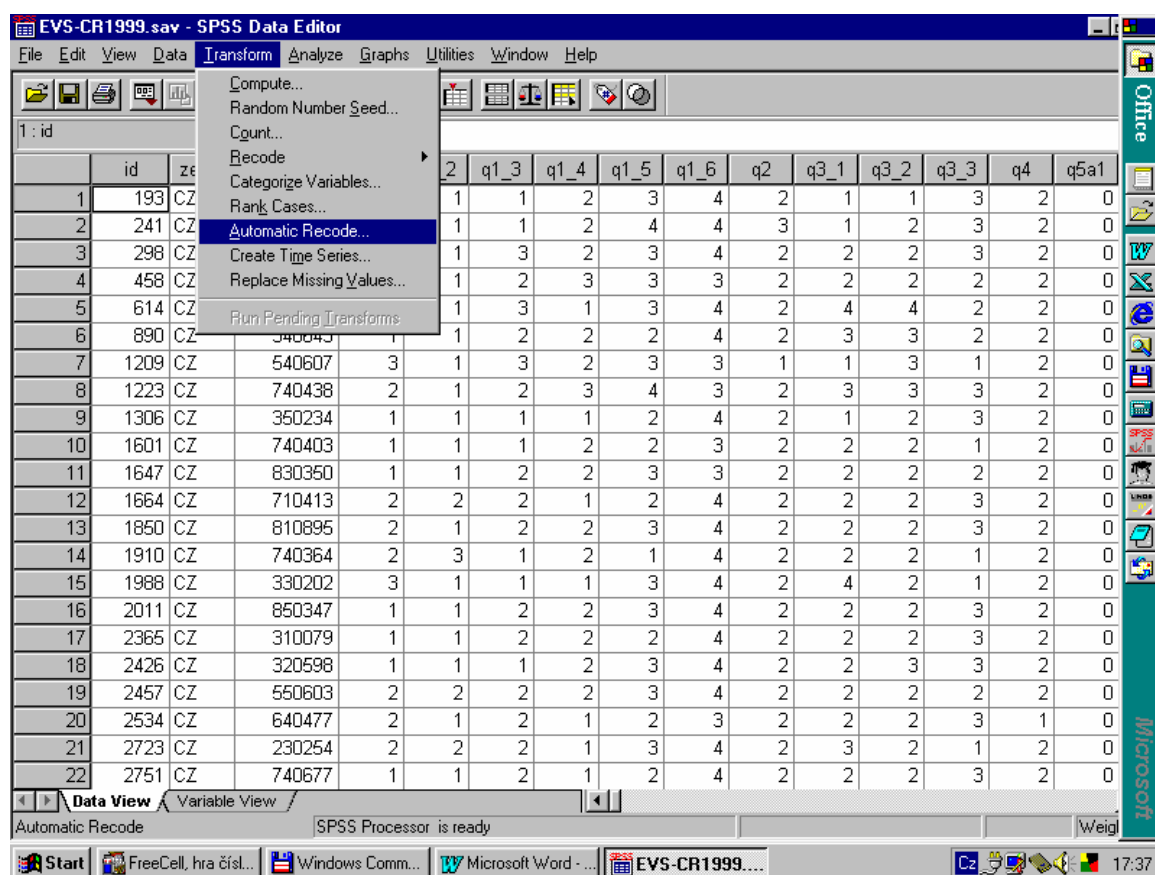


LEKCE 5

UMĚLÉ PROMĚNNÉ

TRANSFORMACE PROMĚNNÝCH

- RECODE
- COMPUTE
- CATEGORIZE VARIABLES
- COUNT
- RANK CASES



RECODE

- Vytváří z variant nominální proměnné či ordinální proměnné obecnější kategorie (kolapsuje obor hodnot dané proměnné).
- Vytváří kategorie spojitěho znaku.
- Odstraňuje nepotřebné kategorie proměnné (například variantu „nevím“).
- Mění pořadí kategorií hodnot proměnných (např. reorientace neboli otočení škály).

Lze to učinit pro všechny případy, nebo jen pro případy určené logickou podmínkou (if).

COMPUTE

- Upravuje hodnotu kódu znaku aritmetickou operací (přičtením konstanty, jejím odečtením, násobením, dělením, logaritmováním,...).
- Počítá ze starých znaků hodnoty nového znaku podle zadaného vzorce (např. vytváří SUMAČNÍ INDEX tak, že sčítá prosté nebo vážené hodnoty stejných variant různých znaků).

Lze to učinit pro všechny případy, nebo jen pro případy určené logickou podmínkou (if).

CATEGORIZE VARIABLES

Obdobně jako RECODE kategorizuje spojitá data. Na rozdíl od RECODE však vytváří ze spojitých dat kategorizovaný znak, jehož kategorie jsou založeny na percentilech, přičemž každá kategorie obsahuje přibližně stejný počet případů.

COUNT

Vytváří novou proměnnou, která pro každý případ sčítá výskyt téže (týchž) hodnoty v určitých proměnných.

Můžeme například mít řadu dichotomických proměnných, identifikujících, zda respondent četl sadu knih (u každého titulu volí z variant 1=četl, 2=nečetl). Pomocí COUNT můžeme vytvořit novou proměnnou, která nás informuje, kolik z těchto knih respondent četl (sečte se kolik je zadaných hodnot, v tomto případě 1).

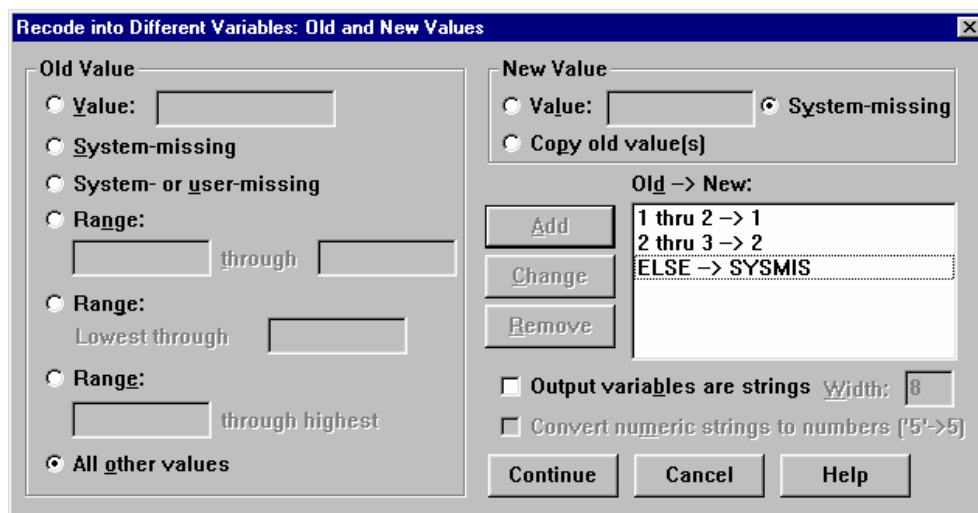
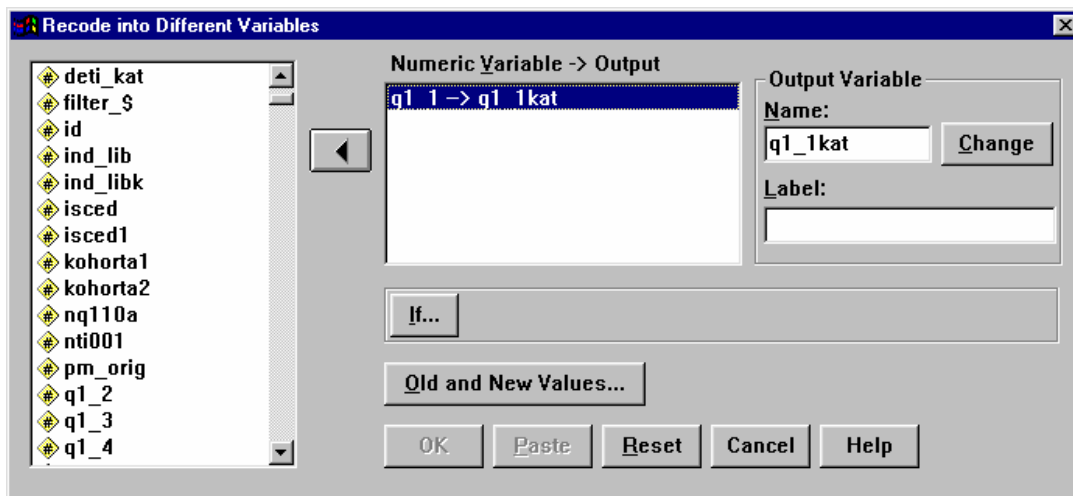
Lze to učinit pro všechny případy, nebo jen pro případy určené logickou podmínkou (if).

RANK CASES

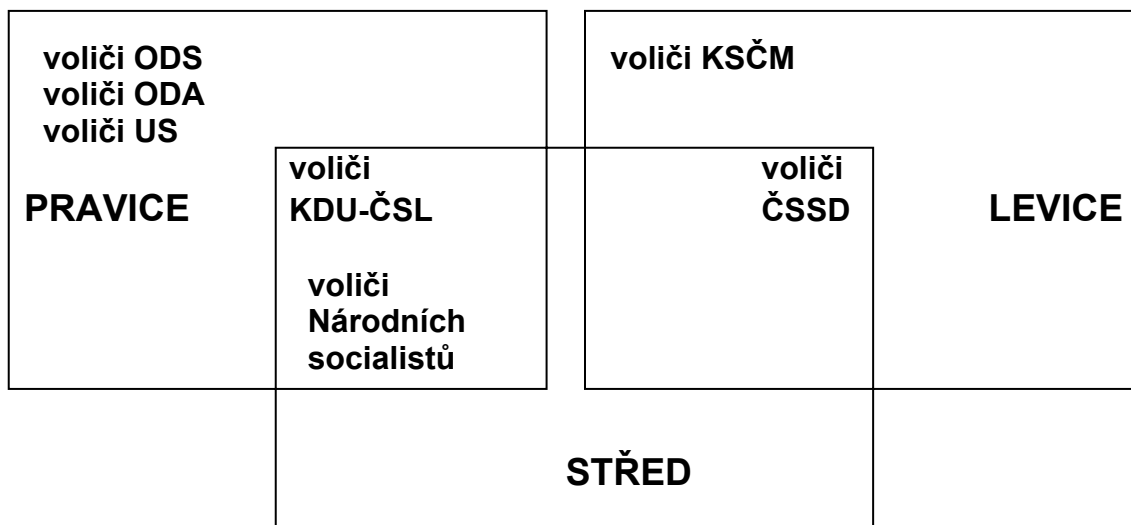
Vytváří novou proměnnou "rank score", která vyjadřuje pořadí případu mezi všemi případy podle hodnoty určité proměnné. Řadit můžeme ve vzestupném i v sestupném pořadí.

- Zvolíme-li (BY LIST) kategorizující proměnnou, pak získáme pořadí pro každou kategorii zvolené proměnné (zvolíme-li např. pohlaví, pak pro muže i pro ženy).
- Zvolíme-li (BY LIST) více kategorizujících proměnných, získáme pořadí pro každou kombinaci jejich kategorií (např. pohlaví a vzdělání: pro ženy se základním vzděláním, muže se základním vzděláním).

TRANSFORMACE PROMĚNNÝCH PŘÍKAZ RECODE



RECODE
KOLAPSOVÁNÍ HODNOT PROMĚNNÝCH
NOMINÁLNÍ PROMĚNNÁ



ORDINÁLNÍ PROMĚNNÁ

A55 POCIT CHUDOBY NA 9 STUP. SKALE

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 2	17	1,7	1,7	1,7
3	51	5,1	5,1	6,8
4	130	13,0	13,0	19,8
5	456	45,6	45,6	65,4
6	149	14,9	14,9	80,3
7	116	11,6	11,6	91,9
8	65	6,5	6,5	98,4
9	16	1,6	1,6	100,0
Total	1000	100,0	100,0	

STARÉ A NOVÉ HODNOTY: 1, 2, 3 = 1 4, 5, 6 = 2 7, 8, 9 = 3

A55KAT TRISTUPNOVA SKALA POCITU CHUDOBY

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 BOHATI	68	6,8	6,8	6,8
2 STRED	735	73,5	73,5	80,3
3 CHUDI	197	19,7	19,7	100,0
Total	1000	100,0	100,0	

KARDINÁLNÍ PROMĚNNÁ

NOVÝ ZNAK:**Věkové skupiny po 5 letech**

1. 18-19 let
 2. 20-24 let
 3. 25-29 let
 4. 30-34 let
 5. 35-39 let
- atd.

NOVÝ ZNAK:**Ekonom. aktivní a neaktivní věk**

1. ekonomicky aktivní (18-60 let)
2. ekonomicky neaktivní (61- let)

Kolik je Vám let?**Možné odpovědi:**

18,...21,22,...30,...
47,...62,63,...70, 71,
...88,... a více let

NOVÝ ZNAK:**Generace**

1. 18-34 let
2. 35-55 let
3. 56-65 let
4. 66 a více let

NOVÝ ZNAK:**Členění věku podle fertility:**

1. 18-24 let
2. 25-29 let
3. 30-39 let
4. 40-44 let
5. 45 a více let

Rekódování proměnné do méně kategorií může být výhodné, pokud proměnná má větší počet hodnot, což činí její rozložení málo interpretovatelné (jednotlivé hodnoty/kategorie jsou případy jen řídce obsazeny) nebo je tento počet vyšší než je optimální pro proceduru, kterou chceme použít (například by při biviační analýze vznikaly příliš rozsáhlé a poto řídce osazené kontingenční tabulky neboli crosstabs – viz příslušný blok, věnovaný základům biviační analýzy).

Protože se však pro jiné typy úloh může hodit původní proměnná s více hodnotami (s níž lze pracovat jako s intervalovou), je NUTNÉ provádět rekódování VŽDY do jiné proměnné a původní proměnnou v datech zachovat!

PŘÍKAZ COMPUTE

JEDNODUCHÝ SOUČET

COMPUTE X4 = X1+X2+X3.

JEDNODUCHÝ ODPOČET

COMPUTE vek = 1998 - rok nar.

Máme proměnnou ROKNAR (rok narození) a chceme mít proměnnou VEK (věk) výzkum se konal v roce 1998.

PŘÍŘAZENÍ ČÍSLA

COMPUTE rok=1998.

Každé proměnné ROK je přiřazena hodnota 1998 pro každý případ. Chceme každému respondentovi přiřadit identifikaci roku sběru materiálu).

SOUČET HODNOT (S PŘIDÁNÍM KONSTANTY)

COMPUTE celek=dil1 + dil2 + dil3 + dil4 + 100.

Výpočet může zahrnovat různé části. Nejprve se sečtou hodnoty jednotlivých znaků DIL1 až DIL4 a pak se přidá 100.

SOUČIN SOUČTU

COMPUTE vysledek = A1 * (A2 + A3).

Příkaz počítá VYSLEDEK jako hodnota znaku A1krát součet hodnoty znaku A2 a znaku A3.

SOUČET Z VÁŽENÝCH POLOŽEK (MAJÍ V NĚM RŮZNOU VÁHU)

COMPUTE predict = .7204 * A1 + .0937 * A2.

Příkaz vynásobí hodnotu znaku A1krát 0.7204, hodnotu znaku A2krát 0.0937, sečte násobky.

LOGARITMOVÁNÍ

COMPUTE logprijem = ln(prijem).

Příkaz zlogaritmuje hodnotu znaku PRIJEM.

MOCNINY

COMPUTE delka**3.

Příkaz k provedení třetí mocniny hodnoty znaku DELKA.

SUMAČNÍ INDEX - MÍRA LIBERALISMU

54. Mohl/a byste mi říci své názory na různá témata. Kam byste svůj názor umístil(a) na těchto škálách?

Q54A

1 2 3 4 5 6 7 8 9 10 -1 -2

Jednotlivci by měli převzít více odpovědnosti, aby se o sebe dokázali postarat

Stát by měl převzít více odpovědnosti, aby bylo o každého postaráno

Neví Neod.

Q54B

1 2 3 4 5 6 7 8 9 10 -1 -2

Nezaměstnaní, by měli mít povinnost přijmout jakoukoli nabídnutou práci, nebo ztratit podporu v nezaměstnanosti

Nezaměstnaní by měli mít právo odmítnout zaměstnání, které nechtějí vykonávat

Neví Neod.

Q54E

1 2 3 4 5 6 7 8 9 10 -1 -2

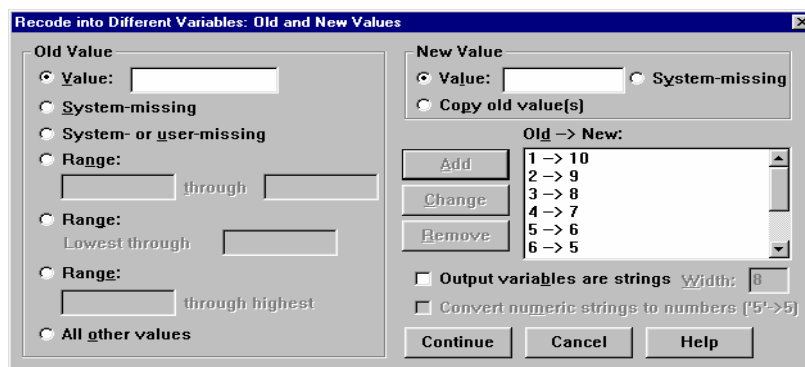
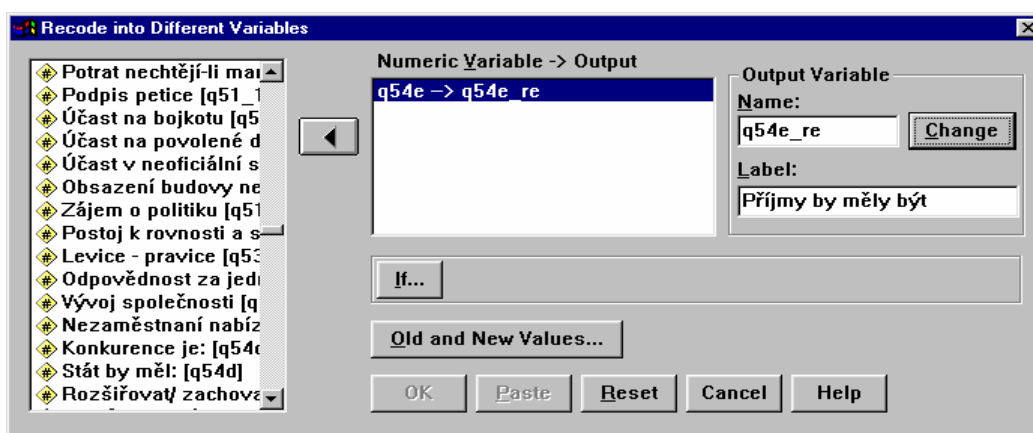
Příjmy by měly být vyrovnanější

Mělo by se více podnítit úsilí jednotlivce

Neví Neod.

Abychom mohli vytvořit sumační index míry „liberalismu“ musíme provést REORIENTACI PROMĚNNÉ Q54E pomocí RECODE:

1=10 2=9 3=8 4=7 5=6 6=5 7=4 8=3 9=2 10=1



výsledek RECODE

	q54 Příjmy by měly být:		q54_re Příjmy by měly být:	
	Count	%	Count	%
1	178	9.5%	133	7.1%
2	171	9.1%	136	7.3%
3	217	11.6%	338	18.0%
4	162	8.7%	181	9.7%
5	219	11.7%	138	7.4%
6	138	7.4%	219	11.7%
7	181	9.7%	162	8.7%
8	338	18.0%	217	11.6%
9	136	7.3%	171	9.1%
10	133	7.1%	178	9.5%
Total	1874	100.0%	1874	100.0%

Teprve pak lze hodnoty dosažené na jednotlivých škálách sečíst

COMPUTE

$IN_LIB = Q54A + Q54B + Q54E$

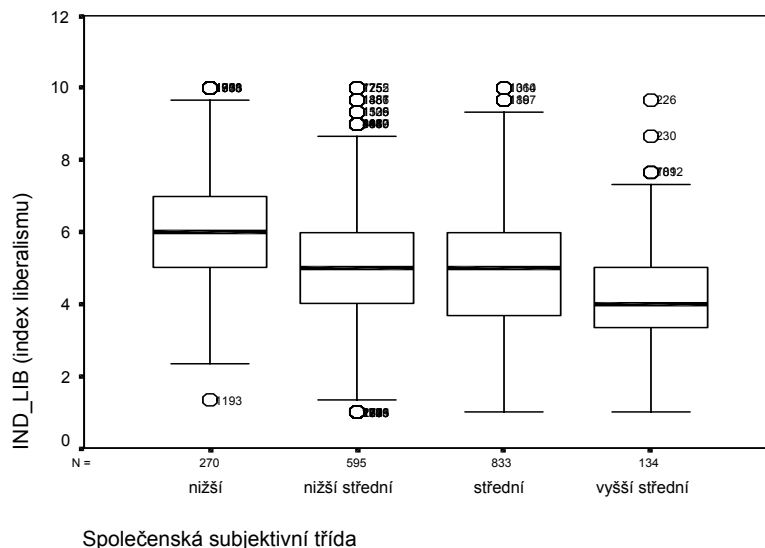
Sumační index: Škála „liberalismus versus paternalismus“ jejíž hodnoty se pohybují v intervalu <3;30>

respektive

$IN_LIB = (Q54A + Q54B + Q54E)/3$

Sumační index: Škála „liberalismus versus paternalismus“ jejíž hodnoty se pohybují v intervalu v <1;10>

S indexem lze dále pracovat, například spočítat jeho střední hodnotu pro různé subjektivní třídy a graficky je zobrazit.

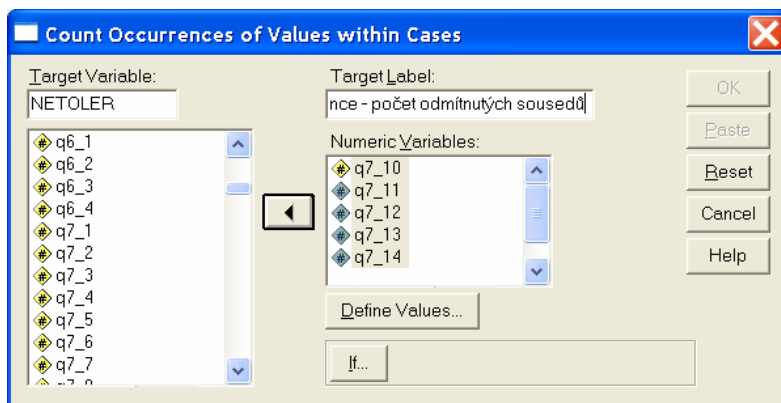


COUNT

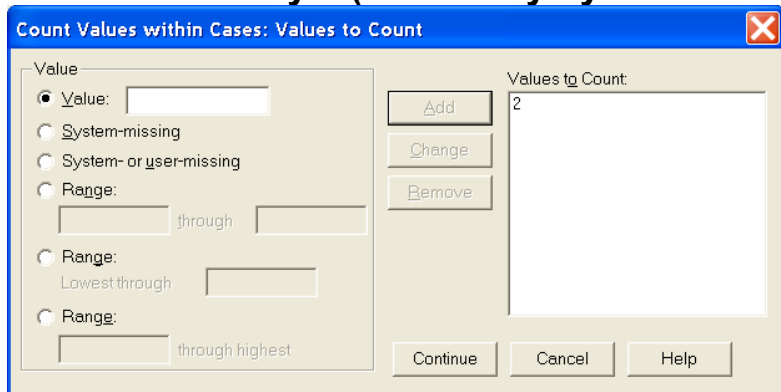
Koho (člena jaké problémové skupiny) by nechtěl mít za souseda:

	Nemocní AIDS	Narkomani	Homosexuálové	Židé	Rómové
	%	%	%	%	%
nezmínil/a	79,4%	26,7%	80,3%	95,6%	60,1%
zmínil	20,6%	73,3%	19,7%	4,4%	39,9%

Míra netolerance vyjádřena počtem problémových skupin, které by respondent nechtěl mít souseda (kolikrát odmítl sousedství: 0-5), se pro každého jedince spočítá pomocí procedury COUNT. S novou proměnnou můžeme provádět opět další operace:



Sčítáme hodnoty 2 (nechtěl by bydlet v sousedství):



Míra netolerance - počet odmítnutých sousedů

	Frequency	Percent	Valid Percent	Cumulative Percent
neodmítl/a žádnou skupinu	307	16,1	16,1	16,1
odmítl/a 1 skupinu	739	38,7	38,7	54,8
odmítl/a 2 skupiny	501	26,3	26,3	81,1
odmítl/a 3 skupiny	217	11,4	11,4	92,5
odmítl/a 4 skupiny	97	5,1	5,1	97,6
odmítl/a všech 5 skupin	47	2,4	2,4	100,0
Total	1908	100,0	100,0	

Mohu si klást nyní řadu otázek: Které sociální či demografické skupiny jsou méně tolerantní, jaká je konkrétní tendence těch, kdo odmítají jen jednu skupinu – jde o určitou skupinu, ?

PŘÍKAZ IF - (vytváření logických typů z více proměnných)

FILE ⇒ NEW ⇒ SYNTAX

The top screenshot shows the SPSS Data Editor window with the 'File' menu open and 'New' selected. The 'Syntax' option is highlighted in the 'New' submenu. The data grid shows variables a11 through a20 and a21.

The bottom screenshot shows the 'Syntax2' editor window with the following IF statement being typed:

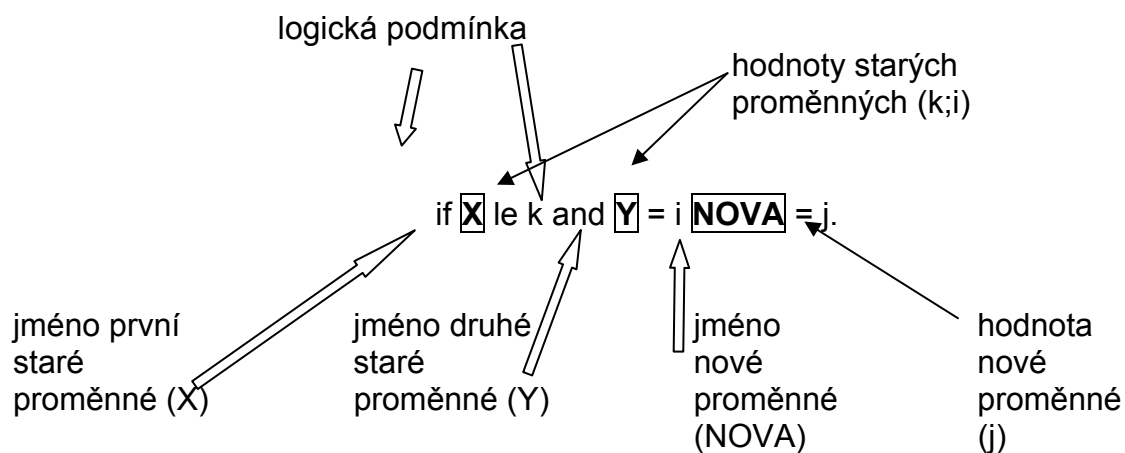
```

1: IF (POHLAVI = 1 AND EKAKT = 1) EKPOHL = 1.
   IF (POHLAVI = 1 AND EKAKT = 2) EKPOHL = 2.
   IF (POHLAVI = 2 AND EKAKT = 1) EKPOHL = 3.
   IF (POHLAVI = 2 AND EKAKT = 2) EKPOHL = 4.
  
```

The data grid in the bottom screenshot shows variables a13 through a20 and a21.

Obecný tvar příkazu

IF (podmínka) JMÉNO NOVÉ PROMĚNNÉ = kód hodnoty, odpovídající podmínce



Nezapomeňte na tečku za každým řádkem.

LOGICKÉ OPERÁTORY

Logická proměnná se vytvoří zápisem logických podmínek jež kombinují jednotlivé varianty použitých dílčích proměnných.

OPERÁTOR		VÝZNAM OPERÁTORU
1. verze	2. verze	
EQ	=	rovno
GT	>	větší než
LT	<	menší než
NE	<>	nerovno
GE	>=	větší nebo rovno
LE	<=	menší nebo rovno

- AND, logický výraz je pravdivý pouze když všechny připojené vztahy jsou pravdivé.
- OR, logický výraz je pravdivý, jestliže je pravdivá některá z připojených relací.
- NOT mění v negaci bezprostředně následující výraz.

Pozor na závorky a jejich počty v příkazu!

MUŽI A ŽENY JEDNOTLIVÝCH VZDĚLANOSTNÍCH SKUPIN

Původní proměnné nabývají hodnot:

SEX 1 = muž, 2= žena

VZD 1 = základní, 2 = střední, 3 = vysokoškolské

IF (SEX EQ 1 AND VZD EQ 1) SEXVZD = 1.

IF (SEX EQ 1 AND VZD EQ 2) SEXVZD = 2.

IF (SEX EQ 1 AND VZD EQ 3) SEXVZD = 3.

IF (SEX EQ 2 AND VZD EQ 1) SEXVZD = 4.

IF (SEX EQ 2 AND VZD EQ 2) SEXVZD = 5.

IF (SEX EQ 2 AND VZD EQ 3) SEXVZD = 6.

1. muž-zákl
2. muž-střed
3. muž-VŠ
4. žena-zákl
5. žena-střed
6. žena-VŠ

SEXVZD

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 muz-zakladni	181	18,1	18,2	18,2
	2 muz-stredni	227	22,7	22,8	41,0
	3 muz-VS	73	7,3	7,3	48,3
	4 zena-zakladni	207	20,7	20,8	69,1
	5 zena-stredni	267	26,7	26,8	96,0
	6 zena-VS	40	4,0	4,0	100,0
	Total	995	99,5	100,0	
Missing	System	5	,5		
Total		1000	100,0		