

# Kvantitativní přístupy v politologii

Kontingenční tabulky – Crosstabs

# Crosstabs

- Specifická metoda zjišťující vzájemné souvislosti u proměnných, kde nemá smysl počítat průměry, rozptyl apod.
- Jedná se většinou o **nominální** znaky (národnost)
- **Ordinální s nízkým počtem znaků** (venkov/městys/město/velké město), příjem, vzdělání
- **Dichotomické** (pohlaví, levák/pravák)
- Hledáme zde vztahy mezi dvěma proměnnými - jak moc jedna ovlivňuje druhou
- Tvrzení: na základě rozložení jedné proměnné, je její distribuce nějak spojena s rozložením druhé proměnné
- Tvrdíme: Rozložení jedné proměnné není náhodné, ale **souvisí s rozložením druhé proměnné**, je vzorováno v závislosti na rozložení hodnot druhé proměnné
- Tyto postupy se nazývají **třídění druhého stupně**, protože rozložení variant jedné proměnné třídíme podle rozložení variant druhé proměnné
- česky vytváříme „kontingenční tabulku“,
- v SPSS jazyce tvoříme křížovou tabulaci – Crosstabulation

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help

Reports  
**Descriptive Statistics**  
 Bayesian Statistics  
 Tables  
 Compare Means  
 General Linear Model  
 Generalized Linear Models  
 Mixed Models  
 Correlate  
 Regression  
 Loglinear  
 Neural Networks  
 Classify  
 Dimension Reduction  
 Scale  
 Nonparametric Tests  
 Forecasting  
 Survival  
 Multiple Response  
 Missing Value Analysis...  
 Multiple Imputation  
 Complex Samples  
 Simulation...  
 Quality Control  
 ROC Curve...  
 Spatial and Temporal Modeling...  
 Direct Marketing

123 Frequencies...  
 Descriptives...  
 Explore...  
**Crosstabs...**  
 TURF Analysis  
 Ratio...  
 P-P Plots...  
 Q-Q Plots...

abulation  
 Rodinný stav

	rozvedený/á	odloučení	svobodný/á	Total		
	11	2	11	161		
	6,8%	1,2%	6,8%	100,0%		
	14	1	49	285		
	4,9%	0,4%	17,2%	100,0%		
	58	15	4	0	21	98
	59,2%	15,3%	4,1%	0,0%	21,4%	100,0%
	97	9	11	0	45	162
	59,9%	5,6%	6,8%	0,0%	27,8%	100,0%
	253	29	26	1	66	375
	67,5%	7,7%	6,9%	0,3%	17,6%	100,0%
	115	11	10	0	43	179
	64,2%	6,1%	5,6%	0,0%	24,0%	100,0%
	198	35	36	3	56	328
	60,4%	10,7%	11,0%	0,9%	17,1%	100,0%
	1018	160	112	7	291	1588
	64,1%	10,1%	7,1%	0,4%	18,3%	100,0%

Rekodovaná q72 K  
 Č  
 K  
 U  
 O  
 jin  
 n  
 Total

Crosstabs...  
 IBM SPSS Statistics Processor is re

# Crosstabs

Rekodovaná q72 \* Rodinný stav Crosstabulation

		Rodinný stav					Total	
		ženatý/ vdaná	vdovec/ vdova	rozvedený/á	odloučení	svobodný/á		
Rekodovaná q72	KSČM	Count	101	36	11	2	11	161
		% within Rekodovaná q72	62,7%	22,4%	6,8%	1,2%	6,8%	100,0%
	ČSSD	Count	196	25	14	1	49	285
		% within Rekodovaná q72	68,8%	8,8%	4,9%	0,4%	17,2%	100,0%
	KDU	Count	58	15	4	0	21	98
		% within Rekodovaná q72	59,2%	15,3%	4,1%	0,0%	21,4%	100,0%
	US	Count	97	9	11	0	45	162
		% within Rekodovaná q72	59,9%	5,6%	6,8%	0,0%	27,8%	100,0%
	ODS	Count	253	29	26	1	66	375
		% within Rekodovaná q72	67,5%	7,7%	6,9%	0,3%	17,6%	100,0%
	jiná	Count	115	11	10	0	43	179
		% within Rekodovaná q72	64,2%	6,1%	5,6%	0,0%	24,0%	100,0%
	nevolil by	Count	198	35	36	3	56	328
		% within Rekodovaná q72	60,4%	10,7%	11,0%	0,9%	17,1%	100,0%
Total		Count	1018	160	112	7	291	1588
		% within Rekodovaná q72	64,1%	10,1%	7,1%	0,4%	18,3%	100,0%

# Trojí procenta

- **Řádková procenta** se počítají tak, že absolutní četnosti v políčku tabulky se dělí celkovým počtem případů příslušného řádku (**zde se většinou umístí závislá proměnná**, tedy ta, která je ovlivňována)
- **Sloupcová procenta** se počítají analogicky, jen s tím rozdílem, že absolutní četnosti v políčku se dělí celkovým počtem případů ve sloupcové kategorii (**zde většinou umístíme nezávislou proměnnou**, tedy tu, která ovlivňuje rozložení druhé proměnné)
- **Celková procenta** získáme tak, že absolutní četnost v políčku dělíme celkovým počtem případů. Ten je uveden v křížovém součtu celkových počtů četností sloupců a řádků

# První krok

- Ověřit podmínky dobré aproximace: (ověření normality)
- 80% očekávaných četností, vyšší než 5 a zbylých 20% ne méně než 2
- Neplatí u čtyřpolní tabulky (použijí Fischerův test, ne chí.kvadrát)
- Nic nemusíme dělat, spočítá to za nás SPSS
- Uvedeme námi zvolená procenta

**Chi-Square Tests**

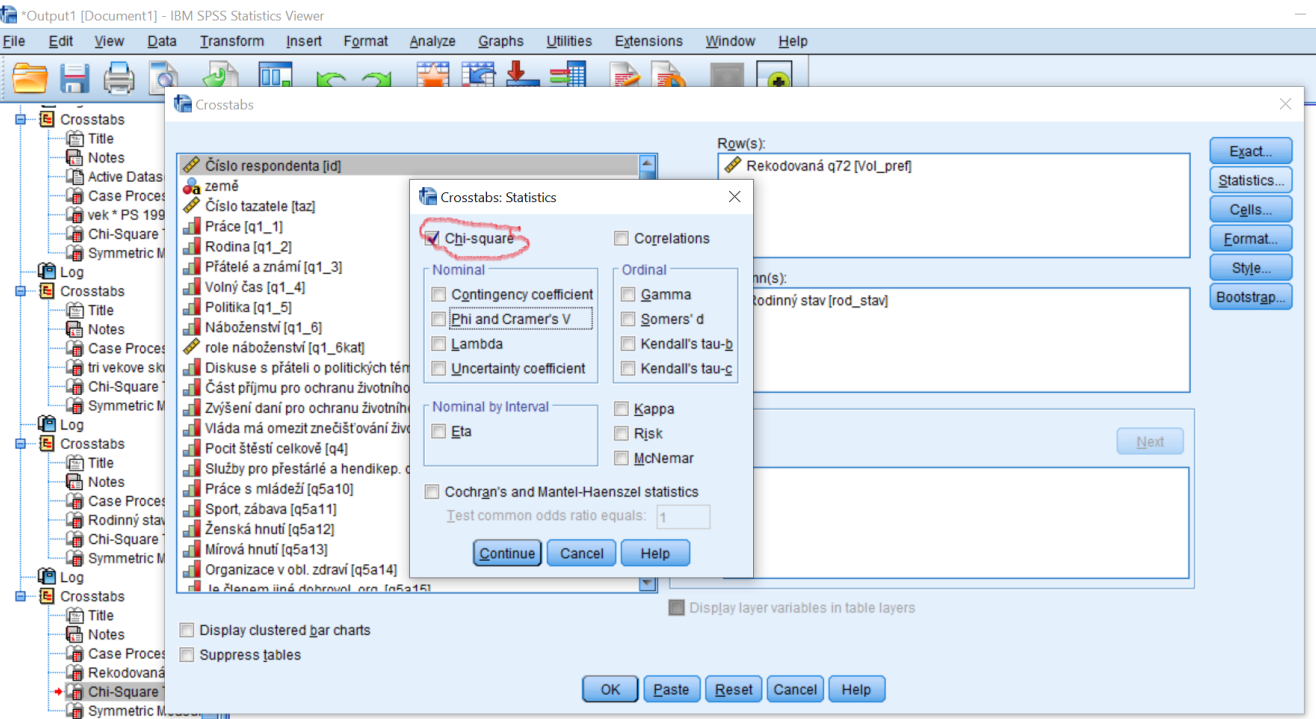
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	79,511 <sup>a</sup>	24	,000
Likelihood Ratio	76,707	24	,000
Linear-by-Linear Association	5,038	1	,025
N of Valid Cases	1588		

a. 7 cells (20,0%) have expected count less than 5. The minimum expected count is ,43.

# Druhý krok

- Výpočet chí-kvadrátu (testovacího kritéria) a **jeho P- hodnoty**
- Důležité jsou **empirické četnosti** (pozorovaná hodnota)
- **Očekávané četnosti** – četnost, která by se objevila, pokud by platila nulová hypotéza nezávislosti
- Výpočet chí, kvadrátu vychází z výpočtu **empirických** (pozorovaných hodnot v políčku tabulky) **četností a očekávaných četností**,
- V každém poli tabulky pak musíme vypočítat rozdíl mezi empirickou a očekávanou četností, ten umocnit na druhou, podělit hodnotou očekávané četnosti a výsledky sečíst: tak získáme **hodnotu chí-kvadrát**.
- Tu následně porovnááme s matematickým modelem rozložení, v tomto případě s modelem chí-kvadrát a zjistíme statistickou významnost – opět to udělá SPSS
- Důležitá vlastnost: **P – hodnota nesmí být vyšší jak 0,05**

# Druhý krok



## Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	79,511 <sup>a</sup>	24	,000
Likelihood Ratio	76,707	24	,000
Linear-by-Linear Association	5,038	1	,025
N of Valid Cases	1588		

a. 7 cells (20,0%) have expected count less than 5. The minimum expected count is ,43.

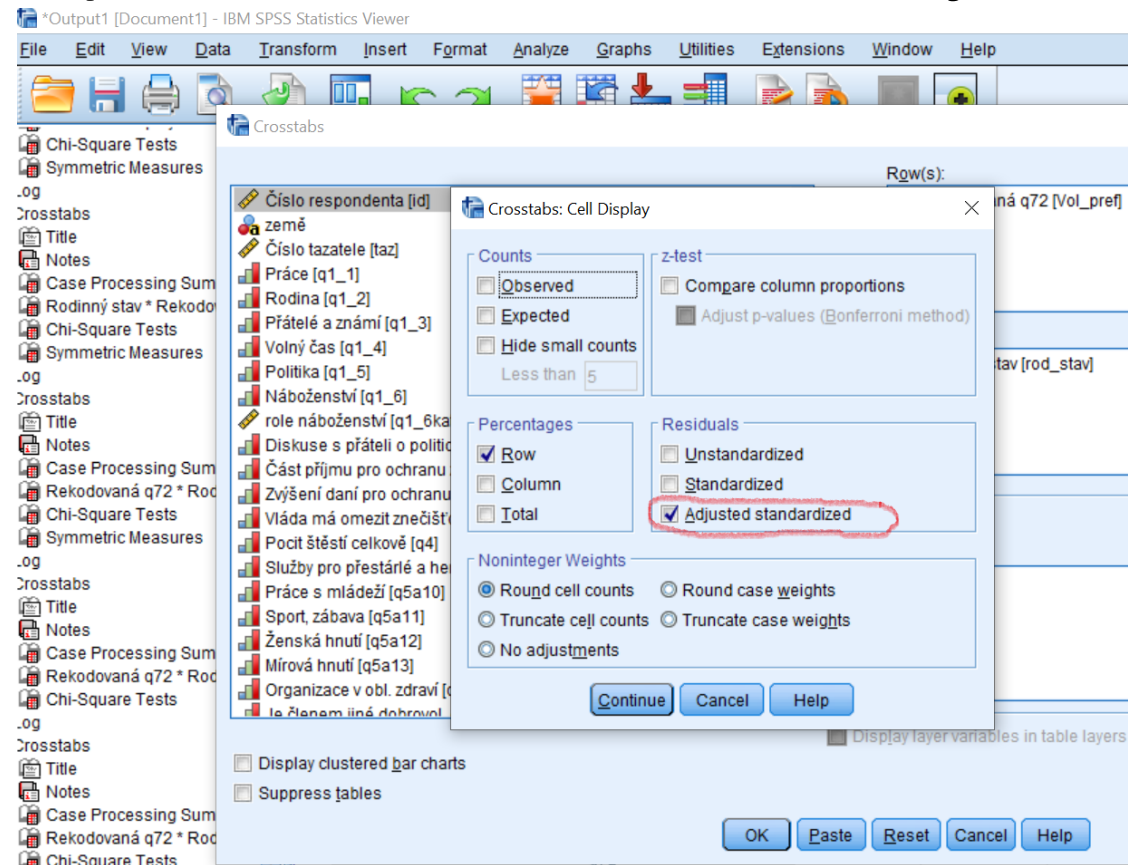


# Třetí krok: Residua

- Spočítání residua – **adjustované** – zobrazují rozdíl mezi pozorovanou a
- rozdíl mezi očekávanou a empirickou (pozorovanou) četností
- **Pokud je jeho hodnota vyšší než „2“ (1,96), rozdíl nevznikl náhodou a jsme si tím s 95% jisti**
- **Analyse – Des.St. – Cross.- Cells**

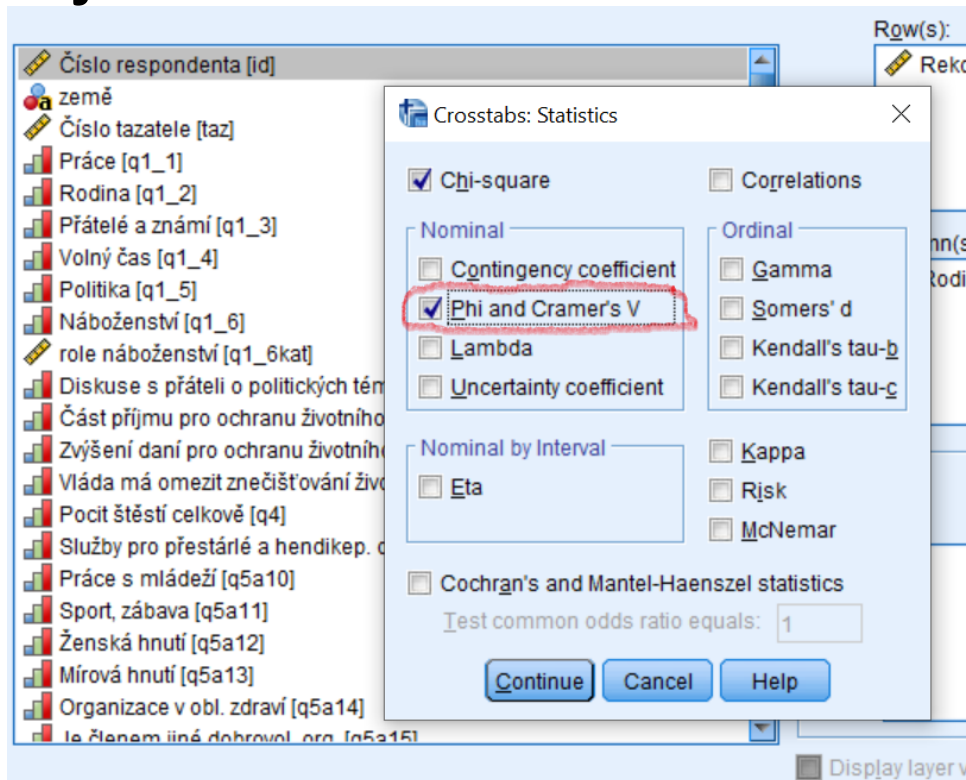
Rekodovaná q72 \* Rodinný stav Crosstabulation

		Rodinný stav						
		ženatý/vdaná	vdovec/ vdova	rozvedený/á	odloučení	svobodný/á	Total	
Rekodovaná q72	KSČM	Count	101	36	11	2	11	161
		Adjusted Residual	-4	5,5	-1	1,6	-4,0	
ČSSD	Count	196	25	14	1	49	285	
		Adjusted Residual	1,8	-8	-1,6	-3	-5	
KDU	Count	58	15	4	0	21	98	
		Adjusted Residual	-1,0	1,8	-1,2	-7	,8	
US	Count	97	9	11	0	45	162	
		Adjusted Residual	-1,2	-2,0	-1	-9	3,3	
ODS	Count	253	29	26	1	66	375	
		Adjusted Residual	1,6	-1,7	-1	-6	-4	
jiná	Count	115	11	10	0	43	179	
		Adjusted Residual	,0	-1,9	-8	-9	2,1	
nevolil by	Count	198	35	36	3	56	328	
		Adjusted Residual	-1,6	,4	3,1	1,5	-7	
Total	Count	1018	160	112	7	291	1588	



# Čtvrtý krok: Koeficient asociace

- **Koeficient  $\phi$**  – kontingenční tabulka má podobu 2x2, vyšší počet kat. než 2
- Sílu závislosti zjistíme koeficientem asociace - **Kramerovo V**.
- je to korelace z nominálních (ordinálních) proměnných. 0 – 1,



**Chi-Square Tests**

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	79,511 <sup>a</sup>	24	,000
Likelihood Ratio	76,707	24	,000
Linear-by-Linear Association	5,038	1	,025
N of Valid Cases	1588		

a. 7 cells (20,0%) have expected count less than 5. The minimum expected count is ,43.

**Symmetric Measures**

		Value	Approximate Significance
Nominal by Nominal	Phi	,224	,000
	Cramer's V	,112	,000
N of Valid Cases		1588	

# Rekapitulace – jak postupovat

- Vytvoříme četnostní tabulku pro obě proměnné, možné sloučení, vynechání, tam kde jsou malé četnosti.
- Zobrazíme kontingenční tabulku ve formátu **sloupcových, řádkových %**
- Vypočítáme chí-kvadrát test a rozhodneme o případné závislosti
- Pokud bude chí - kvadrát významný (0,05), vypočítáme adj. Residua
- Interpretujeme nalezenou závislost
- Popíšeme sílu souvislosti koeficientem asociace – Kramerovo V
- Vzhůru na příklady!

# Odkazy:

- <https://www.youtube.com/watch?v=misMgRRV3jQ>