

## ÚKOL 8

### CVIČENÍ Z ANALÝZY DAT SOCIOLOGICKÉHO VÝZKUMU

Jméno a příjmení:

Imatrik. ročník:

1.

V datech EVS ČR 1999 zjistěte, zdali existuje vztah mezi preferencí svobody, či rovnosti (proměnná q52) a levopravou orientací (trichotomizovaná proměnná l-p). Příklad řešte prostřednictvím CROSSTABS. Před výpočtem rekodejte proměnnou q52 tak, aby nová stupnice zněla: 1.

preferance svobody, 2. ani svoboda, ani rovnost, 3. preference rovnosti.

Po zjištění vztahu otestujte, co se s ním stane, když zavedete třetí proměnnou, a to věk (trichotomizovanou proměnnou VEK-KAT3).

Řešení:

Transform – Recode – Into different variable...

Analyze – Crosstabs... Statistics: Correlations (jde použít i jiné korelační koeficienty pro ordinální proměnné)

L\_P lev. x prav. kateg. (q53) \* Q52REC svoboda, nebo rovnost Crosstabulation

			Q52REC svoboda, nebo rovnost			Total
			1 svoboda	2 ani, ani	3 rovnost	
L_P lev. x prav. kateg. (q53)	1 levice	Count	89	22	128	239
		Row %	37,2%	9,2%	53,6%	100,0%
	2 střed	Count	567	69	349	985
		Row %	57,6%	7,0%	35,4%	100,0%
	3 prav	Count	375	16	109	500
		Row %	75,0%	3,2%	21,8%	100,0%
Total	Count	1031	107	586	1724	
	Row %	59,8%	6,2%	34,0%	100,0%	

Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig. <sup>c</sup>
Interval by Interval	Pearson's R	-,232	,023	-9,889	,000 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	-,234	,023	-9,976	,000 <sup>c</sup>
N of Valid Cases		1724			

a. Not assuming the null hypothesis.

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

c. Based on normal approximation.

Korelace -0,23, pravice má tendenci preferovat svobodu. Po zavedení věku se tento obraz trochu změní (viz tabulky níže):

Analyze – Descriptive statistics - Crosstabs ... Layer 1= vek\_kat3

Korelace byla ve srovnání s bivariačním koeficientem nízká v mladé věkové skupině (což je dáno tím, že v této skupině nejsou rozdíly v názorech na svobodu), byla přibližně stejně vysoká ve střední věkové skupině a zvýšila se ve skupině nejstarší. Tam fungovalo vše, jak má.

L\_P lev. x prav. kateg. (q53) \* Q52REC svoboda, nebo rovnost \* VEK\_KAT1 Vekové kategorie Crosstabulation

VEK_KAT1 Vekové kategorie				Q52REC svoboda, nebo rovnost			Total
				1 svoboda	2 ani, ani	3 rovnost	
1 18-29	L_P lev. x prav. kateg. (q53)	1 levice	Count	14	4	9	27
			Row %	51,9%	14,8%	33,3%	100,0%
	2 střed	Count	166	15	59	240	
		Row %	69,2%	6,3%	24,6%	100,0%	
	3 prav	Count	85	3	31	119	
		Row %	71,4%	2,5%	26,1%	100,0%	
Total			Count	265	22	99	386
			Row %	68,7%	5,7%	25,6%	100,0%
2 30-49	L_P lev. x prav. kateg. (q53)	1 levice	Count	27	4	26	57
			Row %	47,4%	7,0%	45,6%	100,0%
	2 střed	Count	224	31	124	379	
		Row %	59,1%	8,2%	32,7%	100,0%	
	3 prav	Count	145	6	32	183	
		Row %	79,2%	3,3%	17,5%	100,0%	
Total			Count	396	41	182	619
			Row %	64,0%	6,6%	29,4%	100,0%
3 50+	L_P lev. x prav. kateg. (q53)	1 levice	Count	47	13	93	153
			Row %	30,7%	8,5%	60,8%	100,0%
	2 střed	Count	176	23	166	365	
		Row %	48,2%	6,3%	45,5%	100,0%	
	3 prav	Count	145	7	44	196	
		Row %	74,0%	3,6%	22,4%	100,0%	
Total			Count	368	43	303	714
			Row %	51,5%	6,0%	42,4%	100,0%

#### Symmetric Measures

VEK_KAT1 Vekové			Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
1 18-29	Interval by Interval	Pearson's R	-,050	,052	-,972	,332 <sup>c</sup>
	Ordinal by Ordinal	Spearman Correlation	-,055	,052	-1,083	,279 <sup>c</sup>
	N of Valid Cases			386		
2 30-49	Interval by Interval	Pearson's R	-,208	,038	-5,288	,000 <sup>c</sup>
	Ordinal by Ordinal	Spearman Correlation	-,214	,038	-5,431	,000 <sup>c</sup>
	N of Valid Cases			619		
3 50+	Interval by Interval	Pearson's R	-,298	,034	-8,333	,000 <sup>c</sup>
	Ordinal by Ordinal	Spearman Correlation	-,300	,034	-8,406	,000 <sup>c</sup>
	N of Valid Cases			714		

a. Not assuming the null hypothesis.

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

c. Based on normal approximation.

**Zkontrolujte, zdali stejné výsledky přinese také kontrola přes parciální koeficient. U proměnných věk a levice-pravice použijte nekategorizované proměnné – vek a q53. Pokud vyšly výsledky jinak, pokuste se zdůvodnit, proč tomu tak je.**

Analyze – Correlate – Partial – Variables: q52rec, q53. Controlling for: vek

Options: Zero-order correlations

PARTIAL CORRELATION COEFFICIENTS - - -

**Zero Order Partial**

	Q53	Q52REC	VEK
Q53	1,0000		
Q52REC	-,2576**	1,0000	
VEK	-,1359**	,1535**	1,0000

PARTIAL CORRELATION COEFFICIENTS - - -

**Controlling for.. VEK**

	Q53	Q52REC
Q53	1,0000	
Q52REC	-,2418**	1,0000

Výsledky zde vyšly tak, že věk nemá na původní vztah žádný vliv. Tento odlišný výsledek je dán tím, že:

1. Proměnná q52rec není vhodná do výpočtu parciální korelace, protože není intervalová
2. nepracujeme zde s kategorizovanými znaky, ale s původními škálami.

Napadá Vás, kolegové, ještě další příčina?

**3.**

**Zjistěte, jaká je korelace mezi názorem na minulý politický systém (proměnná q61) a na hodnocení současného systému (q60). Je tento vztah ovlivněn levo-pravou politickou orientací respondenta (q53)? Pokud ano, jaký model pro vliv této proměnné platí: zdánlivá korelace, intervenující proměnná, nebo upravený vztah?**

Řešení:

Analyze-Correlate-Bivariate-Pearson

**Correlations**

		Systém řízení naší země funguje:	Pol. systém za komunistického režimu
Systém řízení naší země funguje:	Pearson Correlation	1	-,163**
	Sig. (2-tailed)		,000
	N	1879	1832
Pol. systém za komunistického režimu	Pearson Correlation	-,163**	1
	Sig. (2-tailed)	,000	
	N	1832	1846

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

## Analyze-Correlate-Partial...

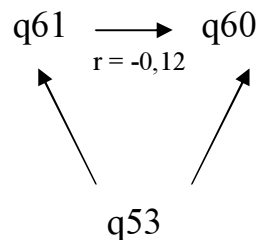
## Correlations

Control Variables			Systém řízení naší země funguje:	Pol. systém za komunistického režimu	Levice - pravice
-none-(a)	Systém řízení naší země funguje:	Correlation	1,000	-,158	,188
		Significance (2-tailed)	.	,000	,000
		df	0	1712	1712
	Pol. systém za komunistického režimu	Correlation	<b>-,158</b>	1,000	-,589
		Significance (2-tailed)	,000	.	,000
		df	1712	0	1712
	Levice - pravice	Correlation	,188	-,589	1,000
		Significance (2-tailed)	,000	,000	.
		df	1712	1712	0
Levice - pravice	Systém řízení naší země funguje:	Correlation	1,000	-,059	
		Significance (2-tailed)	.	,014	
		df	0	1711	
	Pol. systém za komunistického režimu	Correlation	<b>-,059</b>	1,000	
		Significance (2-tailed)	,014	.	
		df	1711	0	

a Cells contain zero-order (Pearson) correlations.

Jelikož po zavedení testové proměnné (q53) se původní korelace výrazně snížila z  $-0,16$  na  $-0,06$ , má levoprává orientace respondenta na původní vztah vliv.

O jaký se jedná model? Mohla by to být intervenující proměnná (viz obr.)



ale vzhledem k tomu, že korelace nezmezila úplně, půjde asi o upravený vztah. Zkontrolujeme tak, že si necháme udělat CROSSTABS třetího stupně q60 a q61, přičemž jako testovou proměnnou zavedeme kategorizovanou q53 (1-p). Nenecháme si udělat tabulku třídění, pouze korelace:

## Symmetric Measures

L_P lev. x prav. kateg. (q53)		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
1 levice	Interval by Interval Pearson's R	-,313	,064	-5,121	,000 <sup>c</sup>
	Ordinal by Ordinal Spearman Correlation	-,334	,062	-5,519	,000 <sup>c</sup>
	N of Valid Cases	244			
2 střed	Interval by Interval Pearson's R	-,046	,033	-1,459	,145 <sup>c</sup>
	Ordinal by Ordinal Spearman Correlation	-,032	,033	-1,003	,316 <sup>c</sup>
	N of Valid Cases	987			
3 prav	Interval by Interval Pearson's R	,036	,044	,801	,423 <sup>c</sup>
	Ordinal by Ordinal Spearman Correlation	,076	,045	1,681	,093 <sup>c</sup>
	N of Valid Cases	488			

a. Not assuming the null hypothesis.

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

c. Based on normal approximation.

Vidíme, že u levice se původní korelace dost zvýšila, zatímco u zbylých dvou kategorií je nízká, navíc s opačnými znaménky.

## 4.

V souboru evropských zemí `dmg-data.sav` spočítejte korelaci pro kojeneckou úmrtnost (**KOJEN\_UM**) a plodností (**TFR= Total Fertility Rate**, průměrný počet dětí na ženu).

Očekávali bychom, že tam, kde je nízká kojenecká úmrtnost, (nízká kojenecká úmrtnost je chápána jako ukazatel dobré životní úrovně země), tam bude i nízká plodnost, neboť vyspělé země až dosud byly charakteristické nízkou plodností. Platí zjištěná korelace i pro východní a západní země (proměnná `z_v`)? Pokud ne, vysvětlete, jak je možné, že korelace nultého řádu byla jiná, než korelace prvního řádu?

Jsou rozdíly v korelacích mezi východem a západem statisticky významné? (Pozor, toto je pouze cvičný výpočet, který by ve skutečné analýze neměl být ze dvou důvodů proveden. **PROČ?**)

## Řešení

Analyze-Correlate-Bivariate-Pearson

## Correlations

		KOJEN_UM	TFR
KOJEN_UM	Pearson Correlation	1	-,079
	Sig. (2-tailed)	,	,656
	N	34	34
TFR	Pearson Correlation	-,079	1
	Sig. (2-tailed)	,656	,
	N	34	34

Nízká korelace nultého řádu, která není statisticky významná.

Analyze-Descriptive statistics-Crosstabs. suppress tables. Layer1 = `z_v...`

## Statistics - Correlations

## Correlations

Z_V			KOJEN_UM	TFR
1	KOJEN_UM	Pearson Correlation	1	-,423
		Sig. (2-tailed)	,	,090
		N	17	17
	TFR	Pearson Correlation	-,423	1
		Sig. (2-tailed)	,090	,
		N	17	17
2	KOJEN_UM	Pearson Correlation	1	,484*
		Sig. (2-tailed)	,	,049
		N	17	17
	TFR	Pearson Correlation	,484*	1
		Sig. (2-tailed)	,049	,
		N	17	17

\*. Correlation is significant at the 0.05 level (2-tailed).

V západních zemích je korelace záporná (čím vyšší kojenecká úmrtnost, tím nižší plodnost), ve východních zemích kladná (čím vyšší kojenecká úmrtnost, tím vyšší plodnost).

Rozdíl mezi korelací nultého řádu a korelacemi prvního řádu spočívá ve velké heterogenitě souboru. Jestliže polovina zemí má středně silnou a kladnou korelaci a druhá polovina zemí středně silnou zápornou, výsledná korelace se vyruší, takže je blízka nule (znaménko tam pak není podstatné).

Rozdíl v korelacích:

### 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:

$r_1 = -0,423$                        $z_1 = -0,454$        $N_1 = 17$

$r_2 = -0,484$                        $z_2 = 0,530$        $N_2 = 17$

$z_1$	$z_2$	$N_1$	$N_2$
-0,454	0,53	17	17

čítatel ( $z_1 - z_2$ )	$(1/N_1 - 3) + 1/N_2 - 3$	jmenovatel	výsledek
-0,984	0,1429	0,3780	<b>-2,60</b>

Pokud je výsledek v intervalu  $<-1,96; +1,96>$ , je rozdíl v koeficientech **statisticky nevýznamný**.

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

Rozdíl vychází statisticky významný. Nelze brát výsledek vážně proto, že (1) nemáme výběrový soubor (data pocházejí z údajů za celou zemi) a (2) proto, že podmínkou je, že  $N$  nesmí být v ani jedné skupině menší než 20. A my máme 17.