

## LEKCE10 PŘÍKLAD

### JAK ODHALIT VLIV TŘETÍ PROMĚNNÉ?

Vztah mezi dvěma proměnnými nemusí vždy vyjadřovat vztah skutečně existující. Např. zjištěná statistická asociace mezi sociální vrstvou a volebními preferencemi může být ovlivněna další proměnnou, a to příjmem respondenta – respondenti, kteří mají vyšší příjmy jsou současně i ve vyšší vrstvě a mají tendenci volit pravici. Nebo tento vztah může mít různou těsnost v různých věkových kategoriích apod. Zkrátka, zjištěním souvislosti (nebo její absence tam, kde bychom ji očekávali) mezi dvěma proměnnými ještě nemáme vyhráno a měli bychom být obezřetní, dokud ji neproověříme testem na možný vliv nějaké třetí proměnné. Řečeno jinými slovy, se vztahem mezi dvěma proměnnými je to jako se vztahem dvou partnerů v životě – může být významně ovlivněn působením třetí osoby.

Způsob, jak vliv třetí proměnné odhalit, je nazván v anglické literatuře termínem *elaboration*, rozpracování. Toto rozpracování vztahu dvou proměnných se zřetelem k působení proměnné třetí činíme prostřednictvím dvou postupů:

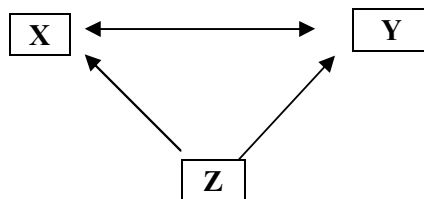
1. Zavedením třetí, testové proměnné do třídění druhého stupně nebo
2. matematicko-statisticky výpočtem tzv. parciálních korelačních koeficientů.

#### Ad 1) Zavedení testové proměnné

Z technického hlediska není zavedení testové proměnné žádný problém. V proceduře *CROSSTABS* přidáme ještě jednu proměnnou (do okénka *layer*), čímž vytvoříme třídění třetího stupně, v němž je tato třetí proměnná v jejích jednotlivých variantách držena konstantní. Jediným limitem pro provedení této operace je dostatečně velký datový soubor. V malých souborech při třídění tří proměnných, zvláště když mají čtyř až pětistupňové stupnice měření, dochází velmi často k tomu, že obsazení jednotlivých políček je malý a na hranici statistické korektnosti. Interpretačně ovšem tato technologie už tak jednoduchá není. Vztah mezi třemi proměnnými může mít totiž interpretačně podobu čtyř následujících modelů.

1. Test na zdánlivou korelaci má následující modelovou podobu (viz model 1).

#### Model 1:



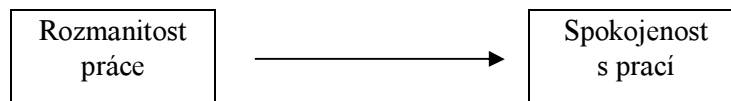
Původně zjištěná korelace mezi X a Y byla způsobena společnou příčinou Z. Korelace XY je tak korelací zdánlivou, falešnou (*spurious*).

Ilustrativní příklad (převzato z Bryman a Cramer, 1997)<sup>1</sup>

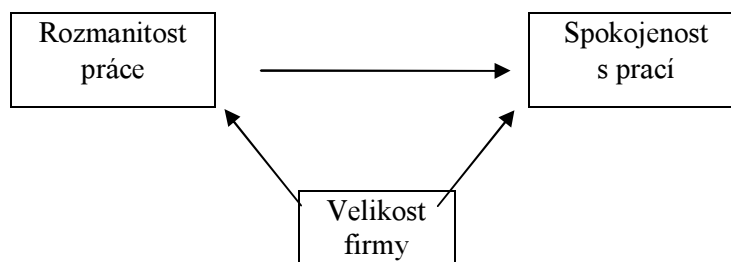
Tab. 10.1: Spokojenost s prací podle rozmanitosti práce (sloupcová %)

Spokojenost s prací	Rozmanitost práce	
	Ano	Ne
Spokojen	80	24
Nespokojen	20	76

Vidíme, že rozmanitost práce (nezávisle proměnná) ovlivňuje spokojenost s prací, že mezi rozmanitostí práce a spokojeností s prací existuje souvislost:



Nemůže být ale tato korelace falešná? Možná, že do tohoto vztahu vstupuje třetí proměnná, již je velikost podniku. Ta může ovlivňovat jak rozmanitost práce, tak také spokojenost s prací. Modelově lze tuto situaci naznačit následovně:



Tab. 2: Spokojenost s prací podle rozmanitosti práce při zavedení testové proměnné velikost podniku (sloupcová %)

Spokojenost s prací	velké podniky		malé podniky	
	rozmanitost práce		rozmanitost práce	
	Ano	Ne	Ano	Ne
Spokojen	95	85	20	9
Nespokojen	5	15	80	91

Tabulka 2 je ukázkou třídění třetího stupně. Všimněte si, že vlastně sestává ze dvou tabulek – v jedné je vztah mezi spokojeností a rozmanitostí práce tabelován v prostředí velkých podniků, v té druhé je v prostředí podniků malých. Srovnáte-li výsledek třídění v tab. 1 s tříděním v tabulce 2, zjistíte, že původní vztah, nalezený v tabulce 1, se nyní podstatně snížil. Zatímco v tab. 1 byl rozdíl mezi spokojenými u práce rozmanité a práce monotónní  $\epsilon = 56\%$  (80-24), v tab. 2 je  $\epsilon$  roven 10 % (95-85) nebo 11 % (20-11).<sup>2</sup> Jelikož 10 % a 11 % je výrazně méně než 56 %, musíme konstatovat, že došlo

<sup>1</sup> Bryman, A., Cramer, D. 1997. *Quantitative Data Analysis with SPSS for Windows*. Routledge.

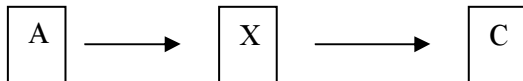
<sup>2</sup> Z lekce 8. již víme, že rozdíl mezi procenty v políčkách tabulky se nazývá epsilon ( $\epsilon$ ).

k výrazné redukci původního vztahu. Takže když byla do původního vztahu zavedena třetí proměnná „velikost podniku“ (řeceno jinými slovy, když je tento vztah kontrolován z hlediska třetí proměnné, která je držena konstantní), vztah mezi spokojeností s prací a její rozmanitostí mizí.

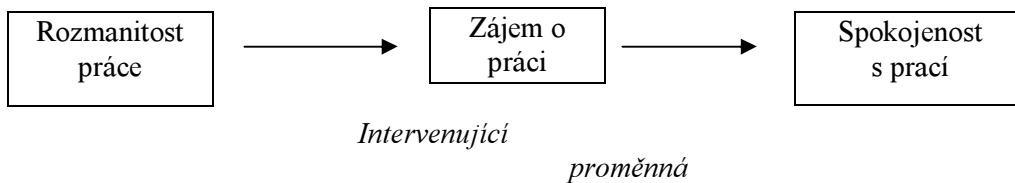
## 2. Test na intervenující proměnnou

Třetí proměnná dostává podobu intervenující proměnné tehdy, když je jednak produktem nezávisle proměnné a současně je příčinou proměnné závislé. Tuto situaci lze modelově znázornit následovně (viz model 2).

**Model 2:**



V našem příkladu by do vztahu mezi rozmanitostí práce a spokojeností s prací, mohla vstoupit třetí proměnná, zájem o práci.



Vztah mezi rozmanitostí práce a spokojeností s prací také není přímý, nýbrž je zprostředkovaný.

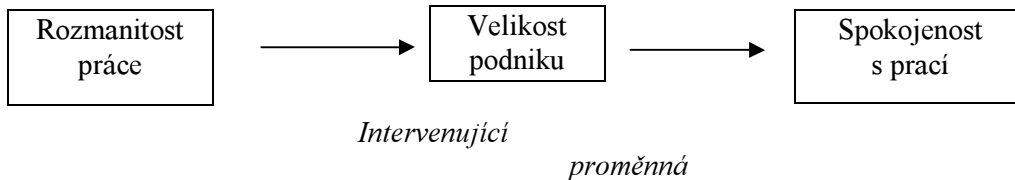
Hledání intervenující proměnné se také nazývá *vysvětlením*. Důvod je prostý: Pokud takovou intervenující proměnnou odhalíme, jsme schopni původní bivariační vztah vysvětlit.

**Tab. 3: Vztah mezi rozmanitostí práce a spokojeností při kontrole proměnné zájem o práci (sloupcová %)**

Spokojenost s prací	O práci:			
	má zájem		nemá zájem	
	rozmanitost práce		rozmanitost práce	
	Ano	Ne	Ano	Ne
Spokojen	93	80	30	10
Nespokojen	7	20	70	90

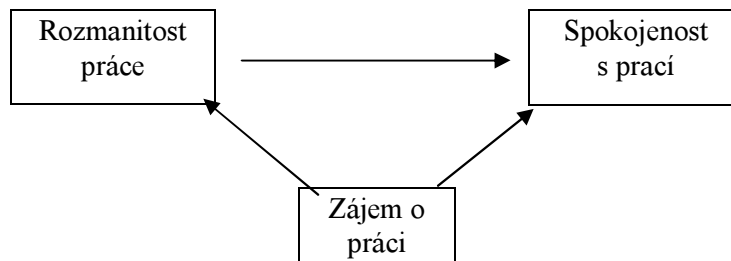
Opět srovnáme příslušné procentuální rozdíly. Zatímco v tab. 1 byl u spokojených rozdíl 56%, v tab. 3 je rozdíl e jednak 13% (93-80), jednak 20% (30-10). Rozdíly tedy nebyly redukovány na nulu, což by naznačovalo, že celý původní vztah byl způsoben proměnnou zájem o práci, avšak došlo ke značnému snížení. Proto tvrdíme, že do vztahu mezi spokojeností s prací a její rozmanitostí intervenuje proměnná zájem o práci. Kdyby rozdíly v tab. 3 zůstaly kolem oněch původních 56%, museli bychom konstatovat, že zájem o práci není intervenující proměnnou.

Ale pozor: výpočet pro odhalení intervenující proměnné je stejný, jako v případě hledání zdánlivé korelace. Jak tedy poznáme, o jaký model se jedná? Návod je prostý. Logickou úvahou. Kdybychom chtěli předpokládat, že v prvním modelu bude velikost firmy intervenující proměnnou, tedy takto:



museli bychom se domnívat, že rozmanitost práce může ovlivnit velikost podniku. Což je pochopitelně nesmysl.

Podobně, kdybychom chtěli model 2 přeskládat tak, abychom testovali vztah na zdánlivou korelaci, tedy takto:

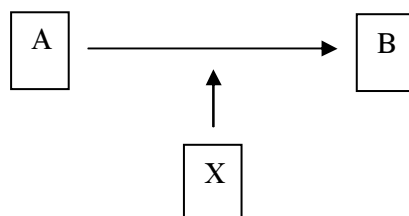


museli bychom se domnívat, že zájem o práci může ovlivnit míru rozmanitosti práce. To je opět nesmysl.

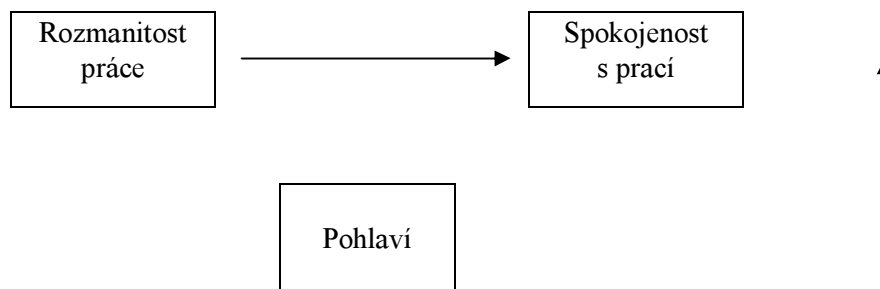
### 3. Test na upravený vztah

Upravený vztah se objevuje tehdy, když po zavedení testové proměnné se původní vztah u jejich některých variant podrží, u jiných vymizí.

#### Model 3:



V příkladu o vztahu mezi rozmanitostí práce a spokojeností s prací by na tento vztah mohlo mít vliv pohlaví zaměstnance.



**Tab. 4: Vztah mezi rozmanitostí práce a spokojeností při kontrole proměnné pohlaví (sloupcová %)**

Spokojenost s prací	Muži		Ženy	
	rozmanitost práce		rozmanitost práce	
	Ano	Ne	Ano	Ne
Spokojen	95	10	57	45
Nespokojen	5	90	43	55

Muži jsou podle dat v tabulce 4 mnohem spokojenější než ženy,  $\epsilon$  je 85% (95-10) u mužů, ale jenom 12% (57-45) u žen. Nastala situace, kdy původní vztah platí jenom pro některé kategorie testovací proměnné, v daném příkladu pro muže.

V případě, jako je tento, tedy když nějaký vztah je různý pro jednotlivé kategorie testovací proměnné, hovoříme o efektu interakce (*interaction effect*).

#### Příklady z reálných dat.

Ve výzkumu Mládež a AIDS<sup>3</sup> jsme našli tento vztah mezi mírou religiozity mladých respondentů (ve věku 16-18 let):

**Tab. 5: Vztah mezi mírou religiozity a zkušeností se sexuálním stykem (řádková %)**

Religiozita	Sexuální styk	
	Ne	Ano
Ano	65	35
Ne	53	47
Celkem	57	43

<sup>3</sup> Viz Rabušic, L. Kepáková, K. 1999. Sexuální chování adolescentů a riziko HIV. *Sociologický časopis*, 35 (2): 161–179.

Ti, kdo nebyli nábožensky založení měli v tomto věku častěji zkušenost se sexuálním stykem než ti, kdo nábožensky založení byli (rozdíl byl 65-53 = 12 %). Jelikož jsme si ale nebyli jisti, zdali do tohoto vztahu neintervenuje velikost místa bydliště, zavedli jsme tuto proměnnou jako proměnnou testovací.

**Tab. 5a: Vztah mezi mírou religiozity a zkušeností se sexuálním stykem (řádková %)**

Religiozita	Brno		Okolní města		Vesnice	
	Sexuální styk Ne	Sexuální styk Ano	Sexuální styk Ne	Sexuální styk Ano	Sexuální styk Ne	Sexuální styk Ano
Ano	<b>64</b>	36	<b>62</b>	38	<b>69</b>	31
Ne	<b>61</b>	39	<b>50</b>	50	<b>45</b>	55
Celkem	62	38	53	47	53	47

Rozdíl mezi věřícími a nevěřícími, kteří ještě neměli zkušenost se sexuálním stykem byl v celém souboru 12% (65-53). V podsouboru mládeže žijící v Brně byl ovšem tento rozdíl jen 3%, v okolních městech 12% a ne vesnicích 24%. Velikost místa bydliště tedy do tohoto vztahu mezi zkušeností religiozitou a zkušeností se sexuálním stykem určitým způsobem vstupuje: neplatí pro respondenty z Brna a naopak je výrazný u respondentů z vesnice.

### Jak zadat třídění třetího stupně v SPSS

Příklad:

Na datech EVS\_ČR1999.sav ověřujeme hypotézu, že postoj k tomu, zdali děti jsou naplněním života člověka, je ovlivněn tím, zdali je respondent věřící, nebo nevěřící (dichotomická proměnná *q28rec*).

Indikátor postoje k dětem zněl:

*Člověk musí mít děti, aby byl jeho život naplněn* (q45\_1REC) a respondent byl dotázán, zdali s tímto výrokem souhlasí nebo ne (předložena byla pětistupňová Lickertova stupnice, ale pro potřeby tohoto příkladu jsme ji trichotomizovali).

Třídění druhého (*Crosstabs*) stupně přineslo tento výsledek:

Q45\_1REC dite naplnenim zivota cloveka \* Q28REC naboz. orientace  
Crosstabulation

			Q28REC naboz. orientace		Total
			1 verici	2 neverici	
Q45_1REC dite naplnenim zivota cloveka	1 souhlas	Count	477	590	1067
		Column %	62,2%	58,9%	60,3%
	2 ani souhlas, ani neshl	Count	126	218	344
		Column %	16,4%	21,8%	19,4%
	3 nesouhlas	Count	164	194	358
		Column %	21,4%	19,4%	20,2%
Total	Count	767	1002	1769	
	Column %	100,0%	100,0%	100,0%	

Vidíme, že jak věřící, tak nevěřící souhlasí s uvedeným výrokiem v podstatě ve stejné míře (62 % věřících a 59 % nevěřících). Potvrzuje to i nulový koeficient asociace – Somersovo D (viz tabulku níže).

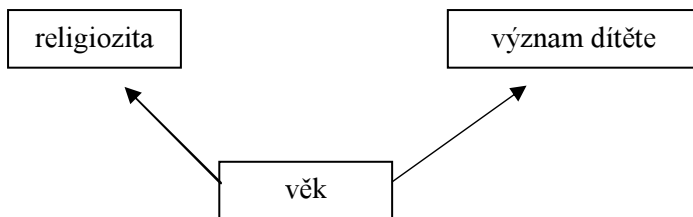
Directional Measures

	Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Ordinal by Somers Symmetric	,017	,023	,756	,450
Ordinal 'd Q45_1REC dite naplnenim zivota cloveka Dependent	<b>,018</b>	,024	,756	,450
Q28REC naboz. orientace Dependent	,016	,021	,756	,450

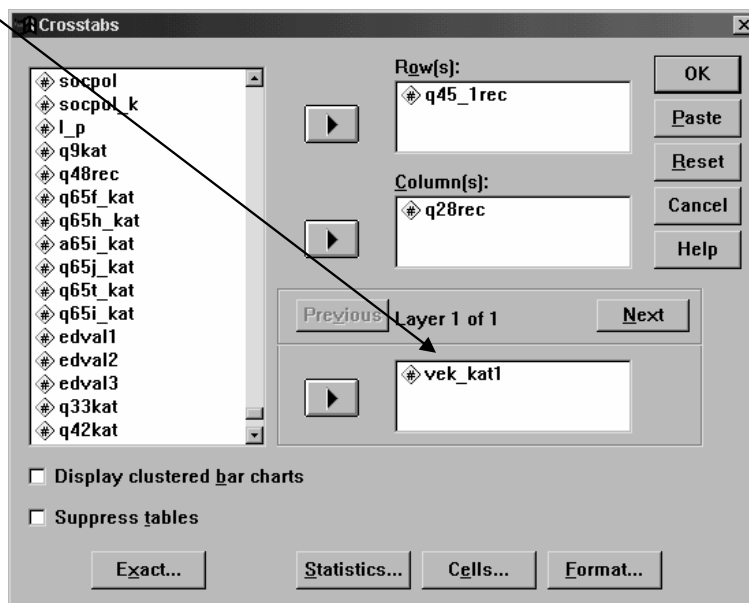
a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Není ale tento vztah ovlivněn působením třetí proměnné, již je věk? Nemůže platit model zdánlivé nekorelace?



K odpovědi na tuto otázku použijeme třídění třetího stupně. V dialogovém okně pro *Crosstabs* přidáme třetí proměnnou.



Výsledek třídění ukazuje, že minimální rozdíl v postoji k významu dítěte se mezi věřícími a nevěřícími zůstal zachován u všech tří věkových skupin. Stejně tak se uchoval i velmi nízký koeficient asociace.

Q45\_1REC dítě naplnením života člověka \* Q28REC naboz. orientace \* VEK\_KAT1 Vekové kategorie  
Crosstabulation

VEK_KAT1 Vekové kategorie				Q28REC naboz. orientace		Total
				1 verici	2 neverici	
1 18-29	Q45_1REC dítě naplnením života člověka	1 souhlas	Count	61	129	190
			Column %	44,5%	49,4%	47,7%
		2 ani souhlas, ani neshl	Count	28	71	99
			Column %	20,4%	27,2%	24,9%
		3 nesouhlas	Count	48	61	109
			Column %	35,0%	23,4%	27,4%
	Total	Count	137	261	398	
	Column %	100,0%	100,0%	100,0%		
2 30-49	Q45_1REC dítě naplnením života člověka	1 souhlas	Count	128	226	354
			Column %	57,1%	54,9%	55,7%
		2 ani souhlas, ani neshl	Count	38	100	138
			Column %	17,0%	24,3%	21,7%
		3 nesouhlas	Count	58	86	144
			Column %	25,9%	20,9%	22,6%
	Total	Count	224	412	636	
	Column %	100,0%	100,0%	100,0%		
3 50+	Q45_1REC dítě naplnením života člověka	1 souhlas	Count	287	235	522
			Column %	71,0%	71,6%	71,3%
		2 ani souhlas, ani neshl	Count	60	46	106
			Column %	14,9%	14,0%	14,5%
		3 nesouhlas	Count	57	47	104
			Column %	14,1%	14,3%	14,2%
	Total	Count	404	328	732	
	Column %	100,0%	100,0%	100,0%		



## Directional Measures

VEK_KAT1 Vekové kategorie				Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
1 18-29	Ordinal by Ordinal	Somers' d	Symmetric	-,080	,048	-1,678	,093
			Q45_1REC dite naplnenim zivota cloveka Dependent	-,097	,057	-1,678	,093
			Q28REC naboz. orientace Dependent	-,069	,041	-1,678	,093
2 30-49	Ordinal by Ordinal	Somers' d	Symmetric	-,004	,038	-,104	,917
			Q45_1REC dite naplnenim zivota cloveka Dependent	-,005	,044	-,104	,917
			Q28REC naboz. orientace Dependent	-,004	,034	-,104	,917
3 50+	Ordinal by Ordinal	Somers' d	Symmetric	-,005	,036	-,134	,893
			Q45_1REC dite naplnenim zivota cloveka Dependent	-,005	,034	-,134	,893
			Q28REC naboz. orientace Dependent	-,005	,037	-,134	,893

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

**Kontrola bivariačního vztahu se zavedením testové proměnné výpočtem parciálního koeficientu**

Ne vždy je možné dělat třídění třetího stupně. Proto je dobré vědět, že stejný účel jako skutečná manipulace s daty splní i jednoduchý výpočet parciálního korelačního koeficientu.

Postup:

1. S pomocí SPSS vypočítáme příslušné bivariační korelační koeficienty mezi třemi proměnnými.
2. S pomocí SPSS vypočítáme parciální koeficient:

Interpretace parciálního koeficientu:

- a) Nastane situace, že vypočtený parciální koeficient má přibližně stejnou hodnotu, jako původní korelace. Co to znamená? Testová proměnná nemá na původní vztah vliv a my si můžeme být jisti, že původní korelace není zdánlivá.
- b) Parciální koeficient je výrazně nižší než původní korelace a je blízký nule. V takovém případě to znamená, že testová proměnná plně vysvětluje původní vztah, který byl zdánlivý.
- c) Parciální koeficient se změní jenom částečně. Pak je testová proměnná jen částečně vhodná k vysvětlení původní korelace.

Pokud nastane případ b, musíme se rozhodnout, o jaký model vysvětlení přijmeme. Zdali model zdánlivé korelace či intervenující proměnné.

**Příklad:** Ve výzkumu mládeže byl Spearmanův koeficient korelace mezi postojem k časnosti zahájení sexuálního života (*mladí lidé v mém věku by ještě neměli mít sexuální styk*) a pohlavím 0,60. Jelikož očekáváme, že tato korelace by mohla být zdánlivá, zavádíme do tohoto vztahu kontrolní proměnnou zkušenost se sexuálním stykem.

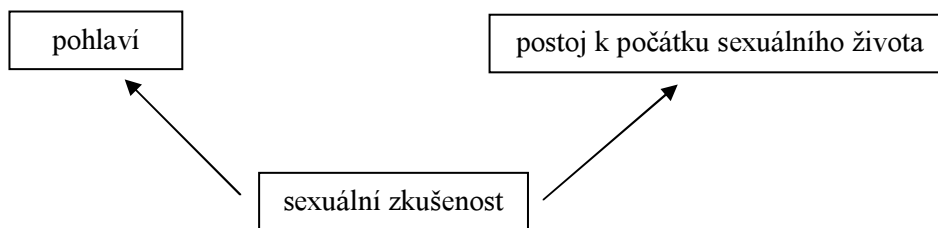
Příslušné vzájemní interkorelace korelace byly:

postoj	pohlaví	zkušenost
(a)	(b)	(c)

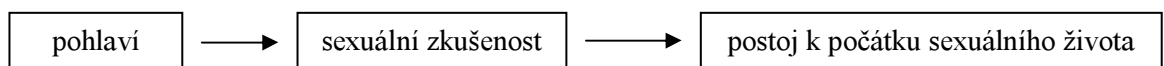
postoj (a) 1,00  
 pohlaví (b) 0,60 1,00  
 zkušenost (c) 0,70 0,80 1,00  
 Vypočtená parciální korelace  $r_{ab.c} = 0,09$

Původní korelace mezi pohlavím a postojem 0,60 se drasticky snížila (parciální korelace je 0,09), takže můžeme směle tvrdit, že se jedná o proměnnou, která má vliv. Teď ještě zbývá rozhodnout, zdali jde o model zdánlivé korelace anebo model intervenující proměnné.

Model zdánlivé korelace by vypadal takto:



Tento model není pro naši situaci adekvátní, neboť sexuální zkušenost nemůže ovlivňovat pohlaví. Musí se tedy o model intervenující proměnné:



### Příklad výpočtu parciální korelace v SPSS

Ve výzkumu EVS1999 byla respondentům položena následující otázka: *Prosím řekněte mi pro každý z následujících výroků, zda dané jednání je vždy ospravedlnitelné, není nikdy ospravedlnitelné, nebo něco mezi tím.*

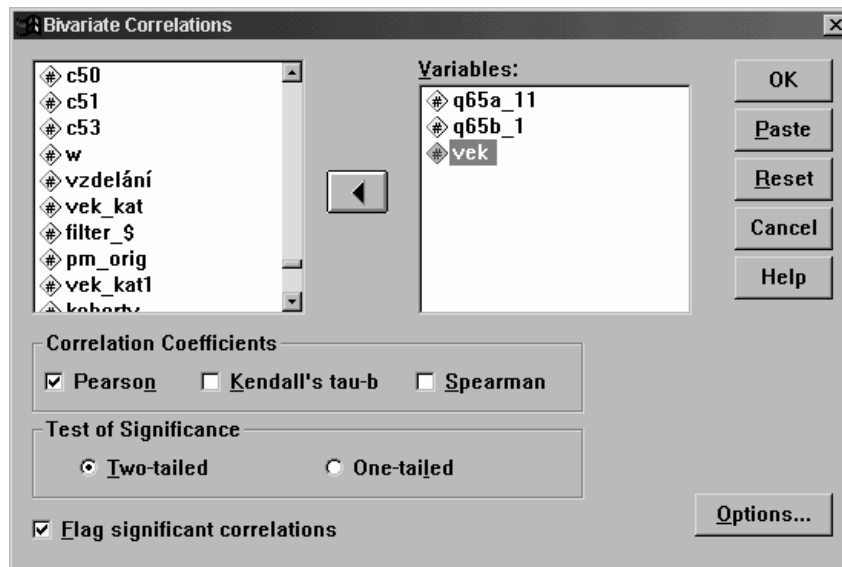
**K Euthanasie (ukončení života nevléčitelně nemocného)**

**L Sebevražda**

Respondentům byla předložena karta a na ní byla stupnice:

1 2 3 4 5 6 7 8 9 10  
 nikdy vždy

Nejdříve nás zajímalo, zdali se postoje k ospravedlnění euthanasie nebo sebevraždy liší podle věku. Jelikož postoje jsou měřeny na stupnici, kterou lze považovat za quasikardinální a věk je kardinální proměnná, můžeme na tuto otázku odpovědět s pomocí výpočtu Pearsonova koeficientu korelace.  
*Analyze — Correlate — Bivariate*



Výsledek výpočtu:

Correlations

		Q65A_11 Euthanasie	Q65B_1 Sebevražda
Q65A_11 Euthanasie	Pearson Correlation		
	Sig. (1-tailed)		
	N		
Q65B_1 Sebevražda	Pearson Correlation	,337*	
	Sig. (1-tailed)	,000	
	N	1775	
VEK	Pearson Correlation	-,218*	-,206*
	Sig. (1-tailed)	,000	,000
	N	1800	1852

\*\* . Correlation is significant at the 0.01 level (1-tailed).

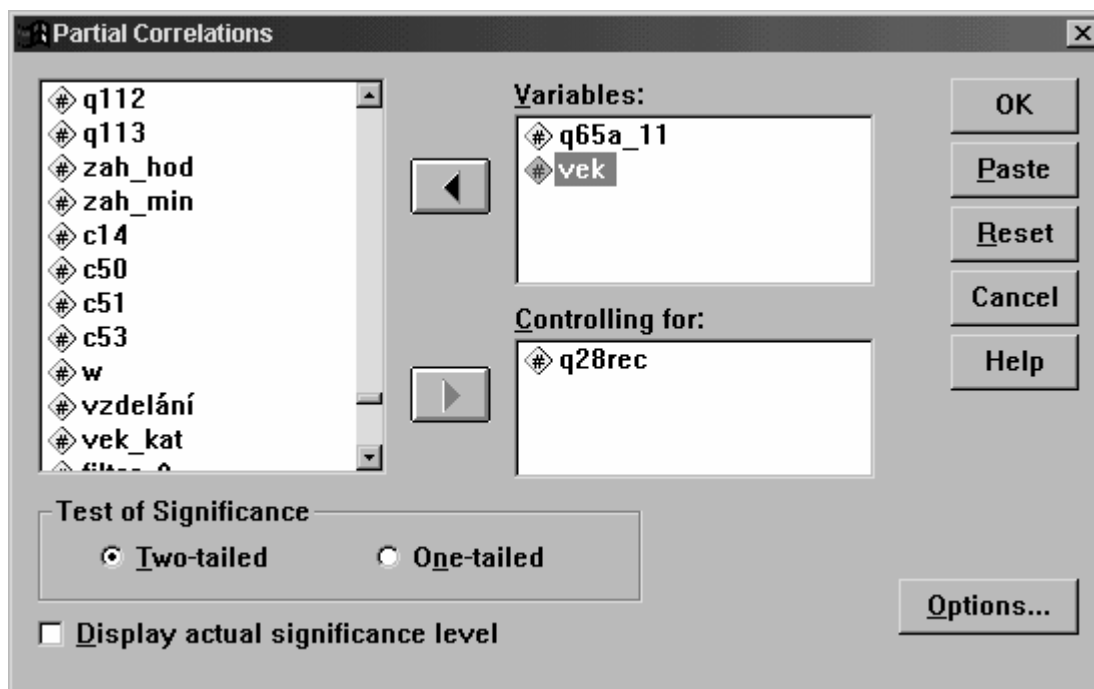
Vidíme, že oba postoje, to je k euthanasii i sebevraždě, jsou věkem ovlivněny. Se zvyšujícím se věkem klesá jak ospravedlnitelnost euthanasie ( $r = -0,22$ ), tak sebevraždy ( $r = -0,21$ ). Starší lidé jsou zkrátka více proti euthanasii a sebevraždě než lidé mladší.

Položme si ale v této souvislosti otázku, zdali tyto postoje nejsou ovlivněny náboženskou orientací, to je zdali do vztahu mezi věkem a postojem k euthanasii či sebevraždě neintervenuje to, zdali je respondent věřící nebo nevěřící:

věk respondenta → náboženská orientace → postoj k euthanasii

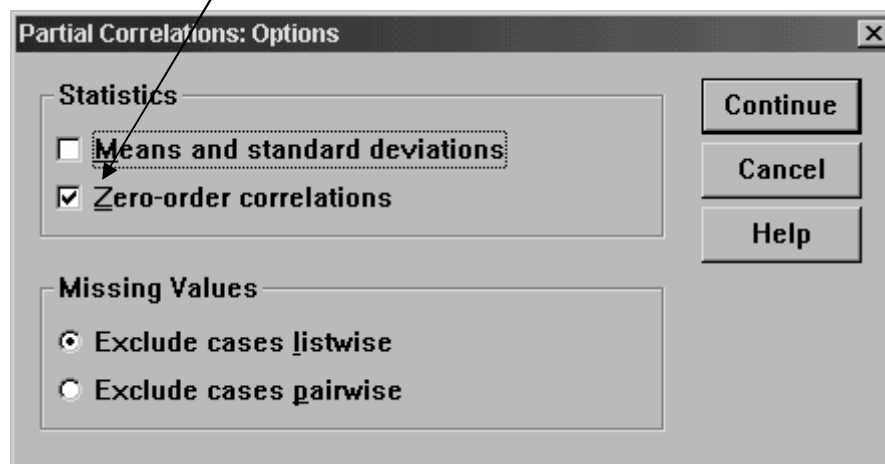
Zjistíme to prostřednictvím výpočtu parciální korelace, kdy do bivariačního vztahu zavedeme proměnnou religiozita (*q28rec*).

*Analyze — Correlate — Partial*



Pozn. Doporučuji nechat políčko *Display actual significance level* nezaškrtnuté.

V *Options* lze zvolit, aby se při výpočtu parciální korelace (říká se mu také korelace prvního řádu) tiskl také koeficient nultého řádu (*zero-order correlations*), to je prostý bivariační koeficient.



Výsledek výpočtu:

**Tab. 6: PARTIAL CORRELATION COEFFICIENTS**

### Zero Order Partial

Q65A_11	VEK	Q28REC
---------	-----	--------

Q65A_11	1,0000		
VEK	-,2125**	1,0000	
Q28REC	,1684**	-,1920**	1,0000

\* - Signif. LE ,05      \*\* - Signif. LE ,01      (1-tailed)

## PARTIAL CORRELATION COEFFICIENTS

Controlling for..      Q28REC

	Q65A_11	VEK
Q65A_11	1,0000	-,1862**
VEK	-,1862**	1,0000

\* - Signif. LE ,05      \*\* - Signif. LE ,01      (1-tailed)

Původní korelace (zero order, nultého řádu) mezi postojem k euthanasii (Q65A\_11) a věkem respondenta byla, jak vidíme z výpočtu,  $-0,21$ . Po zavedení kontrolní proměnné *náboženská orientace* (Q28REC) se původní korelace snížila na  $-0,19$ . Oproti původní hodnotě korelace je to tak malá změna, že můžeme s jistotou tvrdit, že náboženská orientace do vztahu mezi věkem a postojem k euthanasii neintervenuje.

Stejný výsledek získáme i pro postoj k sebevraždě: původní korelace  $r = -0,21$  se snížila pouze na  $r = -0,19$ .

## PARTIAL CORRELATION COEFFICIENT

Zero Order Partial

	Q65B_1	VEK	Q28REC
Q65B_1	1,0000	-,2055**	,0961**
VEK	-,2055**	1,0000	-,2011**
Q28REC	,0961**	-,2011**	1,0000

\* - Signif. LE ,05      \*\* - Signif. LE ,01      (1-tailed)

## PARTIAL CORRELATION COEFFICIENT

Controlling for..      Q28REC

	Q65B_1	VEK
Q65B_1	1,0000	-,1909**
VEK	-,1909**	1,0000

\* - Signif. LE ,05      \*\* - Signif. LE ,01      (1-tailed)