

LEKCE 5

TRANSFORMACE PROMĚNNÝCH

Analýza dat někdy vyžaduje, abychom proměnné, s nimiž pracujeme a které vstupují do analýzy, nějakým způsobem upravili, transformovali. V této kapitole si ukážeme některé ze situací, které transformaci proměnné vyžadují.

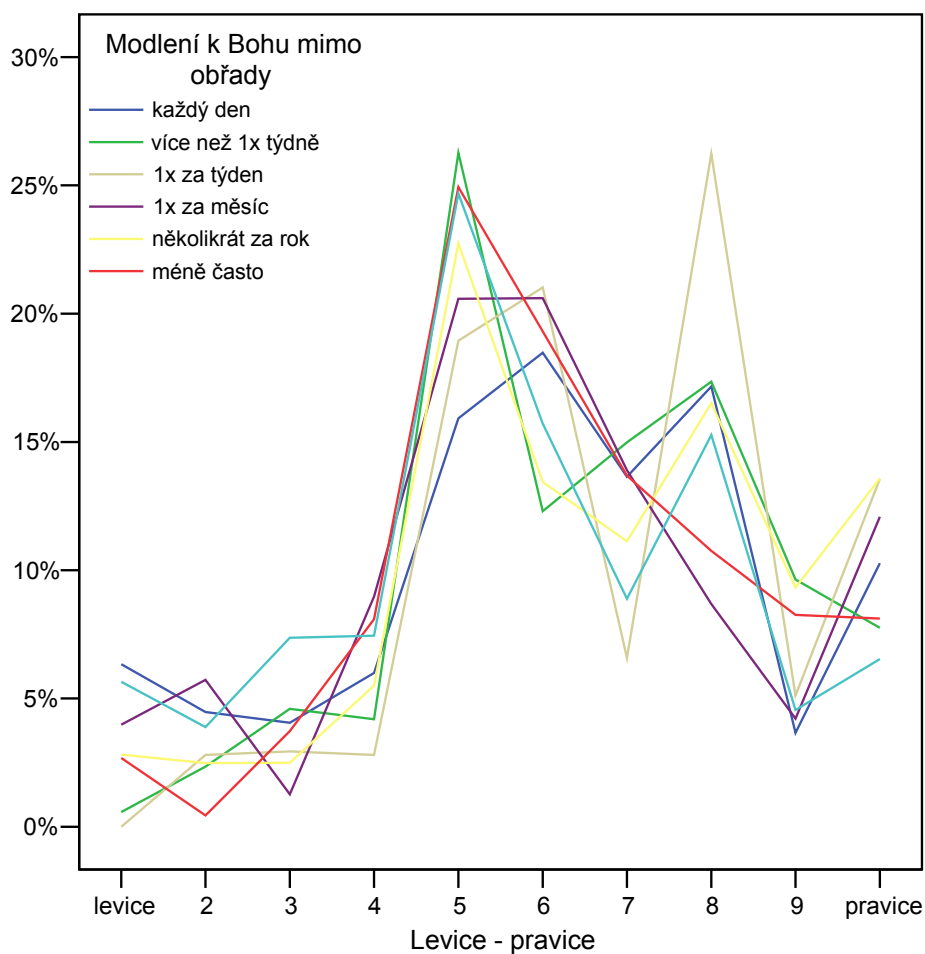
1. Proměnné s mnoha kategoriemi

Někdy mají naše proměnné velký počet kategorií (variant), neboť to jsou znaky, které se takto multivari-
antně vyskytují v reálném životě. Mohou to být jak znaky nominální (např. náboženská denominace), tak
znaky ordinální (levopravá politická orientace měřená na stupnici o deseti bodech) nebo intervalové
(hrubý měsíční příjem respondenta). Při prezentaci takovýchto znaků např. v tabulkách nebo grafech tato
mnohost kategorií může být na překážku. Např. v tab. 4.1 máme uvedenu četnost modlení respondenta
k Bohu podle levopravé politické orientace.

Tab. 4.1: Frekvence modlení se k Bohu podle politické orientace respondenta

		Modlení k Bohu mimo obřady							Total
		každý den	více než 1x týdně	1x za týden	1x za měsíc	několikrát za rok	méně často	nikdy	
1 levice	Count	11	0	0	2	3	5	61	82
	Row %	13,4%	,0%	,0%	2,4%	3,7%	6,1%	74,4%	100,0%
2	Count	7	2	1	4	3	1	42	60
	Row %	11,7%	3,3%	1,7%	6,7%	5,0%	1,7%	70,0%	100,0%
3	Count	7	4	2	1	3	7	80	104
	Row %	6,7%	3,8%	1,9%	1,0%	2,9%	6,7%	76,9%	100,0%
4	Count	10	3	1	6	6	14	81	121
	Row %	8,3%	2,5%	,8%	5,0%	5,0%	11,6%	66,9%	100,0%
5	Count	27	21	10	13	25	45	267	408
	Row %	6,6%	5,1%	2,5%	3,2%	6,1%	11,0%	65,4%	100,0%
6	Count	31	10	11	13	15	35	170	285
	Row %	10,9%	3,5%	3,9%	4,6%	5,3%	12,3%	59,6%	100,0%
7	Count	23	12	4	9	12	24	96	180
	Row %	12,8%	6,7%	2,2%	5,0%	6,7%	13,3%	53,3%	100,0%
8	Count	29	14	14	5	18	19	165	264
	Row %	11,0%	5,3%	5,3%	1,9%	6,8%	7,2%	62,5%	100,0%
9	Count	6	8	3	3	10	15	49	94
	Row %	6,4%	8,5%	3,2%	3,2%	10,6%	16,0%	52,1%	100,0%
10 pravice	Count	17	6	7	8	15	15	71	139
	Row %	12,2%	4,3%	5,0%	5,8%	10,8%	10,8%	51,1%	100,0%
Total	Count	168	80	53	64	110	180	1082	1737
	Row %	9,7%	4,6%	3,1%	3,7%	6,3%	10,4%	62,3%	100,0%

Taková tabulka je v podstatě nepřehledná, zabírá příliš mnoho místa a nepomůže, ani když ji převedeme
do grafické podoby, vědomi si faktu, že obrázek je mnohdy lepší než tabulka (viz obr. 4.1). Navíc u vel-
kých tabulek dochází často k tomu, že mnohá políčka mají nízký počet jednotek, což je i náš případ. Tyto
nízké počty jsou pak zdrojem nekorektních statistických výpočtů.

Obr. 4.1: Modlení se k Bohu podle politické orientace respondenta

Co v takové situaci dělat? Redukovat počet kategorií. To lze provádět dvěma způsoby, substantivně a distribučně (de Vaus 2002:34-38).

a) **substantivní přístup:** kategorie sloučíme podle smyslu.

Proměnnou „modlení se k Bohu“ bychom mohli redukovat do tří kategorií (to je tzv. trichotomizace):

„časté“ (sloučením kategorií každý den, více než 1x za týden a 1x za týden),

„občasné“ (1x za měsíc, několikrát za rok, méně často) a

„nikdy“.

Proměnnou „politická orientace“ bychom mohli také trichotomizovat na „levici“ (sloučením variant 1+2+3), „střed“ (sloučením 4+5+6+7) a „pravici“ (8+9+10). Pro tento způsob transformace neexistuje žádné pravidlo, jak nejlépe kategorie slučovat (říká se tomu také kolapsovat). Záleží vždy na účelu a smyslu analýzy, takže „politickou orientaci“ bychom např. mohli klidně rozdělit do čtyř kategorií: levici (1+2+3), levý střed (4+5), pravý střed (6+7) a pravici (8+9+10).

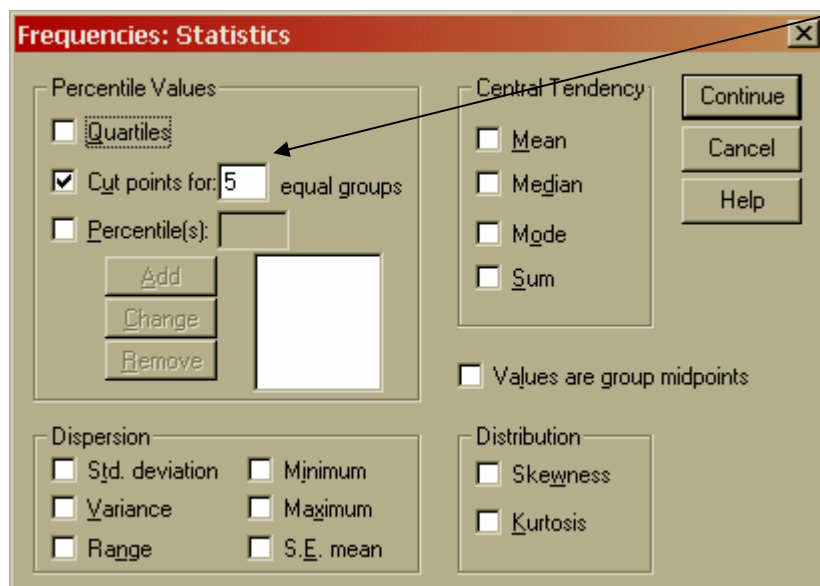
b) **distribuční přístup:** novou kategorizaci provedeme na základě rozdělení rozložení proměnné do početně stejně velkých kategorií. Tato operace má smysl pouze u proměnných ordinálních a intervalových. Budeme chtít např. vytvořit věkové kategorie od mladých, přes mladší střední věk, starší střední věk až po starší věk. Jenže definice toho, co je např. mladý věk a co mladší střední věk se mohou u jednotlivých výzkumníků odlišovat. Pokud ale vytvoříme kategorie na základě distribučního přístupu, rozložení proměnné v našem souboru nám samo o sobě řekne, jak bychom měli jednotlivé skupiny vytvořit. Stejně velké kategorie nalezneme např. s pomocí kumulativních četností, které tiskne procedura *Frequencies*. V tab. 4.2 máme rozložení věku a jelikož chceme vytvořit novou kategorizovanou proměnnou *vek_kat* takovým způsobem aby měla 4 kategorie a v každé byl přibližně stejný počet osob, pak musí každá kategorie obsahovat 25 % jednotek (respondentů). Ve sloupci *Cumulative %* vyhledáme příslušné procento 25., 50., a 75, nebo jemu nejbližší. Dělicí věkové kategorie pak budou: 28 let (24,2 %), 40 let (48,6 %) a

51 let (73,8 %) a příslušné věkové intervaly a jejich kódy budou: 1. mladý věk 18–28 let, 2. mladší střední věk 29–40 let, 3. starší střední věk 41–51 let a 4. starší věk 52–65 let.

Tab. 4.2: Rozložení věku

	Frequency	Percent	Valid Percent	Cumulative %
Valid 18,00	49	3,1	3,1	3,1
19,00	28	1,7	1,7	4,8
20,00	28	1,8	1,8	6,6
21,00	30	1,9	1,9	8,5
22,00	35	2,2	2,2	10,7
23,00	33	2,1	2,1	12,8
24,00	32	2,0	2,0	14,8
25,00	35	2,2	2,2	17,0
26,00	36	2,2	2,2	19,2
27,00	35	2,2	2,2	21,4
28,00	44	2,8	2,8	24,2
29,00	38	2,4	2,4	26,6
30,00	39	2,4	2,4	29,0
31,00	24	1,5	1,5	30,5
32,00	35	2,2	2,2	32,7
33,00	31	2,0	2,0	34,7
34,00	28	1,8	1,8	36,5
35,00	27	1,7	1,7	38,2
36,00	28	1,8	1,8	40,0
37,00	42	2,7	2,7	42,6
38,00	30	1,9	1,9	44,5
39,00	19	1,2	1,2	45,7
40,00	47	2,9	2,9	48,6
41,00	48	3,0	3,0	51,6
42,00	39	2,4	2,4	54,1
43,00	40	2,5	2,5	56,6
44,00	44	2,8	2,8	59,4
45,00	31	1,9	1,9	61,3
46,00	28	1,8	1,8	63,1
47,00	31	1,9	1,9	65,0
48,00	42	2,7	2,7	67,7
49,00	31	2,0	2,0	69,6
50,00	27	1,7	1,7	71,3
51,00	40	2,5	2,5	73,8
52,00	45	2,8	2,8	76,6
53,00	27	1,7	1,7	78,3
54,00	30	1,9	1,9	80,3
55,00	33	2,1	2,1	82,3
56,00	37	2,3	2,3	84,6
57,00	28	1,8	1,8	86,4
58,00	41	2,6	2,6	89,0
59,00	33	2,0	2,0	91,0
60,00	23	1,4	1,4	92,4
61,00	16	1,0	1,0	93,4
62,00	33	2,1	2,1	95,5
63,00	21	1,3	1,3	96,8
64,00	24	1,5	1,5	98,4
65,00	26	1,6	1,6	100,0
Total	1591	100,0	100,0	

Distribuční přístup lze uplatnit ještě precizněji, když necháme SPSS, aby skupiny vytvořil sám na základě tzv. *cut points* neboli dělicích bodů. Chceme-li čtyři stejně početné skupiny, žádáme kvartily, chceme-li pět skupin, hledáme kvintily, chceme-li deset skupin, hledáme decily. Takže pokud budeme chtít z proměnné „příjem domácnosti“ vytvořit kategorizovanou proměnnou s pěti kategoriemi (1. nízký příjem, 2. nižší střední příjem, 3. střední příjem, 4. vyšší střední příjem, 5. vysoký příjem), zadáme pro proměnnou q110a, což je jméno proměnné obsahující příjem domácnosti, v proceduře *Analyze – Descriptive Statistics – Frequencies – Statistics* požadavek, aby dělicí body byly pro pět skupin.



Výpočtem zjistíme, že příjmové kategorie budou mít následující členění:

Statistics

q110a Čistý měsíční rodinný příjem		
N	Valid	1120
	Missing	788
Percentiles	20	9400,00
	40	12000,00
	60	15000,00
	80	20000,00

1. nízký příjem (do 9 400 Kč);
2. nižší střední příjem (9 401–12 000 Kč);
3. střední příjem (12 001–15 000 Kč);
4. vyšší střední příjem (15 001–20 000 Kč);
5. vysoký příjem (20 001 a více). ¹

Víme-li, do jakých skupin máme příjmovou proměnnou rozložit, použijeme transformační proceduru *Recode* a v ní příslušné kategorie vytvoříme. Ukažme si to.

¹ Všimněte si, že hranice jednotlivých intervalů se nepřekrývají. A ani nemohou, neboť to je obecná zásada pro vytváření intervalů. V našem případě bychom tedy nemohli např. definovat nízký příjem intervalem do 9 400 Kč a nižší střední příjem intervalem 9 400–12 000 Kč. Je to z toho důvodu, že pokud bychom měli v souboru respondenta s příjmem 9 400 Kč, nevěděli bychom, zdali mu přiřadit kód 1, nebo 2.

Změna kódů (procedura RECODE)

Příklad P4.1: V příkladu P2.3 v předminulé lekci jsme analyzovali proměnnou, která indikovala roli Boha v osobním životě člověka (*Jak důležitý je Bůh ve Vašem životě?*). Tato proměnná byla měřena na desetibodové stupnici:

1	2	3	4	5	6	7	8	9	10
Vůbec ne důležitý									velmi důležitý

Všimněte si, že orientace stupnice probíhá od nízké důležitosti k důležitosti vysoké, což je správně. Nic by nám sice nebránilo v tom, abychom stupnici orientovali opačně: 1 velmi důležitý a 10 vůbec ne důležitý. Pro statistickou analýzu a především pro výsledky této analýzy je ovšem důležité, aby používané stupnice měly tzv. přirozenou orientaci, tedy aby nízké hodnoty stupnice indikovaly nízké množství vlastnosti a naopak vysoké hodnoty stupnice indikovaly velké množství vlastnosti – řečeno jinak, s narůstající hodnotou stupnice by mělo také narůstat množství vlastnosti. V případě naší stupnice tomu tak je: čím vyšší hodnota proměnné, tím důležitější je Bůh pro život jednotlivce.

Aby bylo možné s tímto znakem pracovat v třídění vyšších stupňů, je třeba stupnici zkrátit. Smysluplná se jeví její trichotomizace, to je její redukce do tří stupňů s následujícími kódy:

1. role Boha důležitá,
2. role Boha ani důležitá, ani nedůležitá,
3. role Boha nedůležitá.²

Nabízí se následující řešení: V původní stupnici první tři stupně (1–3) indikují, že role Boha je důležitá, poslední tři stupně (8–10) naopak, že důležitá není.³ Pozice mezi těmito póly (body 4–7) znamená, že role Boha je i není důležitá. Abychom tuto operaci – vytvořit novou proměnnou s novými kódy, to je rekódovat (neboli *Recode*) původní proměnnou – mohli provést, musíme ji sdělit programu SPSS formalizovaným způsobem. Musíme mu sdělit zhruba následující:

„Jestliže má původní proměnná (*q33*) hodnotu 1, 2 nebo 3, nechť nová proměnná, kterou si nazvěme *q33_kat*⁴, má hodnotu 1. Jestliže původní proměnná *q33* má hodnotu 4, 5, 6 nebo 7, nechť proměnná *q33_kat* má hodnotu 2. A jestliže proměnná *q33* nabývá hodnot v intervalu 8–10, nechť *q33_kat* má hodnotu 3.“

Provedme v SPSS prostřednictvím procedury *Transform – Recode – Into different variables*. Původní proměnnou *q33* necháme zachovanou, *q33_kat* vytvoříme jako novou proměnnou (proto *Recode into different variable*). Nezapomeňte popsat nové varianty znaku! Po provedení *Recode* si vždy novou proměnnou necháme zobrazit na obrazovce (uděláme *Frequencies*) spolu s původní proměnnou – to pro kontrolu, zdali jsme při zadávání rekódovací operace neudělali nějakou chybu. Povšimněte si, že nově vytvořená proměnná se zařadila na úplný konec matice a ne za proměnnou *q33*. Na to je třeba pamatovat, až novou proměnnou budeme potřebovat při dalších analýzách a budeme ji hledat v dialogovém oknu.

Pokud je všechno v pořádku, je třeba celý soubor uložit. Pokud bychom tak neučinili, nová proměnná by se po zavření souboru ztratila a to nechceme, neboť ji budeme možná potřebovat v dalších analýzách.

Rozložení původní proměnné (viz tab. 4.3):

² Poznamenáváme ovšem, že se takto upravená proměnná může hodit k některým typům úloh (např. pro nějakou kontingenční tabulku), ale pro jiné typy úloh je dobré pracovat s původní proměnnou (se kterou lze pracovat jako s intervalovou). Proto je dobré při rekódování *vždy* vytvářet novou proměnnou a původní proměnnou v datech zachovat.

³ Nabízí se ovšem i jiné členění stupnice, jak uvidíme v pozn. 5.

⁴ Tímto jménem si pro naše potřeby (a pro příští analýzy) naznačujeme, že se jedná o kategorizaci proměnné *Q33*.

Tab. 4.3: Rozložení proměnné důležitost Boha v životě**Q33 Bůh - důležitost v životě**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 vůbec ne důležitý	778	40,8	42,1	42,1
	2	192	10,1	10,4	52,5
	3	133	7,0	7,2	59,7
	4	67	3,5	3,6	63,3
	5	205	10,8	11,1	74,5
	6	99	5,2	5,4	79,8
	7	84	4,4	4,6	84,4
	8	79	4,1	4,3	88,7
	9	51	2,7	2,8	91,4
	10 velmi důležitý	158	8,3	8,6	100,0
	Total	1846	96,8	100,0	
Missing	-2 neodpověděl/a	17	,9		
	-1 neví	45	2,3		
	Total	62	3,2		
Total		1908	100,0		

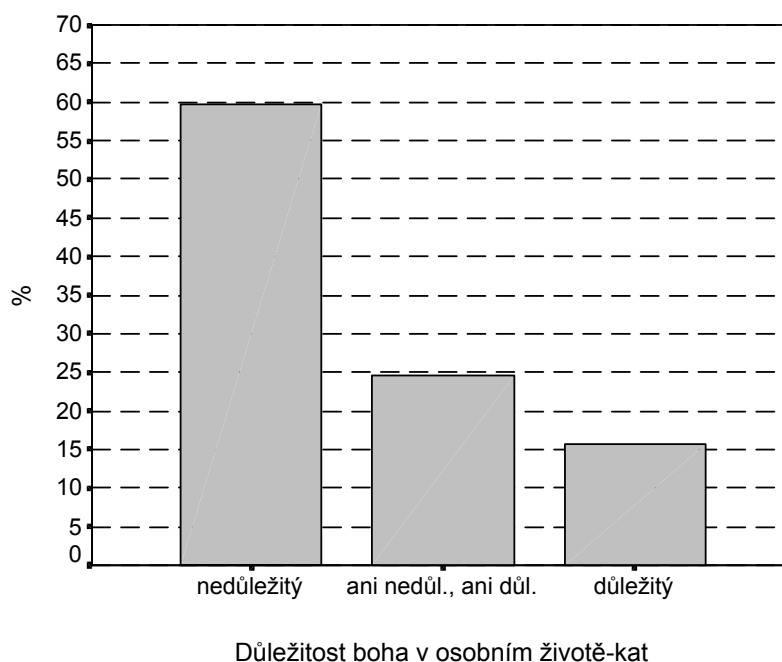
Rozložení nová proměnné je uvedeno v tab. 4.4.

Tab. 4.4: Rozložení proměnné důležitost Boha v životě upravené do tří kategorií**Q33KAT Důležitost boha v osobním životě-kat**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 nedůležitý	1103	57,8	59,7	59,7
	2 ani nedůlež., ani důlež.	455	23,9	24,7	84,4
	3 důležitý	288	15,1	15,6	100,0
	Total	1846	96,8	100,0	
Missing	System	62	3,2		
Total		1908	100,0		

Početní kontrolu rychle zjistíme, že kolapsované hodnoty původní proměnné odpovídají novým kategoriím a že tedy nová proměnná je vytvořena formálně správně. Vidíme, že Bůh je důležitý asi pro 16 % české populace, nedůležitý je pro 60 %. ⁵ Rozložení této proměnné (viz obr. 4.2) není normální, připomíná spíše *L* rozložení, ale tak to v sociálních vědách bývá.

⁵ Pozor, ale: kdybychom kategorii „důležitý“ a „nedůležitý“ definovali jinak, např. pouze hodnotami 1 a 2 původní proměnné na jednom konci stupnice a hodnotami 9 a 10 na konci druhém, byl by podíl lidí, pro něž je Bůh důležitý, 11,4 %. Nedůležitý by byl pro 52,5. Do střední kategorie by se pak umístilo 36,1 % osob.

Obr. 4.2: Rozložení důležitosti Boha v životě jednotlivce – nově kategorizovaná proměnná

2. Změna pořadí kódů

Změna v pořadí kódů stupnice prostřednictvím *Recode* nám někdy umožní proměnit charakter proměnné tím způsobem, že z nominální proměnné získáme proměnnou ordinální. Dejme tomu, že budeme zjišťovat momentální náladu respondenta otázkou „Jak se dnes cítíte?“ a navržené odpovědi budou: 1. velmi dobře, 2. dobře 3. špatně, 4. velmi špatně, 5. nedokážu říci. Proměnná s takovouto stupnicí je proměnnou nominální, neboť pátá varianta odpovědi porušuje ordinalitu předchozích čtyř stupňů. Není proto nic jednoduššího než variantu 5 vsunout mezi variantu 2 a 3 a získat tak běžnou ordinální stupnici se středem. Nové kódy by pak vypadaly následovně: 1. velmi dobře, 2. dobře 3. nedokážu říci., 4. špatně, 5. velmi špatně.

Ne vždy však platí, že tyto neurčité odpovědi typu „nevím“, „nemohu se rozhodnout“, „nedokážu říci“ jsou střední kategorií v ordinální stupnici. Jak připomíná de Vaus (2002), existuje několik výkladů, co mohou tyto odpovědi vlastně indikovat:

- mohou indikovat ambivalentní postoj, a tedy skutečně střední pozici mezi dvěma póly postojové stupnice,
- mohou ale také indikovat to, že respondent skutečně nemá názor, že tedy neví,
- indikují skutečnost, že respondent má názor, který je z jeho pohledu neakceptovatelný, proto jej nevyjádří a uchýlí se k odpovědi „nevím“ (tzv. efekt sociálně požadovaného jednání),
- ukazují na fakt, že respondent nechce vyjádřit svůj skutečný postoj, protože se stydí, není si příliš jist apod.,
- naznačují, že ze soukromých důvodů nechce respondent vyjádřit svůj postoj.

Vždy je třeba před tím, než se rozhodneme tento typ odpovědi vsunout do stupnice a zacházet s ní jako se střední hodnotou stupnice, přemýšlet, která z výše uvedených situací nastala. Ukazuje se, jak napovídá empirická zkušenost, že pouze třetina až polovina těchto odpovědí indikuje ambivalentní postoj. Ve zbývajících počtech případů bychom tedy neměli oprávnění tyto neurčité odpovědi do ordinální stupnice zařazovat a smysluplně s nimi pracovat.

3. Přetočení stupnice (obrácené pořadí kódů)

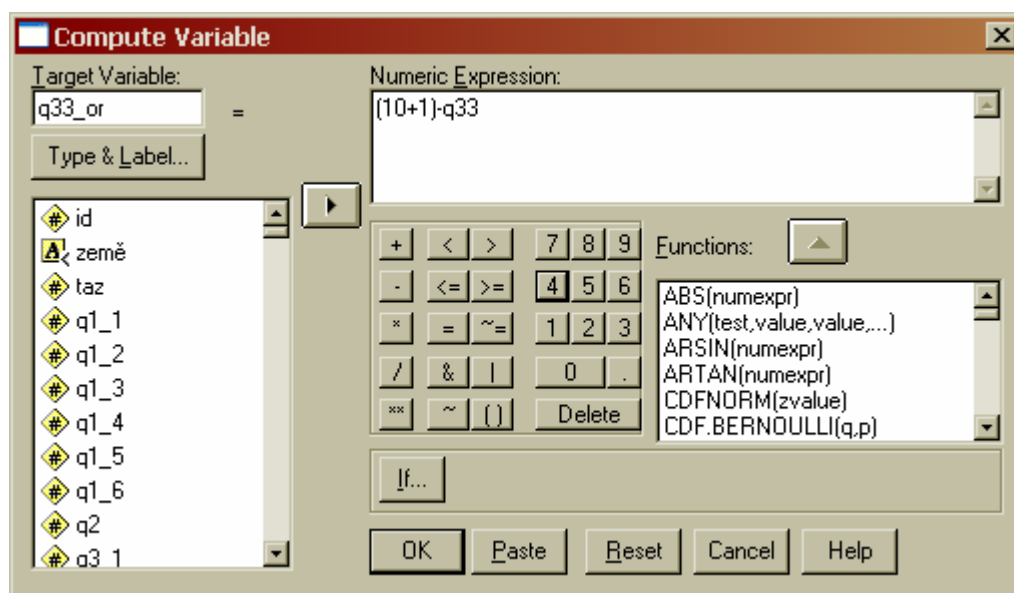
Proceduru *Recode*, jejímž prostřednictvím měníme kódy proměnné a aranžujeme je jiným způsobem, používáme také tehdy, když potřebujeme zcela změnit orientaci stupnice. Taková situace nastává většinou

tehdy, když potřebujeme z několika proměnných vytvořit součtový index. Ten vzniká tak, že sečítáme hodnoty několika proměnných. Abychom tak ale mohli učinit, musí být jejich stupnice všechny orientovány jedním směrem. Což nebývá vždycky, někdy, např. z metodologických důvodů, v baterii otázek záměrně některé stupnice obracíme – snažíme se tak udržet pozornost respondenta při rozhovoru.

Potřebujeme-li zcela obrátit (otočit) pořadí kódů, můžeme samozřejmě použít proceduru *Recode*, ale rychlejší je následující trik, který radí de Vaus (2002). Při ní využijeme transformační proceduru *Transform – Compute*. Kódy proměnné X otočíme podle následující vzorce: (Nejvyšší hodnota stupnice X + 1) – X. Takže např. pokud bychom potřebovali obrátit stupnici naší proměnné ‚důležitost Boha‘ (*q33*), zněl by zápis pro *Compute* takto:

```
COMPUTE q33_or = (10+1)-q33 .
EXECUTE .
```

kde *q33_or* je jméno nové proměnné s přetočenou (nově orientovanou) stupnicí. Zvolíte-li namísto příkazu syntaktického postup klikáním na lišty, bude vypadat příkaz pro tuto operaci takto:



Výsledkem je rozložení četností (viz tab. 4.5), které, jak zjistíte srovnáním s tab. 4.3, je vskutku přetočeným rozložením původní proměnné

Tab. 4.5: Rozložení proměnné důležitost Boha v životě pro přetočení stupnice

Q33_OR Dulezitest Boha-pretocena stupnice					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	158	8,3	8,6	8,6
	2	51	2,7	2,8	11,3
	3	79	4,1	4,3	15,6
	4	84	4,4	4,6	20,2
	5	99	5,2	5,4	25,5
	6	205	10,8	11,1	36,7
	7	67	3,5	3,6	40,3
	8	133	7,0	7,2	47,5
	9	192	10,1	10,4	57,9
	10	778	40,8	42,1	100,0
	Total	1846	96,8	100,0	
Missing	System	62	3,2		
Total		1908	100,0		

4. Vytvoření nové proměnné načítáním hodnot (procedura *Count*)

Příklad P4.2: Ve výzkumu hodnotových orientací EVS ČR 1999 byly respondentům položeny mimo jiné i tyto tři otázky:

q42: *Myslíte si, že žena musí mít děti, aby se splnilo její poslání, nebo to není nutné?*

1. potřebuje děti 2. Není to nutné -1. Neví -2. Neodpověděl(a)

q43: *Souhlasíte, nebo nesouhlasíte s následujícím výrokem? Manželství je zastaralá instituce.*

1. Souhlasí 2. Nesouhlasí -1. Neví -2. Neodpověděl(a)

q44: *Chce-li žena dítě, ale chce je vychovávat sama a žít bez muže, schvalujete to, nebo ne?*

1. Schvaluji 2. Neschvaluji 3. Záleží na okolnostech -1. Neví -2. Neodpověděl(a)

Z těchto tří proměnných je možné vytvořit novou proměnnou, která bude vyjadřovat, do jaké míry je respondent stoupencem nebo odpůrcem role ženy v moderní společnosti. Asi se shodneme v tom, že ti, kdo si myslí, že žena nemusí mít děti, aby se splnilo její poslání (q42, varianta odpovědi 2), a současně se domnívají, že manželství je zastaralá instituce (q43, varianta 1) a současně schvalují to, že žena může mít dítě a vychovávat ho bez muže, když nechce (q44, varianta 1), jsou jistě ve svých postojích blízcí názorům feministického proudu.

Abychom mohli novou proměnnou vytvořit, musíme udělat jeden drobný mezikrok. Feministické postoje u proměnných q43 a q44 jsou indikovány variantou jedna, zatímco u proměnné q42 variantou dvě. Tuto proměnnou proto musíme rekódovat tak, aby i zde byl feministický postoj obsažen ve variantě jedna. Jakmile to provedeme, budeme moci dát SPSS jednoduchý případ: Jestliže respondent odpověděl v otázkách q42, q43 a q44 variantou 1, nechť má v nové proměnné, kterou vytvoříme pod názvem *Feminist*, také hodnotu 1. Všichni ostatní nechť mají hodnotu 2 (a jsou to ti, kteří nebyli ve svých odpovědích tak feministicky consistentní).

Postup

- a) Nejprve rekódujme proměnnou *q42* tak, aby hodnota 1 v původní proměnné nabyvala v rekódované proměnné (označme si ji jako *q42rek*) hodnoty 2 a původní hodnota 2 nabyvala v *q42rek* hodnoty 1.

Q42 Žena musí mít děti, aby splnila poslání

	Frequency	Percent	Valid Percent
Valid 1 ano	795	41,7	44,1
2 není to nutné	1007	52,8	55,9
Total	1803	94,5	100,0
Missing -2 neodpověděl/a	7	,3	
-1 neví	99	5,2	
Total	105	5,5	
Total	1908	100,0	

Q42KAT Potřebuje žena děti?

	Frequency	Percent	Valid Percent
Valid 1 Není nutné	1007	52,8	55,9
2 Potřebuje	795	41,7	44,1
Total	1803	94,5	100,0
Missing System	105	5,5	
Total	1908	100,0	

Transformace proměnné *q42* proběhla úspěšně, správně jsme popsali i varianty znaku *q42kat* (*value labels*). Pro ty, kdo jsou již obeznámeni s možnostmi *Syntax*, uvádíme i syntaktický zápis tohoto příkazu:

```
RECODE
  q42
  (1=2) (2=1) (ELSE=SYSMIS) INTO q42kat .
VARIABLE LABELS q42kat 'Potřebuje žena děti?'.
EXECUTE .
```

Nyní tedy můžeme vytvořit novou proměnnou za použití procedury *Count*. Opakujeme, že v ní musíme dát SPSS v příslušném dialogovém okně příkaz, aby u proměnných q42kat, q43 a q44 napočítal výskyt odpovědi variantou 1.

Syntakticky tento zápis vypadá takto:

```
COUNT
  Feminist = q42kat q43 q44 (1) .
VARIABLE LABELS Feminist 'index z q42kat, q43 a q44' .
EXECUTE .
```

Výsledek:

FEMINIST index z q42kat, q43 a q44

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	593	31,1	31,1	31,1
1	773	40,5	40,5	71,6
2	444	23,3	23,3	94,9
3	98	5,1	5,1	100,0
Total	1908	100,0	100,0	

Jak tuto tabulku číst? 0 znamená, že to jsou respondenti, kteří ani jednou v uvedených tří otázkách neodpověděli variantou 1. Jsou to tedy antifeministé a v našem souboru jich bylo 31 %. Varianta 3 jsou respondenti, kteří ve všech třech indikátorech feminismu odpověděli variantou jedna, jsou tedy stoupenci feministické ideologie. Těch bylo 5 %. Varianta 2 v sobě skrývá respondenty, kteří ve třech otázkách dvakrát vyjádřili feministické stanovisko (ale nevíme, v kterých dvou otázkách to bylo) a varianta 1 jsou respondenti, kteří ve třech otázkách zaujali feministický postoj jenom jedenkrát. Pokud bychom chtěli být v interpretaci poněkud volnější, mohli bychom říci, že k feministickému postoji inklinuje v ČR 28 % populace (sečetli jsme varianty 3 a 2), zatímco nefeministický postoj zaujímají 72 % populace (součet variant 0 a 1). Ihned je ale třeba dodat, „feministický“ pouze z hlediska námi použitých indikátorů.

Proměnná *Feminist* je vlastně sumovaným indexem, tedy proměnnou, která má intervalový charakter. A i když má jenom krátkou, čtyřbodovou stupnici (všimněte si, že nula je zde interpretačně smysluplnou hodnotou), je možné ji popsát i z hlediska příslušných měr, to je středních hodnot a měr variability.

5. Vytvoření nové proměnné počteními operacemi (procedura *Compute*)

V sociologických výzkumech založených na standardizovaném dotazování se rutinně klade otázka na zjištění věku respondenta tak, že je požádán, aby sdělil rok svého narození. Ve výzkumu EVS zněla takto: q85: Můžete mi, prosím, sdělit rok svého narození: 19

(tazatel doplnil do formuláře dotazníku poslední dvojčíslí, takže např. u respondentky narozené v roce 1954 doplnil 54).

Takto vypadá proměnná q85 v datech (ukázka je omezena na prvních patnáct případů (*cases*)).

Procedura *Analyze – Reports – Case Summaries*:

Case Summaries^a

Cases	Variables
	Q85 Rok narození
1	34
2	61
3	35
4	41
5	68
6	54
7	42
8	64
9	54
10	54
11	35
12	35
13	54
14	54
15	76

a. Limited to first 15 cases.

Z této proměnné získáme proměnnou *vekresp* (věk respondenta) tak, že u každého provedeme jednoduchý aritmetický úkon. Poslední dvojčíslí roku narození odečteme od konstanty 99, neboť rok 1999 byl rokem, v němž se výzkum konal. Je to typická úloha pro proceduru *compute*.

Zde je výsledek (opět zobrazujeme jenom prvních 15 případů):

Case Summaries^a

Cases	Variables
	VEKRESP
1	65
2	38
3	64
4	58
5	31
6	45
7	57
8	35
9	45
10	45
11	64
12	64
13	45
14	45
15	23

a. Limited to first 15 cases.

Můžeme zkontrolovat, zdali SPSS počítá správně. Respondent číslo 1 se narodil v roce (19)34. V roce 19(99) mu bylo 65 let.

A pro zájemce uvádíme ještě syntax zápisu této procedury.

```
COMPUTE vekresp = 99 - q85 .
EXECUTE .
```

6. Vytvoření nové proměnné prostřednictvím logických podmínek (procedura *If, Compute If*)

Někdy musíme vytvořit novou proměnnou ne na základě aritmetických operací, nýbrž na základě operací logických. Hezkou ukázkou takového postupu je tvorba postmaterialistické typologie, kterou do sociologie vnesl Ronald Inglehart. Abyste tomuto příkladu dobře porozuměli, budete si muset nejdříve přečíst stat' L. Rabušice: Je česká společnost 'postmaterialistická'? *Sociologický časopis*, 36 (1): 5—22, 2000, především pasáže na str. 4—6. Ve výzkumu EVS byly příslušné otázky formulovány takto:

POKYN: PŘEDLOŽTE LÍSTEK č...

A532: Hodně se dnes hovoří o tom, k jakým cílům by naše země měla směřovat v nejbližších deseti letech. Na tomto lístku jsou uvedeny některé z cílů, jimž by různí lidé dali přednost. Kdybyste si musel(a) vybrat, kterou ze snah na tomto lístku byste označil(a) za nejdůležitější?

(POKYN: VYZNAČTE POUZE JEDNU ODPOVĚĎ VE SLOUPCI „a“ DOLE)

A533: A který z cílů by byl druhý nejdůležitější?

(POKYN: VYZNAČTE POUZE JEDNU ODPOVĚĎ VE SLOUPCI „b“ DOLE)

	A532	A533
(a)	(b)	
První výběr	Druhý výběr	
Udržet pořádek ve státě		1
Dát lidem větší možnost hovořit do důležitých vládních rozhodnutí		2
Bojovat proti růstu cen		3
Bránit svobodu projevu		4
Neví		-1
Neodpověděl(a)		-2

Stupnici Inglehartova materialismu/postmaterialismu vytvoříme prostřednictvím příkazu „COMPUTE (IF)“. Novou proměnnou nazvěme *pm_orig* neboli post-materialismus–originální (původní) stupnice.

```
COMPUTE pm_orig = 2.
IF ((q55 = 1 and q56 = 3) or (q55 = 3 and q56 = 1)) pm_orig = 1.
IF ((q55 = 2 and q56 = 4) or (q55 = 4 and q56 = 2)) pm_orig = 3.
VALUE LABELS pm_orig 1 'Mater' 2 'Mix' 3 'Postma'.
```

A zde je výsledek našeho snažení:

PM_ORIG

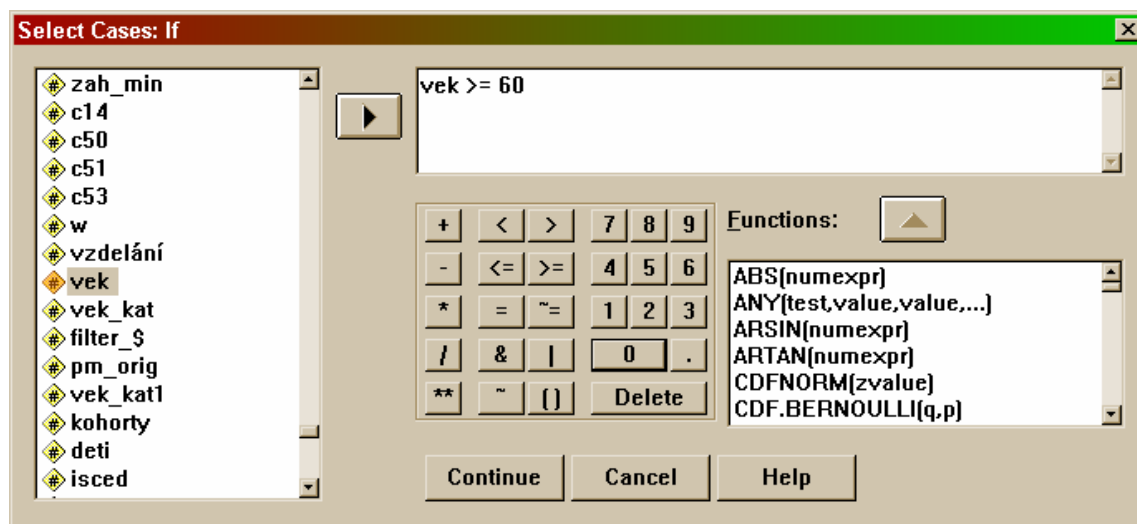
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 Mater	462	24,2	24,2	24,2
2 Mix	1268	66,5	66,5	90,7
3 Postma	178	9,3	9,3	100,0
Total	1908	100,0	100,0	

7. Manipulace s datovým souborem

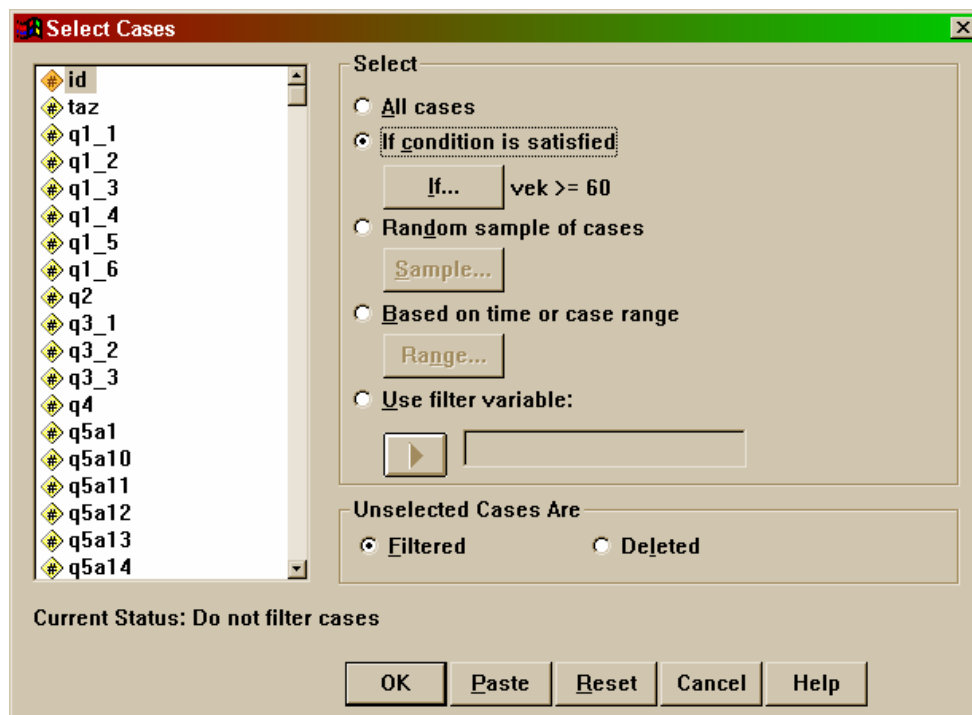
K výše popsaným transformačním procedurám lze také přiřadit manipulaci s datovým souborem – je možné pracovat pouze s podsouborem případů. Např. nás může zajímat analýza lidí ve věku 60 let a starších. K vývěru takového podsouboru použijeme proceduru

Data – Select Cases – If condition is satisfied

Po kliknutí na tlačítko **If...** se objeví dialogové okno, do nějž vepíšeme příslušnou podmínku pro výběr (viz).



Po kliknutí na tlačítko **Continue** si dejte pozor, aby v dialogovém okně, které se objeví, bylo nastaveno, že případy, které nesplňují podmínku (to jsou tedy nevybrané případy neboli **Unselected Cases**) jsou **Filtered** – filtrovány a nikoliv **Deleted** – vymazány (viz obr. níže). Jak napovídá název, filtrované případy zůstávají dále v souboru, pouze se s nimi nepracuje, vymazané případy jsou smazány a zůstávají pouze případy splňující podmínku.



Když si pro kontrolu necháme udělat rozložení takto redukováného souboru, získáme výsledek, který je uveden dole v tabulce (viz). Podmínky lze samozřejmě různě kombinovat, např. bylo by možné získat podsoubor mužů ve věku 60+ let, kteří ještě pracují apod. Někdy mají tyto operace analytický smysl.

VEK

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	60	23	5,0	5,0	5,0
	61	16	3,5	3,5	8,5
	62	33	7,4	7,4	15,8
	63	21	4,6	4,6	20,4
	64	24	5,4	5,4	25,8
	65	26	5,7	5,7	31,5
	66	44	9,7	9,7	41,2
	67	32	7,0	7,0	48,2
	68	27	6,0	6,0	54,2
	69	26	5,7	5,7	60,0
	70	19	4,3	4,3	64,3
	71	22	5,0	5,0	69,2
	72	21	4,6	4,6	73,8
	73	20	4,4	4,4	78,3
	74	26	5,8	5,8	84,0
	75	27	5,9	5,9	90,0
	76	11	2,5	2,5	92,5
	77	7	1,5	1,5	93,9
	78	8	1,9	1,9	95,8
	79	5	1,0	1,0	96,8
	80	4	,8	,8	97,7
	81	2	,5	,5	98,2
	82	1	,3	,3	98,5
	84	4	,9	,9	99,3
	85	2	,4	,4	99,7
	87	1	,2	,2	99,9
	88	0	,1	,1	100,0
Total		453	100,0	100,0	

Pro zvědavé dodáváme, že existuje možnost pracovat v příslušné statistické proceduře jen s některými případy (například jen postmaterialisty) a filtrovat tak pouze dočasně (anglicky temporary). Pro tento způsob práce je nutno před příkazem pro spuštění procedury napsat ještě dva řádky pro určení filtru a jeho dočasnosti (slovko temp). Celkový sled příkazů pro četnosti věku postmaterialistů může tedy být následující:

```
Temp.
Sel if pm_orig = 3
frequency vekresp.
```

Transformace proměnných je při práci s daty poměrně častou operací. Doporučujeme vám, abyste si archivovali všechny příkazy pro úpravu proměnných – to je abyste si ukládali jejich syntax (soubory s příponou .SPS). Druhý z autorů tohoto manuálu to dělá tak, že si do adresáře, v němž má data k nějakému výzkumnému problému, ukládá pravidelně soubor, který si pojmenovává jako *new_variables.SPS*. Obsah tohoto souboru vypadá např. takto:

*Vytvoreni nove promenne Job status.

RECODE

t5

(5=1) (7=3) (8=4) (9=5) (6=6) (10=6) (1 thru 2=1) (3 thru 4=2) (ELSE=SYSMIS) INTO

Job_status.

EXECUTE.

* Nova promenna - Kohabitace.

```
COMPUTE Cohab = 0 .
```

```
EXECUTE .
```

```
IF (m3 = 1 and m4 = 2) Cohab = 1 .
```

```
IF (m6 = 1) Cohab = 0 .
```

```
IF (m4 = 1) Cohab = 0 .
```

```
Variable labels Cohab 'Nesezdane souziti'.
```

```
Value labels Cohab 1 'ano' 2 'ne'.
```

```
EXECUTE .
```

Všimněte si, že každá transformace je nějakým způsobem pojmenovaná. Toto pojmenování se děje tak, že na první pozici v řádku vepíšeme hvězdičku. Pro SPSS to je znamení, že bude následovat text, jemuž nemá věnovat žádnou pozornost, neboť není psán jazykem, jemuž by rozuměl – slouží jako poznámka či komentář k příslušné transformační operaci. U syntaxe pro kohabitaci je, na rozdíl od syntaxe pro Job_status navíc uveden i příkaz pro popis variant proměnné (Value labels). Tak je to správně a mělo by tomu tak být vždy. Vytváříme-li novou proměnnou, nesmíme být líní a musíme si popsat i nově vytvořené varianty. Pokud to neuděláme, za pár týdnů se nám to strašlivě vymstí. Jejich názvy zapomeneme a existující nová proměnná se stane pro další analýzu nepoužitelná.

I v prostředí SPSS tak platí staré dobré přísloví: CO MŮŽEŠ UDĚLAT HNED, NEODKLÁDEJ NA RÁNO.