UK FHS
Historická sociologie
(LS 2011)

Analýza kvantitativních dat II.

2. Vztahy mezi kategorizovanými znaky v kontingenční tabulce

Jiří Šafr

jiri.safr(zavináč)seznam.cz

poslední **aktualizace 23.4. 2011**

Asociace mezi znaky

Asociace (korelace) a kauzalita

- **Asociace (korelace)** neznamená automaticky **kauzální vztah**
- Podmínky kauzality (připomenutí):
- Naměřená korelace
- Časová souslednost (k A došlo před B)

Základní možnosti pro vztah dvou proměnných A x B (opakování)

- **Nominální A** (kategoriální či „kvalitativní“ proměnná) a **nominální B** → procentní podíly (podmíněné pravděpodobnosti) **kontingenční tabulka** (+ chí kvadrát test), znaménkové schéma, koeficient kontingence
- **Dtto ale ordinální** → dtto + pořadové korelace (Sperman, Tab-B)
- **Nominální A** x **kardinální** (číselná) → průměry B v podskupinách A (+ T-test či One-way Anova, 95% konf. intervaly),
  koeficient asociace Eta = míra jednostranné závislosti kvantitativní vysvětlované proměnné na proměnné nominální

Kategoriální data

(nominálními a ordinální znaky)

Asociace v kontingenční tabulce

Kontingenční tabulka

Statistické míry a testování

- **Nezávislost** = oba znaky navzájem neovlivňují v tom, jakých konkrétních hodnot nabývají
- **Homogenita** (shodnost struktury) = očekávané četnosti jsou v políčcích každého řádku ve stejném vzájemném poměru bez ohledu na konkrétní volbu řádku
- → **test dobré shody** = porovnání očekávaných četností v jednotlivých polích tabulky - za předpokladu, že hodnoty obou sledovaných znaků na sobě nezávisí - a skutečných četností.
- Pokud hypotéza nezávislosti (resp. homogenity) platí, má testová statistika přibližně rozdělení **chí kvadrát** o (r-1)(s-1) **stupních volnosti**. Hodnota testové statistiky se tedy porovná s kritickou hodnotou (kvantilem) příslušné hladiny významnosti.

Chí-kvadrát testy: test dobré shody

- Test pro homogenitu distribucí mezi kategoriemi znaku/ů
- test dobré shody = shody relativních četností *ni*/*n* a hypotetických pravděpodobností.
- Pro nominální znaky (i ordinální a kategorizované kardinální)
- Nevyžaduje znalost předchozího rozdělení znaku
- Očekávané frekvence: dle rozložení kategorií 1 znaku nebo v kontingenční tabulce vztah 2 znaků
- Odpovídá na otázku, zda jsou rozdíly mezi empirickými (**pozorovanými** - f_O) četnostmi a teoretickými (**očekávanými** -f_E) četnostmi náhodné nebo ne.

$$\chi^2_{obt} = \sum \frac{(f_O - f_E)^2}{f_E}$$

$$\chi^2 = \sum \frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}}$$

- Počet stupňů volnosti **df** = (r-1) (s-1)

r = počet řádků s = počet sloupců v tabulce

Chí-kvadrát test: příklad 1
Shoda s teoretickými četnostmi (shodné zastoupení kategorií statusu)

Pozorované absolutní četnosti kategorií

velmi nízký	střední	vysoký	
5	10	9	$\sum n_i = n = 24.$

Očekávané (teoretické) četnosti = 24 : 3 = 8.

H0: počet respondentů je ve všech kategoriích stejný

5	10	9	n_i
8	8	8	$\tilde{n}_i$

$$\chi^2 = \sum \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$$

$$\chi^2 = \frac{(5 - 8)^2}{8} + \frac{(10 - 8)^2}{8} + \frac{(9 - 8)^2}{8} = 1{,}74$$

Vypočítanou hodnotu χ^2 porovnáme s kritickou hodnotou z tabulek (viz dále)

Chí-kvadrát test

- **Nulová hypotéza** vyjadřuje očekávání, že **pozorované a očekávané četnosti se neliší**.
- Určení stupňů volnosti df = k - 1 – r
- k - počet kategorií r - počet parametrů předpokládaného rozdělní
- Kritický bod z tabulky statistické významnosti pro Alpha 0,05
- Pokud **vypočítaná $\chi^2 < \chi^2$ kritická** → nelze odmítnout H0 (= četnosti jsou mezi kategoriemi stejné).

Zpět do příkladu

Kritickou hodnotu χ^2 najdeme pro v tabulkách pro zvolenou hladinu významnosti **α** a počtu stupňů volnosti **df** zde obecně

df = k – 1 – r kde **k** počet kategorií znaku a **r** je počet parametrů předpokládaného rozdělení

df = 3 – 1 = 2

Najdeme tabulkovou kritickou hodnotu $\chi^2_{krit} = 5{,}991$

Protože ta je vyšší než námi naměřená $\chi^2 = 1{,}74$
→ rozložení četností odpovídá H0
→ **nemůžeme H0 zamítnout, tj. rozdíly mezi skupinami v populaci nejsou**.

v kontingenční tabulce (pro dva znaky) je počet stupňů volnosti df = (1-1) (s-1)

r = počet řádků s = počet sloupců v tabulce

Chí-kvadrát test nezávislosti

- Nulová hypotéza „o nezávislosti“ odpovídá na otázku, zda jsou rozdíly mezi empirickými-pozorovanými a teoretickými četnostmi náhodné nebo ne.
- Očekávané četnosti lze získat z hodnot v populaci nebo porovnávat s teoretickou hodnotou, např. z jiného výzkumu.
- Nejčastěji třídíme údaje podle dvou nebo více znaků v kontingenční tabulce.
- Lze aplikovat na již existující agregovaná data (publikované tabulky apod.)

- Příklad: porovnání vzdělanostní struktury v kohortě 50-64 a 65-79

Chíkvadrát test

Teoretické hodnoty odjinud než z očekávaných hodnot z dat

Chí-kvadrát test: Příklad 2
(ne)změna v čase

Teoretickou četností zde není poměrové rozložení ale hodnota z předchozí etapy.

Je podle vašeho názoru nabídka kulturních žánrů v našem městě dostatečná?

	Ano	Neví	Ne
Epirická četnost (2010)	65	28	6,7
Teoretická četnost (2007)	60	34	6

$$\chi^2 = \sum \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$$

Chí-kvadr	**1,53**
tabulková hodnota (pro 5 %)	5,99

Vypočítaná hodnota Chisq je **menší než tabulková-kritická hodnota**.
Platí H0 o "nerozdílu„ (rozdíl v četnostech je způsoben náhodnými faktory).

Ukázka v SPSS: porovnání v čase

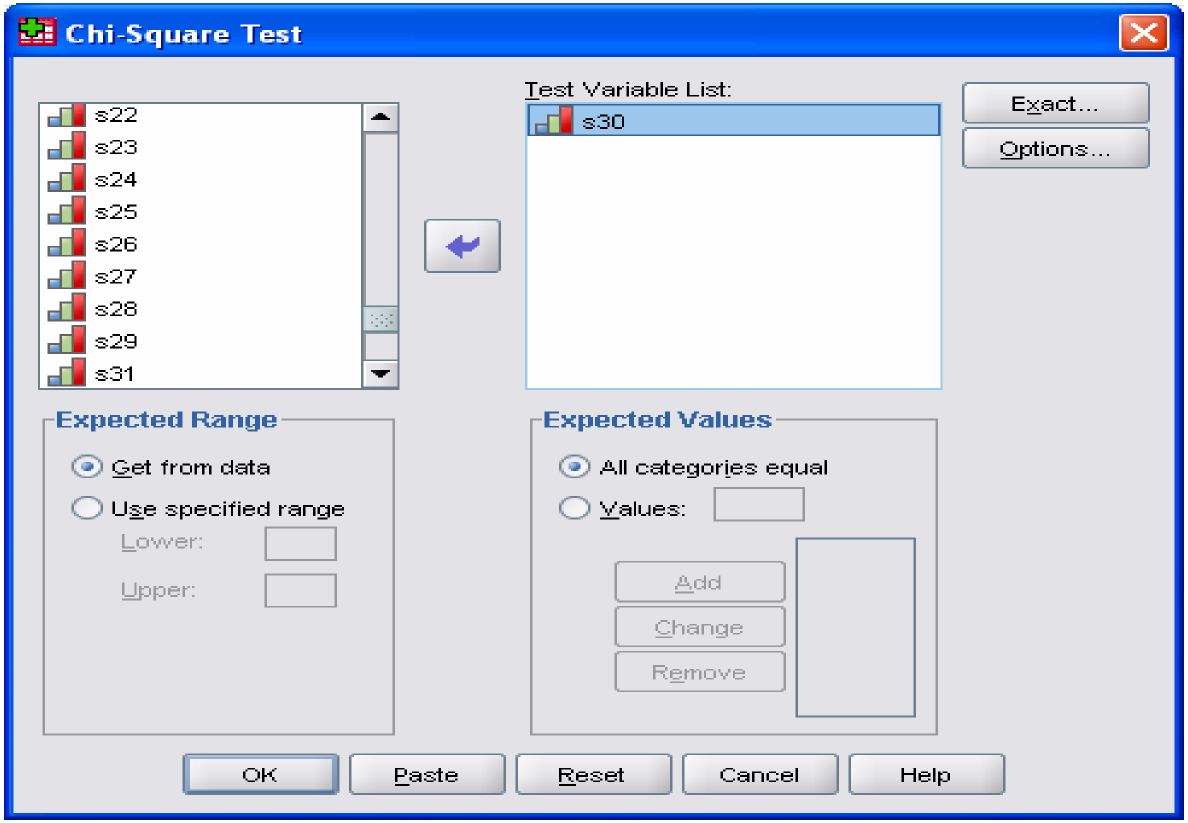
Porovnání proměny vzdělanostní struktury mezi kohortami 50-64 a 65-79 letých. → kohorta 65-79 představuje teoretické-očekávané hodnoty

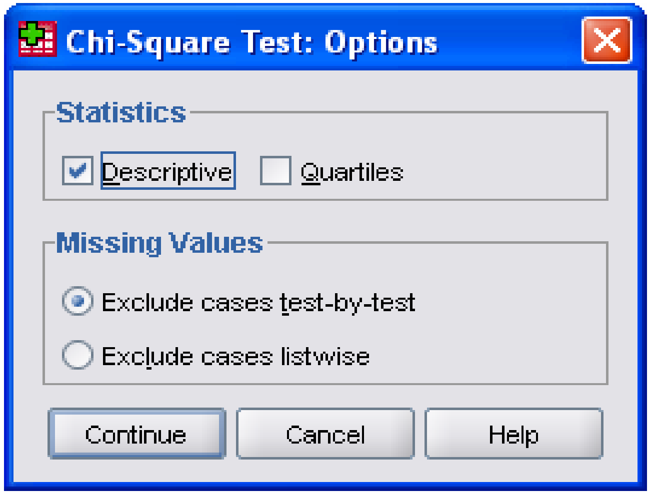
	18 - 29 let	**65 - 79 let**
ZŠ	48	**52**
VYU	165	**135**
SŠ	125	**72**
VŠ	17	**17**
	355	276

NPAR TESTS /CHISQUARE=**vzd4**
/EXPECTED= **52 135 72 17**
/STATISTICS DESCRIPTIVES /MISSING ANALYSIS.

Jiné statistické balíky mají možnost vstupu s tabelárními daty (kontingenční tabulka), v SPSS **pouze jako vážená data (frekvenčních váhy)** viz http://metodykv.wz.cz/syntaxy/**data_input.sps**

ChiSq Test četností (homogenity) kategorií 1 znaku (viz předchozí příklad)





Chisq – SPSS syntax

- *pohlavi - stejne H0: zastoupeni muzu a zen je stejné (EQUAL).
- **NPAR TESTS** /CHISQUARE=s30 /EXPECTED=**EQUAL** /STATISTICS DESCRIPTIVES /MISSING ANALYSIS.
- *pohlaví: **oproti teoretické četnosti** *očekávané-teoretické četnosti zadáváme v /EXPECTED= "CETNOST katg 1" "katg 2" atd. (zde v příkladu vstupujeme s četnostmi blízkými těm empirickým)
- **NPAR TESTS** /CHISQUARE=s30 /EXPECTED=**540 670** /STATISTICS DESCRIPTIVES /MISSING ANALYSIS.

Chíkvadrát test v kontingenční tabulce

Vztahy dvou (a více) znaků v tabulce

Princip testování vztahu 2 a více proměnných

- Většina statistických testů je založena na srovnání naměřené (empirické) distribuce pozorování do polí tabulky s distribucí, jakou bychom obdrželi, kdyby pozorování byla zařazena do polí tabulky náhodně (teoretická četnost).

Příklad: Čtení knih a vzdělání

Count

	vzd3 Vzdělání R (3k.)			
	1 ZŠ-VY	2 SŠ	3 VŠ	Total
1 denně	68	76	29	173
2 několikrát týdně	108	143	28	279
3 několikrát za měsíc	128	98	20	246
4 několikrát za rok nebo méně často	188	106	7	301
5 nikdy	153	48	2	203
Total	645	471	86	1202

Expected Count

	vzd3 Vzdělání R (3k.)			
	1 ZŠ-VY	2 SŠ	3 VŠ	Total
1 denně	**92,8**	67,8	12,4	173,0
2 několikrát týdně	149,7	109,3	20,0	279,0
3 několikrát za měsíc	132,0	96,4	17,6	246,0
4 několikrát za rok nebo méně často	161,5	117,9	21,5	301,0
5 nikdy	108,9	79,5	14,5	203,0
Total	645,0	471,0	86,0	1202,0

Očekávaná četnost pro dané políčko = násobek odpovídajících **marginálních četností** vydělíme celkovou sumou četností

Např. pro f_E11 je 645*173/1202 = **92,8**

$$\chi^2_{obt} = \sum \frac{(f_O - f_E)^2}{f_E}$$

Tabulka 9

Políčko (z tab. 8)	Četnost	Očekávaná četnost $\bar{n}_i$	$n_i - \bar{n}_i$	$(n_i - \bar{n}_i)^2$	$\frac{(n_i - \bar{n}_i)^2}{\bar{n}_i}$
a	18	12,9	5,1	26,00	2,015
b	13	12,1	0,9	0,81	0,056
c	10	16,0	6,0	36,00	2,250
d	23	15,2	7,8	60,84	4,002
e	13	14,1	1,1	1,21	0,085
f	12	18,7	6,7	44,89	2,400
g	11	15,2	4,2	17,64	1,160
h	14	14,1	0,1	0,01	0,000
ch	23	18,7	4,3	18,49	0,988
i	8	16,7	8,7	76,69	4,532
j	16	15,6	0,4	0,16	0,010
k	29	20,6	8,4	70,56	3,425
					$\chi^2 = 20,923$

Příklad: Čtení knih a vzdělání

DF = (5-1)(3-1) = **8** při Alpha 0,05

χ^2_{krit} = 15,507 < naměřená hodnota 112,17 → **zamítáme H0** „o nezávislosti“, tj, že ve čtení nejsou rozdíly mezi vzdělanostními kategoriemi

→ alespoň jedna kategorie se liší od ostatních (tuto skutečnost nalezneme v 95 % případů v celé populaci)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**112,167[a]**	**8**	**,000**
Likelihood Ratio	116,334	8	,000
Linear-by-Linear Association	100,953	1	,000
N of Valid Cases	1202		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 12,38.

Místo porovnání hodnoty testovacího kritéria s kritickými hodnotami se pro rozhodování o nulové hypotéze používá také ***p*-hodnota**, kterou zjisítme pomocí statistického software.

$p < \alpha$ zamítáme H0

$p > \alpha$ nelze zamítnout H0

Kontingenční tabulka
- vyjádření vztahů kategorií

- Statistika **chí kvadrát nevypovídá nic o síle vztahu**, pouze zamítá/nezamítá nulovou hypotézu o závislosti nebo homogenitě na dané hladině významnosti alfa.
- **Pro zjištění síly vztahu** →
  - koeficienty (obdobné korelaci: **CC**),
  - znaménkové schéma – adjustovaná residua
  - **podíl šancí (OR)**,
  - u ordinálních veličin korelační koef. dle pořadí.

  Odlišné testy pro nominální a ordinální proměnné (jedna / obě).

Kontingenční tabulka a testy dobré shody – pozor na:

- Prázdná pole a nízké četnosti v tabulce mohou zkreslit význam koeficientů měřících souvislost.
- Pro použití testů založených na testu dobré shody (test nezávislosti nebo homogenity) je třeba, aby se v tabulce vyskytlo **méně než 20 % políček, v nichž by očekávané (teoretické) četnosti byly menší než 5.** V případě, že se tak stane, můžeme zvážit transformaci — **sloučení některých méně obsazených kategorií** (např. "ano" a "spíše ano").

Chí-kvadrát test: příklad: Kouření marihuany u žáků 9 a 12 třídy.

TABLE 6.14 Marijuana Use

	Grade		
	9th	*12th*	*Total*
None	42	33	75
Marijuana	23	22	45
Total	65	55	120

Because 45 out of 120 of the total sample (9th- and 12th-graders) used marijuana during the past year, the proportion for the total sample is 45/120 = .375. The ex-

Chí-kvadrát test: příklad:

TABLE 6.15 Observed and Expected Frequencies for Marijuana Use

	Grade		
	9th	*12th*	*Total*
None	42 (40.625)	33 (34.375)	75
Marijuana	23 (24.375)	22 (20.675)	45
Total	65	55	120

NOTE: Expected frequencies are in parentheses.

$$\chi^2_{obs} = \sum \frac{(f_O - f_E)^2}{f_E}$$

$df = (\text{Rows} - 1)(\text{Columns} - 1)$, $\quad df = (2 - 1)(2 - 1) = (1)(1) = 1.$

Chí-kvadrát test: příklad

TABLE 6.16 Computation of χ^2_{obt}

Observed (f_O)	Expected (f_E)	$\lvert f_o - f_E \rvert - 0.5$	$(\lvert f_o - f_E \rvert - 0.5)^2$	$\frac{(\lvert f_O - f_E \rvert - 0.5)^2}{f_E}$
42	40.625	.875	0.765625	0.019
33	34.375	.875	0.765625	0.022
23	24.375	.875	0.765625	0.031
22	20.625	.875	0.765625	0.037

NOTE: $\chi^2_{obt} = 0.019 + 0.022 + 0.031 + 0.037 = 0.109$.

For $df = 1$ and $\alpha = .05$, the critical value for χ^2 is 3.84. Our calculated value (χ^2_{obt}) was 0.109. Because the obtained (calculated) value did not exceed the critical value, we would not reject the null hypothesis at $\alpha = .05$.

Chíkvadrát **kritický** z tabulek > Chíkvadrát **dosažený** (naměřený)

→ Ho nelze zamítnout = homogenita mezi kategoriemi

Načtení dat z agregované existující kontingenční tabulky (→ vážení procenty)

věk	vzdělání	volil 1 nevolil	volil 2 volil
1 <49	1 ZŠ+VY	**138**	**92**
	2 SŠ+VŠ	**106**	**218**
2 >50	1 ZŠ+VY	**143**	**257**
	2 SŠ+VŠ	**56**	**175**

Pozice pole v tabulce

věk	vzdělání	volil 0 nevolil	volil 1 volil
1 <49	0 ZŠ+VY	**111**	**112**
	1 SŠ+VŠ	**121**	**122**
2 >50	0 ZŠ+VY	**211**	**212**
	1 SŠ+VŠ	**221**	**222**

```
****nacteni kontingencni tabulky aneb
sekundarni analyza (ČR, ISSP 2007).
DATA LIST LIST/vek vzdel volil freq.
VAL LAB vzdel 1 "ZŠ+VY" 2 "SŠ+VŠ" /
vek 1 "<49" 2 ">50" / volil 1
"nevolil" 2 "volil".
BEGIN DATA
1 1 1 138
1 1 2 92
1 2 1 106
1 2 2 218
2 1 1 143
2 1 2 257
2 2 1 56
2 2 2 175
END DATA.
FORMATS vek vzdel volil freq (f8).
WEIGHT by freq.
CROSS vzdel by volil by vek.
CROSS vzdel by volil.
```

Syntax:
crosstab_data_input.sps

Adjustovaná residua

a znaménkové schéma

Adjustovaná residua Znaménkové schéma

- CROSSTABS: Adj. standardised (v SPSS / PSPP)

Adjustovaná residua =

- Residuum v daném políčku tabulky (= Pozorovaná (observed) minus Očekávaná (expected) hodnota) dělené odhadem vlastní standardní chyby. Standardizovaný residuál je vyjádřen v jednotkách směrodatné odchylky nad nebo pod průměrem.

Znaménkové schéma → jednoduchá vizualizace

- 'kde abs(z) >= **3.29** nahradí +++ resp. ---,
- 'kde abs(z) >= **2.58** nahradí ++ resp. --,
- 'kde abs(z) >= **1.96** nahradí + resp. -.

Znaménkové schéma: Znaménka

Struktura adjustovaných residuí může skrývat působení nějakých latentních faktorů, které jsou přímo neměřitelné, ale které se v dané asociační struktuře projevují.

Jde o **latentní vlivy**, na které můžeme usuzovat pouze na základě takto zjištěného vnějšího projevu. V praxi je struktura charakterizována, např. tzv. **znaménkovým schématem** (s volbou hranic pro znaménka: -, + = významné na hladině **0,05**; **--**, **++** = na **0,01**; **---**, **+++** = na **0,001**). Rozlišujeme:

- **simultánní inferenci**, → postihuje významnou strukturu toku **jako celku** (implementováno v SPSS v Asresid),
- **testování postupně všech jednotlivých polí** → struktura znamének označuje významnost těchto jednotlivých proudů.

  Zde je schéma znamének v tabulce bohatší, protože prokázat statistickou vlastnost jednoho dílčího proudu bez ohledu na chování ostatních vyžaduje podstatně méně odchylné skóry než přijetí statisticky prokazatelného závěru o šedesáti **dílčích proudech současně**, tj. přijetí pravděpodobnostně spolehlivého závěru o tom, že všechny označené proudy jsou statisticky významně specifické (slabší nebo silnější) a tudíž jejich struktura může být interpretována jako systematicky vznikající **celistvý tok**.

ZS je běžná rutina československých sociologů, umožňuje názorně pracovat se strukturou asociací v kontingenční tabulce. Je logickým krokem v analýze interakčních vazeb mezi kategoriemi řádků a sloupců.

[Řehák, Mánek 1991]

Opět příklad: Čtení knih a vzdělání:

absolutní četnosti, sloupcová %, adjustovaná residua

q1_d Jak často - Čtení knih * vzd3 Vzdělání (3k.) Crosstabulation

			vzd3 Vzdělání (3k.)			
			1 ZŠ+VY	2 SŠ	3 VŠ	Total
q1_d Jak často - Čtení knih	1 denně	Count	68	76	29	173
		% within vzd3 Vzdělání (3k.)	10,5%	16,1%	33,7%	14,4%
		Adjusted Residual	-4,1	1,4	5,3	
	2 několikrát týdně	Count	108	143	28	279
		% within vzd3 Vzdělání (3k.)	16,7%	30,4%	32,6%	23,2%
		Adjusted Residual	-5,7	4,7	2,1	
	3 několikrát za měsíc	Count	128	98	20	246
		% within vzd3 Vzdělání (3k.)	19,8%	20,8%	23,3%	20,5%
		Adjusted Residual	-,6	,2	,7	
	4 několikrát za rok nebo méně často	Count	188	106	7	301
		% within vzd3 Vzdělání (3k.)	29,1%	22,5%	8,1%	25,0%
		Adjusted Residual	3,5	-1,6	-3,8	
	5 nikdy	Count	153	48	2	203
		% within vzd3 Vzdělání (3k.)	23,7%	10,2%	2,3%	16,9%
		Adjusted Residual	6,8	-5,0	-3,7	
Total		Count	645	471	86	1202
		% within vzd3 Vzdělání (3k.)	100,0%	100,0%	100,0%	100,0%

Chi-Square Tests

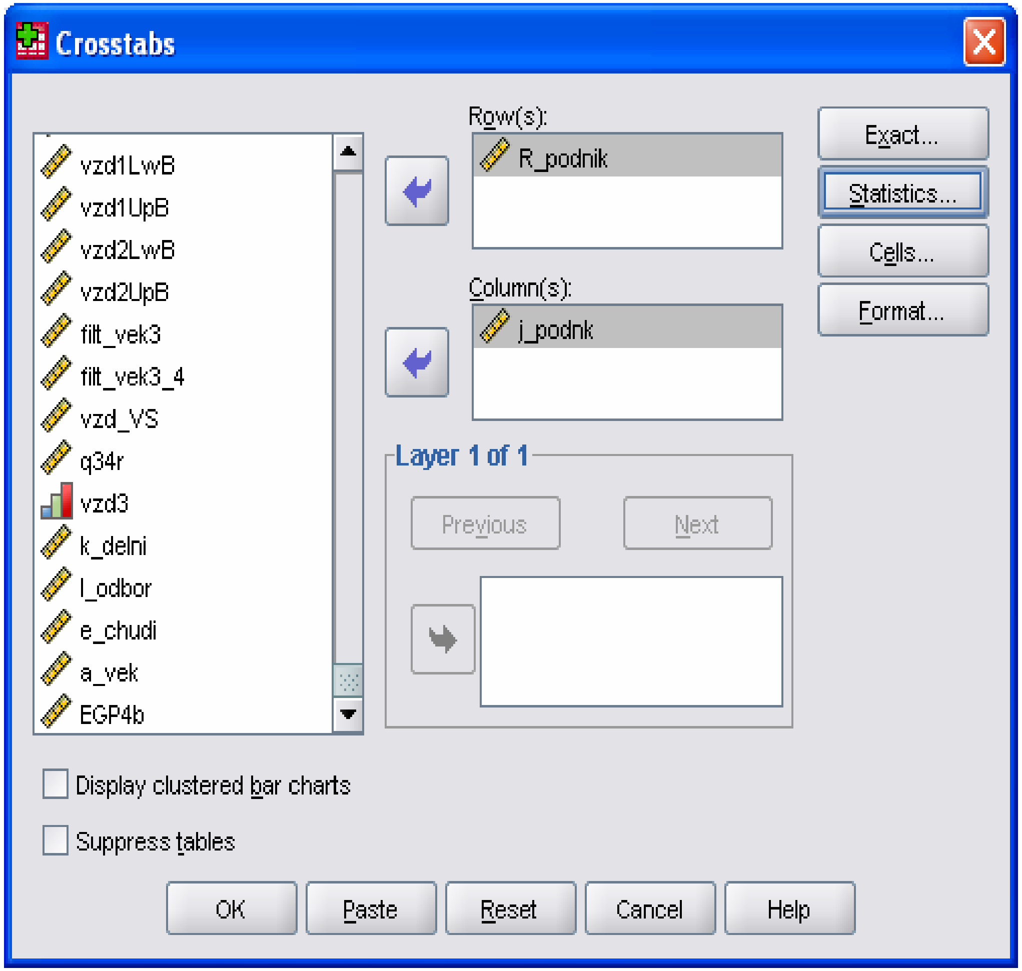
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	112,167[a]	8	,000
Likelihood Ratio	116,334	8	,000
Linear-by-Linear Association	100,953	1	,000
N of Valid Cases	1202		

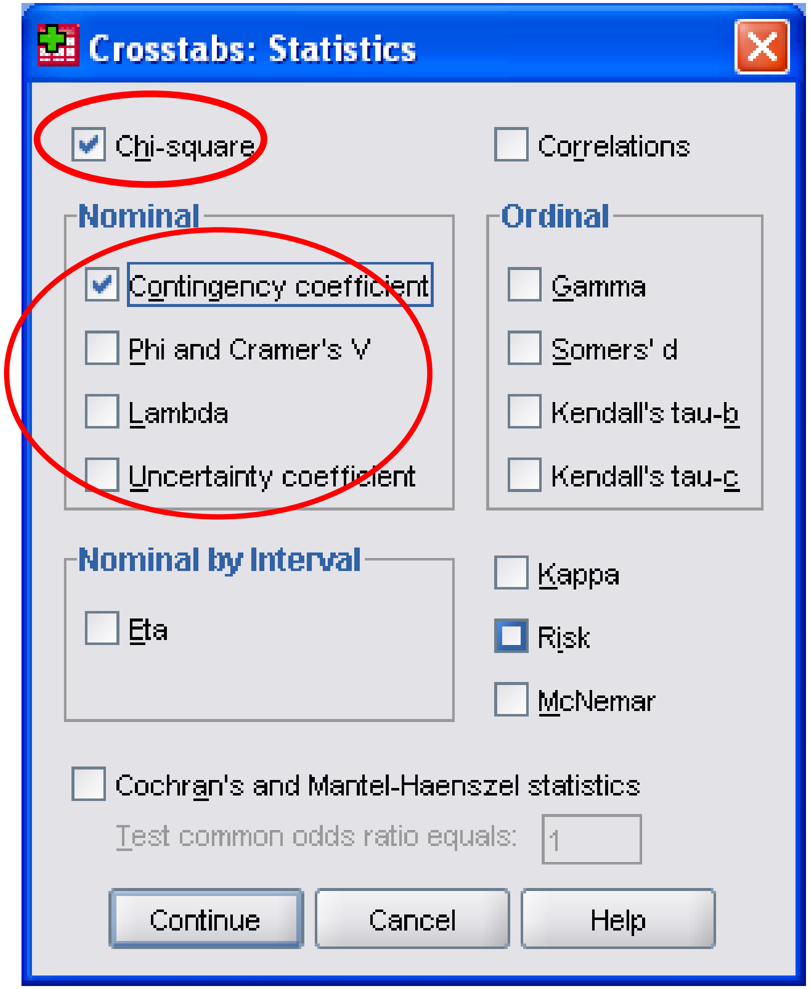
a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 12,38.

Znaménkové schéma

Jak často - Čtení knih	Vzdělání (3k.)		
	ZŠ+VY	SŠ	VŠ
1 denně	---	o	+++
2 několikrát týdně	---	+++	+
3 několikrát za měsíc	o	o	o
4 několikrát za rok/ méně často	+++	o	---
5 nikdy	+++	---	---

CROSSTABS: zadání Chíkvadrátu





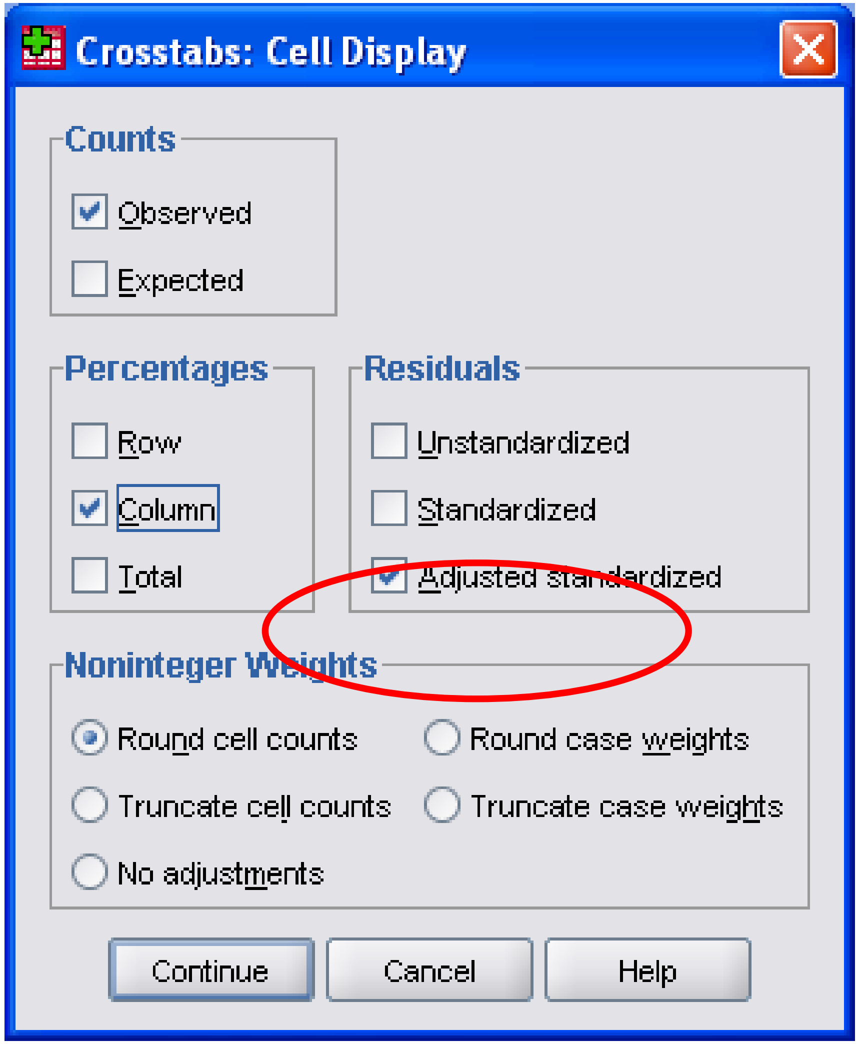
CROSSTABS: zadání **adjustovaných residuí** pro znaménkové schéma

- Samotné znaménkové schéma musíme následně vytvořit ručně z tabulky (dle hodnot z 3.29 2.58 1.96) a nebo použít skript

www.spss.cz/sc_znamenkoveschema.htm

Vzdělání * Pohlaví Crosstabulation

			Pohlaví		Total
			muž	žena	
Count	Vzdělání	základní	30	41	71
		vyučen	90	67	157
		maturita	91	113	204
		VŠ	44	26	70
	Total		255	247	502
Sign Scheme	Vzdělání	základní	o	o	
		vyučen	+	-	
		maturita	-	+	
		VŠ	+	-	



Procvičit v SPSS

0. kontrola absolutních četností v jednotlivých polích → transformace (sloučení)
1. správně orientovaná procenta
2. chíkvadrát test nezávislosti (tabulky jako celku)
3. adjustovaná residua a znaménkové schéma k detekování významných odchylek

Úkol:

- Pohlaví a volil v 2006
- Náboženské vyznání x Volil 2006
- Náboženské vyznání x Velikost bydliště
- Náboženské vyznání x Velikost bydliště x Volil 2006

Poměr šancí - ODDS RATIO

Poměr šancí - ODDS RATIO (OR)

- OR ukazuje asociaci v kontingenčních tabulkách
- **šance** (O) = poměr pravděpodobnosti jedné možnosti p1 (událost nastala) ke druhé p2(událost nenastala) (šance nebo také riziko)
- OR = **poměr dvou šancí** (odds)

- OR = $\mathbf{f_{11} * f_{22} / f_{12} * f_{21}} = \dfrac{p_1/(1-p_1)}{p_2/(1-p_2)}$

Poměr šancí (OR)

- **OR - podíl šancí** výskytu (rizika výskytu) pro dvě různé hodnoty dvou proměnných.
- OR: A k B a B k A jsou komplementární, vždy však s opačným směrem 1:3 =0,33 a 1/0,33 =3
- **O** je kladné číslo, kdy: >1 vyšší šance a <1 nižší šance

- OR není citlivé na marginální distribuce (změníme-li velikost *n* o konstantu, OR zůstávají stejné)
- Používá se také přirozený logaritmus poměru šance LOR <∞;∞>

ODDS RATIO - příklad

		VŠ - vzdělání		
		0	1	Total
Volil 2006	0 ne	**424**	**19**	443
	1 ano	**674**	**68**	742
Total		1098	87	1185

$$\mathbf{OR = f_{11}\, f_{22} \,/\, f_{12}\, f_{21}}$$

f11	f12
f21	f22

OR = (424*68)/(19*674) = **2,25**

U vysokoškoláků je v porovnání s ostatními 2,25x vyšší šance, že půjdou volit.

Úkol

- Procvičit v SPSS
- 2 x 2 tabulky
- Pohlaví a volil v 2006
- Pohlaví a Vzdělání

n x n

- Velikost bydliště x Vzdělání

→ sloučení nebo vybraná pole tabulky

Vyloučení vlivu třetího jevu

→ Třídění 3 stupně

- Kontingenční tabulka A x B x C
- Příklad: **pohlaví** x **volil** x **VŠ**

- Parciální korelace

- Multivariační metody (např. regresní analýza, ANOVA)

Elaborace

Třídění 3 stupně

aneb

kontrola pro další faktor

Vícerozměrná analýza: třídění třetího stupně

Analyzujeme souběžně vztahy mezi několika proměnnými (nejčastěji více nezávislých – vysvětlujících znaků).

Princip je stejný jako u dvourozměrné analýzy.

Vícerozměrná analýza: třídění třetího stupně

Church Attendance, Sex, and Age

"How often do you attend religious services?"

	Under 40		40 and Older	
	Men	Women	Men	Women
	Rozdíl **9 %**		Rozdíl **16 %**	
Weekly	21%	30%	34%	50%
Less often	79	70	66	50
100% =	(270)	(332)	(317)	(414)
	100 %		100 %	

Jak mezi muži tak ženami starší lidé chodí do kostela častěji než mladí.

V každé věkové kategorii ženy navštěvují kostel častěji než muži.

Podle tabulky, pohlaví má nepatrně větší efekt na chození do kostela než věk.

Věk a pohlaví mají nezávislý vliv na chození do kostela. **Uvnitř** každé kategorie nezávislé proměnné odlišné vlastnosti té druhé přesto ovlivňují jednání.

Podobně obě nezávislé proměnné mají **kumulativní efekt** na jednání. Starší ženy chodí nejčastěji a mladí muži nejméně často.

[Babbie 1997: 391, tabulka 15-9]

Zjednodušení předchozí tabulky:

	Percent Who Attend Weekly	
	Men	Women
Under 40	21	30
	(270)	(332)
40 and older	34	50
	(317)	(414)

100 %

→ 70 % méně často

Ukazujeme pouze pozitivní kategorie znaku („do kostela chodí týdně).

Při tom neztrácíme žádný údaj. Četnosti v závorkách uvádí procentní základ, z něj lze dopočítat podíl nezobrazené kategorie.

[Babbie 1997: 391: Table 15-10]

Příklad I.: **Nepravá souvislost**
1. bivariátní vztah (třídění 2.st.)

	Preferované jídlo:		Celkem:
Zbožnost:	HAMBURGY	KAVIÁR	
vysoká	78% 780	25% 125	905
nízká	22% 220	75% 375	595
Celkem: N	100% 1000	100% 500	1500

LAMBDA = .420

Zdroj: [Disman 1993: 219-223]

2. Při kontrole vlivu vzdělání (třídění 3 st.)

Osoby s nízkým vzděláním

	Preferované jídlo:		Celkem:
Zbožnost:	HAMBURGY	KAVIÁR	
vysoká	78% 624	78% 156	780
nízká	22% 176	22% 44	220
Celkem: N	100% 800	100% 200	1000

LAMBDA = 0

2. Při kontrole vlivu vzdělání (třídění 3 st.)

Osoby s vysokým vzděláním

	Preferované jídlo:		Celkem:
Zbožnost:	HAMBURGY	KAVIÁR	
vysoká	25% 50	25% 75	125
nízká	75% 150	75% 325	475
Celkem: N	100% 200	100% 400	600

LAMBDA = 0

Zdroj: [Disman 1993: 219-223]

Příklad II.: **Potlačená souvislost (nepravá nezávislost)**

1. bivariátní vztah (třídění 2.st.)

	Balení A	Balení B	Celkem:
asi by koupil	40% 80	40% 160	240
asi ne	60% 120	60% 240	360
Celkem:	100% 200	100% 400	600

LAMBDA = 0

Zdroj: [Disman 1993: 225-227]

2. s kontrolou pohlaví (třídění 3 st.)

muži

	Balení A	Balení B	Celkem:
asi by koupil	40% 40	40% 40	80
asi ne	60% 60	60% 60	120
Celkem:	100% 100	100% 100	100

LAMBDA = 0

ženy

	Balení A	Balení B	Celkem:
asi by koupil	100% 100	20% 60	160
asi ne	0% 0	80% 240	240
Celkem:	100% 100	100% 300	400

LAMBDA = .625

Kontrola 3 faktoru odhalila **potlačenou souvislost** (nepravou nezávislost) mezi dvěma proměnnými

Příčina zkreslení → **vztah mezi dvěma proměnnými existuje pouze v části populace**

Testování/ kontrola vlivu dalšího faktoru

- Vytvořením **samostatných tabulek podle kategorií třetí proměnné** je testovaný faktor (třetí proměnná) udržován na konstantní hodnotě.

  → souvislost mezi původními proměnnými je **očištěna od zkreslujícího vlivu** této další proměnné.

Testování vlivu dalšího faktoru

- Porovnáme intenzitu souvislosti v původní tabulce se souvislosti zjištěnou v nových tabulkách s kontrolou 3 faktoru .
- Když v nových tabulkách **souvislost** mezi původními daty **zmizí**/ je podstatně **oslabena → souvislost v původní tabulce je funkcí třetího faktoru**

Třídění 3 st.: kontrola vlivu 3 proměnné: interpretace a uspořádání tabulky

	Základní vzdělání			Střední vzdělání			Vysokoškolské vzdělání		
	< 39 let	40-59	> 60 let	< 39 let	40-59	> 60 let	< 39 let	40-59	> 60 let
Volil	**18%**	**24%**	**32%**	**36%**	**34%**	**49%**	**40%**	**50%**	**70%**
Nevolil	82	76	68	64	66	51	60	50	30
Celkem	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
N	(109)	(202)	(45)	(97)	(271)	(139)	(27)	(62)	(50)

Ptáme se:

1. Nacházíme rozdíly v X (věk) a Y (volil) uvnitř kategorií kontrolní proměnné Z (vzdělání)? Porovnáme s tabulkou třídění 2. st. Pro X a Y.

2. Jsou rozdíly mezi krajními kategoriemi X (věk) v rámci kategorií kontorlní proměnné Z (vzdělání) stejné?

Vztahy mezi X-Y a (Z)

Moderace a mediace

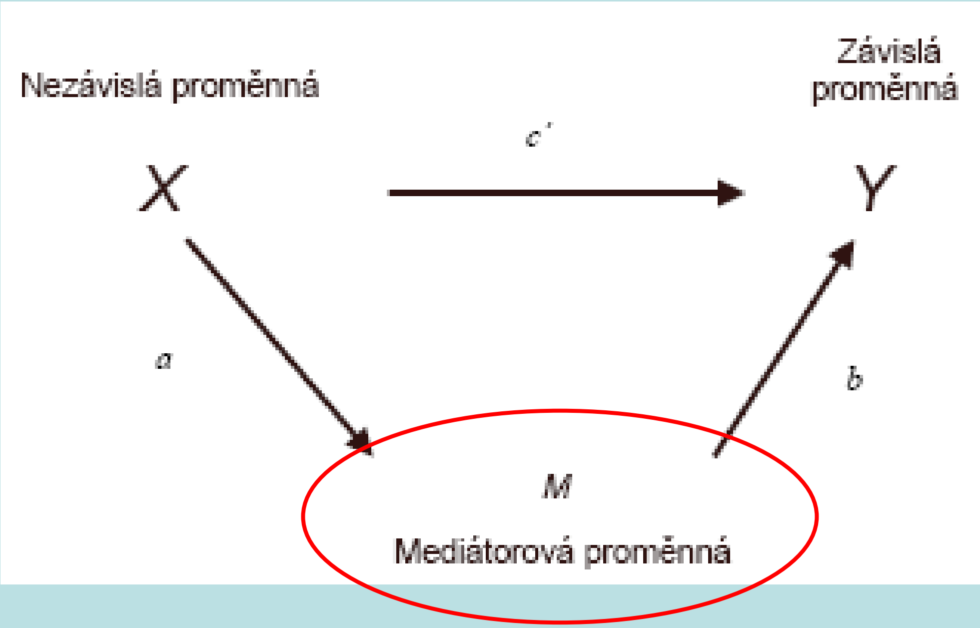
úvod

Vztah X-Y a Z: **moderace** a **mediace**

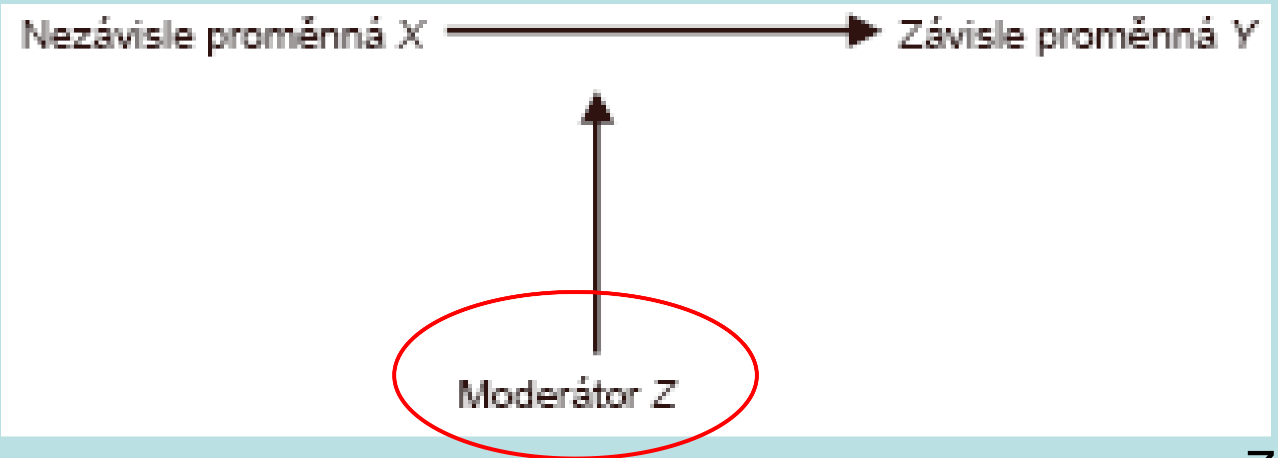
- **Mediátor** propojuje příčinu a následek.
  - Příčinna ovlivnuje mediátorovou proměnnou a ta pak působí na závislou proměnnou Y.
- **Moderátor** modifikuje přímé působení nezávislé X na závislou proměnnou Y.
  - Stálá vlastnost (např. kontextuální proměnná jako charakterisika okolí) modifikuje příčinnou závislost.

[Hendl 2010].

Vztah X-Y a Z: **moderace** a **mediace**



Mediátor



Moderátor

Zdroj: [Hendl 2010: 3, 6]

Literatura

- Disman, M. (1993): *Jak se vyrábí sociologická znalost.* Praha: Karolinum
- Babbie, E. (1995). *The Practice of social Research*. 7th Edition. Belmont: Wadsworth
- Hendl, J. 2010. „Analýza působení mediátorových a moderátorových proměnných“ *Informační Bulletin České statistické společnosti* 21(1): 1- 15.

Vícenásobné výběrové otázky - Multiple response

- ála: *„Ze seznamu vyberte 3 položky, které považujete za* …“
- Není doslova tříděním 1.stupně protože
- Odpovědi jsou na sobě specifickým způsobem závislé
- Častou chybou je pak třídění 1. st. pomocí formálně zavedených znaků: 1 znak = 1. zatržená hodnota, 2 znak = 2 zatržená hodnota atd. → Nedává interpretační smysl.
- správně: **počet voleb pro položku**, **procento z počtu voleb**, **procento z počtu respondentů**
- Lze dále zkonstruovat dichotomické proměnné pro každou položku
- V SPSS: *Analyze* → ***Multiple response*** → 1. *Define Variable Sets* a pak 2. *Frequencies* případně *Crosstabs*