

Analýza závislosti: nominální a ordinální data

Petr Ocelík

MVZ4003 Úvod do kvantitativních metod

Opakování

- K čemu používáme **t-test**(y)?

Opakování

- Co nám říká **p-hodnota** (statistická významnost)?

Osnova

- Kontingenční tabulky
- Chí-kvadrát (χ^2) test nezávislosti
- Kendallův korelační koeficient tau (t)

Míry závislosti

- Existuje **mnoho měr závislosti** (MZ), korelační koeficienty jsou pouze podmnožinou MZ
- MZ měří **přítomnost** nebo/a **sílu závislosti** mezi proměnnými
- MZ typicky nabývají hodnot v intervalech $\langle 0,1 \rangle$ nebo $\langle -1,1 \rangle$

→ **neomezujeme se na korelaci**, či dokonce Pearsonovo r

- Korelace neimplikuje kauzalitu (více příští týden)
- Kauzalita může být založena na různých typech závislosti

Míry závislosti

úrovně měření	míry
nominální ordinální	chí-kvadrát test nezávislosti
ordinální	Kendallův korelační koeficient tau
metrická	Pearsonův korelační koeficient r

Příklad

- **Inferenční problém:** existuje závislost mezi pohlavím a kouřením?

Kontingenční tabulky

- KT obsahují hodnoty vytvořené **klasifikací** (cross-tabulating) **pozorování** (případů) v závislosti na hodnotách dvou (či více) proměnných

Datová matice

pozorování	pohlaví	kuřák
Pavel	muž	ne
Lukáš	muž	ne
Jana	žena	ano
Sára	žena	ano
Jan	muž	ne
František	muž	ano
Lída	žena	ano
Jarmila	žena	ne

Kontingenční tabulka

	muž	žena	celkem
ano	1 (25 %)	3 (75 %)	4
ne	3 (75 %)	1 (25 %)	4
celkem	4 (100 %)	4 (100 %)	8

Kontingenční tabulky

- **Kontingenční tabulky** slouží k zobrazení vztahu nejméně dvou kategorických proměnných
- Kategorické proměnné: binární, multinominální (> 2 kategorie), ordinální
- **Řádky** reprezentují hodnoty jedné proměnné (např. ne/kuřák)
- **Sloupce** reprezentují hodnoty druhé proměnné (např. pohlaví)
- **Pole** obsahují počet a/nebo podíl pozorování vykazující hodnotu příslušného řádku a sloupce

	muž	žena	celkem
kuřák	1 (25 %)	3 (75 %)	4
nekuřák	3 (75 %)	1 (25 %)	4
celkem	4 (100 %)	4 (100 %)	8

- (Konvence: sloupce reprezentují nezávislou proměnnou, řádky závislou proměnnou.)

		pohlaví		celkem
		muž	žena	
kuřák	ano	1 (25 %)	3 (75 %)	4
	ne	3 (75 %)	1 (25 %)	4
celkem		4 (100 %)	4 (100 %)	8

		proměnná 2		celkem
		1 (ano)	0 (ne)	
proměnná 1	1 (ano)	a	b	a + b
	0 (ne)	c	d	c + d
celkem		a + c	b + d	a + b + c + d

počet případů

okrajové (marginální) **hodnoty**: řádkové a sloupcové součty

Babbie 1997

Intuice

- **Inferenční problém:** existuje závislost mezi pohlavím a kouřením?
- Pokud tato **závislost neexistuje (H_0)**, jaké rozdělení hodnot v tabulce bychom **očekávali**?

	muž	žena	celkem
kuřák	?	?	
nekuřák	?	?	
celkem			

výběr (N=80) zahrnuje:

40 kuřáků, 40 nekuřáků
40 mužů, 40 žen

Očekávané hodnoty

- **Očekávané hodnoty (E):** hodnoty za předpokladu, že proměnné jsou nezávislé
 - Reference vůči pozorovaným hodnotám
 - Výpočet: $E = (\text{součet řádku} * \text{součet sloupce}) / \text{celkový součet}$
- Např.: $E(\text{muži/ano}) = (40 * 40) / 80 = 20$

Kontingenční tabulka (pozorované hodnoty)

	muž	žena	celkem
ano	10	30	40
ne	30	10	40
celkem	40	40	80

Tabulka očekávaných hodnot

	muž	žena	celkem
ano	20	20	40
ne	20	20	40
celkem	40	40	80

Chí-kvadrát test nezávislosti

- **Neparametrický test** (χ^2) hodnotící **nezávislost dvou a více kategorií** na základě srovnání rozdílu jejich **pozorovaných a očekávaných četností**
- $\chi^2 = \frac{\text{signal}}{\text{baseline}}$; $\chi^2 = \frac{\text{pozorované četnosti} - \text{očekávané četnosti}}{\text{očekávané četnosti}}$
- **Očekávané četnosti:** hodnoty, které bychom pozorovali, pokud jsou proměnné vzájemně nezávislé
- **H0:** Proměnná X a proměnná Y *jsou nezávislé*
- **HA:** Proměnná X a proměnná Y *jsou závislé*

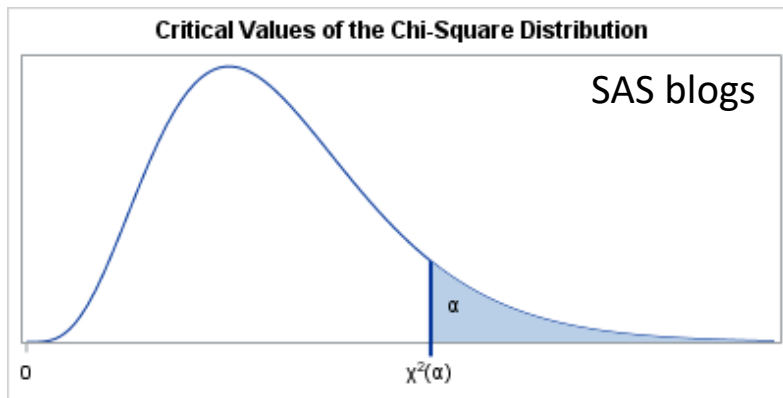


Předpoklady

- Chí-kvadrát měří přítomnost (významnost) závislosti, **nikoli její sílu**
- Přípustná pouze **celá pozitivní čísla** (counts)
- **Striktní:** očekávaná četnost E pro žádnou z buněk by neměla být < 5
- **Relaxovaná:** žádná buňka $E < 2$ a alespoň 80 % buněk s $E > 5$
- **Řešení:**
 - pokud možné a smysluplné, slučování řádků/sloupců
 - Yatesova korekce (Yates' correction for continuity)
 - užití Fisherova exaktního testu

Hypotéza o nezávislosti proměnných

- **H₀**: Proměnná X a proměnná Y *jsou nezávislé*
- **H_A**: Proměnná X a proměnná Y *jsou závislé*
- **Kritická hodnota (KH)** testovací statistiky (χ^2) udává hranici mezi statisticky ne/významným výsledkem testu při zvolené hladině významnosti α
- Testovací statistika $\chi^2 \Rightarrow$ **KH ($\alpha = 0.05$)?**



← vždy užíváme jednosměrný test

- **p-hodnota**: pravděpodobnost, že testovací statistika nabývá pozorované (nebo extrémnější) hodnoty při platnosti H₀

Chí-kvadrát hodnota

- **Chí-kvadrát** χ^2 měří **agregovanou** (celá tabulka) **odchylku pozorovaných** hodnot O od **očekávaných** hodnot E
 - **Agregovaná míra:** nelze říct, do jaké míry přispívají jednotlivé buňky
 - **Citlivost vůči odlehlým hodnotám** (čtverce odchylek)

$$\chi^2 = \sum (O - E)^2 / E$$

	muž	žena	celkem
ano	10	30	40
ne	30	10	40
celkem	40	40	80

Chí-kvadrát hodnota

Výpočet **očekávané hodnoty** (E) = (součet řádku * součet sloupce) / N

$$E_A = (40 * 40) / 80 = 20$$

$$E_B = (40 * 40) / 80 = 20$$

$$E_C = (40 * 40) / 80 = 20$$

$$E_D = (40 * 40) / 80 = 20$$

Kontingenční tabulka

	muž	žena	celkem
ano	A 10	C 30	40
ne	B 30	D 10	40
celkem	40	40	N 80

Tabulka očekávaných hodnot

	muž	žena	celkem
ano	A 20	C 20	40
ne	B 20	D 20	40
celkem	40	40	N 80

Chí-kvadrát hodnota

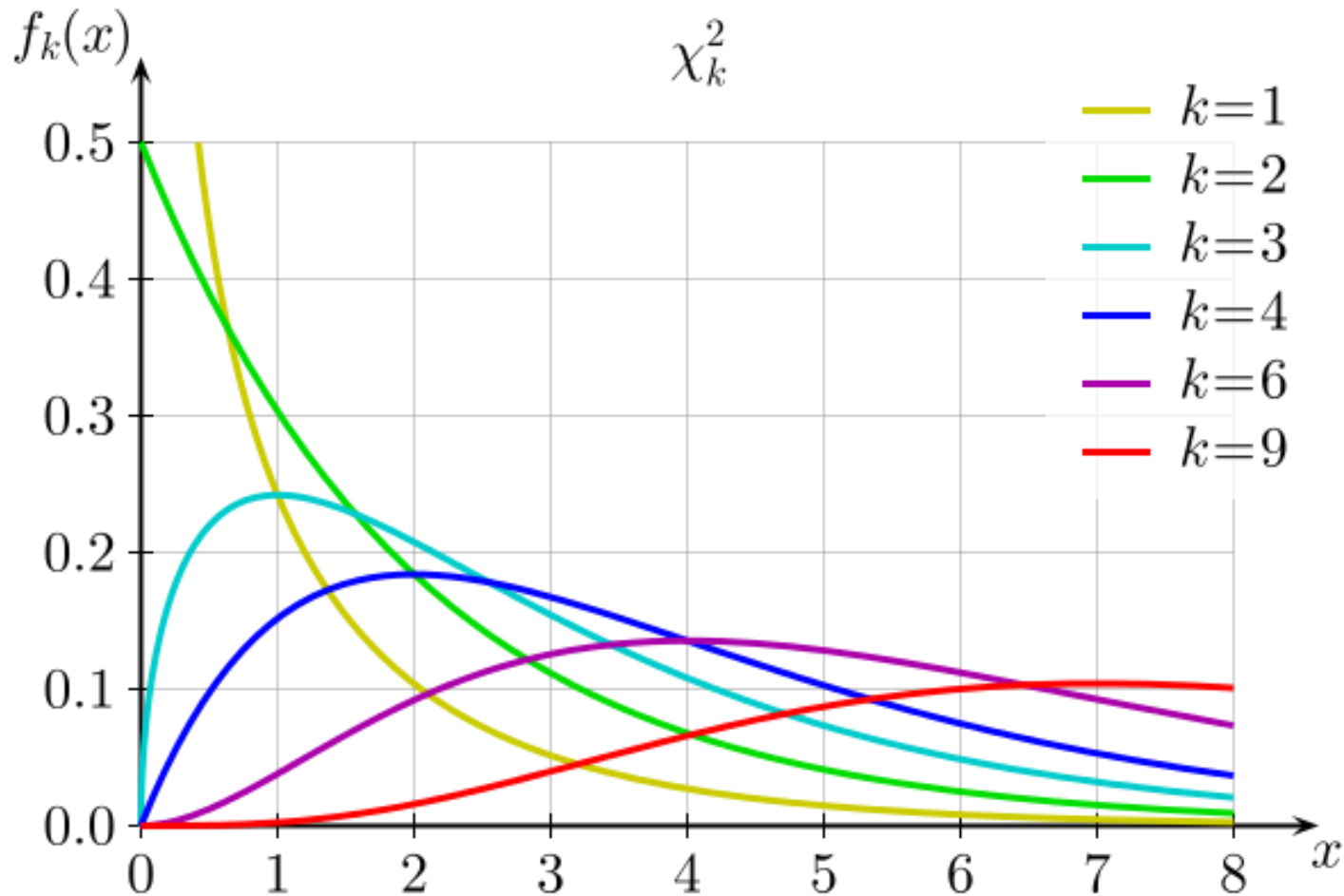
$$\chi^2 = \sum (O - E)^2 / E$$

	muž	žena	celkem
ano	10	30	40
ne	30	10	40
celkem	40	40	80

	muž	žena	celkem
ano	20	20	40
ne	20	20	40
celkem	40	40	80

$$\begin{aligned} & (10 - 20)^2/20 + (30 - 20)^2/20 + (30 - 20)^2/20 + (10 - 20)^2/20 = \\ & \quad 5 + \quad 5 + \quad 5 + \quad 5 = 20 \end{aligned}$$

Chí-kvadrát rozdělení



Stupně volnosti

- **Stupně volnosti (k):** počet hodnot ve výpočtu, které jsou nezávislé
- Pro χ^2 celkový **počet buněk**, jejichž hodnoty se **mohou měnit** aniž by se měnily **okrajové hodnoty**

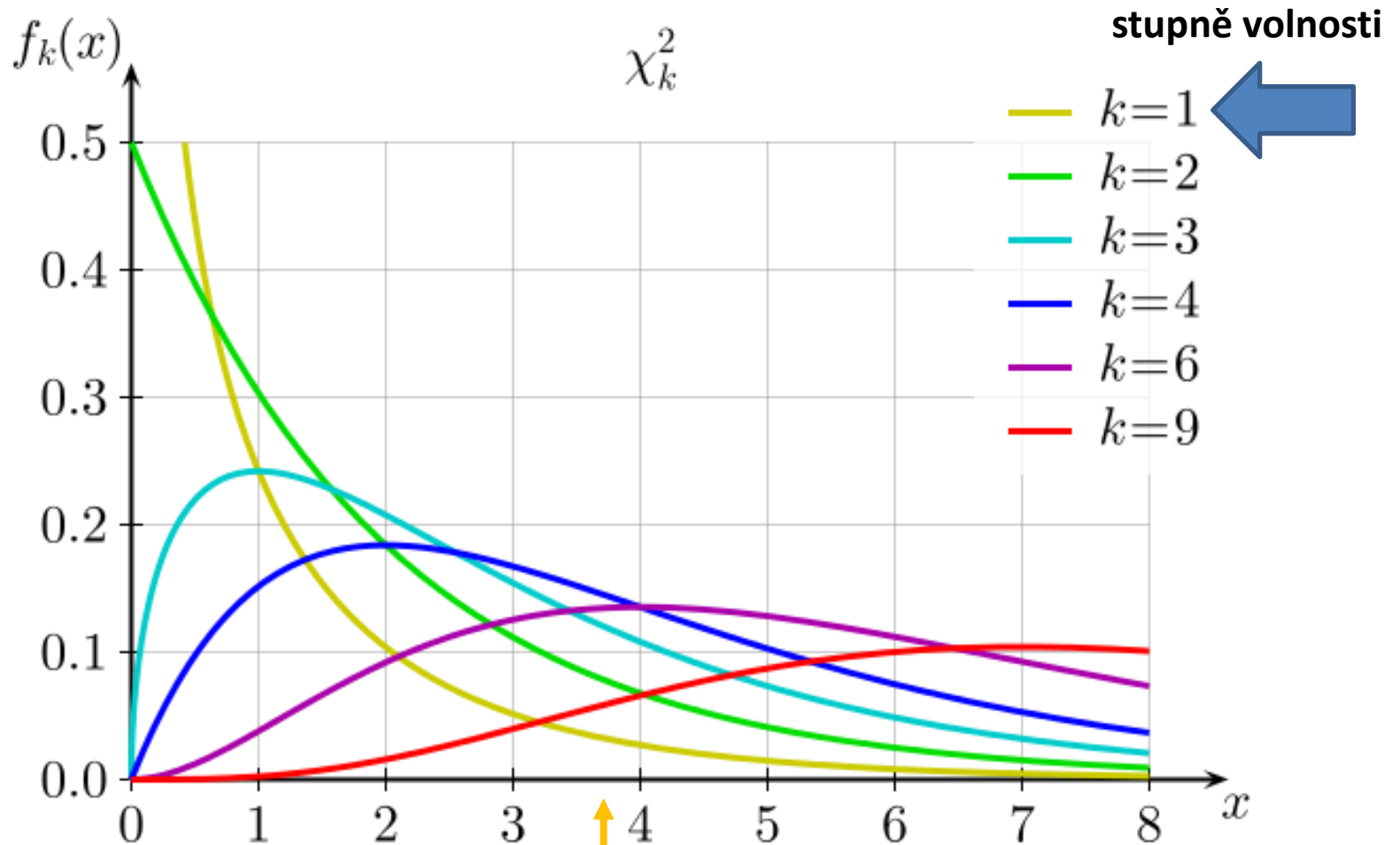
– Výpočet pro 2x2 kontingenční tabulku:

$$k = (\text{počet řádků} - 1) * (\text{počet sloupců} - 1)$$

$$k = (2 - 1) * (2 - 1) = \mathbf{1}$$

v	α					
	0.100	0.050	0.025	0.010	0.005	0.001
1	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
31	41.4217	44.9853	48.2319	52.1914	55.0027	61.0983
63	77.7454	82.5287	86.8296	92.0100	95.6493	103.4424
127	147.8048	154.3015	160.0858	166.9874	171.7961	181.9930
255	284.3359	293.2478	301.1250	310.4574	316.9194	330.5197
511	552.3739	564.6961	575.5298	588.2978	597.0978	615.5149
1023	1081.3794	1098.5208	1113.5334	1131.1587	1143.2653	1168.4972

$p = 0.75$



$\chi^2 = 20 > \text{KH} (\alpha = 0.05) = 3.84$

wikipedia.org

Rozhodnutí o H_0

- **Příklad:** chí-kvadrát hodnota = **20** (2x2 tabulka); sv = **1**
- Kritická hodnota ($\alpha = 0.05$) = **3.84**
- Protože χ^2 hodnota = **20** > ($\alpha = 0.05$) = **3.84**, analogicky
- Protože p-hodnota < hladina významnosti **0.05 (5 %)**
- **Ponecháváme H_0 :** Proměnná A a proměnná B *jsou nezávislé*
- A nepřijímáme **H_A :** Proměnná A a proměnná B *jsou závislé*
- Tj. mezi pohlavím a kouřením existuje souvislost

Síla závislosti

- Chí-kvadrát měří **přítomnost (statistickou významnost)** závislosti, **nikoli její sílu (věcnou významnost)**
- Sílu vztahu je možné určit pomocí **Phi koeficientu (2x2)** či **Cramérova V (> 2x2)**

- **Phi (abs)** = $\sqrt{\frac{\chi^2}{N}}$; N = velikost výběru

Náš příklad:

$$\text{Phi} = \sqrt{20 / 80} = 0.5$$

- **V <0,1>** = $\sqrt{\frac{\chi^2}{N(k-1)}}$; N = velikost výběru, k = min(nrows, ncols)

Phi (abs hodnota)	síla závislosti
0	nulová
0.01-0.19	nicotná
0.20-0.29	slabá
0.30-0.39	středně silná
0.40-0.69	silná
> 0.69	velmi silná

Cramerovo V	síla závislosti
0.01-0.05	nicotná
0.06-0.10	slabá
0.11-0.15	středně silná
0.16-0.25	silná
> 0.25	velmi silná

Míry závislosti

úrovně měření	míry
nominální ordinální	chí-kvadrát test nezávislosti
ordinální	Kendallův korelační koeficient tau
metrická	Pearsonův korelační koeficient r

Kendallův korelační koeficient tau

- **Kendalovo tau** (τ) užívané pro ordinální data (např. Lickertovy škály)
- **Neparametrická** míra závislosti mezi dvěma proměnnými
- Vhodná rovněž pro malé vzorky a vzorky s mnoha hodnotami se stejným pořadím (ranking)
- **Hodnoty v rozmezí $\langle -1, 1 \rangle$**
 - Úplný souhlas (proměnné jsou bez výjimky totožně řazeny) = 1
 - Úplný nesouhlas (proměnné jsou bez výjimky opačně řazeny) = -1
 - Neřazený vztah = 0
- **KT vyjadřuje míru souhlasu (concordance) mezi dvěma ordinálními proměnnými**
 - τ_a nebere v potaz svázaná pořadí (tied values)
 - τ_b bere v potaz svázaná pořadí
- **Např.:** existuje řazená závislost mezi úrovní příjmu (X) a důvěrou v demokracii (Y)?

Hypotéza o závislosti mezi proměnnými

- **H₀**: Korelační koeficient je nulový nebo negativní, *mezi proměnnými X a Y je nulová nebo negativní pořadová závislost*; $\tau \leq 0$
- **H_A**: Korelační koeficient je pozitivní, *mezi proměnným X a Y je pozitivní pořadová závislost*; $\tau > 0$
- **Kritická hodnota** (KH) testovací statistiky (τ) udává hranici mezi statisticky ne/významným výsledkem testu při zvolené hladině významnosti α
- Testovací statistika $\tau \Rightarrow$ **KH ($\alpha = 0.05$)?**
- **p-hodnota**: pravděpodobnost, že testovací statistika nabývá pozorované (nebo extrémnější) hodnoty při platnosti H₀

případy (N)	X: příjem	Y: důvěra
A	1 (nízký)	1 (nízká)
B	2 (střední)	1 (nízká)
C	2 (střední)	2 (střední)
D	3 (vysoký)	3 (vysoká)

- Existuje $n*(n - 1)/2$ neorientovaných párů; tj. $4*(4-1)/2 = 6$
- Tedy: (A,B), (A,C), (A,D), (B,C), (B,D), (C,D)
- **Souhlas (concordance):** $X_i > X_j$ AND $Y_i > Y_j$; or: $X_i < X_j$ AND $Y_i < Y_j$
- **Nesouhlas (discordance):** $X_i > X_j$ AND $Y_i < Y_j$; or: $X_i < X_j$ AND $Y_i > Y_j$
- **Svázaná pořadí (tied values):** $X_i = X_j$ OR $Y_i = Y_j$
 - Pár (A,B) = **svázané pořadí**; $X_A < X_B$ & $Y_A = Y_B$
 - Pár (A,C) = **souhlas**; $X_A < X_C$ & $Y_A < Y_C$
 - Pár (A,D) = **souhlas**; $X_A < X_D$ & $Y_A < Y_D$
 - Pár (B,C) = **svázané pořadí**; $X_B = X_C$ & $Y_B < Y_C$
 - Pár (B,D) = **souhlas**; $X_B < X_D$ & $Y_B < Y_D$
 - Pár (C,D) = **souhlas**; $X_C < X_D$ & $Y_C < Y_D$

případy	X: příjem	Y: důvěra
A	1 (nízký)	1 (nízká)
B	2 (střední)	1 (nízká)
C	2 (střední)	2 (střední)
D	3 (vysoký)	3 (vysoká)

- Máme $n*(n - 1)/2$ neorientovaných párů; tj. $4*(4-1)/2 = 6$
 - Pár (A,B) = **svázané pořadí**; $Y_A = Y_B$
 - Pár (A,C) = souhlas; $X_A < X_C$ & $Y_A < Y_C$
 - Pár (A,D) = souhlas; $X_A < X_D$ & $Y_A < Y_D$
 - Pár (B,C) = **svázané pořadí**; $X_B = X_C$
 - Pár (B,D) = souhlas; $X_B < X_D$ & $Y_B < Y_D$
 - Pár (C,D) = souhlas; $X_C < X_D$ & $Y_C < Y_D$

$\tau_a = (\# \text{ souhlasných párů} - \# \text{ nesouhlasných párů}) / \# \text{ všech párů}$

$$\tau_a = (n_c - n_d) / ((n * (n - 1)) / 2)$$

$$\tau_a = (4 - 0) / ((4 * (4 - 1)) / 2) = 4 / 6 = \mathbf{0.66}$$

- Máme $n*(n - 1)/2$ neorientovaných párů; tj. $4*(4-1)/2 = 6$

- Pár (A,B) = svázané pořadí; $Y_A = Y_B$
- Pár (A,C) = souhlas; $X_A < X_C$ & $Y_A < Y_C$
- Pár (A,D) = souhlas; $X_A < X_D$ & $Y_A < Y_D$
- Pár (B,C) = svázané pořadí; $X_B = X_C$
- Pár (B,D) = souhlas; $X_B < X_D$ & $Y_B < Y_D$
- Pár (C,D) = souhlas; $X_C < X_D$ & $Y_C < Y_D$

	X: příjem	Y: důvěra
A	1 (nízký)	1 (nízká)
B	2 (střední)	1 (nízká)
C	2 (střední)	2 (střední)
D	3 (vysoký)	3 (vysoká)

$$\tau_b = (n_c - n_d) / \text{sqrt}((N - n_1) * (N - n_2))$$

$N = (n * (n - 1))/2$; celkový # párů

$n_1 = t_1 * (t_1 - 1)/2$; $t_1 = \#$ svázaných párů pro první proměnnou

$n_2 = t_2 * (t_2 - 1)/2$; $t_2 = \#$ svázaných párů pro druhou proměnnou

$n_1 = 2 * (2 - 1)/2 = 1$ (příjem: střední/střední)

$n_2 = 2 * (2 - 1)/2 = 1$ (postoj: nesouhlas/nesouhlas)

$$\tau_b = (4 - 0) / \text{sqrt}((6 - 1)*(6 - 1)) = 4 / \text{sqrt}(25) = 4 / 5 = \mathbf{0.8}$$

	Nominal α					
n	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	-	-	-	-
5	0.800	0.800	1.000	1.000	-	-
6	0.600	0.733	0.867	0.867	1.000	-
7	0.524	0.619	0.714	0.810	0.905	1.000
8	0.429	0.571	0.643	0.714	0.786	0.857
9	0.389	0.500	0.556	0.667	0.722	0.833
10	0.378	0.467	0.511	0.600	0.644	0.778
11	0.345	0.418	0.491	0.564	0.600	0.709
12	0.303	0.394	0.455	0.545	0.576	0.667
13	0.308	0.359	0.436	0.513	0.564	0.641
14	0.275	0.363	0.407	0.473	0.516	0.604
15	0.276	0.333	0.390	0.467	0.505	0.581
16	0.250	0.317	0.383	0.433	0.483	0.567
17	0.250	0.309	0.368	0.426	0.471	0.544
18	0.242	0.294	0.346	0.412	0.451	0.529
19	0.228	0.287	0.333	0.392	0.439	0.509
20	0.221	0.274	0.326	0.379	0.421	0.495
21	0.210	0.267	0.314	0.371	0.410	0.486
22	0.203	0.264	0.307	0.359	0.394	0.472
23	0.202	0.257	0.296	0.352	0.391	0.455
24	0.196	0.246	0.290	0.341	0.377	0.449
25	0.193	0.240	0.287	0.333	0.367	0.440
26	0.188	0.237	0.280	0.329	0.360	0.428
27	0.179	0.231	0.271	0.322	0.356	0.419
28	0.180	0.228	0.265	0.312	0.344	0.413
29	0.172	0.222	0.261	0.310	0.340	0.404

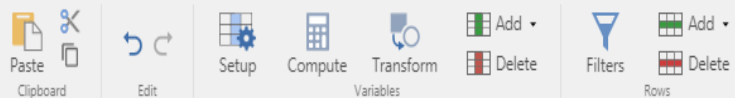
- n = počet pozorovaných párů
- Pro **dvoustranné testy** = $\alpha/2$

Rozhodnutí o H_0

- **Příklad:** Kendallovo tau $\tau_b = 0.8$ ($n = 6$ pozorovaných párů)
- Kritická hodnota ($\alpha = 0.05$) = **0.73**
- Protože τ_b hodnota = **0.8** > KH ($\alpha = 0.05$) = **0.73**, analogicky
- **Odmítáme H_0 :** Korelační koeficient je nulový nebo negativní, *pořadová závislost mezi proměnnou X a Y není nebo je negativní*;
 H_0 : $\tau \leq 0$
- A přijímáme **H_A :** Korelační koeficient je pozitivní, *mezi proměnnou X a Y existuje pozitivní pořadová závislost*; **H_A :** $\tau > 0$
- Tj. mezi úrovní příjmů a úrovní důvěry v demokracii je souvislost

Transformace proměnných

Data Analyses



TRANSFORMED VARIABLE

● TRANSFORM used by 1

religiosity3_tsf

ordinal Variable suffix

+ Add recode condition

- f_x if \$source == "Low" use 1 × ↑
- f_x if \$source == "Mid" use 2 × ↓
- f_x else use 3

Measure type Ordinal

	relig_prot	relig_high	relig_low	religiosity3	religiosity...	romney20
1	50.0	31.3	39.3	Low	1	5
2	79.3	55.7	14.3	High	3	6
3	78.6	52.3	18.3	High	3	6
4	43.3	36.6	33.3	Mid	2	5
5	37.8	34.5	36.3	Low	1	5
6	46.1	33.5	39.3	Low	1	4
7	32.1	30.5	40.3	Low	1	4
8	52.0	35.2	33.3	Mid	2	5
9	51.2	37.6	30.3	Mid	2	4
10	70.3	47.9	20.3	High	3	5
11	42.2	31.4	40.3	Low	1	2
12	58.1	41.3	31.3	Mid	2	4
13	39.6	45.1	34.3	High	3	6
14	47.9	38.0	32.3	Mid	2	4
15	61.8	42.7	28.3	Mid	2	5
16	61.4	45.1	26.3	High	3	5
17	67.4	45.4	24.3	High	3	6
18	56.2	53.3	17.3	High	3	5
19	27.4	26.5	46.3	Low	1	5

Chí-kvadrát test nezávislosti v Jamovi



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R



Modules

	abort_ran...	abortion_...	adv_or_n	x12	cig_tax1	
1	Less restr	35		2.000	HiTax	
2	Mid	20		0.425	LoTax	
3	More restr	4		1.150	MidTax	
4	More restr	5		2.000	HiTax	
5	Less restr	49	1	0.870	MidTax	
6	Mid	25	1	0.840	MidTax	
7	Less restr	45	1	3.400	HiTax	
8	Mid	30	1	1.600	MidTax	
9	Mid	26	1	1.339	MidTax	
10	More restr	9		0.370	LoTax	
11	Less restr	42		3.200	HiTax	
12	Less restr	37		1.360	MidTax	
13	Mid	22		0.570	LoTax	
14	Less restr	36	11.7	30.0	1.980	HiTax
15	More restr	7	8.1	22.5	0.995	MidTax
16	More restr	11	10.2	29.5	0.790	LoTax
17	More restr	17	8.5	21.0	0.600	LoTax
18	More restr	1	6.9	21.4	0.360	LoTax
19	Less restr	40	16.4	38.2	2.510	HiTax
20	Less restr	43	16.0	35.7	2.000	HiTax
21	Mid	31	9.6	26.9	2.000	HiTax
22	Mid	18	9.4	24.6	2.000	HiTax
23	Mid	28	10.3	31.5	1.600	MidTax
24	More restr	8	9.5	25.2	0.170	LoTax
25	More restr	15	7.1	19.6	0.680	LoTax
26	Less restr	41	8.3	27.4	1.700	MidTax
27	Mid	27	8.8	26.5	0.450	LoTax
28	More restr	12	6.7	25.8	0.440	LoTax
29	More restr	6	8.8	27.4	0.640	LoTax
30	Mid	32	11.2	32.0	1.680	MidTax
31	Less restr	46	12.9	34.5	2.700	HiTax
32	Less restr	38	10.4	25.3	1.660	MidTax
33	Less restr	39	7.6	21.8	0.800	LoTax

One Sample Proportion Tests

2 Outcomes

Binomial test

N Outcomes

 χ^2 Goodness of fit

Contingency Tables

Independent Samples

 χ^2 test of association

Paired Samples

McNemar test

Log-Linear Regression



Data

Analyses



Exploration



T-Tests



ANOVA



Regression



Frequencies




Factor



R




Modules

Contingency Tables 

- pop_18_24_10
- prcapinc
- region
- relig_import
- religiosity
- reppct_m
- rtw
- secularism
- secularism3
- seniority_sen2
- to_0004
- to_0408
- trnout00



Rows

obama_win12 

Columns

south 

Counts (optional)



Layers

> Statistics

> Cells

 χ^2 Tests

	Value	df	p
χ^2	.	.	.
N	.		

Contingency Tables

Contingency Tables

obama_win12	south		Total
	Nonsouth	South	
No	12	12	24
Yes	22	4	26
Total	34	16	50

 χ^2 Tests

	Value	df	p
χ^2	6.87	1	0.009
N	50		

References

[1] The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.

[2] R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.

Contingency Tables



region
 relig_import
 religiosity
 repct_m
 rtw
 secularism
 secularism3
 seniority_sen2
 to_0004
 to_0408
 tmout00
 tmout04
 unemploy

Rows
 obama_win12

Columns
 south

Counts (optional)

Layers

> Statistics

∨ Cells

Counts

- Observed counts
 Expected counts

Percentages

- Row
 Column
 Total

Contingency Tables

Contingency Tables

obama_win12		south		Total
		Nonsouth	South	
No	Observed	12	12	24
	Expected	16.3	7.68	
	% within column	35.3 %	75.0 %	
	% of total	24.0 %	24.0 %	
Yes	Observed	22	4	26
	Expected	17.7	8.32	
	% within column	64.7 %	25.0 %	
	% of total	44.0 %	8.0 %	
Total	Observed	34	16	50
	Expected	34.0	16.00	
	% within column	100.0 %	100.0 %	
	% of total	68.0 %	32.0 %	

 χ^2 Tests

	Value	df	p
χ^2	6.87	1	0.009
N	50		














References

[1] The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.


[2] R Core Team (2019). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://www.r-project.org/>

Contingency Tables




-  religiosity
-  reppct_m
-  rtw
-  secularism
-  secularism3
-  seniority_sen2
-  to_0004
-  to_0408
-  trnout00
-  trnout04
-  unemploy
-  union04
-  union07

Rows

→ obama_win12 

Columns

→ south 

Counts (optional)

→ 

Layers

→

Statistics

Tests

- χ^2
- χ^2 continuity correction
- Likelihood ratio
- Fisher's exact test

Nominal

- Contingency coefficient
- Phi and Cramer's V

Comparative Measures (2x2 only)

- Log odds ratio
- Odds ratio
- Relative risk
- Confidence intervals

Interval %

Ordinal

- Gamma
- Kendall's tau-b

> Cells

Contingency Tables

Contingency Tables

		south		
obama_win12	Nonsouth	South	Total	
No	12	12	24	
Yes	22	4	26	
Total	34	16	50	

 χ^2 Tests

	Value	df	p
χ^2	6.87	1	0.009
χ^2 continuity correction	5.37	1	0.020
Likelihood ratio	7.09	1	0.008
Fisher's exact test	0.189		0.015
N	50		

[3]

Nominal

	Value
Contingency coefficient	0.348
Phi-coefficient	0.371
Cramer's V	0.371

References

Kendallovu tau v Jamovi



Data

Analyses



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R



Modules

DATA VARIABLE

abortion_rank12

Description

 Continuous Ordinal Nominal ID

Data type Integer

Levels

Retain unused levels

	abortion_rank12	adv_or_more	ba_or_more	cig_tax12	cig_tax1
1	Less restr	35	9.0	26.6	2.000
2	Mid	20	7.7	22.0	0.425
3	More restr	4	6.1	18.9	1.150
4	More restr	5	9.3	25.6	2.000
5	Less restr	49	10.7	29.9	0.870
6	Mid	25	12.7	35.9	0.840
7	Less restr	45	15.5	35.6	3.400
8	Mid	30	11.4	28.7	1.600
9	Mid	26	9.0	25.3	1.339
10	More restr	9	9.9	27.5	0.370
11	Less restr	42	9.9	29.6	3.200
12	Less restr	37	7.4	25.1	1.360
13	Mid	22	7.5	23.9	0.570
14	Less restr	36	11.7	30.6	1.980
15	More restr	7	8.1	22.5	0.995
16	More restr	11	10.2	29.5	0.790
17	More restr	17	8.5	21.0	0.600
18	More restr	1	6.9	21.4	0.360
19	Less restr	40	16.4	38.2	2.510

Data **Analyses**

Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R



Modules

DATA VARIABLE

gun_rank11

Description

 Continuous Ordinal Nominal ID

Data type Integer

Levels

1

2

3

4

5

Retain unused levels

m_adva...	govt_wor...	gun_rank3	gun_rank11	gun_scale...	hr_cons_r...	hr.
1	12.2	28.0	Less restr	50	0	200.000
2	14.6	17.5	Mid	17	14	151.714
3	-1.4	17.6	Less restr	39	4	132.500
4	-3.5	15.5	Less restr	50	0	155.571
5	14.9	14.9	More restr	1	81	274.288
6	-2.4	15.7	More restr	15	15	161.571
7	17.0	15.9	More restr	5	58	338.200
8	15.9	16.1	Mid	18	13	273.000
9	4.5	14.5	Less restr	41	3	165.320
10	-4.7	17.8	Mid	22	8	180.077
11	24.0	27.8	More restr	6	50	377.500
12	4.3	14.0	Mid	25	7	227.800
13	27.6	15.6	Less restr	47	2	155.500
14	16.7	14.1	More restr	9	35	245.474
15	-4.3	11.5	Less restr	39	4	140.778
16	16.3	17.2	Less restr	39	4	43.000
17	1.5	17.5	Less restr	47	2	197.333
18	-1.3	19.9	Less restr	47	2	126.286
19	20.0	15.8	More restr	3	65	372.400



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R











Modules



	abort_ran...	abortion...	Correlation Matrix	_more	cig_tax12	cig_tax1
1	Less restr	35	Linear Regression	26.6	2.000	HiTax
2	Mid	20	Logistic Regression	22.0	0.425	LoTax
3	More restr	4	2 Outcomes	18.9	1.150	MidTax
4	More restr	5	Binomial	25.6	2.000	HiTax
5	Less restr	49	N Outcomes	29.9	0.870	MidTax
6	Mid	25	Multinomial	35.9	0.840	MidTax
7	Less restr	45	Ordinal Outcomes	35.6	3.400	HiTax
8	Mid	30		28.7	1.600	MidTax
9	Mid	26		25.3	1.339	MidTax
10	More restr	9	9.9	27.5	0.370	LoTax
11	Less restr	42	9.9	29.6	3.200	HiTax
12	Less restr	37	7.4	25.1	1.360	MidTax
13	Mid	22	7.5	23.9	0.570	LoTax
14	Less restr	36	11.7	30.6	1.980	HiTax
15	More restr	7	8.1	22.5	0.995	MidTax
16	More restr	11	10.2	29.5	0.790	LoTax
17	More restr	17	8.5	21.0	0.600	LoTax
18	More restr	1	6.9	21.4	0.360	LoTax
19	Less restr	40	16.4	38.2	2.510	HiTax
20	Less restr	43	16.0	35.7	2.000	HiTax
21	Mid	31	9.6	26.9	2.000	HiTax
22	Mid	18	9.4	24.6	2.000	HiTax
23	Mid	28	10.3	31.5	1.600	MidTax
24	More restr	8	9.5	25.2	0.170	LoTax
25	More restr	15	7.1	19.6	0.680	LoTax
26	Less restr	41	8.3	27.4	1.700	MidTax
27	Mid	27	8.8	26.5	0.450	LoTax
28	More restr	12	6.7	25.8	0.440	LoTax
29	More restr	6	8.8	27.4	0.640	LoTax
30	Mid	32	11.2	32.0	1.680	MidTax
31	Less restr	46	12.9	34.5	2.700	HiTax
32	Less restr	38	10.4	25.3	1.660	MidTax
33	Less restr	39	7.6	21.8	0.800	LoTax

Correlation Matrix



-  adv_or_more
-  ba_or_more
-  cig_tax12
-  conserv_advantage
-  conserv_public
-  dem_advantage
-  govt_worker
-  hr_cons_rank11



-  abortion_rank12
-  gun_rank11

Correlation Coefficients

- Pearson
 Spearman
 Kendall's tau-b

Additional Options

- Report significance
 Flag significant correlations
 Confidence intervals
 Interval %

Hypothesis

- Correlated
 Correlated positively
 Correlated negatively

Plot

- Correlation matrix
 Densities for variables
 Statistics

Correlation Matrix

Correlation Matrix

		abortion_rank12	gun_rank11
abortion_rank12	Kendall's Tau B	—	—
	p-value	—	—
gun_rank11	Kendall's Tau B	-0.353	—
	p-value	< .001	—

Note. H_a is negative correlation

References

- [1] The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.

Seminář