

LEKCE 9

JAK ODHALIT VLIV TŘETÍ PROMĚNNÉ (ELABORACE)

Výzkumné zjištění, že mezi nezávisle proměnnou X a závisle proměnnou Y existuje vztah, je samozřejmě důležitým výsledkem. Nicméně samo o sobě by nás ve většině případů ještě nemělo uspokojit a nemělo by nás vést k ukončení analýzy. Musíme si totiž položit, jak nás nabádá de Vaus (2002), ještě několik otázek:

1. Jaká je povaha tohoto vztahu, je kauzální nebo ne?
2. Pokud je tento vztah kauzální, je přímý, nebo nepřímý (to je když X ovlivňuje Y prostřednictvím třetí proměnné)?
3. Pokud je tento vztah nepřímý, jakým mechanismem proměnná X ovlivňuje proměnnou Y?
4. Pokud je vztah mezi X a Y nekauzální povahy, jakou funkcí se dá modelovat?

Řečeno jinými slovy, 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 bivariační souvislosti (nebo zjištěním její absence tam, kde bychom ji očekávali) ještě nemáme vyhráno a měli bychom být obezřetní, dokud ji neproověříme testem na možný vliv nějaké třetí proměnné. Se vztahem mezi dvěma proměnnými je to totiž jako se vztahem dvou partnerů v životě – může být významně ovlivněn působením třetí osoby (a většinou si nemůžeme být nikdy zcela jisti).

Způsob jak odhalit vliv třetí proměnné rozpracovali už v roce 1950 Patricia Kendall s Paulem Lazarsfeldem. Technika byla nazvána *elaboration*, což lze překládat jak rozpracování, precizace nebo elaborace. Definice tohoto způsobu analýzy by mohla znít: Elaborační analýza spočívá v zavedení třetí proměnné do vztahu mezi dvěma proměnnými a ve zhodnocení jejího působení. To umožňuje hlubší porozumění původnímu bivariačnímu vztahu.

Než si jednotlivé postupy předvedeme (nechali jsme se v těchto pasážích výrazně inspirovat vynikající učebnicí Davida de Vause: *Analyzing Social Science Data*, Sage, 2002), je třeba zavést příslušnou terminologii. Vztahu mezi dvěma proměnnými X a Y se v elaboraci říká vztah nultého řádu (*zero-order*). Po zavedení třetí proměnné, které se říká testová proměnná (*test variable*) a značí se písmenem Z, vznikne vztah první řádu (*first-order*).

Vztahy mezi dvěma proměnnými a proměnnou testovou mohou nabývat různých podob – viz **Modely vztahů mezi třemi proměnnými** v souboru *elaborace.ppt* (na dokumentovém serveru).

ZPŮSOB ELABORACE

Elaborace se obvykle provádí prostřednictvím dvou postupů:

- I.
 - a) Zavedením třetí, testové proměnné do vztahu nultého řádu (v proceduře *Crosstabs*) – jinými slovy vytvořením podmíněných tabulek
 - b) výpočtem podmíněných korelací;
- II.
 - a) sestavením parciálních tabulek nebo
 - b) výpočtem parciálních korelací.

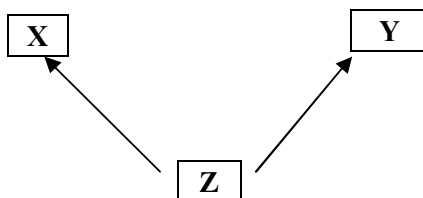
AD IA: VYTVÁŘENÍM PODMÍNĚNÝCH TABULEK

Při vytváření podmíněných tabulek pracujeme s procedurou *Crosstabs*. Můžeme získat následující variantu výsledků.

1. Podmíněné vztahy jsou všechny mnohem slabší, než původní a nejsou statisticky signifikantní.

Interpretace: Vztah mezi X a Y je buď (a) falešný, nebo Z je (b) intervenující proměnnou (nepřímý vztah). Modelově má následující podobu:

Model (a):



Model (b):



Příklad: Nalezli jsme jasný vztah mezi pohlavím a příjmem. Tento vztah (nultého řádu) vypadá následovně: viz tab. 9. 1.

Tab. 9.1. Vztah mezi pohlavím a příjmem

Příjem	Pohlaví	
	Muži	Ženy
nízký	25 %	52 %
střední	29 %	34 %
vysoký	36 %	14 %
Celkem	600 (100 %)	400 (100 %)

Vidíme, že s nízkým platem bylo pouze 25 % mužů, ale 52 % žen ($\epsilon = 27\%$).¹ Naopak s vysokým příjmem bylo 36 % mužů a 14 % žen. Tento vztah mezi pohlavím a příjmem je potvrzen i korelačním koeficientem, gama je $-0,49$ (čím více žena, tím nižší příjem). Zavedeme testovou proměnnou, že zde hraje nějakou roli počet odpracovaných hodin. Po zavedení počtu odpracovaných hodin (viz tab. 9.2.)

Tab. 9.2. Vztah mezi počtem odpracovaných hodin, pohlavím a příjmem

Počet odpracovaných hodin týdně	Příjem	Pohlaví	
		Muži	Ženy
Méně než 15 hodin	nízký	69	72
	střední	24	22
	vysoký	7	6
		100 %	100 %
15–35 hodin	nízký	26	29
	střední	56	52
	vysoký	18	19

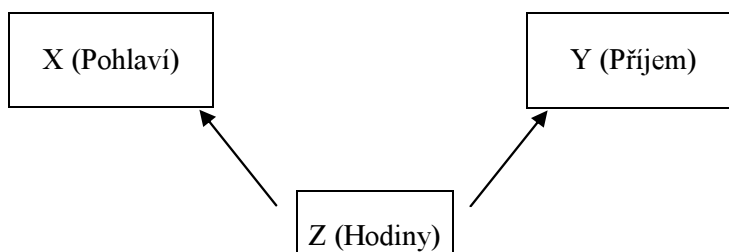
¹ Z lekce 8. již víme, že rozdíl mezi procenty v políčkách tabulky se nazývá epsilon (ϵ).

		100 %	100 %
35 a více hodin	nízký	17	21
	střední	29	28
	vysoký	54	51
		100%	100%

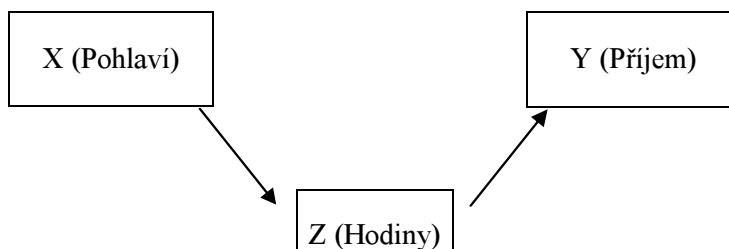
Původní poměrně vysoké rozdíly zmizely (např. ϵ pro nízký příjem ve skupině méně než 15 hodin je 3 %), což značí, že na vztah mezi pohlavím a příjmem má nějakým způsobem vliv počet odpracovaných hodin. Jak ale poznáme, který zdali máme pro vysvětlení použít modelu (a), nebo modelu (b)?
Postup:

1. Namalujte si diagramy obou modelů a místo symbolů použijte reálná jména proměnných.
2. Zkontrolujte, zda směr působení proměnných odpovídá logice nebo časovému sledu působení proměnných.

Zdánlivý vztah:



Z jako intervenující proměnná:



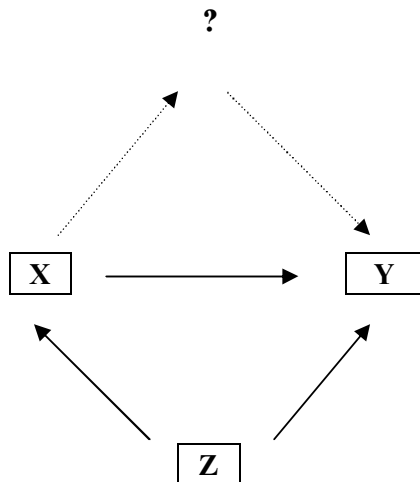
Jelikož musí platit logická a časová sekvence ve vztahu mezi proměnnými, je v našem případě zřejmé, že nemůže platit model zdánlivého vztahu, neboť odpracované hodiny (Z) nemohou mít vliv na pohlaví respondenta. Proto přijímáme model intervenující proměnné.

Ne vždy jsou výsledky tak jednoznačné. Jaké mohou ještě nastat výsledky?

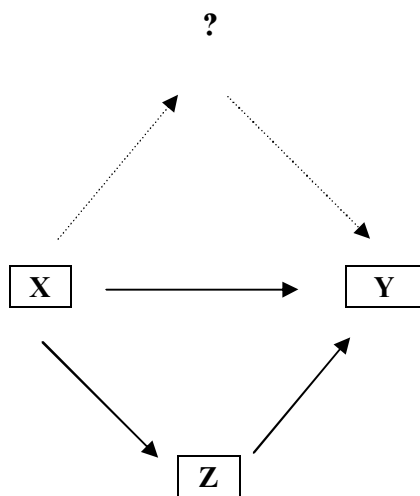
2: Podmíněné vztahy jsou slabší, avšak jsou statisticky významné.

Interpretace: Vztah mezi X a Y je částečně zdánlivý a částečně nepřímý. Navíc je i přímý, což může být způsobeno existencí vztahu nebo také působením další proměnné

Model: Zdánlivý a přímý vztah



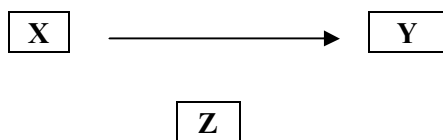
Model: Intervenující proměnná a přímý vztah



5. Původní a podmíněné vztahy se vůbec nezmění.

Interpretace: Původní vztah je ryzí, není způsoben působením testové proměnné.

Model:



Příklad z výzkumných dat:

Na datech EVS ČR 1999 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 výrokiem souhlasí nebo ne (předložena byla pětistupňová Lickertova stupnice, ale pro potřeby tohoto příkladu jsme ji trichotomizovali).

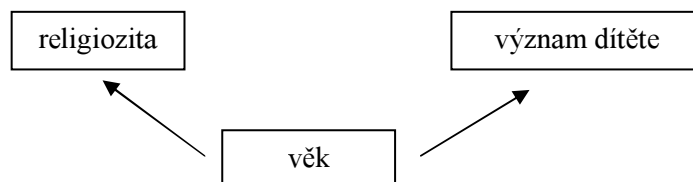
Vztah nultého řádu měl tuto podobu (procedura *Crosstabs*):

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%	

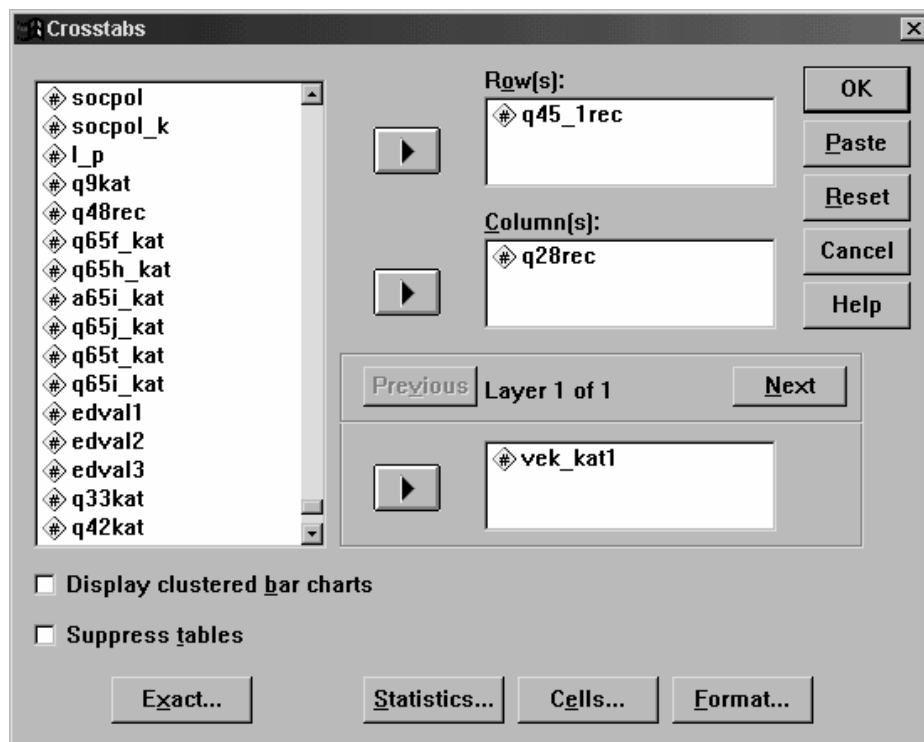
Vidíme, že jak věřící, tak nevěřící souhlasí s uvedeným výrokiem v podstatě ve stejné míře (62 % věřících a 59 % nevěřících, $\varepsilon = 3$ %).

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



Zavedme proto třetí proměnnou věk. 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 (věkové kategorie) 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 malé a na hranici statistické korektnosti. V dialogovém okně pro *Crosstabs* přidáme třetí proměnnou (viz obrázek příslušného dialogového okna níže). Statistiky pro výpočet korelačního koeficient už máme nastaveny z předchozího běhu programu pro vztah nultého řádu.



Výsledek třídění ukazuje, že minimální rozdíl v postoji k významu dítěte se mezi věřícími a nevěřícími zůstal zachován i v podmíněných tabulkách u všech tří věkových skupin.

Q45_1REC dite naplnenim zivota cloveka * 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 dite naplnenim zivota cloveka	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 dite naplnenim zivota cloveka	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 dite naplnenim zivota cloveka	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%	

4. Vztah mezi X a Y je různý v kategoriích proměnné Z

Interpretace: Vztah je specifický pro některé podskupiny. Interakční efekt

Model:

$$\begin{array}{l} Z_1 \quad X \quad \longrightarrow \quad Y \\ Z_2 \quad X \quad \quad \quad \quad Y \end{array}$$

Příklad z výzkumných dat:

Ve výzkumu Mládež a AIDS² jsme našli tento vztah mezi mírou religiozity mladých respondentů (ve věku 16-18 let):

Tab. 9. 3: Vztah mezi mírou religiozity a zkušeností se sexuálním stykem (řádková %)

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

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

Tab. 9.3a: Vztah mezi mírou religiozity a zkušeností se sexuálním stykem (řádková %)

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

Rozdíl mezi věřícími a nevěřícími, kteří ještě neměli zkušenost se sexuálním stykem byl v celém souboru $\epsilon = 12\%$ ($65 - 53$). V podsouboru mládeže žijící v Brně byl ovšem tento rozdíl jen $\epsilon = 3\%$ ($64 - 61$), v okolních městech $\epsilon = 12\%$ ($62 - 50$) a na vesnicích $\epsilon = 24\%$ ($69 - 45$). 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.

Ne vždy je ale výhodné pracovat s tabulkami třídění pro tři proměnné. Výhodnější může někdy být použit místo tabulek k analýze podmíněných korelačních koeficientů.

AD) IB: UŽITÍ PODMÍNĚNÝCH KORELAČNÍCH KOEFICIENTŮ

Podmíněné korelační koeficienty jsou korelace vypočítané z podmíněných třídění. Při odhalování, zdali se změnila původní korelace postupujeme tak, že nejdříve vypočítáme příslušný korelační koeficient nul-

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

tého řádu a poté spočítáme stejný korelační koeficient pro jednotlivé kategorie testové proměnné. Tyto korelace srovnáme.

Někdy může nastat např. takováto situace: Viděli jsme, že korelace nultého řádu mezi pohlavím a příjmem byla $-0,49$. Pokud spočítáme podmíněné korelace pro jednotlivé kategorie počtu týdně odpracovaných hodin, zjistíme, že (viz tab. 9. 4)

Tab. 9. 4. Korelace nultého řádu pro vztah mezi pohlavím a příjmem a podmíněné korelace po zavedení testové proměnné počet odpracovaných hodin týdně

korelace nultého řádu	gamma = -0,49, sig. < 0.001
Méně než 15 hodin	gamma = -0,07, sig. = ns.
15–35 hodin	gamma = -0,03, sig. = ns.
35 a více hodin	gamma = -0,07, sig. = ns.

Podmíněné korelace jsou výrazně nižší než korelace původní, všechny jsou navíc nesignifikantní (ns). Takový výsledek říká, že korelace nultého řádu je korelací nepravou, že je způsobena působením proměnné počet odpracovaných hodin. Z předchozích analýz již víme, že pro tuto situaci platí model zdánlivé (falešné) korelace.

Zavedením testové proměnné soubor svým způsobem standardizujeme. Např. v prvním řádku tab. 9.4 jsou pouze respondenti, jejich týdenní pracovní doba byla stejně dlouhá, maximálně 15 hodin – jak u mužů, tak u žen. Pokud by počet odpracovaných hodin do původního vztahu neintervenoval, musela by být podmíněná korelace přibližně stejně vysoká jako korelace původní. Totéž pak platí i pro ostatní kategorie testové proměnné.

Podívejme se nyní na výsledky analýzy vztahu mezi postoje k dítěti jako naplnění života a náboženskou orientací. Původní korelace byla nízká a statisticky nevýznamná, Somersovo $d = 0,02$ (viz tabulku níže).

Directional Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Somers Symmetric	,017	,023	,756	,450
Ordinal 'd Q45_1REC dítě naplnením života cloveka Dependent	,018	,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.

Po zavedení testové proměnné věk jsme spočítali podmíněné korelace mezi postojem k dítěti a náboženskou orientací v jednotlivých věkových kategoriích. Všechny byly nízké (viz tabulku níže) a statisticky nevýznamné, takže lze konstatovat, že na původní vztah věk respondenta nepůsobí.

Directional Measures

VEK_KAT1 Vekové kategorie				Value	Asymp. Std. Error ^a	Approx. T ^b	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.

Někdy se ovšem může stát, že podmíněné korelace vyjdou poněkud nižší, avšak ne o tak mnoho nižší, abychom mohli učinit jednoznačný závěr.

Příklad:

Zajímá nás, zdali existuje vztah mezi proměnnou politická orientace respondenta (levice, střed pravice), a postojem k výroku, že chudí jsou nás chudými proto, že jsou líní (souhlas, nesouhlas). Korelace nultého řádu měřená pořadovým koeficientem Tau_c vychází 0,18 (sig <0.001). To značí, že lidé pravicově orientovaní souhlasí s leností, jako příčinou chudoby, zatímco lidé levicově orientovaní s touto příčinou nesouhlasí. Můžeme se ale domnívat, že na tento vztah působí proměnná důvěra v lidi (lidem je možné důvěřovat; lidem není možné důvěřovat). Po zavedení testové proměnné důvěra v lidi jsme našli následující výsledky (viz tab 9.5)

Tab. 9.5. Podmíněné korelace mezi politickou orientací a leností jako příčinou chudoby a důvěrou v lidi

korelace nultého řádu	$Tau_c = 0,18$, sig. <0.001
(1) lidem je možné důvěřovat	$Tau_c = 0,13$, sig. <0.005
(2) lidem není možné důvěřovat	$Tau_c = 0,19$, sig. <0.001

Nastala situace, kdy podmíněná korelace se v jedné kategorii testové proměnné poněkud snížila než korelace původní, v druhé kategorii se lehce zvýšila. U těch respondentů, kteří lidem nedůvěřují, je vliv levicové orientace na nesouhlas s výrokem, že chudí jsou u nás chudými, protože jsou líní, nejsilnější.

Otázkou u tohoto výsledku je, zdali je rozdíl mezi těmito dvěma podmíněnými koeficienty statisticky významný. Postup která nám to umožní říci, navrhuje Pallant (2001):

- Musíme kromě podmíněných korelačních koeficientů znát také velikost příslušných skupin: v první skupině bylo 422 osob, ve druhé 1303 osob – pozor, je zde podmínka, že N nesmí v žádné skupině klesnout pod 20!).
- Podle transformační tabulky (viz tab. 9.6) převedeme korelační koeficienty na z skóry. Ve sloupci pro korelační koeficienty nalezneme hodnotu, která nejlépe odpovídá vypočtené hodnotě, a k ní odpovídající z skór. V našem případě má jedna korelace hodnotu 0,13, nejbližší hodnota r je 0,15 a příslušný z skór také 0,15. Druhá korelace je 0,19, nejbližší korelace je 0,20 a z skór také 0,20.

Obr. 9.6: Tabulka pro převod korelačních koeficientů na z skóry

r	z _r	r	z _r	r	z _r	r	z _r	r	z _r
.000	.000	.200	.203	.400	.424	.600	.693	.800	1.099
.005	.005	.205	.208	.405	.430	.605	.701	.805	1.113
.010	.010	.210	.213	.410	.436	.610	.709	.810	1.127
.015	.015	.215	.218	.415	.442	.615	.717	.815	1.142
.020	.020	.220	.224	.420	.448	.620	.725	.820	1.157
.025	.025	.225	.229	.425	.454	.625	.733	.825	1.172
.030	.030	.230	.234	.430	.460	.630	.741	.830	1.188
.035	.035	.235	.239	.435	.466	.635	.750	.835	1.204
.040	.040	.240	.245	.440	.472	.640	.758	.840	1.221
.045	.045	.245	.250	.445	.478	.645	.767	.845	1.238
.050	.050	.250	.255	.450	.485	.650	.775	.850	1.256
.055	.055	.255	.261	.455	.491	.655	.784	.855	1.274
.060	.060	.260	.266	.460	.497	.660	.793	.860	1.293
.065	.065	.265	.271	.465	.504	.665	.802	.865	1.313
.070	.070	.270	.277	.470	.510	.670	.811	.870	1.333
.075	.075	.275	.282	.475	.517	.675	.820	.875	1.354
.080	.080	.280	.288	.480	.523	.680	.829	.880	1.376
.085	.085	.285	.293	.485	.530	.685	.838	.885	1.398
.090	.090	.290	.299	.490	.536	.690	.848	.890	1.422
.095	.095	.295	.304	.495	.543	.695	.858	.895	1.447
.100	.100	.300	.310	.500	.549	.700	.867	.900	1.472
.105	.105	.305	.315	.505	.556	.705	.877	.905	1.499
.110	.110	.310	.321	.510	.563	.710	.887	.910	1.528
.115	.116	.315	.326	.515	.570	.715	.897	.915	1.557
.120	.121	.320	.332	.520	.576	.720	.908	.920	1.589
.125	.126	.325	.337	.525	.583	.725	.918	.925	1.623
.130	.131	.330	.343	.530	.590	.730	.929	.930	1.658
.135	.136	.335	.348	.535	.597	.735	.940	.935	1.697
.140	.141	.340	.354	.540	.604	.740	.950	.940	1.738
.145	.146	.345	.360	.545	.611	.745	.962	.945	1.783
.150	.151	.350	.365	.550	.618	.750	.973	.950	1.832
.155	.156	.355	.371	.555	.626	.755	.984	.955	1.886
.160	.161	.360	.377	.560	.633	.760	.996	.960	1.946
.165	.167	.365	.383	.565	.640	.765	1.008	.965	2.014
.170	.172	.370	.388	.570	.648	.770	1.020	.970	2.092
.175	.177	.375	.394	.575	.655	.775	1.033	.975	2.185
.180	.182	.380	.400	.580	.662	.780	1.045	.980	2.298
.185	.187	.385	.406	.585	.670	.785	1.058	.985	2.443
.190	.192	.390	.412	.590	.678	.790	1.071	.990	2.647
.195	.198	.395	.418	.595	.685	.795	1.085	.995	2.994

TABLE 11.1

Transformation of r to z

Source: McCall (1990); originally from Edwards, A. L. (1967). *Statistical methods* (2nd edition). Holt, Rinehart and Winston.

c) Dosadíme hodnoty z skóru a N do vzorce:

$$z_{\text{obs}} = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} = \frac{0,13 - 0,20}{\sqrt{\frac{1}{422 - 3} + \frac{1}{1033 - 3}}} = \frac{-0,07}{\sqrt{0,00234 + 0,00097}} = -0,12$$

Pokud vypočtená hodnota z_{obs} leží v intervalu mezi $-1,96$ a $+1,96$, nemůžeme zamítnout nulovou hypotézu, že mezi dvěma korelačními koeficienty není žádný rozdíl. Pouze pokud vypočtená hodnota leží mimo tento interval, pak se korelační koeficienty od sebe statisticky významně odlišují. Náš výsledek $-0,12$ je uvnitř intervalu „nevýznamnosti“, proto nulovou hypotézu nemůžeme zamítnout a musíme konstatovat, že rozdíl mezi korelačními koeficienty ve skupině, kteří si myslí, že lidem je možné důvěřovat, a ve skupině, kteří si to nemyslí, je statisticky nevýznamný.

Abychom vám nahradili skutečnost, že SPSS tuto proceduru neumí, přikládáme excelovskou tabulku, jejímž prostřednictvím příslušný výpočet velmi rychle provedete. Stačí dvakrát kliknout pravým tlačítkem

myši na text v tabulce a naskočí vám excelovské prostředí. Pak už jen vepište do příslušných sloupečků hodnoty z_1 a z_2 a N_1 a N_2 a dostanete rovnou výsledek.. Vyzkoušejte si to.

Výpočet statistické signifikance rozdílu mezi dvěma korelačními koeficienty

(viz J. Pallant, 2001:128)

Vzorec:

$$Z_{obs} = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

Ukázka výpočtu:

Kor. koef. 1 = -0,220 $z_1 = -0,224$ $N_1 = 184$

Kor. koef. 2 = -0,394 $z_2 = -0,418$ $N_2 = 250$

z_1	z_2	N_1	N_2
-0,224	-0,418	184	250

čítatel ($z_1 - z_2$)	$(1/N_1 - 3) + 1/N_2 - 3$	jmenovatel	výsledek
0,194	0,0096	0,0978	1,98

Pokud je výsledek v intervalu $\langle -1,96; +1,96 \rangle$, je rozdíl v koeficientech **statisticky nevýznamný**.

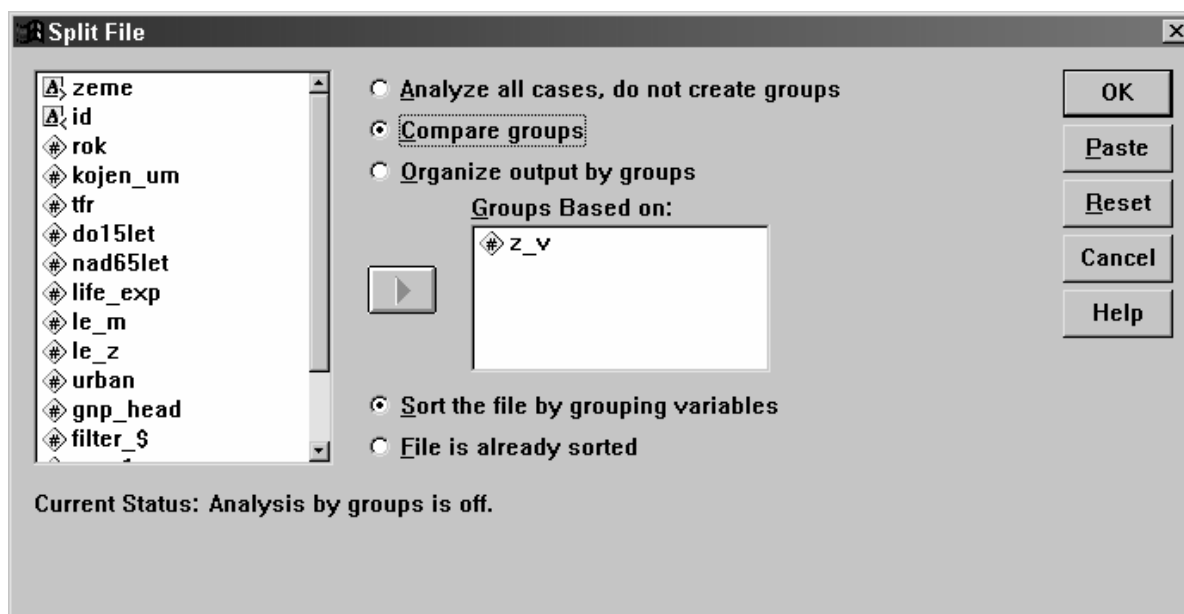
Náš výsledek leží mimo tento interval, rozdíl je tedy statisticky signifikantní

Podmíněné korelace je možné také vypočítat tak, že použijte proceduru *Data – Split file*. Výhodný je tento postup především když pracujete s intervalovými proměnnými. Ukažme si vše na příkladu.

V souboru demografických dat (viz soubor *dmg-data.sav*) spočítáme korelaci mezi ekonomickou vyspělostí evropských zemí (indikovanou údajem o hrubém domácím produktu v dolarech na hlavu) a naději dožití žen. Máme totiž tezi, že čím více je země ekonomicky vyspělá, tím vyšší je pravděpodobnost, že se lidé dožijí vyššího věku. A skutečně, Pearsonovo r má hodnotu 0,83, což je korelace vysoká. Položme si ale otázku, zdali se tato korelace udrží i v případě, že ji budeme počítat zvlášť pro západoevropské a zvlášť pro východoevropské země. Soubor proto rozdělíme na dva podsoubory s pomocí proměnné z_v (tato proměnná má kód 1 pro země západní a kód 2 pro země východní), tak že v dialogovém okně pro *Split file* (viz obrázek níže) zakliknete příkaz *Compare groups* a do okénka vlepíte proměnnou, pro jejíž skupiny chcete počítat korelační koeficienty³ – v našem případě proměnnou z_v . Po kliknutí na OK se nastaví rozdělení souboru, takže všechny výpočetní operace nyní budou probíhat odděleně pro podsoubor zemí západních a východních. Nezapomeňte po skončení analýzy vrátit nastavení ve *Split file* pro práci s celým souborem (kliknutím do kroužku *Analyze all cases, do not create groups*).

³ Práci s podsoubory, které získáme procedurou *Split file*, můžeme samozřejmě použít i v jiných případech, než je výpočet korelačních koeficientů, např. při práci s průměry nebo při tříděních.

Nový výpočet korelačního koeficientu přinesl tento výsledek:



Correlations

			GNP_HEAD GNP na hlavu v US \$ (1998)	LE_Z
Z_V				
1	zapadni	GNP_HEAD GNP na hlavu v US \$ (1998)	Pearson Correlation Sig. (2-tailed) N	
		LE_Z	Pearson Correlation Sig. (2-tailed) N	,427 ,077 18
2	vychodni	GNP_HEAD GNP na hlavu v US \$ (1998)	Pearson Correlation Sig. (2-tailed) N	
		LE_Z	Pearson Correlation Sig. (2-tailed) N	,764** ,001 15

** . Correlation is significant at the 0.01 level (2-tailed).

Vidíme, že původní korelace (0,83) se proměnila a je odlišná na západě a na východě. Zatímco v západních zemích je 0,43, v zemích východních je 0,76. Což znamená, že ve vyspělých západních zemích, které mají vysoký hrubý domácí produkt na hlavu (v průměru, jak lze z dat v souboru spočítat, mají 23 300 US dolarů na hlavu), je souvislost mezi životní úrovní a úmrtností žen mnohem nižší, než v zemích východních, které jsou charakteristické v průměru mnohem nižší životní úrovní (2 600 US dolarů na hlavu).

Zajímavý výsledek, ne?

Ad IIa) Zatím vynecháno (viz de Vaus 2002:333–336)

AD IIB) VYUŽITÍ PARCIÁLNÍCH KOEFICIENTŮ

Ne vždy je možné dělat třídění třetího stupně, např. tehdy, když proměnné mají větší počet kategorií. Proto je dobré vědět, že stejný účel jako tabulkové třídění dat splní i jednoduchý výpočet parciálního korelačního koeficientu.

Při použití programu SPSS narazíme ovšem na jeden problém, že SPSS umí vypočítat parciální koeficient pouze pro intervalové znaky (počítá totiž Pearsonovo r). Existují ovšem i způsoby, jak z něj získat parciální korelace pro znaky ordinální. Začneme pro jednoduchost parciální korelací pro znaky intervalové.

Postup:

1. S pomocí SPSS vypočítáme příslušný Pearsonův lineární korelační koeficient korelační koeficienty pro vztah dvou proměnných, jejichž analýza nás zajímá.
2. Poté vypočítáme parciální koeficient a srovnáme tuto hodnotu s hodnotou korelace nultého řádu.

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 změnil 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. Zda-li model zdánlivé korelace či intervenující proměnné.

Příklad: Ve výzkumu mládeže byl Pearsonův koeficient korelace mezi postojem k časnosti zahájení sexuálního života (souhlas či nesouhlas s výrokem *mladí lidé v mém věku by ještě neměli mít sexuální styk*) a pohlavím 0,60 (je možné Pearsonovco r počítat, neboť obě proměnné mají dichotomickou povahu). Jelikož očekáváme, že tato korelace by mohla být zdánlivá, zavádíme do tohoto vztahu kontrolní proměnnou zkušenost se sexuálním stykem (zdali již respondent měl či neměl pohlavní styk). Příslušné vzájemné interkorelace korelace byly:

Tab. 9.7: Interkorelace mezi pohlavím, postojem k věku, kdy by měli mít mladí sexuální zkušenost a existencí zkušenosti s pohlavním stykem

	postoj (a)	pohlaví (b)
postoj (a)	1,00	
pohlaví (b)	0,60	1,00
zkušenost (c)	0,70	0,80

Vypočtená parciální korelace $r_{ab.c} = 0,09^4$

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, zda-li jde o model zdánlivé korelace anebo model intervenující proměnné.

⁴ Výraz $r_{ab.c}$ čteme: korelační koeficient mezi proměnnými a a b při vyloučení vlivu c . Je-li za tečkou pouze jedna proměnná, počítáme parciální koeficient prvního řádu, budou-li tam dvě tečky, jedná se o parciální koeficient druhého řádu atd.

Výsledek výpočtu:

Correlations

		Q65A_11 Euthanasie	VEK
Q65A_11 Euthanasie	Pearson Correlation		
	Sig. (2-tailed)		
	N		
VEK	Pearson Correlation	-,217*	
	Sig. (2-tailed)	,000	
	N	1799	

** . Correlation is significant at the 0.01 level

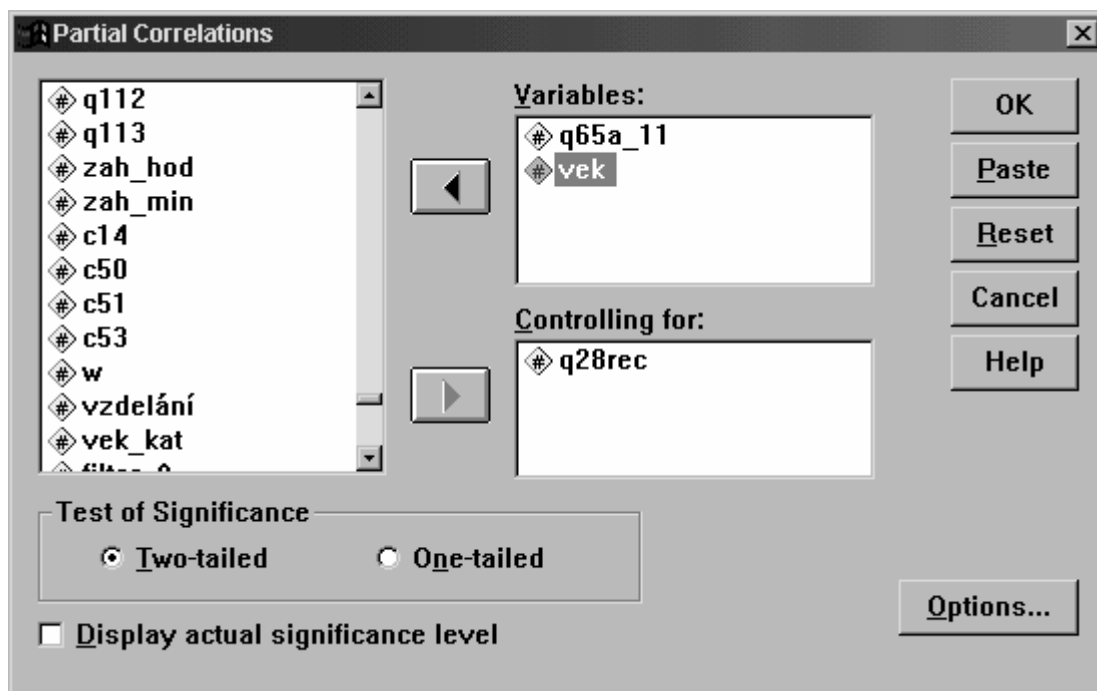
Vidíme, že postoj k euthanasii je věkem ovlivněn. Se zvyšujícím se věkem klesá ospravedlnitelnost euthanasie ($r = -0,22$). Starší lidé jsou zkrátka více proti euthanasii než lidé mladší.

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

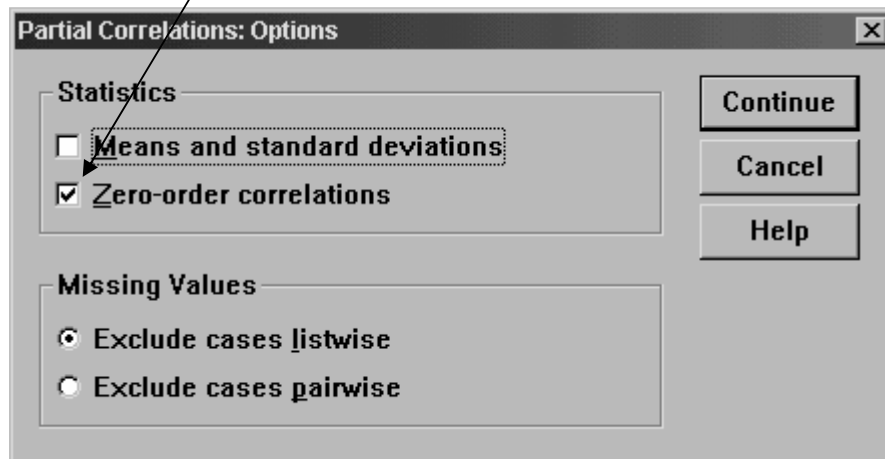
Zjistíme to prostřednictvím výpočtu parciální korelace, kdy do bivariačního vztahu zavedeme proměnnou religioziota ($q28rec$) – je to dichotomická proměnná, takže můžeme směle pracovat s Pearsonovými koeficienty.

Analyze — Correlate — Partial



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

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



Výsledek výpočtu:

Tab. 9.8: Výsledky výpočtu parciálního koeficientu

ZERO ORDER PARTIALS

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

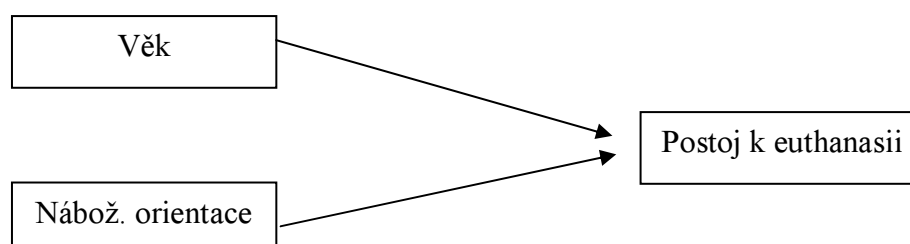
* - Signif. LE ,05 ** - Signif. LE ,01 (1-tailed)

PARTIAL CORRELATION COEFFICIENTS

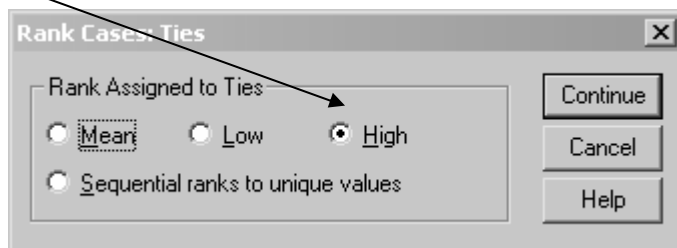
Controlling for..		Q28REC
	Q65A_11	VEK
Q65A_11	1,0000	-,1862**
VEK	-,1862**	1,0000

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

Původní korelace (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 vztah mezi postojem k euthanasii a věkem respondenta existuje. Náš model intervenující proměnné tedy můžeme směle opustit. Namísto něj, zdá se, je možné přijmout model dvojí příčiny.



Až dosud jsme analyzovali vztah mezi proměnnými, které měly intervalový charakter. Co ale dělat v situaci, kdy máme proměnné jasně ordinální? Pro ně nelze Pearsonovo r použít. Nabízíme malý trik. Ordinální proměnné, které chcete analyzovat prostřednictvím parciálních koeficientů, rankujte (to je nechte jednotlivým hodnotám připsat pořadí, jaké zajímají v souboru). V proceduře *Rank* volte pořadí u stejných hodnot tak, aby se jim připsala hodnota toho nejvyššího pořadí (v tlačítku *Ties* zaklikněte volbu *High*).



Při parciálních korelacích pak pracujte s rankovanými proměnnými. Vzorec pro výpočet Pearsonova koeficientu dává pro rankované hodnoty takové výsledky, které jsou velmi blízké hodnotám Spearmanova koeficientu pořadového koeficientu korelace.

* * *

Jak se snažila ukázat tato lekce, při hledání statistických vztahů bychom se neměli spokojit pouze s bivariačním výsledkem. Vždy, když to má smysl, se pokoušejte zavádět třetí proměnné a zjišťujte, zdali původní vztah „vydržel“, nebo byl modifikován. Prohloubíte tím své poznání a přinesete výsledky, které budou postaveny na solidním základě.