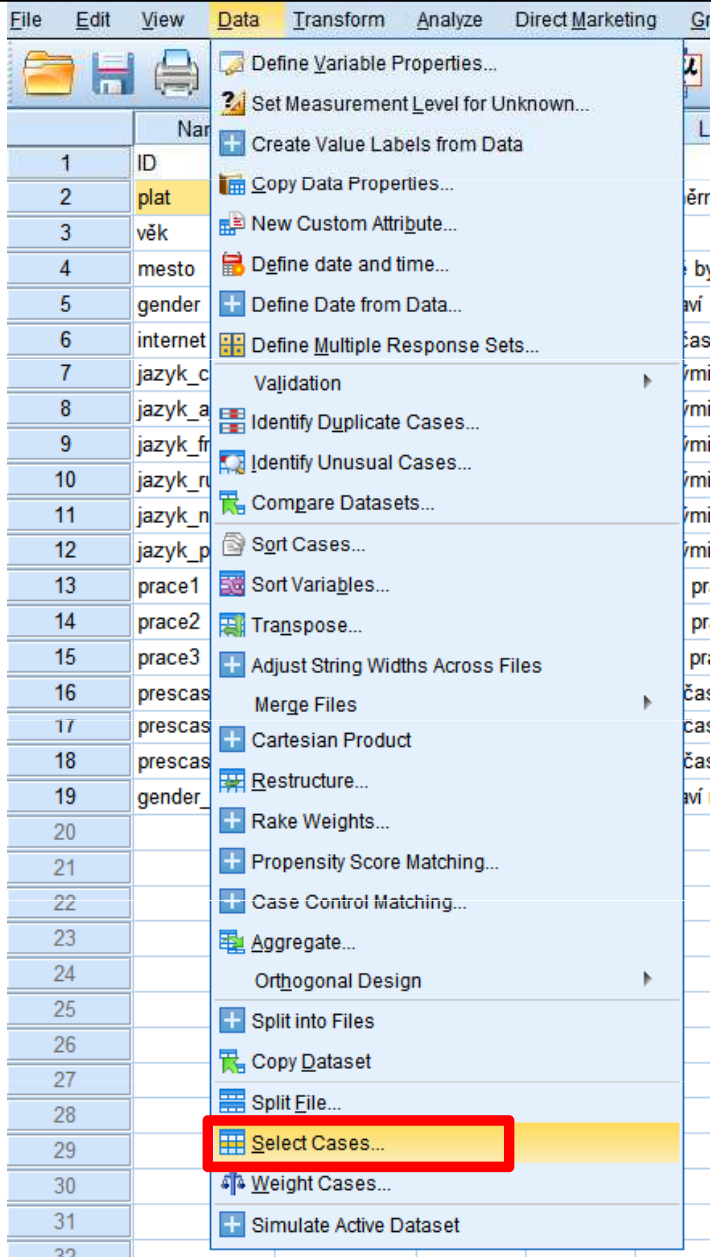
Liší se medián důvěry k TV Nova napříč věkovými kategoriemi?

... aneb Jak si rozdělit výsledky podle zvolené proměnné?

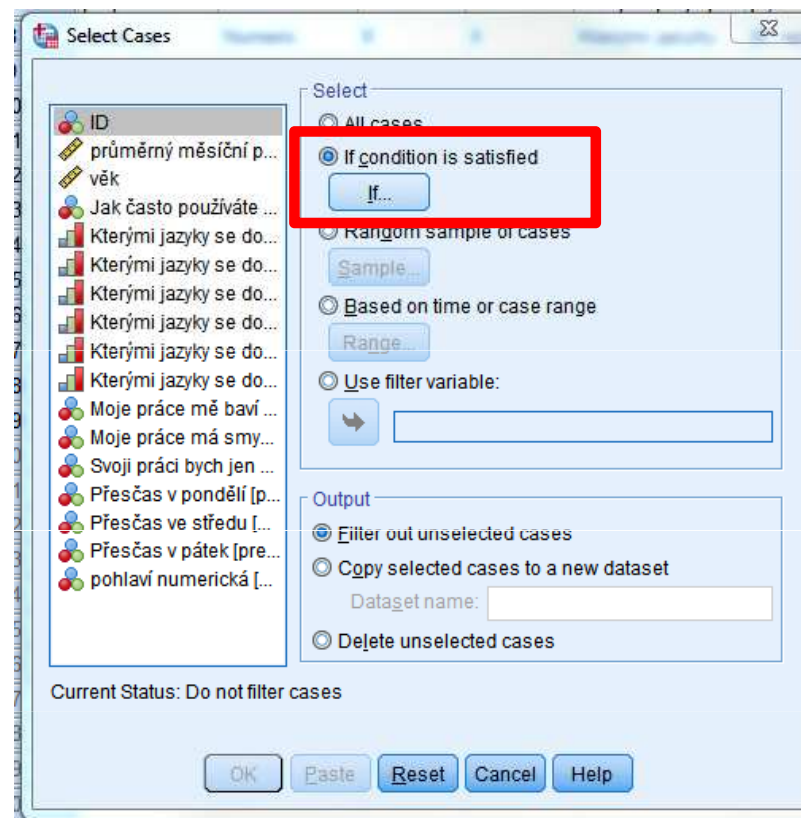
Výběr případů (Select Cases)

- výběr takových případů v datasetu, které splňují zadané podmínky
 - např. pouze ženy
- SPSS poté provádí všechny další analýzy a příkazy **pouze pro vybranou podskupinu**
 - např. průměrný věk žen

Data → Select Cases → nejčastěji **If condition is satisfied**



Výběr případů (Select Cases)



Výběr případů (Select Cases)

Příklad: Chceme zobrazit medián důvěry pro nejmladší věkovou kategorii.

1. If → zadáme podmínku vek_kat = 1 (numerická proměnná)
2. potvrdíme zadání
3. vyjedeme si medián pomocí Analyze → Descriptive Statistics → Frequencies
4. výsledky se zobrazí jen pro kategorii 1, tedy nejmladší skupinu

Výběr případů (Select Cases)

- v Data View máme vyškrtnuté nevyužité případy
- pokud chceme opět pracovat s celým vzorkem, podmínku zrušíme výběru v Select Cases zrušíme:

Data → Select Cases → All cases

	R1	DVM_9	R2	R12_1	R12_2	R12_3	VAL1_1	
1	70	2	1	4	-3	4	5	
2	69	2	2	4	3	4	5	
3	53	3	2	4	3	2	8	
4	60	2	2	4	3	4	5	
5	39	-3	2	5	5	4	10	
6	54	2	1	4	2	3	1	
7	25	3	1	2	3	4	2	
8	32	4	1	2	1	1	2	
9	43	3	1	5	-3	1	10	
10	70	3	2	5	3	1	5	
11	18	3	1	4	3	2	9	
12	63	2	2	3	3	3	8	
13	65	4	1	3	3	3	10	
14	39	4	1	3	1	5	12	
15	72	3	1	3	2	4	5	
16	27	2	1	4	4	3	10	
17	75	4	1	5	3	1	10	
18	42	4	1	3	2	3	10	
19	75	3	2	-3	1	5	5	
20	41	4	1	2	2	3	2	
21	31	3	1	4	4	3	7	
22	54	1	2	1	2	1	5	
23	28	1	1	4	-3	1	10	
24	39	3	2	3	2	4	5	

Split File

- rozdělení datasetu dle zadané proměnné
 - výsledek v okně Output se rozdělí dle zadaných podskupin
1. Data → Split File → **Compare Groups**
 2. výběr proměnné, podle které se output rozdělí do podskupin (např. vek_kat)

Pro práci s celým datasetem zrušíme rozdělení pomocí varianty **Analyze all cases, do not create groups.**

Cvičení: Split File

Použijte operaci Split File, abyste odpověděli na otázku:

Liší se medián důvěry k TV Nova napříč věkovými kategoriemi?