

Analýza závislosti

Petr Ocelík

MVZ4003 Úvod do kvantitativních metod

16. listopadu 2021

Osnova

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

Míry závislosti

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

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

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

Míry závislosti

úroveň měření	míry
nominální	chí-kvadrát test nezávislosti
ordinální	chí-kvadrát test nezávislosti Kendallův korelační koeficient tau
spojitá	Pearsonův korelační koeficient ró

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	5 (33 %)	15 (43 %)	20
nekuřák	10 (67 %)	20 (57 %)	30
celkem	15 (100 %)	35 (100 %)	50

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

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

		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

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.: $E(\text{muži/ano}) = (4 * 4) / 8 = 2$

Kontingenční tabulka

	muži	ženy	celkem
ano	1	3	4
ne	3	1	4
celkem	4	4	8

Tabulka očekávaných hodnot

	muži	ženy	celkem
ano	2	2	4
ne	2	2	4
celkem	4	4	8

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**
- Sílu vztahu je možné určit pomocí **Phi koeficientu** či **Cramérova V**

df^*	<i>small</i>	<i>medium</i>	<i>large</i>
1	.10	.30	.50
2	.07	.21	.35
3	.06	.17	.29
4	.05	.15	.25
5	.04	.13	.22



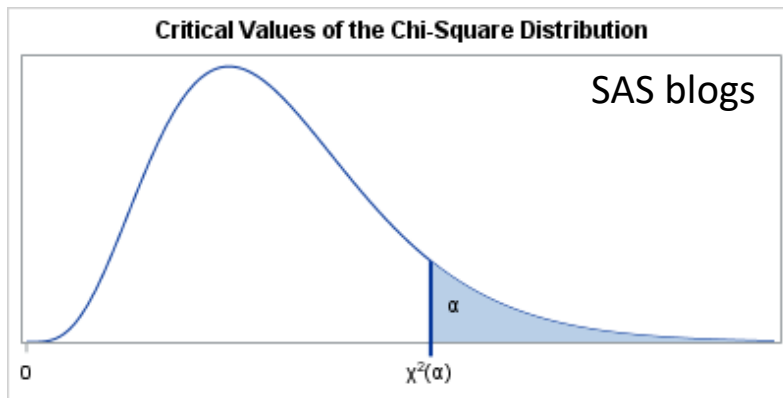
2 x 2 tabulka

Zaointz 2000

- 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)

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

	muži	ženy	celkem
ano	5	15	20
ne	10	20	30
celkem	15	35	50

Chí-kvadrát hodnota

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

$$E_A = (20 * 15) / 50 = 6$$

$$E_B = (30 * 15) / 50 = 9$$

$$E_C = (20 * 35) / 50 = 14$$

$$E_D = (30 * 35) / 50 = 21$$

Kontingenční tabulka

	muži	ženy	celkem
ano	A 5	C 15	20
ne	B 10	D 20	30
celkem	15	35	N 50

Tabulka očekávaných hodnot

	muži	ženy	celkem
ano	A 6	C 14	20
ne	B 9	D 21	30
col-total	15	35	N 50

Chí-kvadrát hodnota

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

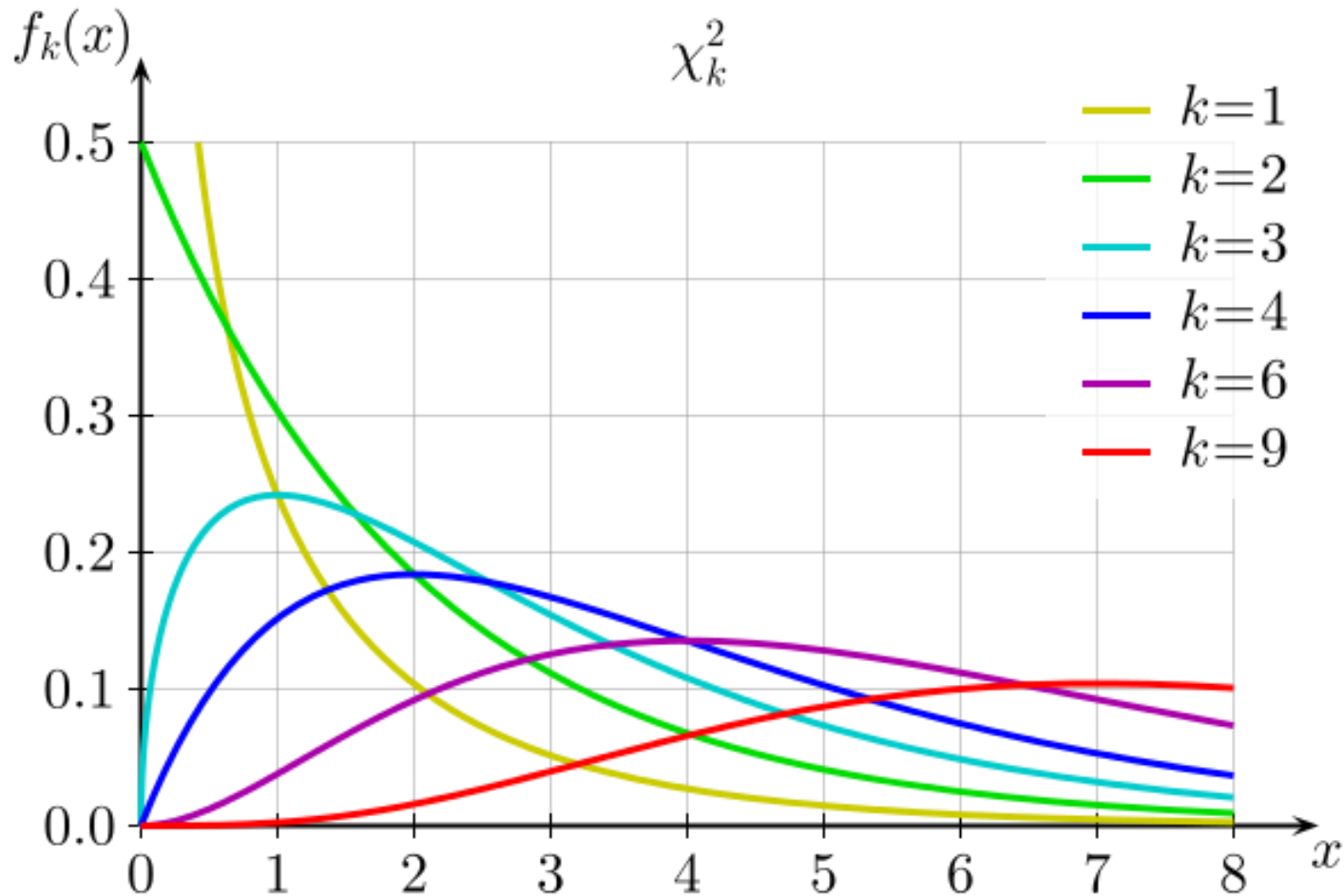
	muži	ženy	celkem
ano	5	15	20
ne	10	20	30
celkem	15	35	50

	muži	ženy	celkem
ano	6	14	20
ne	9	21	30
celkem	15	35	50

$$(5 - 6)^2/6 + (10 - 9)^2/9 + (15 - 14)^2/14 + (20 - 21)^2/21 =$$

$$0.167 + 0.111 + 0.071 + 0.047 = \mathbf{0.39}$$

Chí-kvadrát rozdělení



Stupně volnosti

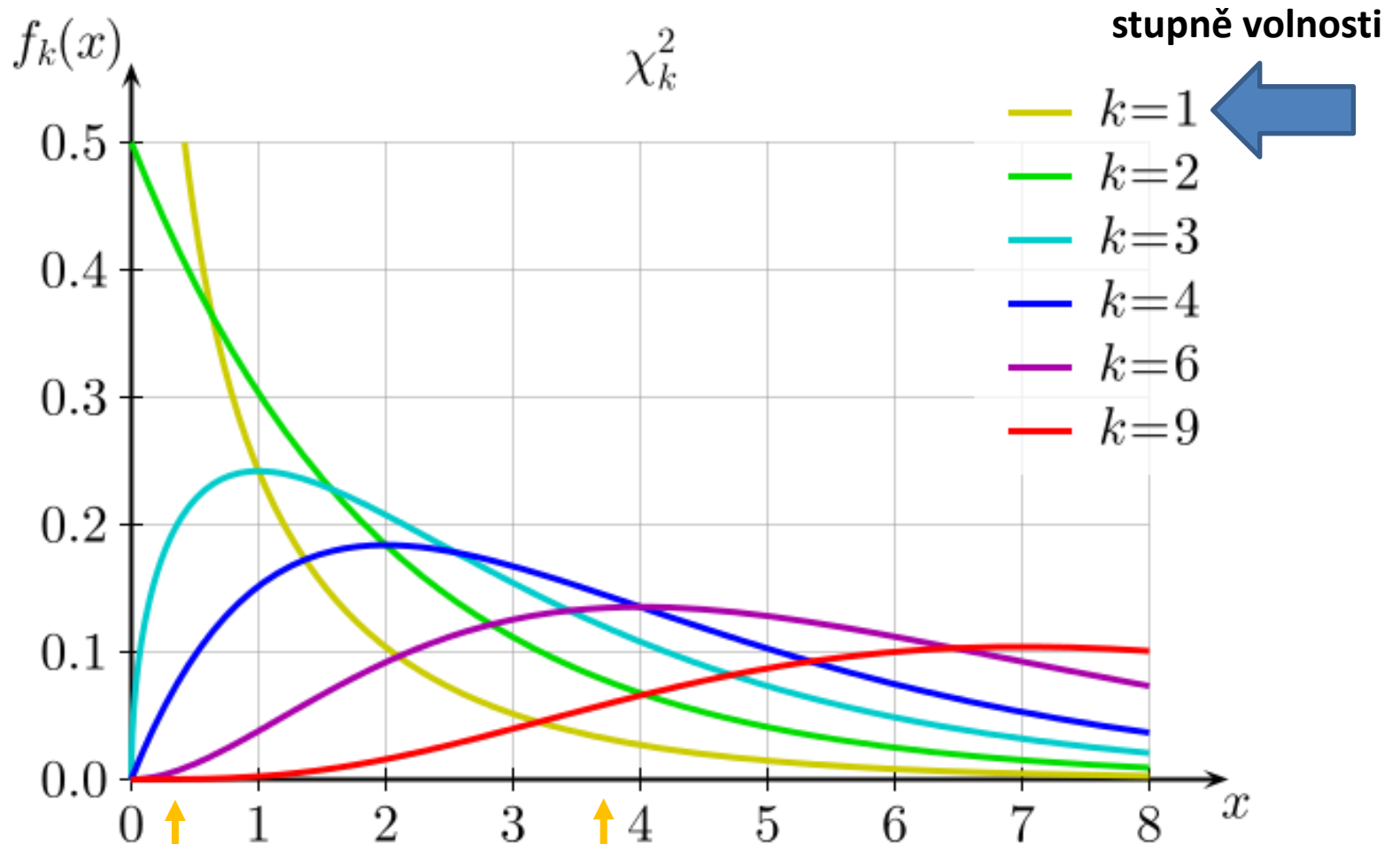
- **Stupně volnosti (k):** počet hodnot ve výpočtu, které jsou nezávislé
- Výpočet pro 2x2 kontingenční tabulku:

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

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

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

$p = 0.75$



$\chi^2 = 0.39 < KH (\alpha = 0.05) = 3.84$

wikipedia.org

Rozhodnutí o H_0

- **Příklad:** chí-kvadrát hodnota = **0.39** (2x2 tabulka); $sv = 1$
- Kritická hodnota ($\alpha = 0.05$, $sv = 1$) = **3.84**
- Protože χ^2 hodnota = **0.39** < **3.84** ($\alpha = 0.05$, $sv = 1$), analogicky
- Protože **p-hodnota = 0.75 (75 %) > 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 není souvislosti

Míry závislosti

úroveň měření	míry
nominální	chí-kvadrát test nezávislosti
ordinální	chí-kvadrát test nezávislosti Kendallův korelační koeficient tau
spojitá	Pearsonův korelační koeficient ró

Kendallův korelační koeficient tau

- **Kendalovo tau** (τ) užívané pro ordinální data (např. Lickertovy škály)
- **Neparametrická** míra závislosti mezi dvěma proměnnými
- Vhodná rovněž pro malé výběry a výběry 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 ($\alpha = 0.05$, $n = 6$) = **0.6**
- Protože τ_b hodnota = **0.8 > 0.6** KH ($\alpha = 0.05$, $n = 6$), 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 řazená souvislost

Kendallův korelační koeficient tau

hodnota Kendallova tau	interpretace
0.07 (0.07-0.20)	slabá
0.21 (0.21-0.34)	střední
0.35 (≥ 0.35)	silná

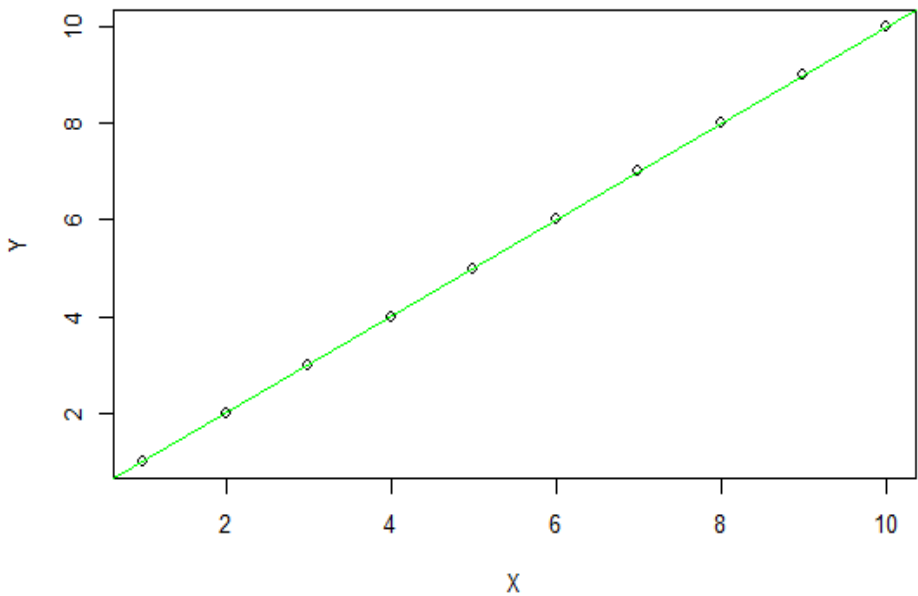
Míry závislosti

úroveň měření	míry
nominální	chí-kvadrát test nezávislosti
ordinální	chí-kvadrát test nezávislosti Kendallův korelační koeficient tau
spojitá	Pearsonův korelační koeficient ró

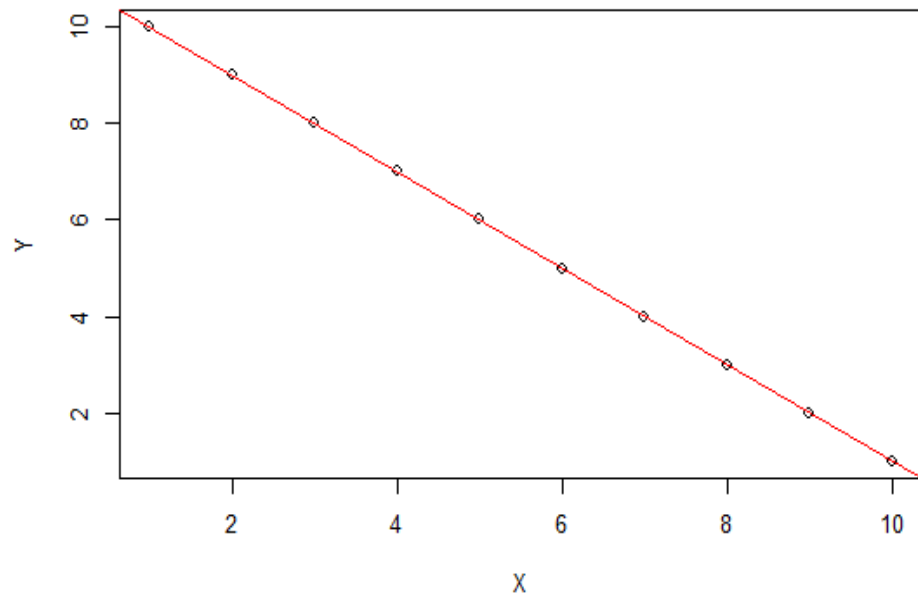
Pearsonův korelační koeficient r

- Pearsonův korelační koeficient (r)
- Pearsonovo r měří **sílu a směr lineární závislosti mezi dvěma spojitými proměnnými**
- Pearsonovo r nabývá hodnot $\langle -1, 1 \rangle$
 - Dokonale pozitivní lineární vztah = 1
 - Dokonale negativní lineární vztah = -1
 - Žádný lineární vztah = 0
- Hodnota r není závislá na jednotkách proměnných

r=1

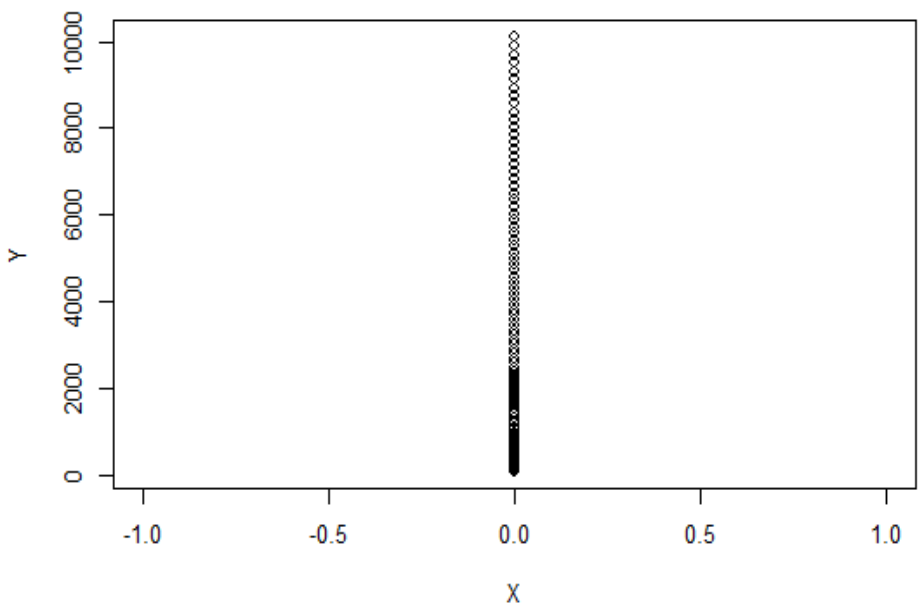


r=-1

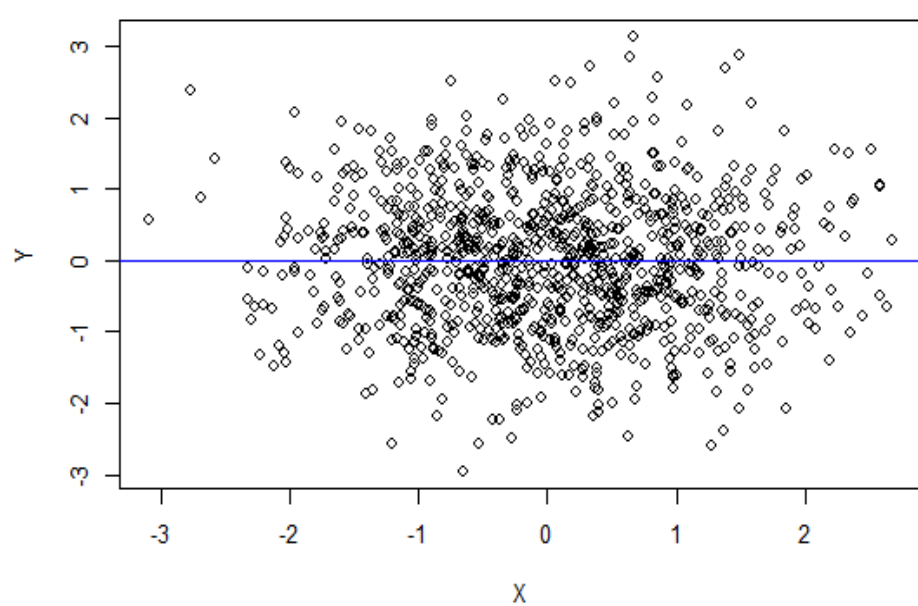


<http://guessthecorrelation.com/>

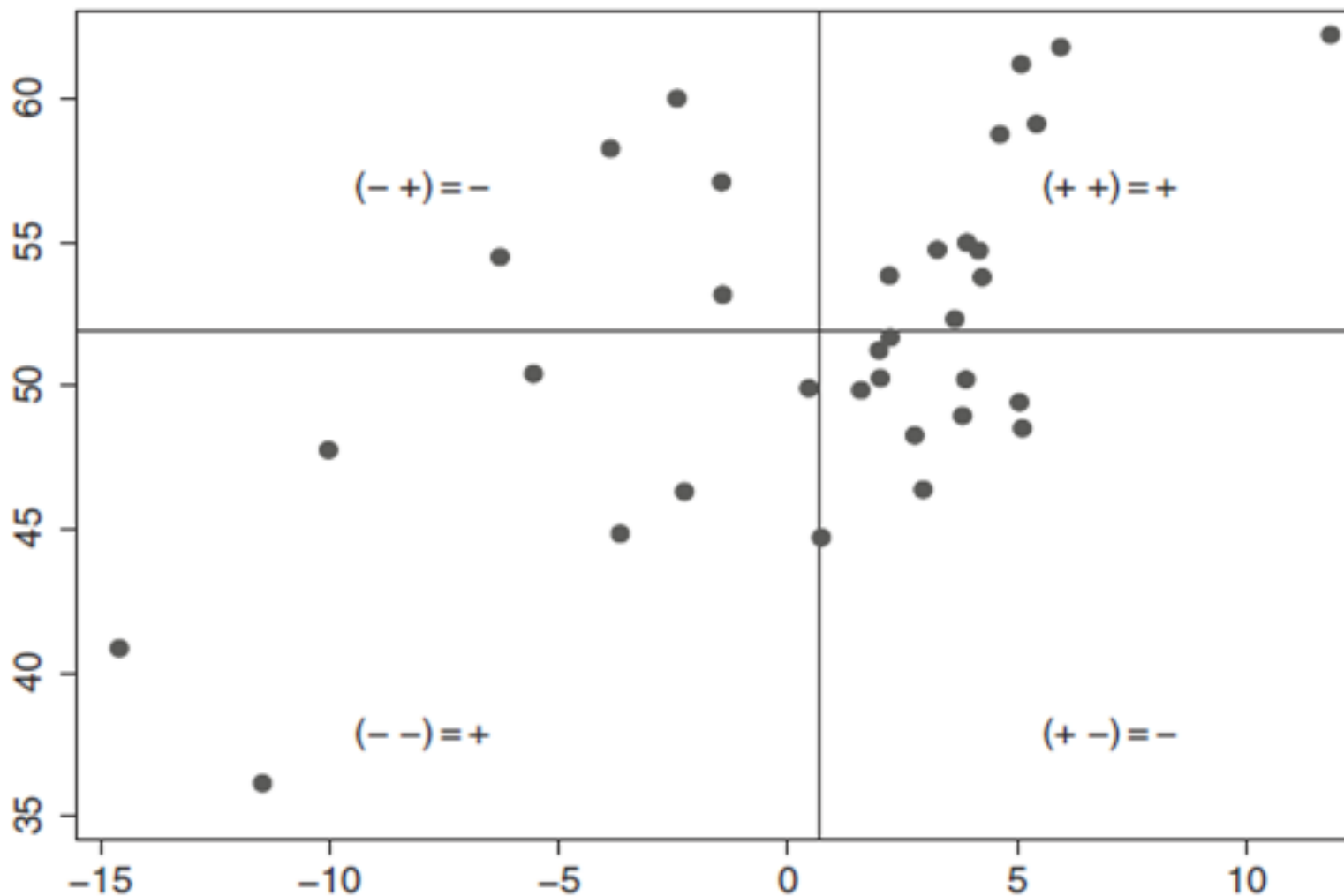
r=0

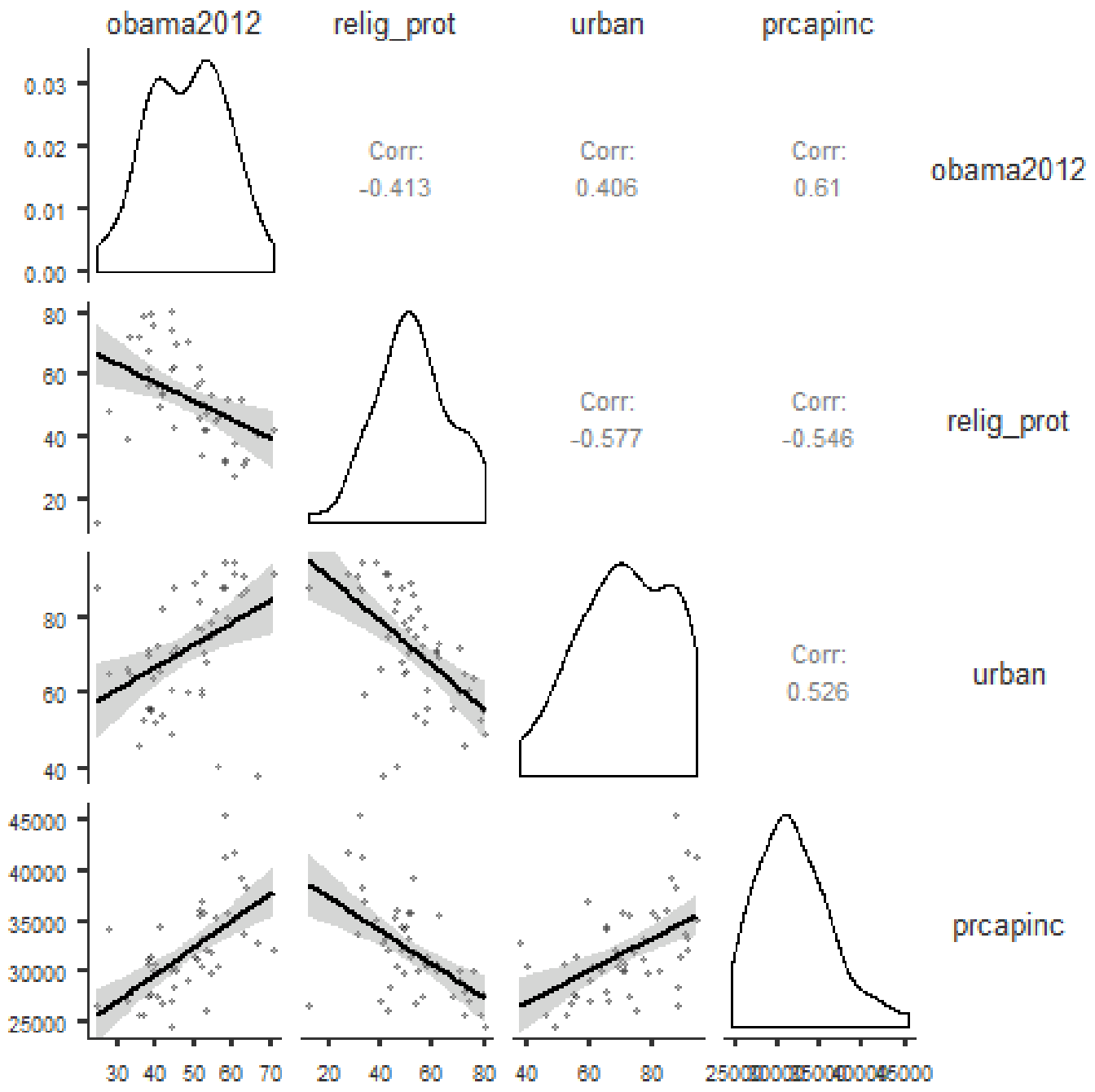


r=0

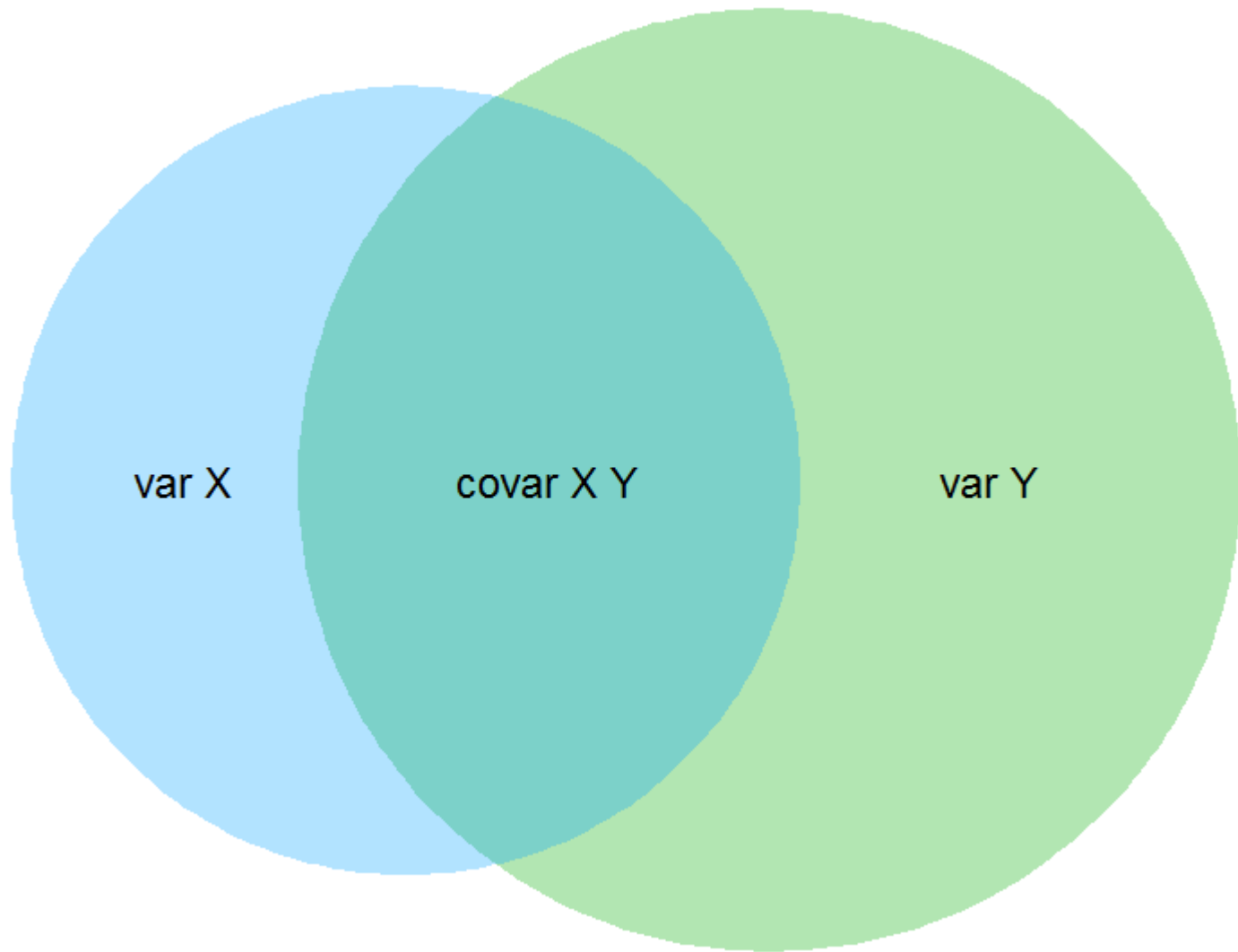


- Statisticky významná korelace = 0.574

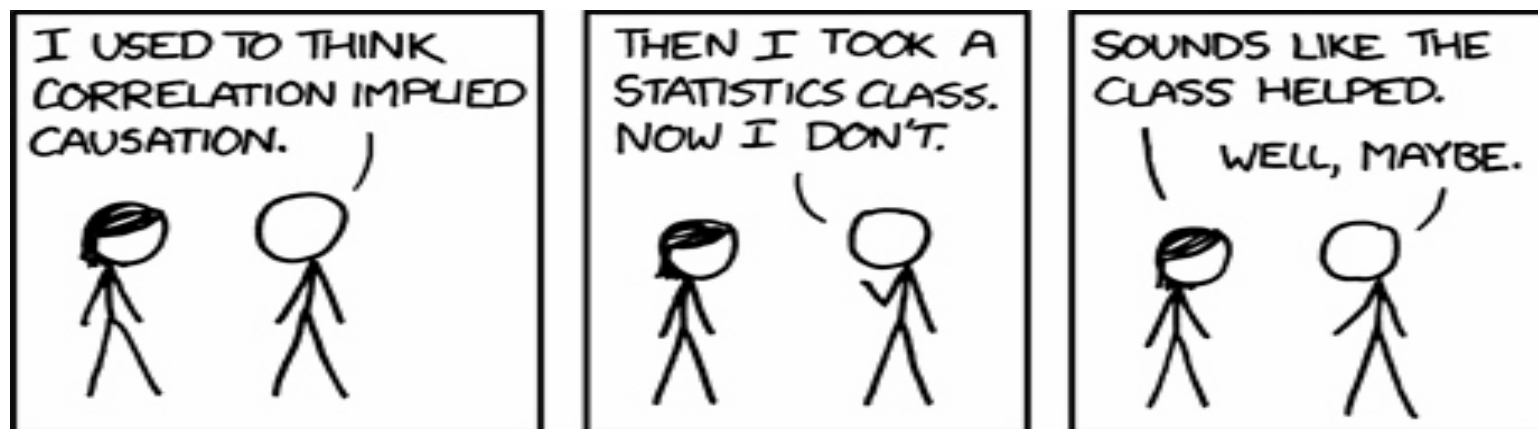




- $r = \text{kovariance } X, Y / \text{kombinovaný rozptyl } X, Y$



Korelace a kauzalita



- Korelace je nezbytnou (v SAD) nikoli však postačující podmínkou kauzality (viz Kellstedt and Whitten 2013)

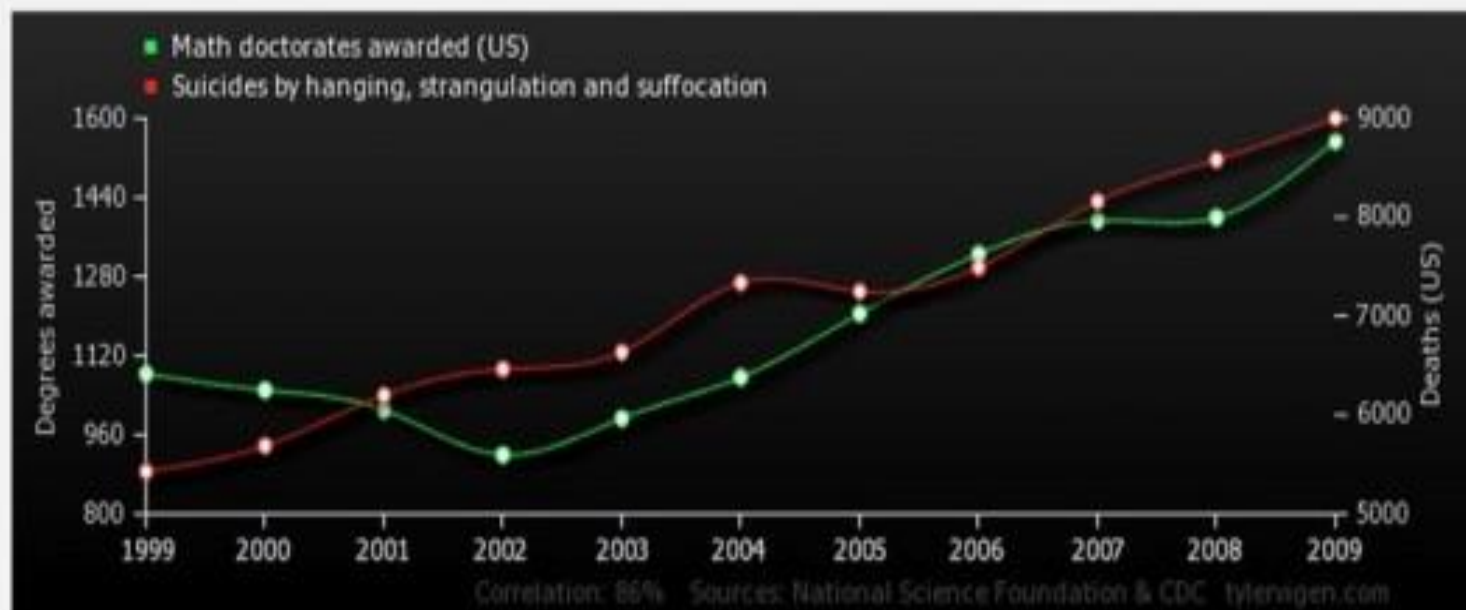
Obecné vzorce

1. X způsobuje Y a Y způsobuje X (**obousměrný vztah**):
 - Vyšší úroveň demokracie (X) zvyšuje ekonomický růst (Y) a vyšší ekonomický růst (Y) zvyšuje úroveň demokracie (X)
 2. Y způsobuje X (**obrácený vztah**):
 - Vyšší ekonomický růst (Y) zvyšuje úroveň demokracie (X)
 3. Vztah X a Y jako důsledek **společné příčiny**:
 - Korelace mezi vyšším ekonomickým růstem (X) a nižší úrovní konfliktu (Y) je způsobena vyšší úrovní demokracie (Z)
 4. Mezi X a Y je náhodná korelace (**koincidence**):
 - „Nesmyslné korelace“
- **Korelace sama o sobě nedokáže kontrolovat žádný z těchto efektů!**

Math doctorates awarded (US)

correlates with

Suicides by hanging, strangulation and suffocation



	<u>1999</u>	<u>2000</u>	<u>2001</u>	<u>2002</u>	<u>2003</u>	<u>2004</u>	<u>2005</u>	<u>2006</u>	<u>2007</u>	<u>2008</u>	<u>2009</u>
<i>Math doctorates awarded (US)</i> <i>Degrees awarded (National Science Foundation)</i>	1,083	1,050	1,010	919	993	1,076	1,205	1,325	1,393	1,399	1,554
<i>Suicides by hanging, strangulation and suffocation</i> <i>Deaths (US) (CDC)</i>	5,427	5,688	6,198	6,462	6,635	7,336	7,248	7,491	8,161	8,578	9,000

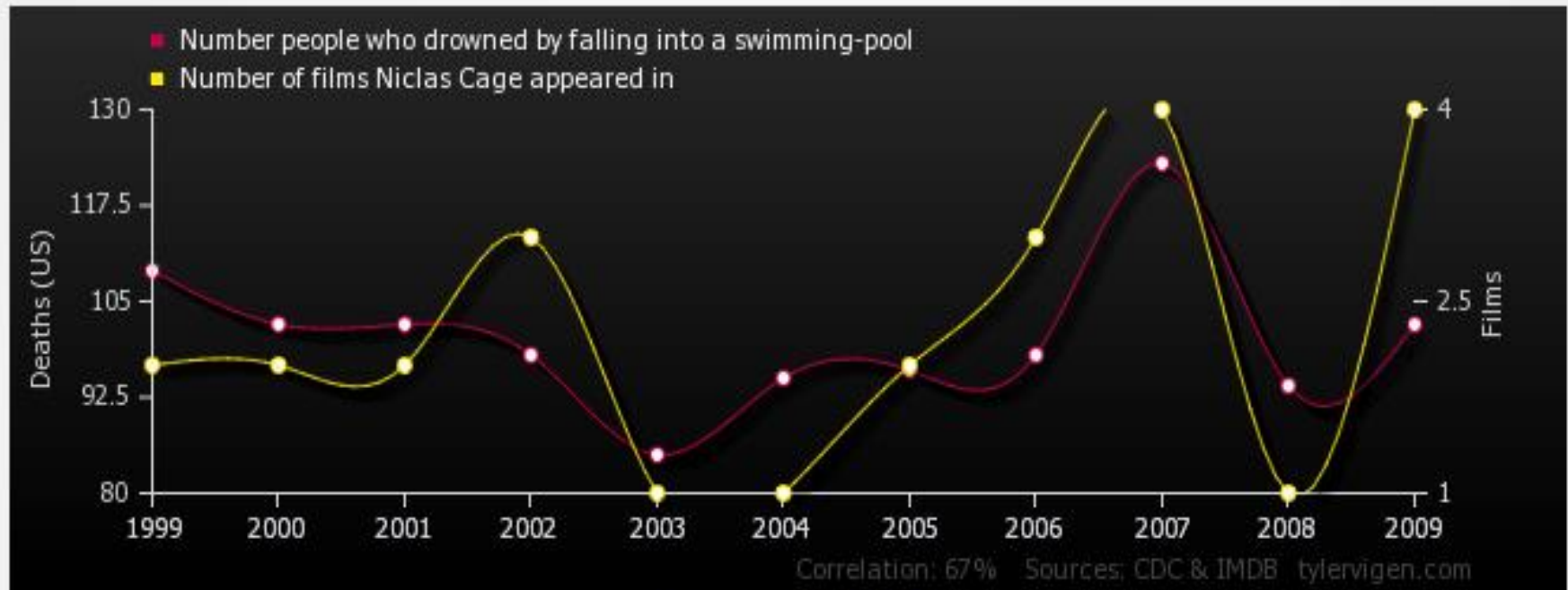
Correlation: 0.860176

[Permalink](#) - Not interesting

Number people who drowned by falling into a swimming-pool

correlates with

Number of films Nicolas Cage appeared in



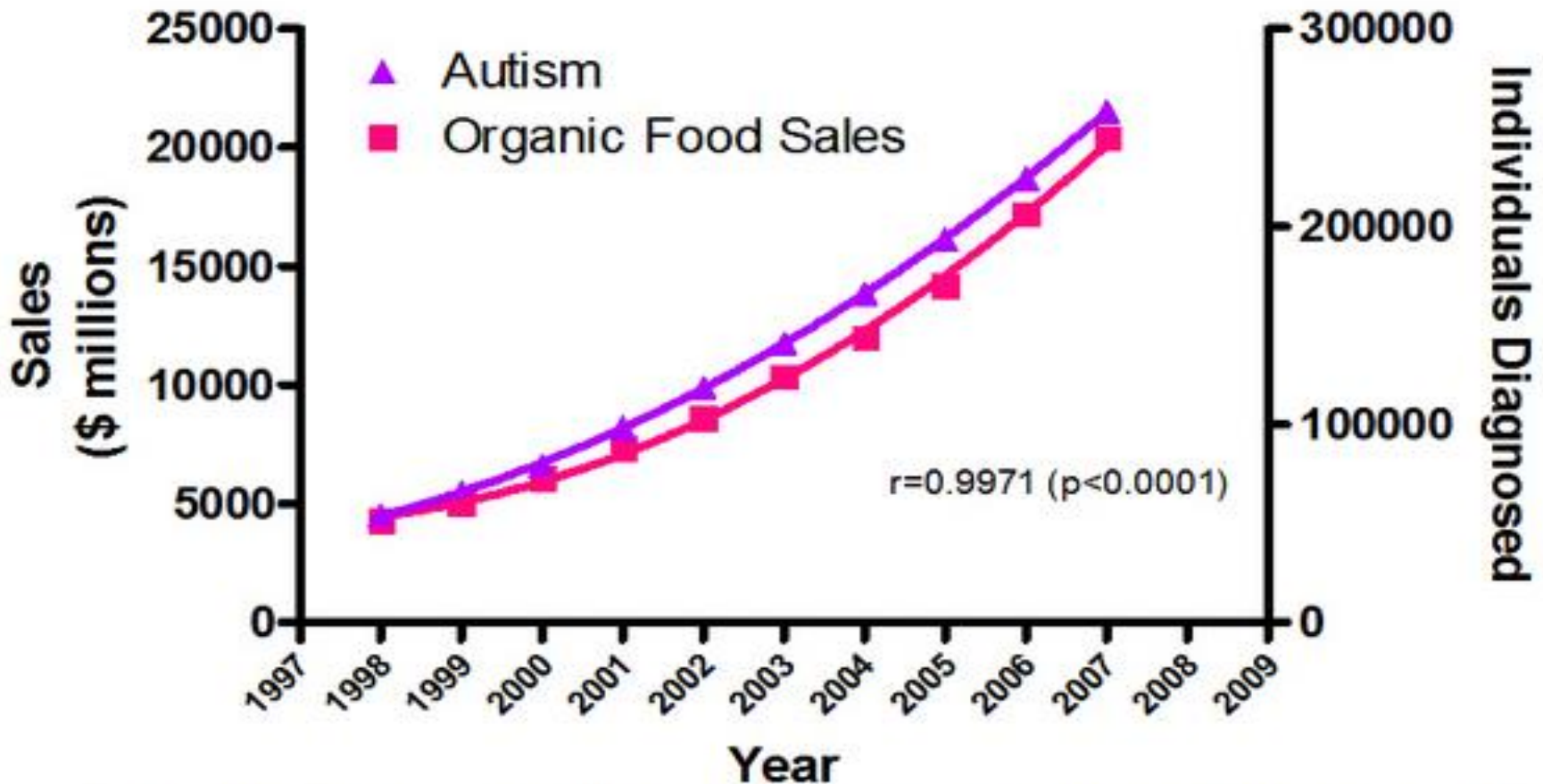
[Upload this image to imgur](#)

	<u>1999</u>	<u>2000</u>	<u>2001</u>	<u>2002</u>	<u>2003</u>	<u>2004</u>	<u>2005</u>	<u>2006</u>	<u>2007</u>	<u>2008</u>	<u>2009</u>
<i>Number people who drowned by falling into a swimming-pool Deaths (US) (CDC)</i>	109	102	102	98	85	95	96	98	123	94	102
<i>Number of films Nicolas Cage appeared in Films (IMDB)</i>	2	2	2	3	1	1	2	3	4	1	4

Correlation: 0.666004

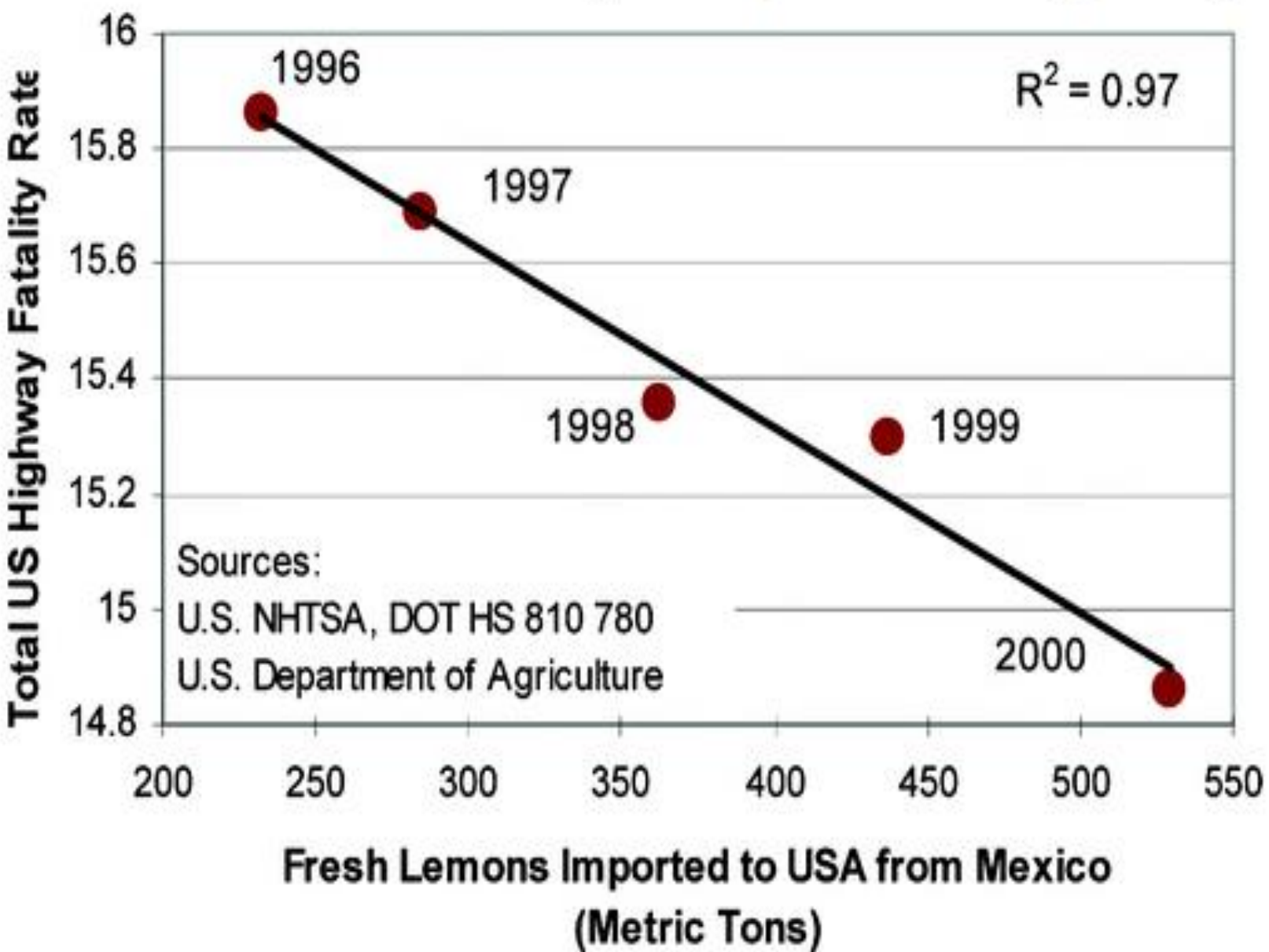
4. Eating organic food causes autism.

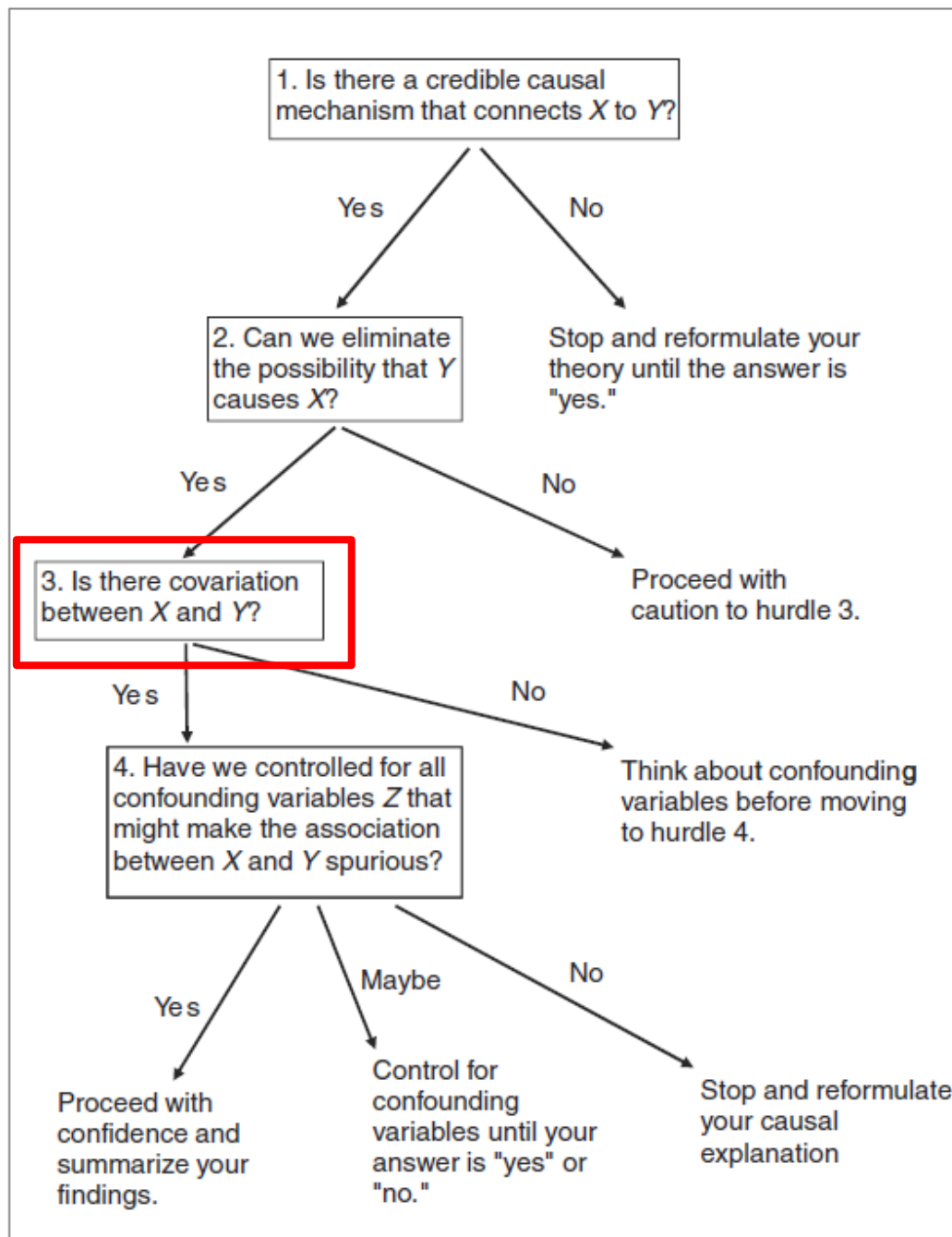
The real cause of increasing autism prevalence?



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act"

7. Mexican lemon imports prevent highway deaths.





Test hypotézy: Pearsonovo r

- H_0 : Mezi proměnnými X a Y *neexistuje* vztah lineární závislosti (korelace r)
- H_A : Mezi proměnnými X a Y *existuje* vztah lineární závislosti (korelace r)

- $H_0: r = 0 ; r \geq 0 ; r \leq 0$
- $H_A: r \neq 0$ (oboustranná hypotéza)
- $H_A: r > 0$ (jednostranná hypotéza, pozitivní korelace)
- $H_A: r < 0$ (jednostranná hypotéza, negativní korelace)

Test hypotézy: Pearsonovo r

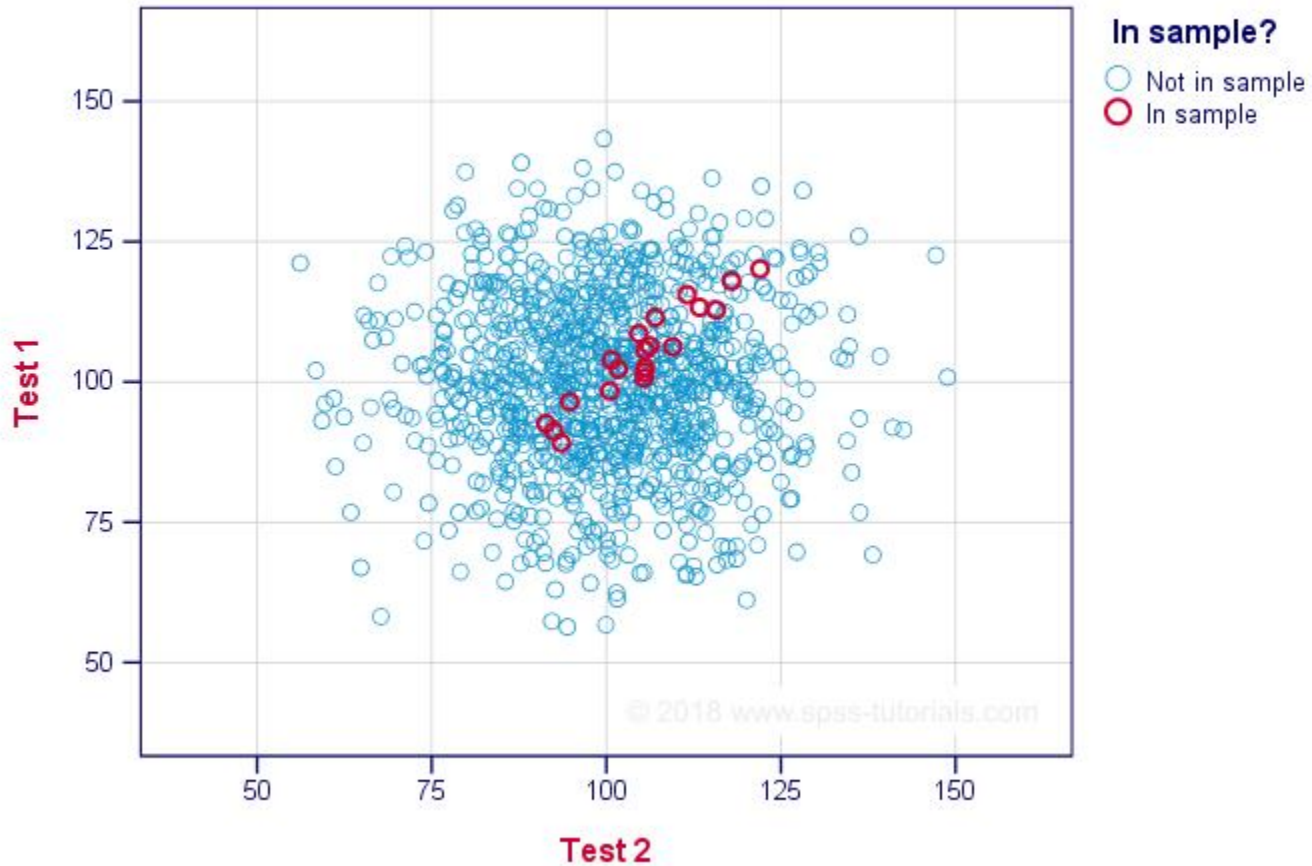
- H_0 : Mezi proměnnými X a Y *neexistuje* vztah lineární závislosti (korelace r)
- H_A : Mezi proměnnými X a Y *existuje* vztah lineární závislosti (korelace r)
- $H_0: r = 0 ; r \geq 0 ; r \leq 0$
- $H_A: r \neq 0$ (oboustranná hypotéza)
- $H_A: r > 0$ (jednostranná hypotéza, pozitivní korelace)
- $H_A: r < 0$ (jednostranná hypotéza, negativní korelace)

H_0 vždy zahrnuje nulu; v literatuře se pro jednostranné testy objevují dvě varianty zápisu: se znaménkem nerovnosti, nebo bez něj

Test hypotézy: Pearsonovo r

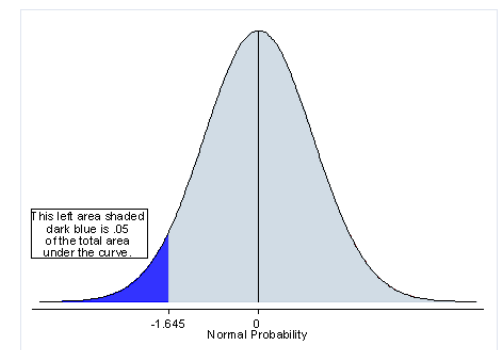
Sample 2 | N = 20

Sample Correlation = 0.95



Test hypotézy: Pearsonovo r

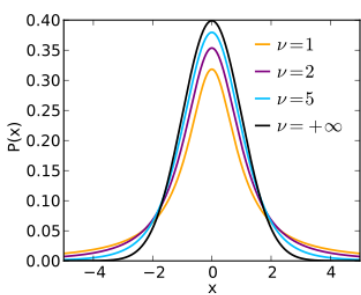
- **Teoretický rámec:** volební chování je do určité míry určeno sociokulturními štěpnými liniemi (Norris & Inglehart 2019; Lipset & Rokkan 1967)
- **H₀:** Mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) *není* korelace; $r(x, y) \geq 0$
- **H_A:** Mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) *je negativní* korelace; $r(x, y) < 0$



Test hypotézy: Pearsonovo r

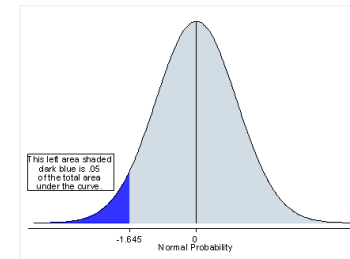
- **H₀**: Mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) *není* korelace; $r(x, y) \geq 0$
- **H_A**: Mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) *je negativní* korelace; $r(x, y) < 0$
- **Data**: 50 pozorování (malý výběr pro účely prezentace)
- Stanovení **stupňů volnosti** pro r : $n - 2$, tj. $50 - 2 = 48$
- Stanovení **hladiny testu** = 0.05 (5 %) a kritické hodnoty **t** pro jednostrannou hypotézu $(0.05, 48) = -1.677$

Appendix: Critical Values Tables



Degrees of Freedom (<i>df</i>)	80%	90%	95%	98%	99%
41	1.303	1.683	2.020	2.421	2.701
42	1.302	1.682	2.018	2.418	2.698
43	1.302	1.681	2.017	2.416	2.695
44	1.301	1.680	2.015	2.414	2.692
45	1.301	1.679	2.014	2.412	2.690
46	1.300	1.679	2.013	2.410	2.687
47	1.300	1.678	2.012	2.408	2.685
48	1.299	1.677	2.011	2.407	2.682
49	1.299	1.677	2.010	2.405	2.680
50	1.299	1.676	2.009	2.403	2.678
51	1.298	1.675	2.008	2.402	2.676
52	1.298	1.675	2.007	2.400	2.674
53	1.298	1.674	2.006	2.399	2.672
54	1.297	1.674	2.005	2.397	2.670
55	1.297	1.673	2.004	2.396	2.668
56	1.297	1.673	2.003	2.395	2.667
57	1.297	1.672	2.002	2.394	2.665
58	1.296	1.672	2.002	2.392	2.663
59	1.296	1.671	2.001	2.391	2.662
60	1.296	1.671	2.000	2.390	2.660
61	1.296	1.670	2.000	2.389	2.659
62	1.295	1.670	1.999	2.388	2.657
63	1.295	1.669	1.998	2.387	2.656
64	1.295	1.669	1.998	2.386	2.655
65	1.295	1.669	1.997	2.385	2.654
66	1.295	1.668	1.997	2.384	2.652
67	1.294	1.668	1.996	2.383	2.651
68	1.294	1.668	1.995	2.382	2.650

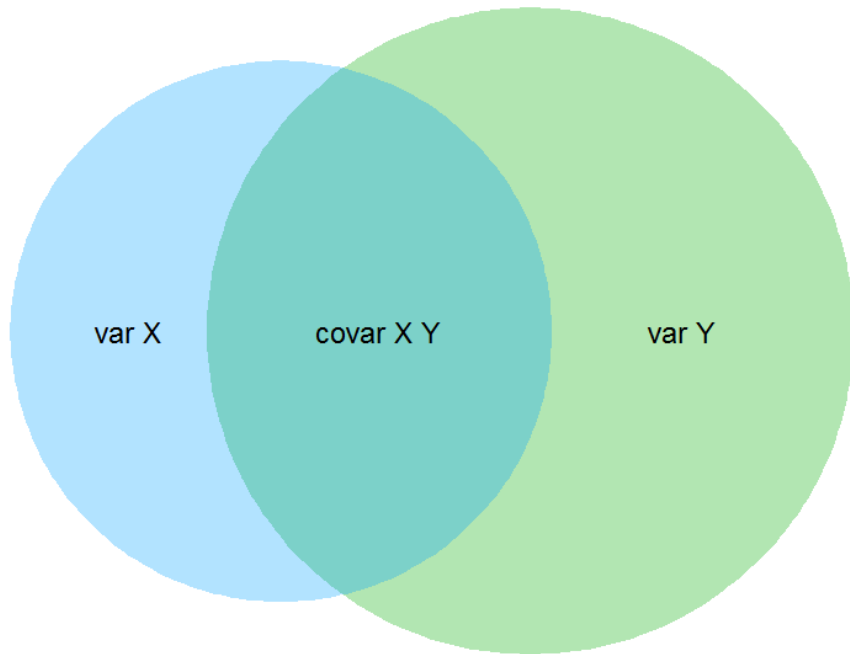
90% hodnota pro oboustranný test, tj. 95% pro jednostranný



V tabulce jsou pouze pozitivní hodnoty t . Protože t rozdělení je symetrické, je nadbytečné uvádět tutéž hodnotu pro negativní a pozitivní t . Pro $H_A: r < 0$ tedy t předradíme znaménko minus.

74	1.293	1.666	1.993	2.378	2.644
75	1.293	1.665	1.992	2.377	2.643
76	1.293	1.665	1.992	2.376	2.642
77	1.293	1.665	1.991	2.376	2.641
78	1.292	1.665	1.991	2.375	2.640
79	1.292	1.664	1.990	2.374	2.640
80	1.292	1.664	1.990	2.374	2.639
81	1.292	1.664	1.990	2.373	2.638
82	1.292	1.664	1.989	2.373	2.637
83	1.292	1.663	1.989	2.372	2.636

- $r = \text{kovariance } X, Y / \text{kombinovaný rozptyl } X, Y$



$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$

- Korelace mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) : $r = -0.413$

Testovací statistika t

- Abychom zjistili, zda se r významně liší od nuly, uijeme **t-test** pro korelační koeficient r .
- *Je hodnota r významně odlišná od předpokládaného populačního průměru 0 ($H_0: r \geq 0$)?*

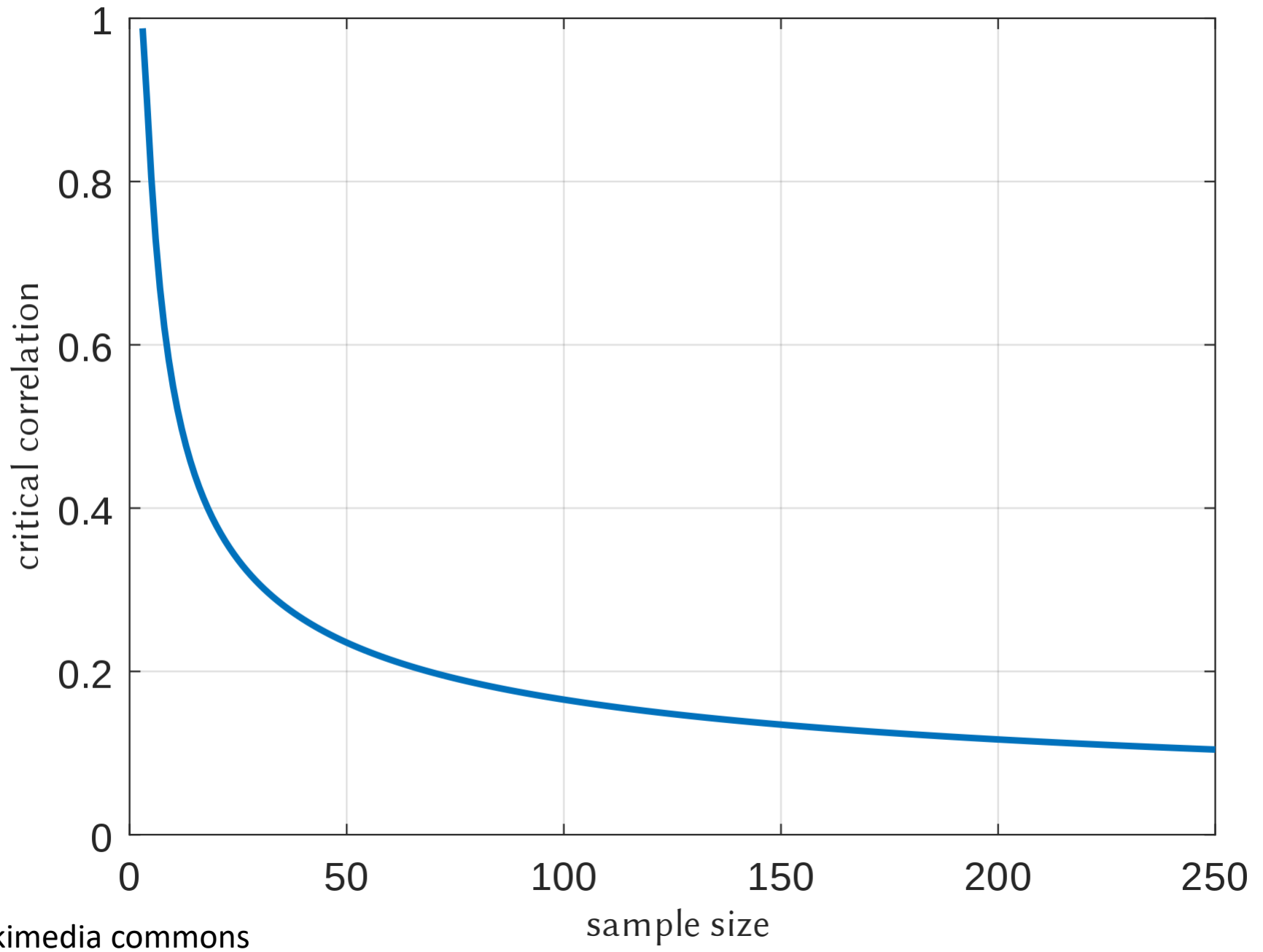
- $t = \frac{\text{signal}}{\text{noise}} ; t = \frac{r * \sqrt{n - 2}}{\sqrt{1 - r^2}} ; n = \text{velikost vzorku}$

- $t = \frac{-0.413 * \sqrt{50 - 2}}{\sqrt{1 - (-0.413)^2}} = -3.14$

- **t-hodnota** Pearsonova r je **testovací statistikou**

Rozhodnutí o H_0

- **H_0 :** Mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) *není* korelace; $r(x, y) \geq 0$
- **H_A :** Mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) *je negativní* korelace; $r(x, y) < 0$
- Pearsonovo $r = -0.413$
- Testovací statistika $t = -3.14$; kritická hodnota t ($\alpha = 0.05$) = -1.677
- Protože $t = -3.14 < -1.677$ ($\alpha = 0.05$, $sv = 48$), **odmítáme H_0 : $r \geq 0$** a přijímáme **H_A : $r < 0$** .
- **p-hodnota** = 0.003 (tj. 0.3 %) značí pravděpodobnost pozorování dané, či ještě extrémnější, hodnoty **testovací statistiky** ($t = -3.14$) při **platnosti H_0** .
- Tj. mezi podporou prezidentství Trumpa (X) a přijatelností ekonomické migrace (Y) je, na 5% hladině st. významnosti, lineární závislost (korelace r).



wikimedia commons

sample size

Pearsonův korelační koeficient r

hodnota Pearsonova r	interpretace
0.00–0.19	velmi slabá
0.20–0.39	slabá
0.40–0.59	mírná
0.60–0.79	silná
0.80–1.00	velmi silná

Ovšem vždy závislé na kontextu!

Pearsonovo r v Jamovi



Data

Analyses



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R



Modules

Descriptives

Descriptives



abort_rank3
 abortion_rank12
 adv_or_more
 ba_or_more
 cig_tax12
 cig_tax12_3
 conserv_advantage
 conserv_public
 dem_advantage

Variables



Split by

 Frequency tables

> | Statistics

> | Plots










Descriptives

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/>.

Descriptives



-  popchnng0010
-  popchnngpct
-  pot_policy
-  prochoice
-  prolife
-  relig_cath
-  relig_high
-  relig_low
-  religiosity3




Variables

-  obama2012
-  relig_prot



Split by

 Frequency tables 

Statistics

Sample Size

 N Missing

Percentile Values

 Quartiles

 Cut points for equal groups

Dispersion

 Std. deviation Minimum

 Variance Maximum

 Range S. E. Mean

Central Tendency

 Mean

 Median

 Mode

 Sum

Distribution

 Skewness

 Kurtosis

Normality

 Shapiro-Wilk

Plots

Descriptives

Descriptives

	obama2012	relig_prot
N	50	50
Missing	0	0
Mean	48.2	52.4
Median	50.2	51.6
Standard deviation	10.3	14.8
Minimum	24.7	12.3
Maximum	70.5	80.3
25th percentile	40.0	42.6
50th percentile	50.2	51.6
75th percentile	55.4	61.7










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/>.

Descriptives




-  popchnng0010
-  popchnngpct
-  a_pot_policy
-  prochoice
-  prolife
-  relig_cath
-  relig_high
-  relig_low
-  a_religiosity3

Variables

-  obama2012
-  relig_prot

Split by

 Frequency tables 

Statistics

Sample Size

 N Missing

Percentile Values

 Quartiles
 Cut points for equal groups

Dispersion

 Std. deviation Minimum
 Variance Maximum
 Range S. E. Mean

Central Tendency

 Mean
 Median
 Mode
 Sum

Distribution

 Skewness
 Kurtosis

Normality

 Shapiro-Wilk

Plots

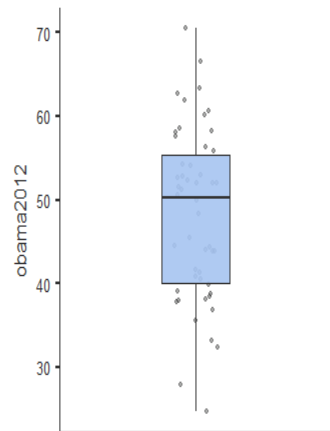
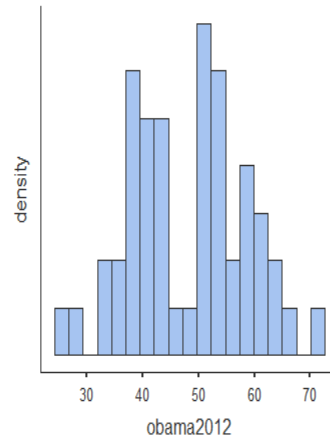
Histograms 
 Histogram
 Density
Box Plots 
 Box plot
 Violin
 Data

Jittered ▾

Bar Plots 
 Bar plot
Q-Q Plots 
 Q-Q

Plots

obama2012



relig_prot

Descriptives



- popchn0010
- popchnpct
- pot_policy
- prochoice
- prolife
- relig_cath
- relig_high
- relig_low
- religiosity3

- Variables
- obama2012
 - relig_prot

Split by

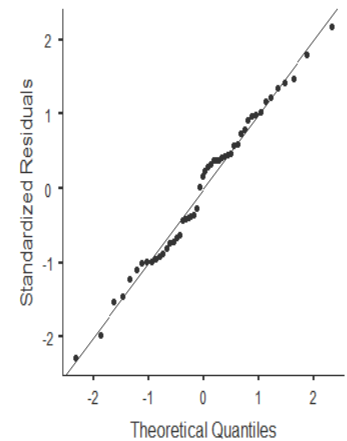
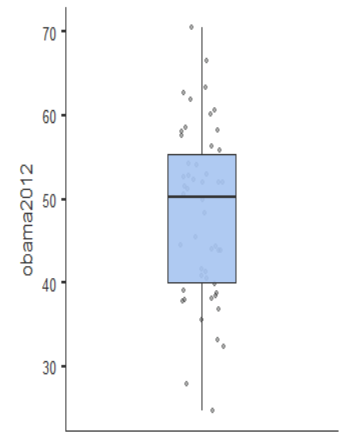
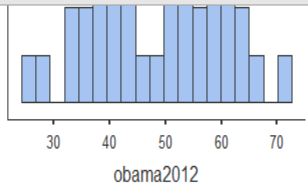
Frequency tables

Statistics

- Sample Size**
- N
 - Missing
- Percentile Values**
- Quartiles
 - Cut points for equal groups
- Dispersion**
- Std. deviation
 - Minimum
 - Variance
 - Maximum
 - Range
 - S. E. Mean
- Central Tendency**
- Mean
 - Median
 - Mode
 - Sum
- Distribution**
- Skewness
 - Kurtosis
- Normality**
- Shapiro-Wilk

Plots

- Histograms**
- Histogram
 - Density
- Box Plots**
- Box plot
 - Violin
 - Data
- Bar Plots**
- Bar plot
- Q-Q Plots**
- Q-Q
- Jittered





Descriptives

- popchng0010
- popchngpct
- pot_policy
- prochoice
- prolife
- relig_cath
- relig_high
- relig_low
- religiosity3

Frequency tables

Statistics

Sample Size

N Missing

Central Tendency

Mean Median Mode Sum

Percentile Values

Quartiles

Cut points for 4 equal groups

Dispersion

Std. deviation Minimum Variance Maximum Range S. E. Mean

Distribution

Skewness Kurtosis

Normality

Shapiro-Wilk

Plots

Histograms

Histogram Density

Box Plots

Box plot Violin Data

Q-Q Plots

Q-Q

Jittered

Correlation Matrix

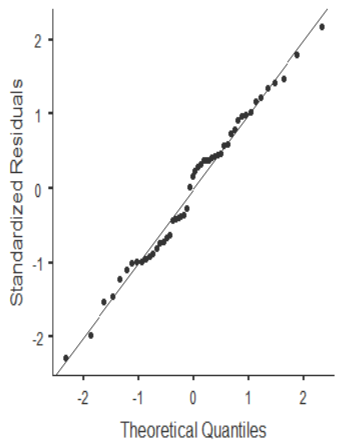
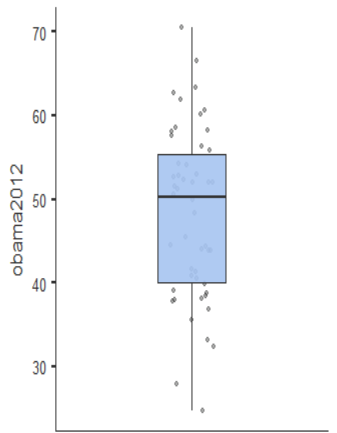
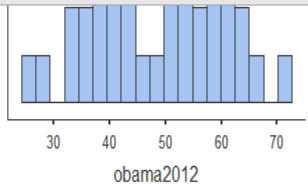
Linear Regression

Logistic Regression

2 Outcomes Binomial

N Outcomes Multinomial

Ordinal Outcomes



Correlation Matrix



ns_or_more
 pop2000
 pop2010
 pop2010_hun_thou
 popchg0010
 popchnpct
 relig_cath
 relig_high
 relig_low

→

obama2012
 relig_prot

Correlation Coefficients

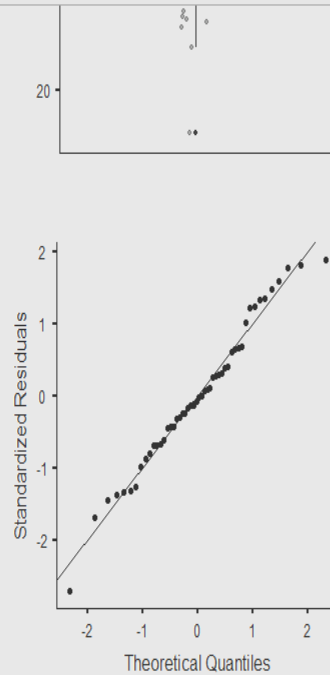
Additional Options

- Pearson
 Spearman
 Kendall's tau-b
- Report significance
 Flag significant correlations
 Confidence intervals
 Interval %

Hypothesis

Plot

- Correlated
 Correlated positively
 Correlated negatively
- Correlation matrix
 Densities for variables
 Statistics



Correlation Matrix

Correlation Matrix

		obama2012	relig_prot
obama2012	Pearson's r	—	—
	p-value	—	—
	95% CI Upper	—	—
	95% CI Lower	—	—
relig_prot	Pearson's r	-0.413**	—
	p-value	0.001	—
	95% CI Upper	-0.197	—
	95% CI Lower	-1.000	—

Note. H_a is negative correlation

Note. * $p < .05$, ** $p < .01$, *** $p < .001$, one-tailed



Correlation Matrix



- modpct_m
- obama08
- over64
- pop_18_24
- pop_18_24_10
- reppct_m
- to_0004
- to_0408
- trout00

- obama2012
- relig_prot
- urban
- prcapinc

Correlation Coefficients

Pearson

Spearman

Kendall's tau-b

Additional Options

Report significance

Flag significant correlations

Confidence intervals

Interval %

Hypothesis

Correlated

Correlated positively

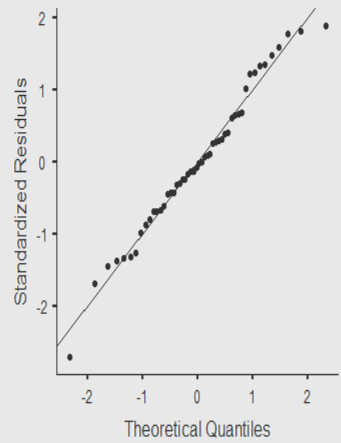
Correlated negatively

Plot

Correlation matrix

Densities for variables

Statistics



Correlation Matrix

Correlation Matrix

		obama2012	relig_prot	urban	prcapinc
obama2012	Pearson's r	—			
	p-value	—			
	95% CI Upper	—			
	95% CI Lower	—			
relig_prot	Pearson's r	-0.413**	—		
	p-value	0.003	—		
	95% CI Upper	-0.152	—		
	95% CI Lower	-0.620	—		
urban	Pearson's r	0.406**	-0.577***	—	
	p-value	0.003	<.001	—	
	95% CI Upper	0.615	-0.356	—	
	95% CI Lower	0.144	-0.737	—	
prcapinc	Pearson's r	0.610***	-0.546***	0.526***	—
	p-value	<.001	<.001	<.001	—
	95% CI Upper	0.759	-0.315	0.701	—
	95% CI Lower	0.400	-0.715	0.290	—

Note. * p < .05, ** p < .01, *** p < .001

Correlation Matrix



- modpct_m
- obama08
- over64
- pop_18_24
- pop_18_24_10
- reppct_m
- to_0004
- to_0408
- trout00

→

- obama2012
- relig_prot
- urban
- prcapinc

Correlation Coefficients **Additional Options**

- Pearson
 - Spearman
 - Kendall's tau-b
 - Report significance
 - Flag significant correlations
 - Confidence intervals
- Interval %

Hypothesis

- Correlated
- Correlated positively
- Correlated negatively

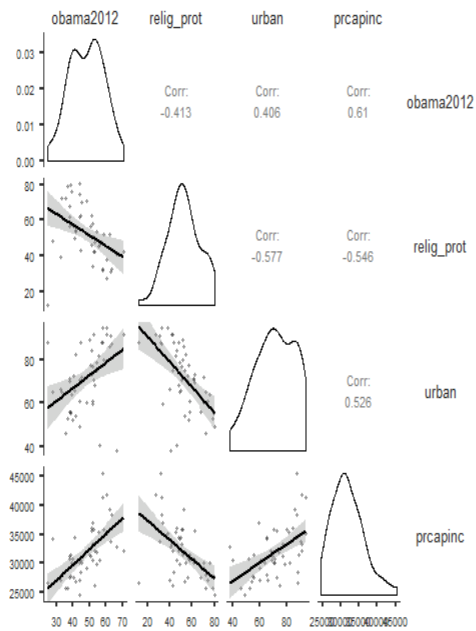
Plot

- Correlation matrix
- Densities for variables
- Statistics

	95% CI Lower	—		
relig_prot	Pearson's r	-0.413**	—	
	p-value	0.003	—	
	95% CI Upper	-0.152	—	
	95% CI Lower	-0.620	—	
urban	Pearson's r	0.406**	-0.577***	—
	p-value	0.003	<.001	—
	95% CI Upper	0.615	-0.356	—
	95% CI Lower	0.144	-0.737	—
prcapinc	Pearson's r	0.610***	-0.546***	0.526***
	p-value	<.001	<.001	<.001
	95% CI Upper	0.759	-0.315	0.701
	95% CI Lower	0.400	-0.715	0.290

Note. * p < .05, ** p < .01, *** p < .001

Plot



Seminář

- Otevřete „MVZn4003_08_analyza_zavislosti_2_seminar.pdf“ a postupujte dle instrukcí

Pearsonovo r: výpočet

- Dvě proměnné: X a Y

X	Y
1	0
2	1
1	4
6	8
7	4

- (1) výpočet rozptylu proměnných
- $mean(x) = 3.4; mean(y) = 3.4$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

X	(x - m)	dev.	dev.^2	Y	(y - m)	dev.	dev.^2
1	(1 - 3.4)	-2.4	5.76	0	(0 - 3.4)	-3.4	11.56
2	(2 - 3.4)	-1.4	1.96	1	(1 - 3.4)	-2.4	5.76
1	(1 - 3.4)	-2.4	5.76	4	(4 - 3.4)	0.6	0.36
6	(6 - 3.4)	2.6	6.76	8	(8 - 3.4)	4.6	21.16
7	(7 - 3.4)	3.6	12.96	4	(4 - 3.4)	0.6	0.36
sum	0	0	33.2	sum	0	0	39.2

- $s^2(X) = 33.2 / 4 = 8.3; s^2(Y) = 39.2 / 4 = 9.8$

- (2) výpočet **kovariance proměnných**
- **Kovariance** je součet produktů odchylek proměnných vydělený $n-1$

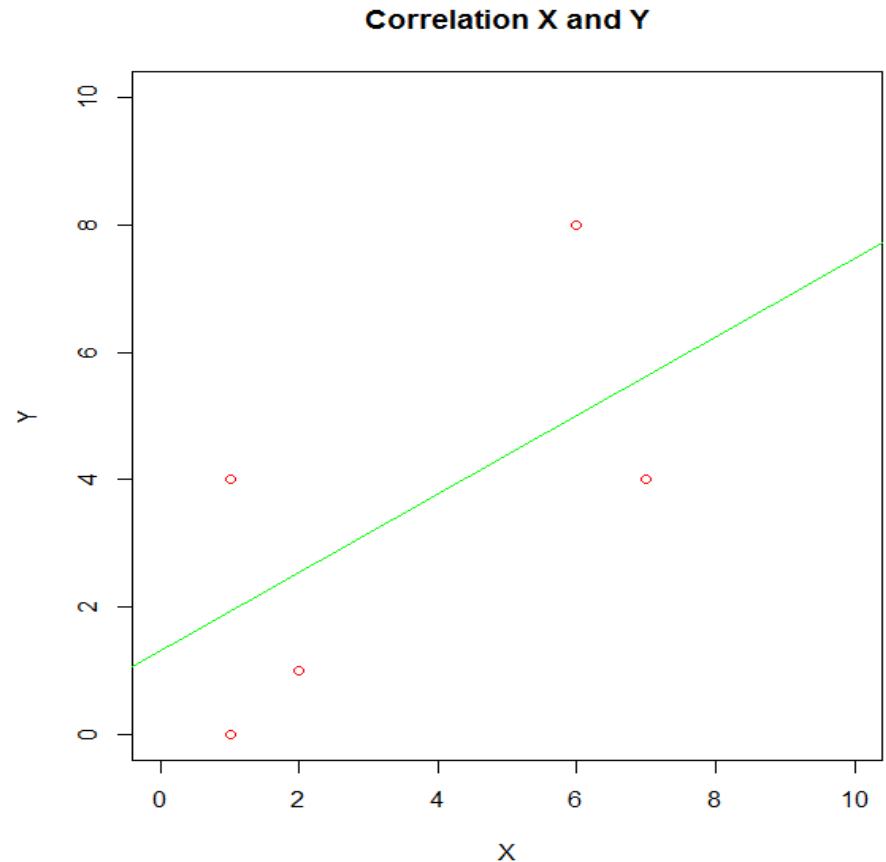
(x - m)	(y - m)	cross-prod.
(1 - 3.4)	(0 - 3.4)	8.16
(2 - 3.4)	(1 - 3.4)	3.36
(1 - 3.4)	(4 - 3.4)	-1.44
(6 - 3.4)	(8 - 3.4)	11.96
(7 - 3.4)	(4 - 3.4)	2.16
0	0	24.2

$$\text{cov}(X, Y) = 24.2 / 4 = \mathbf{6.05}$$

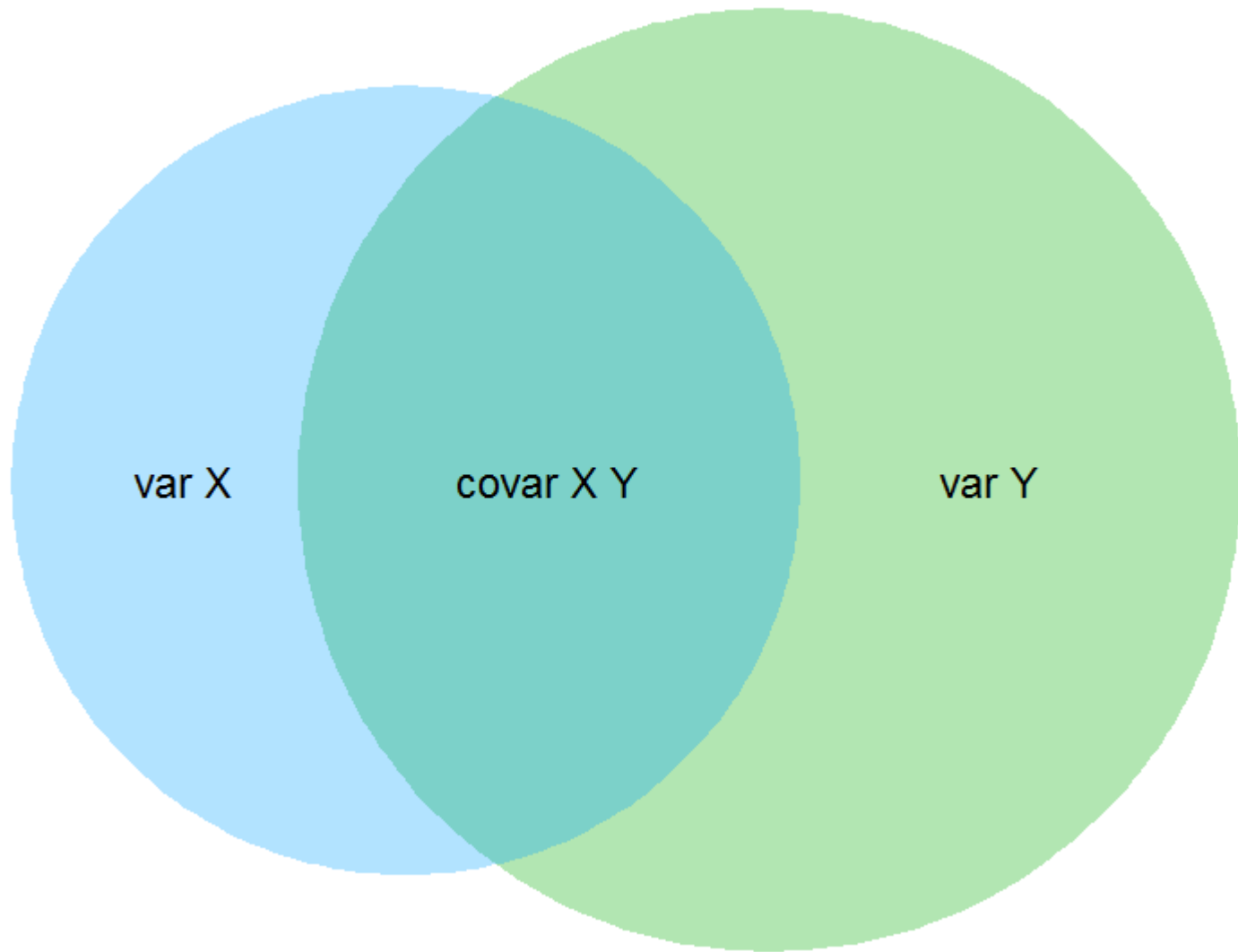
$$\text{COV}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- (3) Kovarianci vydělíme odmocninou produktu rozptylů X a Y
 - $r = \text{cov}(X, Y) / \text{sqrt}(\text{var}(X) * \text{var}(Y))$
 - $r = 6.05 / \text{sqrt}(8.3 * 9.8) = \mathbf{0.67}$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$



- $r = \text{kovariance } X, Y / \text{kombinovaný rozptyl } X, Y$



Critical Values for Pearson's Correlation Coefficient

DF	Proportion in ONE Tail					
	.25	.10	.05	.025	.01	.005
	Proportion in TWO Tails					
	.50	.20	.10	.05	.02	.01
1	.7071	.9511	.9877	.9969	.9995	.9999
2	.5000	.8000	.9000	.9500	.9800	.9900
3	.4040	.6870	.8054	.8783	.9343	.9587
4	.3473	.6084	.7293	.8114	.8822	.9172
5	.3091	.5509	.6694	.7545	.8329	.8745
6	.2811	.5067	.6215	.7067	.7887	.8343
7	.2596	.4716	.5822	.6664	.7498	.7977
8	.2423	.4428	.5494	.6319	.7155	.7646
9	.2281	.4187	.5214	.6021	.6851	.7348
10	.2161	.3981	.4973	.5760	.6581	.7079
11	.2058	.3802	.4762	.5529	.6339	.6835
12	.1968	.3646	.4575	.5324	.6120	.6614
13	.1890	.3507	.4409	.5140	.5923	.6411
14	.1820	.3383	.4259	.4973	.5742	.6226
15	.1757	.3271	.4124	.4821	.5577	.6055
16	.1700	.3170	.4000	.4683	.5425	.5897
17	.1649	.3077	.3887	.4555	.5285	.5751
18	.1602	.2992	.3783	.4438	.5155	.5614
19	.1558	.2914	.3687	.4329	.5034	.5487
20	.1518	.2841	.3598	.4227	.4921	.5368
21	.1481	.2774	.3515	.4132	.4815	.5256
22	.1447	.2711	.3438	.4044	.4716	.5151
23	.1415	.2653	.3365	.3961	.4622	.5052
24	.1384	.2598	.3297	.3882	.4534	.4958
25	.1356	.2546	.3233	.3809	.4451	.4869
26	.1330	.2497	.3172	.3739	.4372	.4785
27	.1305	.2451	.3115	.3673	.4297	.4705
28	.1281	.2407	.3061	.3610	.4226	.4629
29	.1258	.2366	.3009	.3550	.4158	.4556
30	.1237	.2327	.2960	.3494	.4093	.4487
31	.1217	.2289	.2913	.3440	.4032	.4421
32	.1197	.2254	.2869	.3388	.3972	.4357
33	.1179	.2220	.2826	.3338	.3916	.4296
34	.1161	.2187	.2785	.3291	.3862	.4238
35	.1144	.2156	.2746	.3246	.3810	.4182
36	.1128	.2126	.2709	.3202	.3760	.4128
37	.1113	.2097	.2673	.3160	.3712	.4076
38	.1098	.2070	.2638	.3120	.3665	.4026
39	.1084	.2043	.2605	.3081	.3621	.3978
40	.1070	.2018	.2573	.3044	.3578	.3932
41	.1057	.1993	.2542	.3008	.3536	.3887
42	.1044	.1970	.2512	.2973	.3496	.3843
43	.1032	.1947	.2483	.2940	.3457	.3801
44	.1020	.1925	.2455	.2907	.3420	.3761
45	.1008	.1903	.2429	.2876	.3384	.3721
46	.0997	.1883	.2403	.2845	.3348	.3683
47	.0987	.1863	.2377	.2816	.3314	.3646
48	.0976	.1843	.2353	.2787	.3281	.3610
49	.0966	.1825	.2329	.2759	.3249	.3575
50	.0956	.1806	.2306	.2732	.3218	.3542

Transformace proměnných



Data

Analyses



TRANSFORMED VARIABLE

● TRANSFORM

used by 1

religiosity3_tsf

ordinal

Variable suffix

+ Add recode condition

- if \$source == "Low" use 1
- if \$source == "Mid" use 2
- 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



	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



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

	south		Total
	nonsouth	South	
obama_win12			
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

		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