

UK FHS
Historická sociologie a Řízení a supervize
(2011, 2012, 2013, 2014)

Analýza kvantitativních dat I./II.
& Praktikum elementární analýzy dat

Kontingenční tabulky

– analýza kategoriálních dat:

**Třídění 3. stupně – úvod do
elaborace a míry asociace**

Jiří Šafr

[jiri.safr\(zavináč\)seznam.cz](mailto:jiri.safr(zavináč)seznam.cz)

poslední aktualizace 5/6/2014

Obsah

- **Třídění třetího stupně v kontingenční tabulce**
(úvod do vícerozměrné analýzy)
- **Testování/ kontrola vlivu dalšího faktoru (3. proměnná) → Elaborace**
- **Uspořádání tabulky a její interpretace**
- **Dva příklady proč provádíme elaboraci:**
 1. **Nepravá souvislost** (falešná asociace/korelace)
 2. **Potlačená – skrytá souvislost**
- **Třídění 3. stupně v SPSS → CROSSTABS**
- **Míry asociace a pořadové korelace v kontingenční tabulce** (úvod)
- **Typy kontingenčních tabulek se 3 proměnnými a míry asociace/korelace**
- **Příklad elaborace: Vzdělanostní aspirace žáků**

Vícerozměrná analýza: třídění třetího stupně v kontingenční tabulce

**→ detailnější popis a
elaborace (rozkrytí vztahů)
(úvod)**

Třídění třetího stupně v kontingenční tabulce

- Analyzujeme souběžně vztahy mezi několika proměnnými (nejčastěji více nezávislých – vysvětlujících znaků).
- Princip je stejný jako u dvourozměrné analýzy.
- Cílem **třídění dat 3. (a vyššího) stupně** je v zásadě:
 - **Detailnější deskripce** (v pod/podskupinách)
 - **Elaborace vztahů** → hledání kauzálních vztahů, hlubších souvislostí, rozlišení podstatných a falešných vztahů, kontrola vlivu 3 proměnné ($X \leftrightarrow Y / Z$)
- Platí pro třídění vyššího stupně obecně, tj. i pro průměry v podskupinách a lineární závislosti (bodové grafy, korelace, regrese), zde si to nejprve vysvětlíme na kontingenční tabulce.

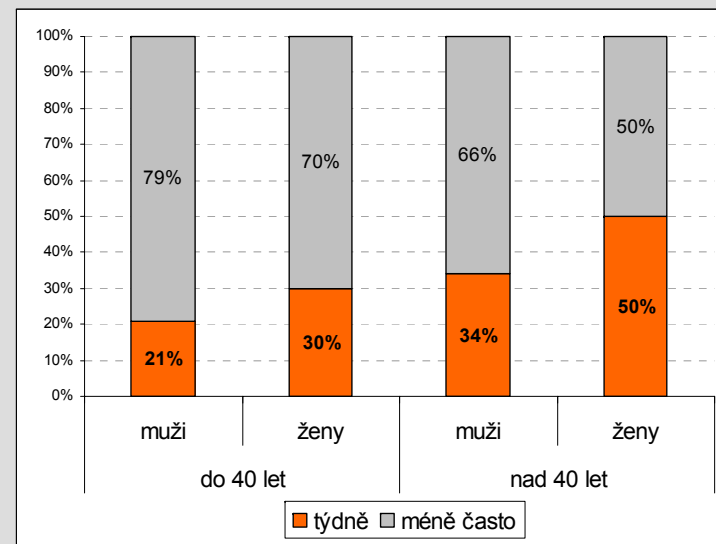
Princip vícerozměrné analýzy: třídění 3. stupně (2x2x2 tabulka)

Jak často navštěvujete bohoslužby?

	do 40 let		nad 40 let	
	muži	ženy	muži	ženy
týdně	21%	30%	34%	50%
méně často	79	70	66	50
100% =	(587)	(746)	(587)	(746)

Rozdíl 9 % bodů

Rozdíl 16 % bodů



Zdroj: General Social Survey, NORC.

Závislá proměnná: *Chození do kostela souběžně podle 2 nezávislých: Věk, Pohlaví*

Jak mezi muži tak ženami starší lidé chodí do kostela častěji než mladí (tj. s věkem roste religiozita).

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 do kostela nejčastěji, zatímco mladí muži nejméně často.**

Zdroj: [Babbie 1997: 391-392]

Zjednodušení předchozí tabulky:

	muži	ženy
do 40 let	21 (270)	30 (332)
nad 40 let	34 (317)	50 (414)

→ 70 % méně často
dopočet do **100 %**

Ukážeme **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.

Třídění 3. stupně (2x2x2 tabulka) → deskripce/explorace

Propadají více studenti „kolejáci“ – muži nebo „kolejáci“ – ženy?

Muži	Kolej	Jinde	Celkem
propadl	4%	19%	17%
nepropadl	96%	81%	83%
Celkem	100%	100%	100%

Rozdíl 15
procentních
bodů

Ženy	Kolej	Jinde	Celkem
propadla	30%	31%	30%
nepropadla	70%	69%	70%
Celkem	100%	100%	100%

Rozdíl 1
procentního
bodu

V porovnání s mužskými protějšky studentky bydlící na koleji propadají častěji. Ale je jich stejný podíl jako u těch studentek, co bydlí jinde (tzn. vliv koleje na prospěch se u žen zřejmě neuplatňuje; u mužů je pozitivní: „kolejáci“ – muži jsou u zkoušky úspěšnější a zároveň nejúspěšnější ze všech).

Úvod do elaborace

Třídění 3 stupně

aneb

kontrola pro další faktor

Testování/ kontrola vlivu dalšího faktoru (3. proměnná) → **Elaborace**

- Vytvořením **samostatných tabulek podle kategorií třetí proměnné** je testovaný faktor (třetí proměnná) udržován na konstantní hodnotě.
→ souvislost mezi původními proměnnými je **očistěna od zkreslujícího vlivu** této další proměnné.

Třídění 3 st.: kontrola vlivu 3 proměnné: interpretace a uspořádání (2x3x3) tabulky

Souvisí účast ve volbách s věkem, i při kontrole vlivu vzdělání?

U **ordinálních** nezávislých proměnných porovnáváme **procentní rozdíly** krajních kategorií odděleně mezi kategoriemi kontrolního faktoru.

	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)

Rozdíly mezi krajními kategoriemi věku v procentních bodech:

14 %

13 %

30 %

Ptáme se:

Zatímco v případě ZŠ a SŠ jsou rozdíly mezi nejmladšími a nejstaršími stejné, tak u VŠ je rozdíl větší.
→ Vzdělání tedy do vztahu mezi volební účastí a věkem částečně intervenuje.

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.

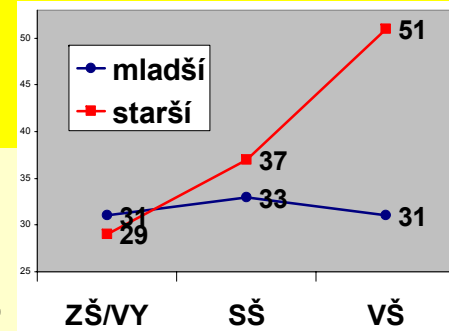
2. Jsou **rozdíly mezi krajními kategoriemi X** (věk) v rámci kategorií kontrolní proměnné Z (vzdělání) **stejné**?

Interakční a aditivní efekt

- Efekt 1 na 2 proměnnou závisí na 3 proměnné
- **Interakční efekt:** Dvě proměnné navzájem interagují a **vytváří u 3 proměnné jiný výsledek než by měla každá zvlášť**
- Při absenci interakčního efektu lze uvažovat o **aditivním efektu**, kdy vlivy jsou v principu podobné ale podél kategorií jedné proměnné **zesilují/ oslabují**
- Třetí možností pochopitelně je, že žádný společný efekt dvou nezávislých proměnných není a uplatňuje se vliv pouze jedné z nich.

Interakční a aditivní efekt

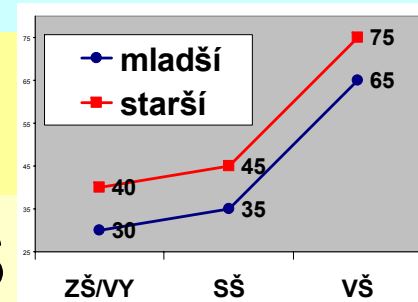
Interakční efekt – efekt jedné proměnné na druhou závisí na hodnotě třetí proměnné



Dopočet do 100 % je % Nevolil

VOLIL	ZŠ	vzdělání		SŠ	VŠ	
mladí		31	↔	33	↔	31
starší		29	→	37	→	51

Odlíšný vliv věku v kategoriích vzdělání: u Mladých žádný rozdíl, u Starších se % Volení zvyšuje s vyšším vzděláním. Nejvyšší volební účast je u starších vysokoškoláků.



Aditivní efekt – efekty obou proměnných se navzájem přidávají

VOLIL	ZŠ	SŠ	VŠ
mladí	30	35	65
starší	40	45	75

Stejný rozdíl mezi katg. věku v katg. vzdělání

Podobný vliv věku v kategoriích vzdělání, pouze na „odlišné hladině“

Odhalení vlivu 3. proměnné pomocí asociačních koeficientů

- Rychlou identifikaci vlivu 3. proměnné můžeme provést pomocí asociačních koeficientů spočítaných zvlášt' v jejích kategoriích.
- pro nominální znaky: Lambda, Phi, Cramérovo V, Koeficient kontingence
- pro ordinální znaky: ordinální korelace (Kendaullovo Tau-B a Tau-C, Spermanův korelační koeficient)

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** nebo je podstatně **oslabena** → **souvislost v původní tabulce je funkcí třetího faktoru**
- Dále je uvedeno, jak odhalit skrytý vztah rychle pomocí asociačních koeficientů v podskupinách 3-kontrolního faktoru (pro nominální znaky Lambda, Phi, CramV a ordinální korelace)
- V AKD II. si pak také ukážeme jak tabulku standardizovat (převážít) podle faktoru Z, tj. jako kdyby všichni v kategoriích X měli stejné podíly v kategoriích Z (např. stejné vzdělání).

Proč provádíme elaboraci?

1. Zjistit a popsat **interakční (aditivní) efekty**
a přitom můžeme objevit

2. **Nepravou souvislost (falešnou asociace/korelace)**

3. **Potlačenou – skrytou souvislost**

Cílem je **čistý vztah mezi dvěma proměnnými v tabulce s kontrolou vlivu 3 faktorů**

Následující dva příklady to osvětlí

Koeficienty asociace (jako např. zde Lambda) jsou vysvětleny dále, viz také

http://metodykv.wz.cz/spss2_tabulky.ppt .

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

$$\text{LAMBDA} = .420$$

Zdroj: [Disman 1993: 219-223]

Na první pohled silná souvislost, ale ...

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**

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

LAMBDA = 0

Souvislost zmizí jakmile kontrolujeme vliv vzdělání → faktor v pozadí mající vliv jak na zbožnost tak preferované jídlo.

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: 219-223]

Na první pohled žádná souvislost, ale ...

2. s kontrolou pohlaví (třídění 3 st.)

muži

ženy

	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

	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 faktorů odhalila **potlačenou souvislost** (nepravou nezávislost) mezi dvěma proměnnými

Příčina zkreslení → **vztah mezi dvěma proměnnými existuje pouze v části populace (u žen)**

Třídění 3. stupně v CROSSTABS v SPSS

- Kategoriální (vysvětlovaná-závislá) × kategoriální (vysvětlující- nezávislá) × **kategoriální kontrolní faktor**

CROSSTABS *var1-závislá* **BY** *var2-nezávislá*
BY *var3-kontrolní*.

- Procenta zůstávají **sloupcová** = **COLUMN** a nebo **řádková** = **ROW**

CROSSTABS *var1-závislá* **BY** *var2-nezávislá*
BY *var3-kontrolní* **/CELLS** **COL**.

Nebo otočeně: pořadí proměnných je obráceně a počítáme řádková %:

CROSSTABS *var2-nezávislá* **BY** *var1-závislá*
BY *var3-kontrolní* **/CELLS** **ROW**.

- Zadání CROSSTABS pro 3 znaky je opět podobné jako v MEANS:

MEANS *var1-závislá-číselná* **BY** *var2-nezávislá-kategoriální*
BY *var3-kontrolní-kategoriální*
/CELLS **MEAN** **STDDEV** **COUNT**.

Pozor na **absolutní četnosti** při třídění vyššího stupně

- Při třídění 3. a vyššího stupně vždy bedlivě **kontrolujte absolutní počty v jednotlivých polích tabulky**, zejména u malých souborů.

CROSSTABS *var1* BY *var2* BY *var3*
/CELLS COL COUNT.

- Pokud jsou četnosti v tabulkách velmi malé, pak je jejich interpretace ze statistického i věcného hlediska v podstatě bezcenná.

Míry asociace / korelace v kontingenční tabulce

pro kategoriální znaky

Míry asociace v kontingenční tabulce

- Při interpretaci i měření souvislosti je důležité, zda jsou jedna nebo obě proměnné nominální nebo ordinální.
- Základním nástrojem analýzy jsou vždy **procentní rozdíly**.
- Navíc můžeme měřit míru těsnosti vzájemného vztahu pomocí:
- pro **nominální** znaky koeficientů **asociace** (Kontingenční koeficient, Cramérovo V, Lambda atd.).
- pro **ordinální** znaky navíc (kromě koeficientů asociace) koeficientů **pořadové korelace** (Spermanovo Rho, Gamma, Kendallovu Tau B, ..).

Zadání nominálních asociací a pořadových korelací v SPSS uvádíme dále; podrobně viz 2. *Korelace a asociace: vztahy mezi kardinálními/ ordinálními znaky* na http://metodykv.wz.cz/AKD2_korelace.ppt

Pokud máme výběrová data (vzorek z populace), pak bychom měli testovat statistickou významnost koeficientů asociace/korelace (to se naučíme v AKD II.).

- K jednoduché analýze kontingenční tabulky také používáme např.: **odds ratio** = poměry šancí (→ vzájemně podmíněné pravděpodobnosti)

Podrobně viz 5. *Poměry šancí (Odds Ratio)* http://metodykv.wz.cz/AKD2_odds_ratio.ppt

míry rozptýlení, např. **Index nepodobnosti (Δ)**

Viz 9. *Míry variability: variační koeficient a další indexy* http://metodykv.wz.cz/AKD2_variacni_koef.ppt

Míry asociace (pro nominální proměnné)

- Obecně pro **koeficienty asociace** platí:
 - Mají rozpětí **0 = žádná souvislost až 1 = dokonalá souvislost** mezi znaky.
 - V principu říkají kolik – jaký podíl variability jedné proměnné lze vysvětlit pomocí druhé. Ale pozor, „vysvětlení“ je třeba chápat ve smyslu redukce statistického rozptýlení dat, nikoliv ve smyslu kauzální interpretace. [Řehák, Řeháková 1986: 250]
 - **Nevyjadřují směr asociace** (jako tomu je v případě korelací, nicméně některé koeficienty asociace jsou asymetrické (directional), tj. musíme definovat, která proměnná je závislá a které nezávislá).
- **Kontingenční koeficient C (CC)**
Nejjednodušší na výpočet. Ale nepoužívejte je, tam kde porovnáváte míru asociace mezi tabulkami s různým počtem kategorií.
- **Cramér's V** (CV nebo Cr) obecně ho lze doporučit (ale má také nedostatky)
- Pokud jsou **obě proměnné dichotomické** (2×2 tabulka) používáme **Phi koeficient** (pro 2×2 tabulku je stejný jako CV)
- **Lambda Λ** (symetrická/ asymetrická) měří procentní zlepšení odhadu jedné proměnné na základě hodnot jiné proměnné (oboustranné – symetrická nebo pouze predikující závislou proměnnou – asymetrická)
- Všechny tyto koeficienty jsou k dispozici v SPSS pomocí CROSSTABS (viz dále)

Pozor: pokud nenaměříme korelaci, mezi znaky stále ještě může být (nominální) asociace.

- Pokud není přítomná ordinální závislost – korelace, tak to automaticky neznamena statistickou nezávislost. Znamená to pouze, že není ordinálně uspořádaný vztah (~ linearita). Stále mezi znaky ale může být **asociace**, tj. vzájemný spoluvýskyt hodnot je např. kumulován do jednoho políčka tabulky (nebo několika políček mimo diagonálu resp. bez jakéhokoliv jiného „trendu“).
- Tuto situaci indikuje **signifikantní koeficient asociace** (např. Cramerovo V) zatímco **ordinální korelace je přibližně nulová** (např. Gamma).
- **Pouze absence nominální závislosti – asociace znamená (celkovou) statistickou nezávislost.** (např. CV = 0)
- → **spočítejte oba typy koeficientů: asociace** (Cramer's V atd.) i **ordinální korelace** (Gamma atd.) a **porovnejte je.**

Míry asociace v kontingenční tabulce a Elaborace

- Míry asociace/korelace využíváme také při elaboraci
 - tj. v třídění dat 3. stupně (vč. popisných cílů analýz).
- Jsou asociace **v podskupinách podle 3. kontrolní proměnné** v zásadě stejné?
A nebo se liší jejich intenzita, či dokonce v případě korelací i směr souvislosti?

Míry asociace v třídění (2) a 3. stupně v CROSSTABS

- V rámci CROSSTABS můžeme spočítat míry asociace a korelace pro proměnné $Y \times X$ (bivariálně) a navíc i **odděleně v kategoriích kontrolního faktoru $Z \rightarrow$** což nám pomůže rychle posoudit interakce a zhodnotit „falešné“ vlivy.
- Pro **nominální znaky** (Y, X, Z -kontrolní faktor) **koeficienty asociace** (mají hodnoty 0-1):

```
CROSSTABS var1 BY var2 BY var3-kontrolní /CELLS COL  
/STATISTICS CC PHI.
```

Koeficienty asociace: **CC** = Kontingenční koeficient, **PHI** = *Cramérovo V* (+ ekvivalent pro dichotomické znaky *Phi*); jsou zde k dispozici i další koeficienty asociace a korelace (např. *Lambda*).

- Pro **ordinální znaky** (A, B) a **nominální/ordinální kontrolní faktor** (C) navíc krom asociací i **pořadové korelace** (hodnoty -1–0–1 \rightarrow **směr**):

```
CROSSTABS var1 BY var2 BY var3-kontrolní /CELLS COL  
/STATISTICS CC PHI GAMMA CORR BTAU.
```

Korelační koeficienty: **GAMMA** = Goodman&Kruskalovo *Gamma*, **BTAU** = Kendaullovo *Tau B*, **CORR** = Spermanovo *Rho* (+ Pearsonův korel. koef. R pro kardinální znaky)

- **Pozor, neměříme-li korelaci, neznamená to, že mezi znaky nemusí být silná závislost – asociace.** Navíc u ordinálních znaků nám porovnání korelací a koeficientů asociace může napovědět o (nelineární) povaze vztahu.
- Poznámka: v případě průměrů v podskupinách v MEANS lze počítat koeficient(y) **Eta²** (pro kardinální \times nominální znak):

```
MEANS var1-závislá-číselná BY var2-nezávislá-kateg. BY var3-kontrolní-kategoriální  
/CELLS MEAN STDDEV COUNT /STATISTICS ANOVA.
```

Více o koeficientech asociace a korelace v 2. *Korelace a asociace: vztahy mezi kardinálními/ ordinálními znaky* na http://metodykv.wz.cz/AKD2_korelace.ppt

Pořadové korelace pro ordinální znaky - třídění 2. stupně

			vzd4 Vzdělání				
			1 ZŠ	2 VYUČ	3 SŠ	4 VŠ	Total
prijem4 Příjem-osobní - kvartily (s24)	1 I. do 7 tis.	Count	67	95	94	10	266
		% within vzd4 Vzdělání	62,6%	26,4%	28,9%	19,6%	31,6%
	2 II. 7-9 tis.	Count	28	115	51	9	203
		% within vzd4 Vzdělání	26,2%	31,9%	15,7%	17,6%	24,1%
	3 III. 9-15 tis.	Count	8	107	95	11	221
		% within vzd4 Vzdělání	7,5%	29,7%	29,2%	21,6%	26,2%
	4 IV. nad 15 tis.	Count	4	43	85	21	153
		% within vzd4 Vzdělání	3,7%	11,9%	26,2%	41,2%	18,1%
Total	Count		107	360	325	51	843
	% within vzd4 Vzdělání		100,0%	100,0%	100,0%	100,0%	100,0%

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	,241	,029	8,114	,000
	Gamma	,343	,040	8,114	,000
N of Valid Cases		843			

Pro výběrová data navíc musíme nejprve testovat statistickou hypotézu, že koeficient není roven nule (tj. je nenulový i v celé populaci a nejen v našem vzorku). O tom ale až v AKD II.

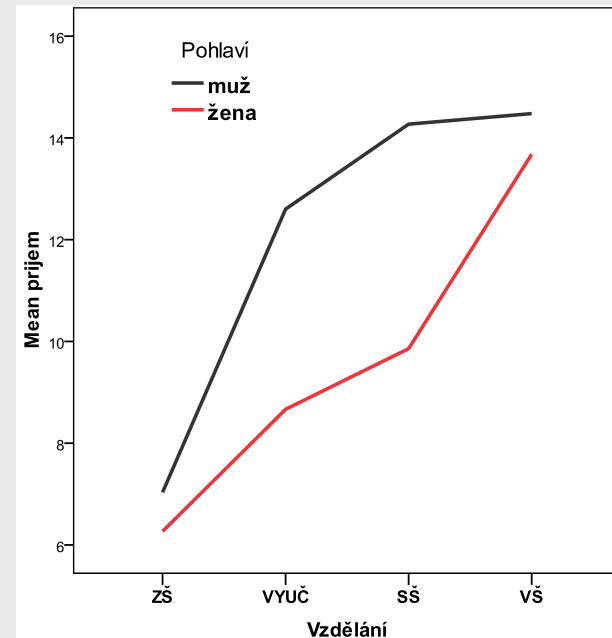
a. Not assuming the null hypothesis.

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

Pořadové korelace pro ordinální znaky v třídění

3. stupně (odděleně pro muže a ženy) → pohlaví [s30] je **kontrolní faktor**

s30 Pohlaví		vzd4 Vzdelání	vzd4 Vzdelání				Total	
			1 ZŠ	2 VYUČ	3 SŠ	4 VŠ		
1 muž	příjem4 Příjem-osobní - kvartily (s24)	1 I. do 7 tis.	Count	16	21	25	2	64
		% within vzd4 Vzdelání	55,2%	12,1%	18,8%	8,7%	17,8%	
	2 II. 7-9 tis.	Count	7	47	19	7	80	
		% within vzd4 Vzdelání	24,1%	27,0%	14,3%	30,4%	22,3%	
	3 III. 9-15 tis.	Count	2	66	33	4	105	
		% within vzd4 Vzdelání	6,9%	37,9%	24,8%	17,4%	29,2%	
	4 IV. nad 15 tis.	Count	4	40	56	10	110	
		% within vzd4 Vzdelání	13,8%	23,0%	42,1%	43,5%	30,6%	
Total		Count	29	174	133	23	359	
		% within vzd4 Vzdelání	100,0%	100,0%	100,0%	100,0%	100,0%	
2 žena	příjem4 Příjem-osobní - kvartily (s24)	1 I. do 7 tis.	Count	51	74	69	8	202
		% within vzd4 Vzdelání	65,4%	39,8%	35,9%	28,6%	41,7%	
	2 II. 7-9 tis.	Count	21	68	32	2	123	
		% within vzd4 Vzdelání	26,9%	36,6%	16,7%	7,1%	25,4%	
	3 III. 9-15 tis.	Count	6	41	62	7	116	
		% within vzd4 Vzdelání	7,7%	22,0%	32,3%	25,0%	24,0%	
	4 IV. nad 15 tis.	Count	0	3	29	11	43	
		% within vzd4 Vzdelání	,0%	1,6%	15,1%	39,3%	8,9%	
Total		Count	78	186	192	28	484	
		% within vzd4 Vzdelání	100,0%	100,0%	100,0%	100,0%	100,0%	



CROSSTABS příjem4 BY vzd4 BY s30 /STATISTICS GAMMA BTAU.

Symmetric Measures

s30 Pohlaví			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
1 muž	Ordinal by Ordinal	Kendall's tau-b	,191	,048	3,930	,000
		Gamma	,276	,068	3,930	,000
	N of Valid Cases			359		
2 žena	Ordinal by Ordinal	Kendall's tau-b	,281	,038	7,183	,000
		Gamma	,406	,052	7,183	,000
	N of Valid Cases			484		

U žen má vzdělání na příjem o něco větší efekt, ale celkově ženy vydělávají bez ohledu na vzdělání méně (viz též graf s průměry příjmu).

Pokud je min. jedna proměnná **multi-nominální**

- Princip je stejný jako u ordinálních znaků, ale **nemůžeme počítat korelace**, **pouze koeficienty asociace** (Kontingenční koeficient, Cramérovo V, Lambda atd.). Pokud je nominální pouze 3. kontrolní proměnná (a ostatní ordinální), pak korelace počítat a vzájemně je porovnávat lze.
- **Při interpretaci procentních rozdílů u nominálních znaků musíme brát v úvahu všechny kategorie závislé proměnné i nezávislých proměnných.** Jednodušší je to, pokud je alespoň některá ordinální.
- Ideální je, pokud máme závislou proměnnou dichotomickou nebo ordinální.
- Pokud je závislá proměnná dichotomická, tak jde o ekvivalent porovnávání průměrů v pod/podskupinách.

Typy kontingenčních tabulek se 3 proměnnými a míry asociace/korelace

Vždy lze míru asociace vyjádřit pomocí koef. asociace

- **2×2×2** (podobně **2×2×3n**) – všechny dichotomické → koeficienty asociace a bodově biseriální korelace nebo tetrachorické korelace
- **2×3o×3n** nebo **2×3o×2** – závislá dichotomická, nezávislá ordinální, kontrolní nominální → **pořadové korelace** ve skupinách kontrolního faktoru (bez možnosti posouzení trendu asociace/korelace).
- **2×3n×3o** – závislá dichotomická, nezávislá nominální, kontrolní ordinální → **pouze koeficienty asociace** (lze posuzovat trend v asociacích mezi kategoriemi kontrolního faktoru)
- **3o×3o×3o** (podobně i **2×2×3o**) – všechny ordinální → **pořadové korelace** (lze posuzovat trend v korelacích mezi kategoriemi kontrolního faktoru) + koeficient **parciální korelace**

Platí i pro více kategorií než 3.

Příklady pro bivariální asociace/korelace v konting. tabulce

2x2

		vek50plus		Total
		<49	>50	
Čte knihy >týdně	<měsíc	65,5%	59,8%	62,5%
	>týdně	34,5%	40,2%	37,5%
Total		100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,059
	Cramer's V	,059
	Contingency Coefficient	,059

2x3nominální

		Náboženská orientace			Total
		Římskokatolické	Katolické_ostat	žádné	
Čte knihy >týdně	<měsíc	58,1%	51,7%	65,0%	62,3%
	>týdně	41,9%	48,3%	35,0%	37,7%
Total		100,0%	100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,081
	Cramer's V	,081
	Contingency Coefficient	,081
N of Valid Cases		1155

2x3ordinální

		vek3			Total
		19-29	30-49	50+	
Čte knihy >týdně	<měsíc	64,7%	66,0%	59,8%	62,5%
	>týdně	35,3%	34,0%	40,2%	37,5%
Total		100,0%	100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,060
	Cramer's V	,060
	Contingency Coefficient	,060
Ordinal by Ordinal	Gamma	,096
	Spearman Correlation	,054

3x3o

		vek3			Total
		19-29	30-49	50+	
Čte knihy, 3k	>rok/nikdy	45,9%	43,6%	39,6%	41,9%
	<měsíc	18,8%	22,5%	20,2%	20,6%
	>týdně	35,3%	34,0%	40,2%	37,5%
Total		100,0%	100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,068
	Cramer's V	,048
	Contingency Coefficient	,068
Ordinal by Ordinal	Gamma	,084
	Spearman Correlation	,058

Pro tabulky větší než 2x2 lze vždy Cramérovo V a Kontingenční koeficient.

Příklad elaborace,
ke kterému se ještě vrátíme v AKD II.
při standardizaci (vážení) v tabulce podle faktoru

**Vzdělanostní aspirace žáků
8.-9.tříd ZŠ mezi
dvěma typy škol,
při kontrole vlivu vzdělání rodičů**

Vzdělanostní aspirace žáku **podle typu ZŠ** a vzdělání rodičů

Aspirace žáka	Typ školy (lokalita)		Celkem
	Sociálně vyloučená lokalita	Majoritní populace	
max. Vyučení	42%	32%	36%
min. SŠ-maturita	58%	68%	64%
Celkem	100%	100%	100%

Cramérovo $V = 0,101$

Na SŠ s maturitou aspiruje na školách v sociálně vyloučených lokalitách méně dětí než na školách v lokalitách s majoritní populací.

Aspirace žáka	Vzdělání rodičů (vyšší)		Celkem
	ZŠ/VYUČ	SŠ/VŠ	
max. Vyučení	55%	30%	35%
min. SŠ-maturita	45%	70%	65%
Celkem	100%	100%	100%

Cramérovo $V = 0,208$

Na aspirace má také vliv vzdělání rodičů: ve vzdělanějších rodinách jsou aspirace dětí vyšší.

Vliv vzdělání rodiny ($CV = 0,21$) je větší než v případě typu školy ($CV = 0,10$).

Zároveň ale ...

Vzdělání rodičů	Typ školy (lokalita)		Celkem
	Sociálně vyloučená lokalita	Majoritní populace	
max. Vyučení	25%	14%	19%
min. SŠ-maturita	75%	86%	81%
Celkem	100%	100%	100%

Cramérovo V = 0,148

Na školy v sociálně vyloučených lokalitách chodí více dětí s nižším vzděláním rodičů.

Vzdělanostní aspirace žáků podle typu školy a vzdělání rodičů: uspořádání tabulek

Výstup z SPSS pro kontingenční tabulku 3. stupně vypadá takto (panely dle vzdělání rodičů jsou pod sebou → lze přehodit pomocí Pivoting Trays):

PA1_Vzd_SS min.SŠ vzdělání-aspirace dle profese 1 * lok2 Typ školy * vzd_rod2 Vzdělání rodičů (vyšší) Crosstabulation

% within lok2 Typ školy

		lok2 Typ školy			
		1 Soc. vyloučená lokalita	2 Majoritní pop.	Total	
vzd_rod2 Vzdělání rodičů (všší)	PA1_Vzd_SS min.SŠ vzdělání-aspirace dle profese 1				
1 ZŠ/VYUČ	0 ZŠ/VY	54,5%	55,6%	55,0%	
	1 SŠ/VŠ	45,5%	44,4%	45,0%	
	Total	100,0%	100,0%	100,0%	
2 SŠ/VŠ	0 ZŠ/VY	37,3%	25,0%	30,3%	
	1 SŠ/VŠ	62,7%	75,0%	69,7%	
	Total	100,0%	100,0%	100,0%	

Výstup z SPSS upravíme pro klasický formát, kde kontrolní proměnná (vzdělání rodičů) je v panelech

Aspirace žáka	Vzdělání rodičů	
	ZŠ/VYUČ	SŠ/VŠ
	Soc.vylouč. lokalita	Majoritní populace
SŠ/VŠ	45%	44%
ZŠ/VY	55%	56%
	Soc.vylouč. lokalita	Majoritní populace
SŠ/VŠ	63%	75%
ZŠ/VY	37%	25%

Ideální je úsporný 3-dimenzionální formát tabulky:

- vynechána jsou % pro aspirace na ZŠ/VY (tvoří dopočet do 100 %)
- Otočeno o 90 st.

Podíl žáků aspirujících min. na maturitu podle typu školy a vzdělání rodičů

Typ školy	Vzdělání rodičů	
	ZŠ/VY	SŠ/VŠ
Soc.vylouč. lokalita	45%	63%
Majoritní populace	44%	75%

Vzdělanostní aspirace žáků podle typu školy a vzdělání rodičů: **interpretace**

Aspirace žáka	Vzdělání rodičů			
	ZŠ/VYUČ		SŠ/VŠ	
	Soc.vylouč. lokalita	Majoritní populace	Soc.vylouč. lokalita	Majoritní populace
SŠ/VŠ	45%	44%	63%	75%
ZŠ/VY	55%	56%	37%	25%
Cramérovo V	0,010		0,133	

- Rozdíl mezi vzděláním rodičů (s/bez maturity) → mezi panely tabulky:
 - ve školách ze sociálně vyloučených lokalit: $45 - 63 = 18$ % bodů
 - ve školách z lokalit s majoritní populací : $44 - 75 = 31$ % bodů
- Rozdíly mezi typy škol jsou v kategoriích vzdělání rodičů odlišné: pro nižší vzdělání (ZŠ/VY) nezáleží na typu školy (koeficient asociace CrV=0), zatímco u vyššího vzdělání rodičů (SŠ/VŠ) je ve školách z lokalit s majoritní populací vyšší podíl zájmu o maturitní obory (ZŠ/VY) (CrV=0,13).
- → **interakční efekt vzdělání rodičů a typu školy**: nejvyšší aspirace mají žáci z výše vzdělaných rodin a zároveň v prostředí škol z lokalit s majoritní populací.

A k třídění 3. stupně v kontingenční tabulce se ještě vrátíme v AKD II.

→ **6. Elaborace, standardizace/vážení**

http://metodykv.wz.cz/AKD2_kontg_tab_standardizace.ppt

→ **4. Testování hypotéz a asociace v kontingenčních tabulkách**

http://metodykv.wz.cz/AKD2_hypotezy1.ppt

Literatura

- Babbie, E. 1995. **Elaboration Model**. (kapitola 16)
Pp. 395-412 in *The Practice of social Research. 7th Edition*. Belmont: Wadsworth.
- Disman, M. 1993. *Jak se vyrábí sociologická znalost*. Praha: Karolinum. Pp. 217-282 Kapitola 9. „**Všechno je jinak aneb vícerozměrná analýza.**“
- Kapr, J., Z. Šafář. 1969. *Sociologie nebo zdravý rozum?* Praha: Mladá fronta.
- Kreidl, M. 2000. *Podklady ke kurzu Analýza kvantitativních dat*. FSV UK, LS 2000-2001.
- Řehák, J., B. Řeháková. 1986. *Analýza kategorizovaných dat v sociologii*. Praha: Academia.
- Treiman, D. J. 2009. *Quantitative data analysis: doing social research to test ideas*. San Francisco: Jossey-Bass.
- Šafr, J., B. Kalný. 2012. „Vzdělanostně–profesní aspirace žáků z rodin ohrožených sociálním vyloučením.“ Pp. 699–710 in M. Bargel, E. Janigová, E. Jarosz. (eds.). *Dilemata sociální pedagogiky v postmoderním světě*, Brno: Institut mezioborových studií Brno.
- de Vaus, D., A. (1985) 2002. *Surveys in Social Research, Fifth Edition*. St Leonards NSW: Allen & Unwin / London: Routledge.