

Analýza závislosti: intervalová a poměrová data (Pearsonovo r)

Petr Ocelík

MVZn4003 Úvod do kvantitativních metod

8. prosince 2020

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 $\langle -1,1 \rangle$

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

- Korelace neimplikuje kauzalitu
- Kauzalita může být založena na různých typech závislosti

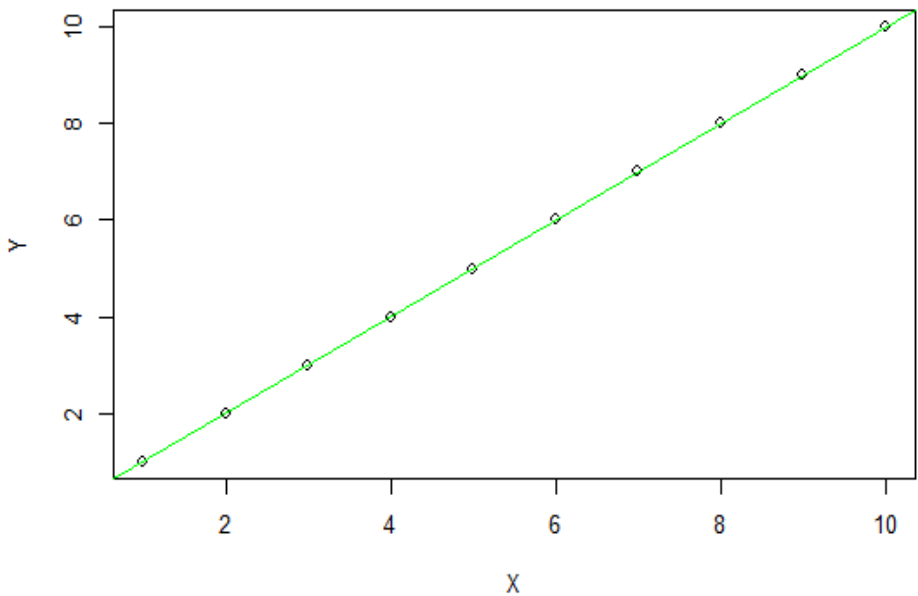
Míry závislosti

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

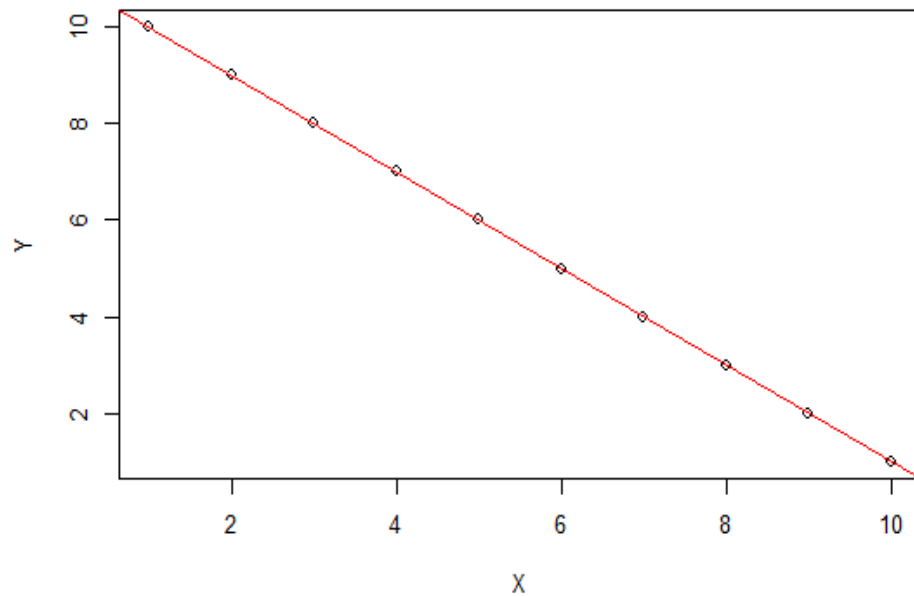
Pearsonův korelační koeficient r

- 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

$r=1$

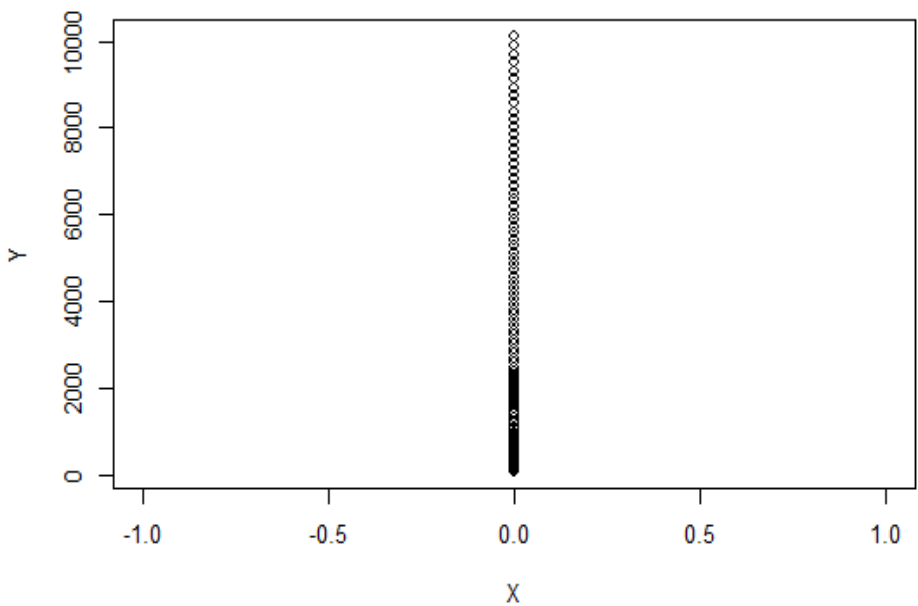


$r=-1$

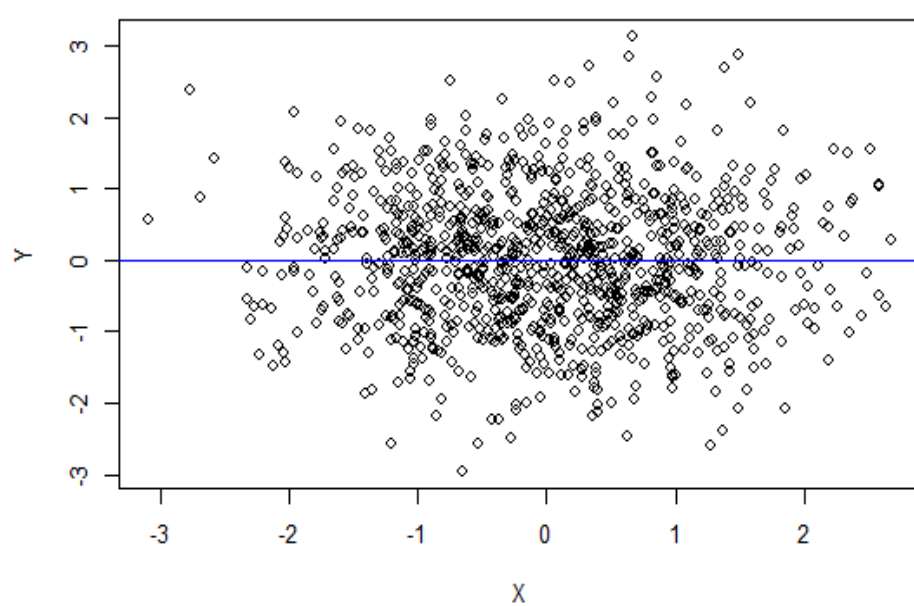


<http://guessthecorrelation.com/>

$r=0$



$r=0$



- Kellstedt & Whitten (2013) příklad korelace volebního zisku inkumbenta (vládní politické strany) a změny HDP během předcházejícího volebního období.
- Statisticky významná korelace = 0.574

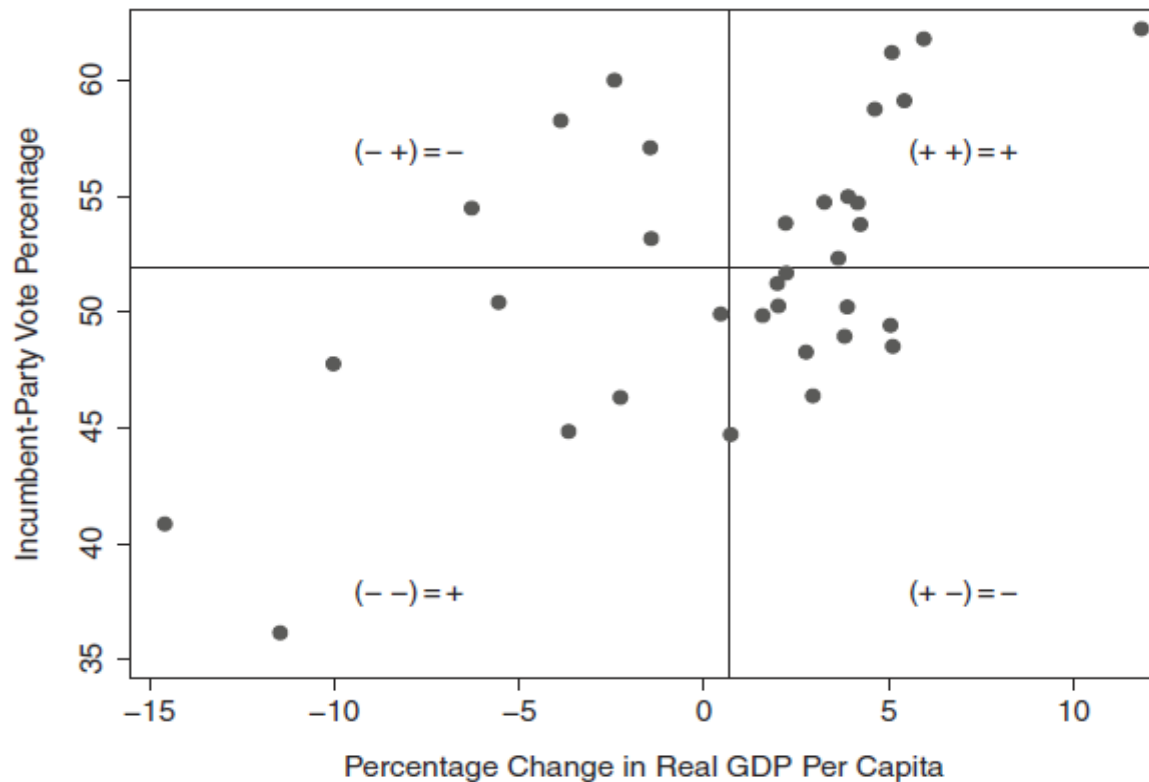
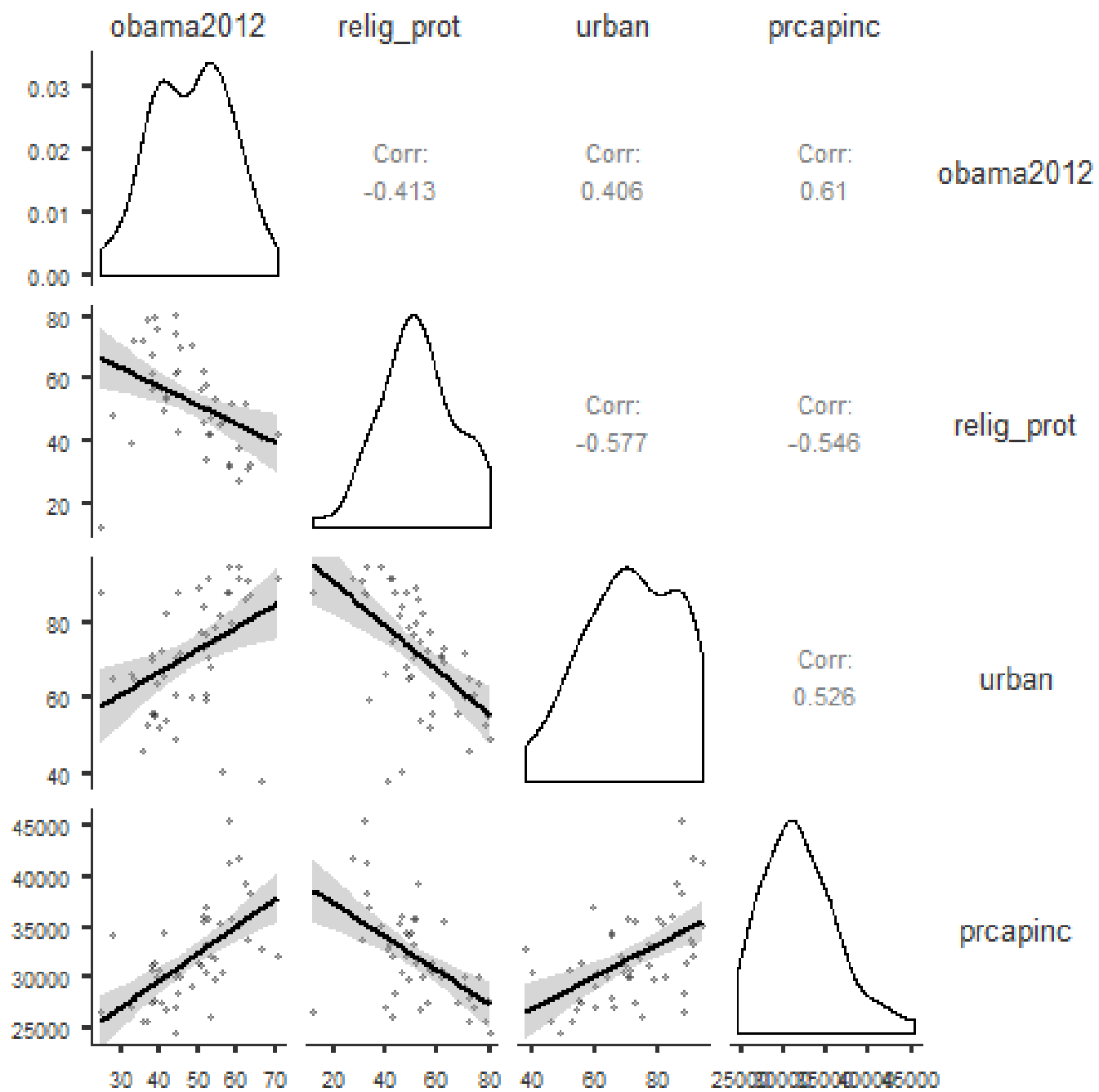


Figure 7.4. Scatter plot of change in GDP and incumbent-party vote share with mean-delimited quadrants.

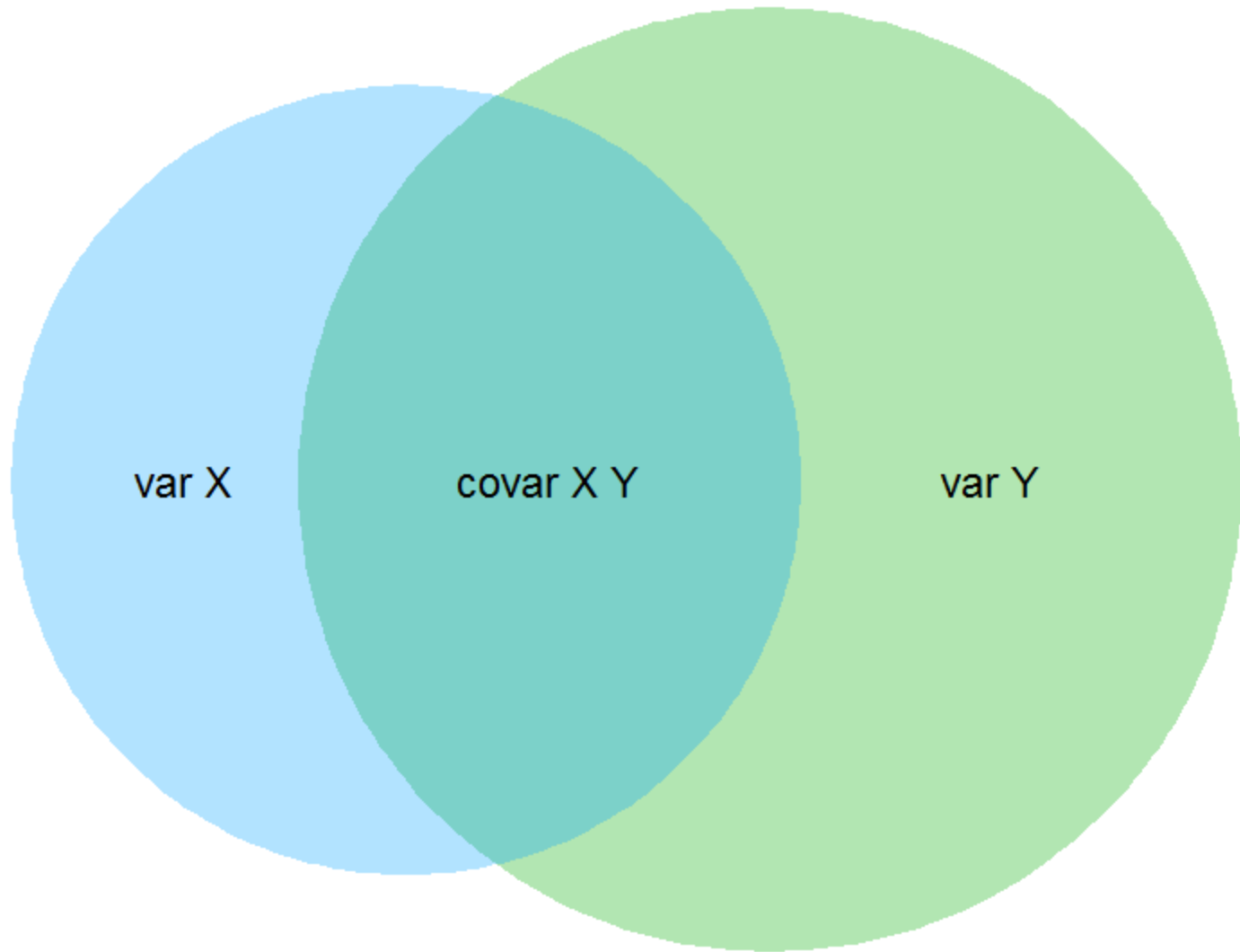


Pearsonův korelační koeficient r

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!

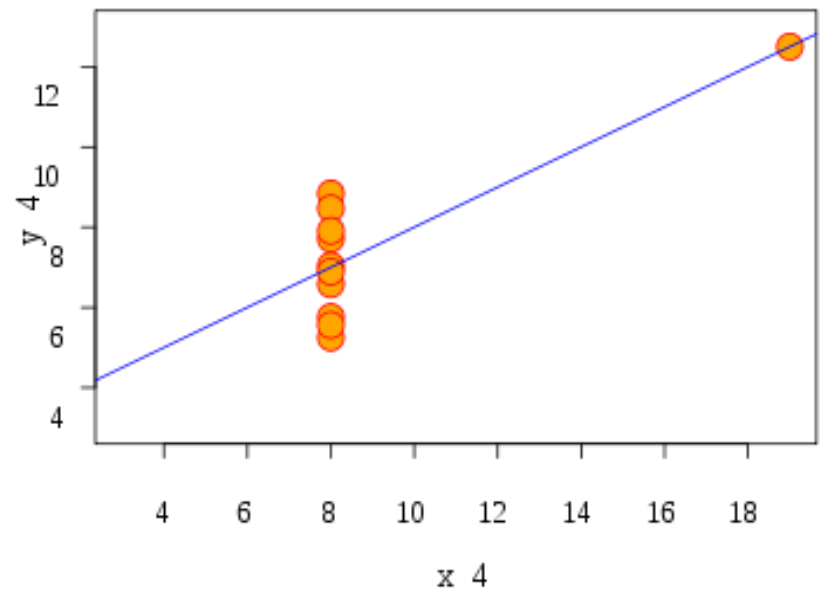
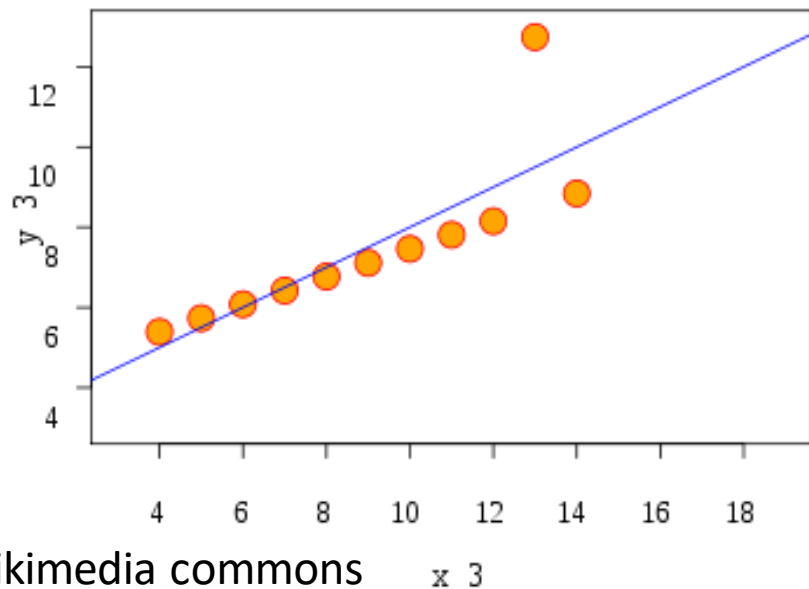
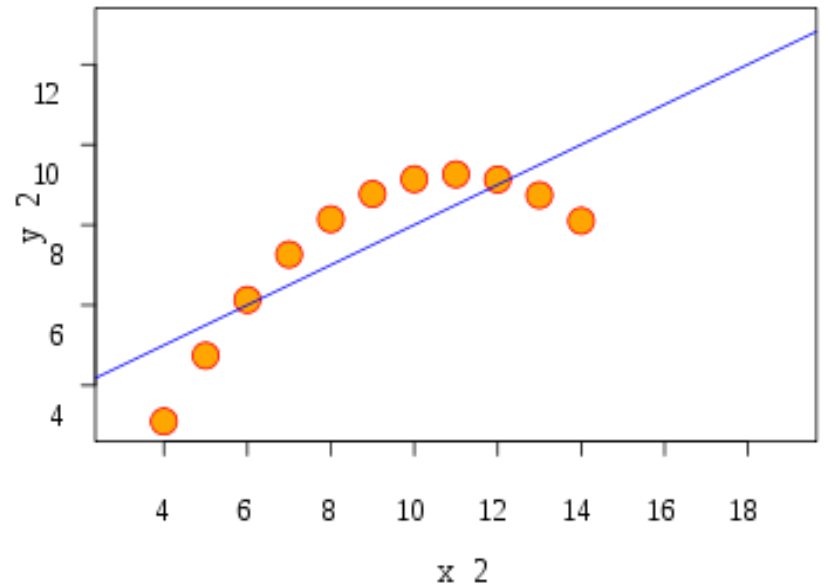
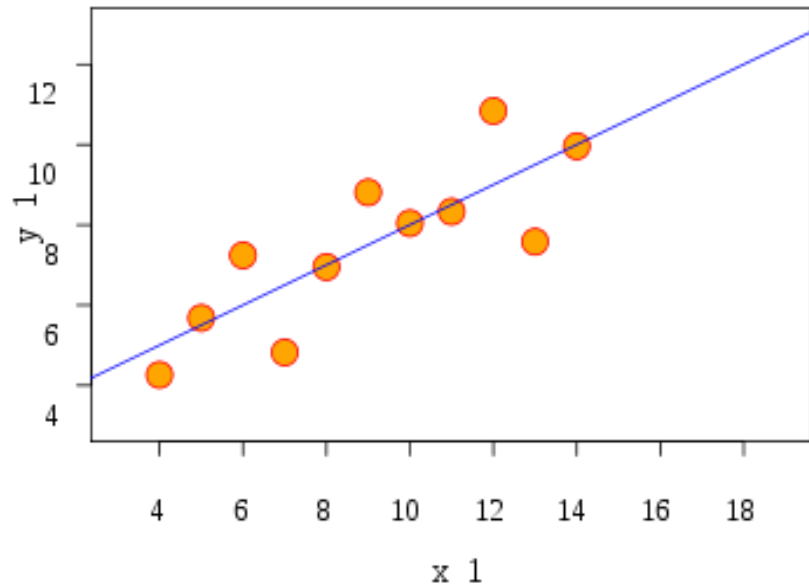
- $r = \text{kovariance } X, Y / \text{kombinovaný rozptyl } X, Y$

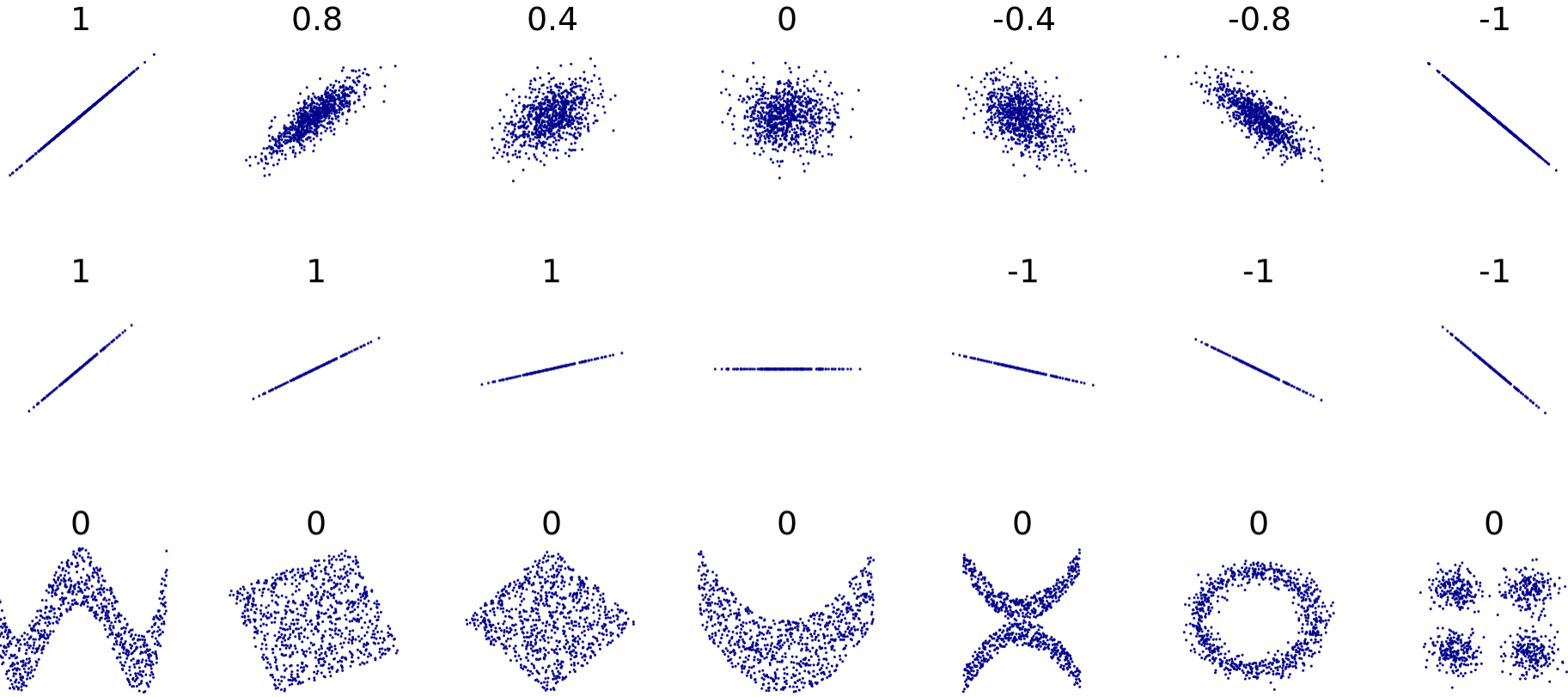


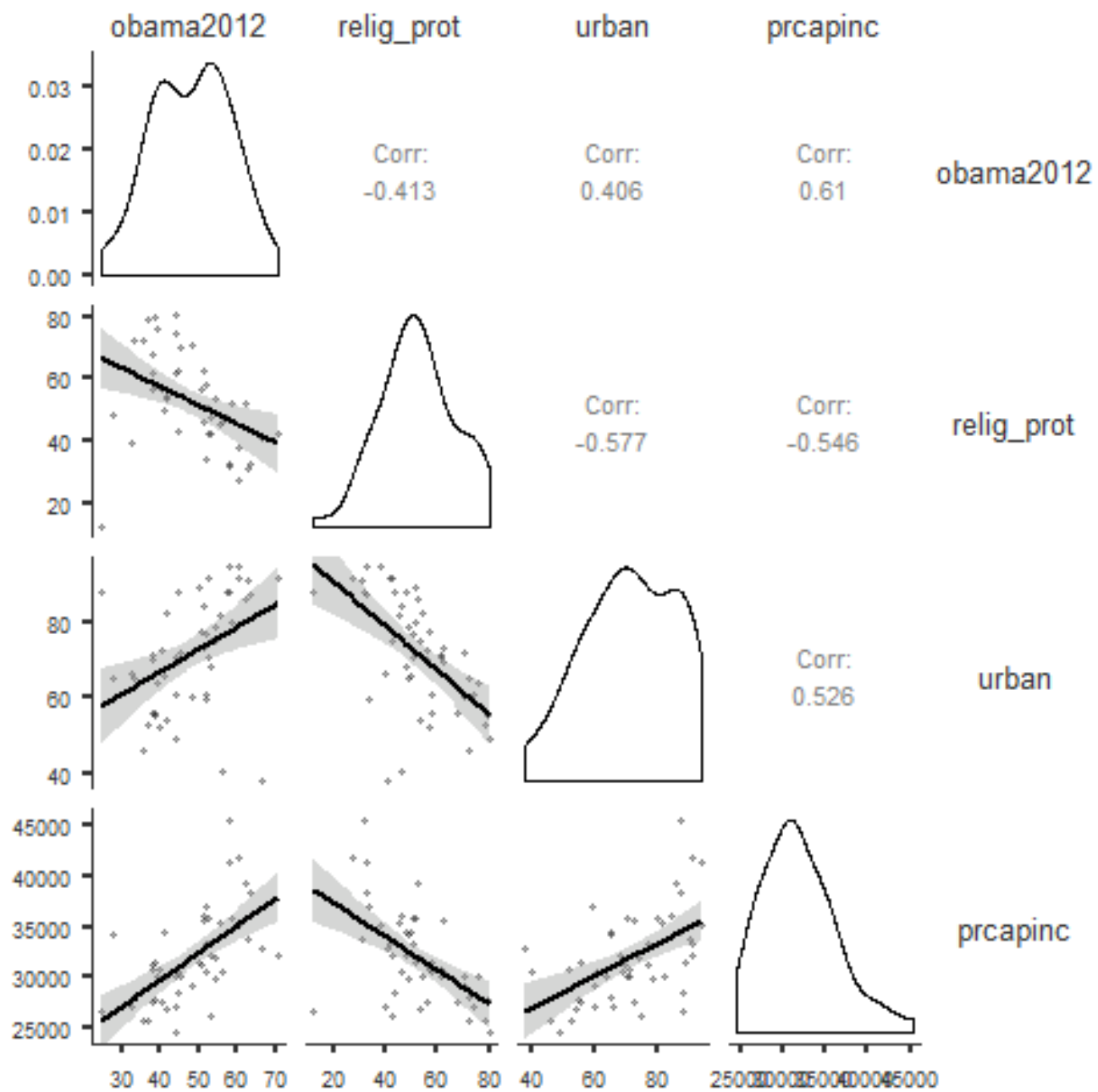
Pearsonovo r: předpoklady a omezení

- Metrická úroveň (intervalová nebo poměrová) měření
- **Lineární vztah mezi proměnnými X a Y**
- Homoskedasticita (nezávislost rozptylu)
- Citlivost vůči odlehlým hodnotám (outliers)

Anscombe's quartet

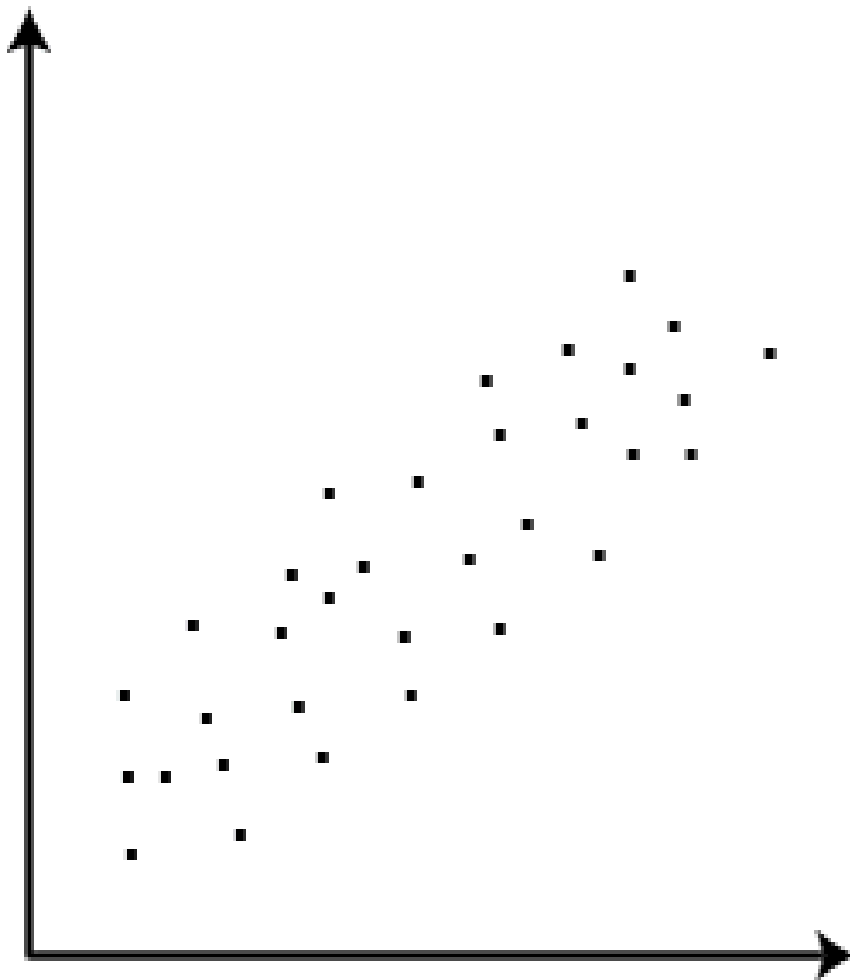




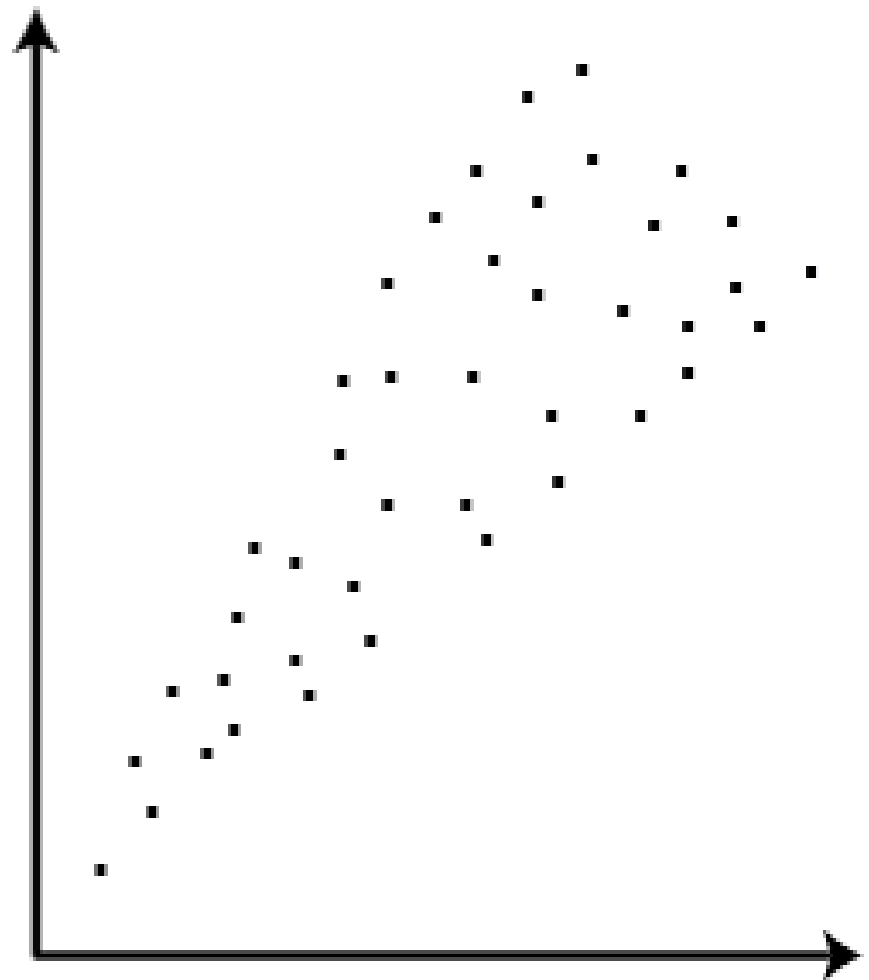


Pearsonovo r: předpoklady a omezení

- Metrická úroveň (intervalová nebo poměrová) měření
- Lineární vztah mezi proměnnými X a Y
- **Homoskedasticita (nezávislost rozptylu)**
- Citlivost vůči odlehlým hodnotám (outliers)



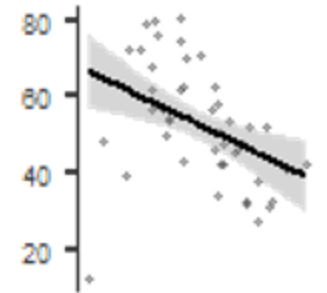
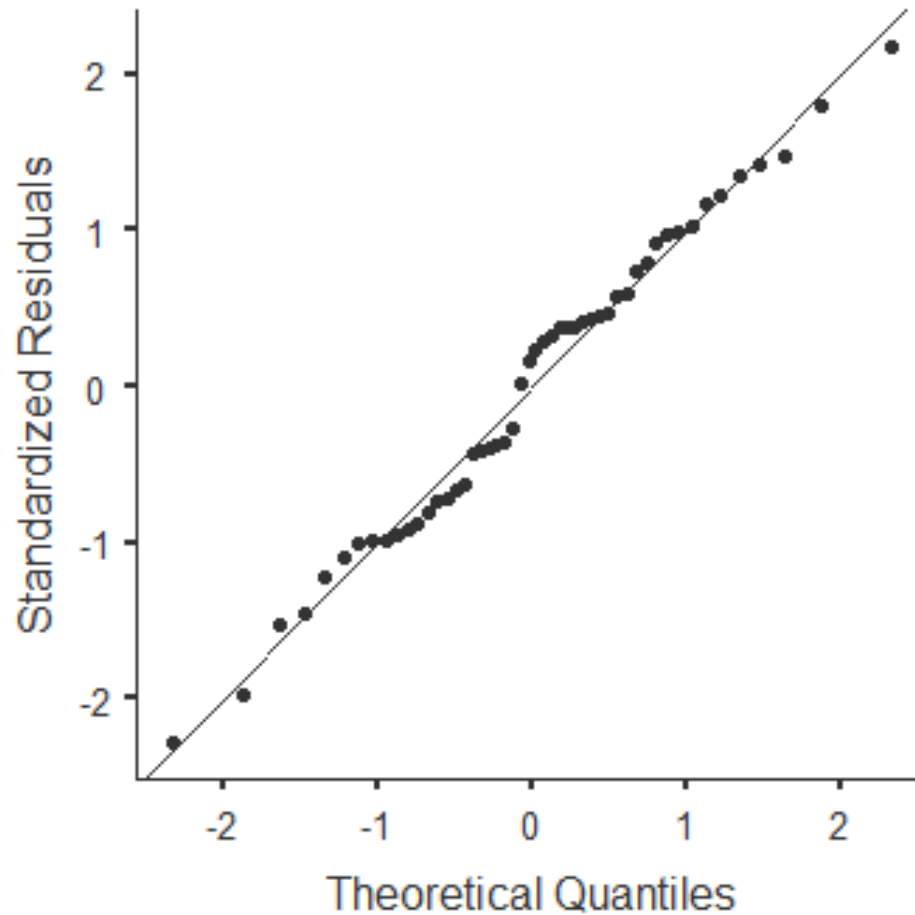
Homoscedasticity



Heteroscedasticity

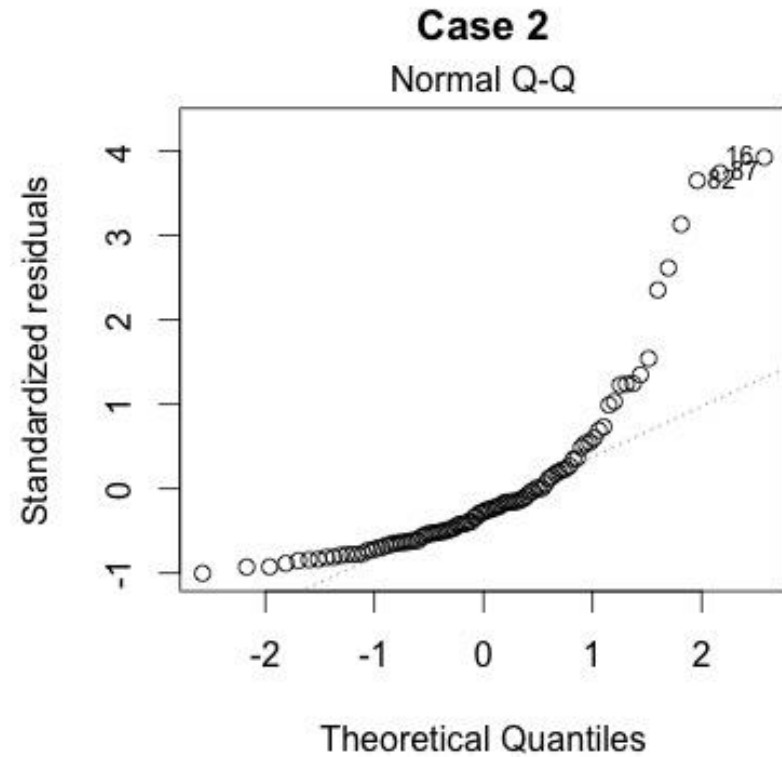
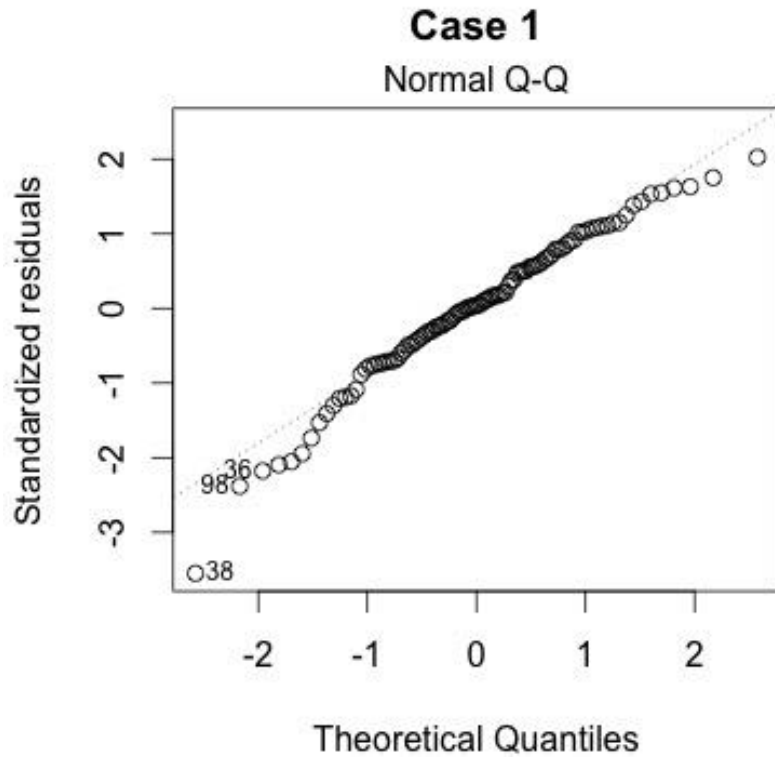


Kvantilový graf (Q-Q plot)



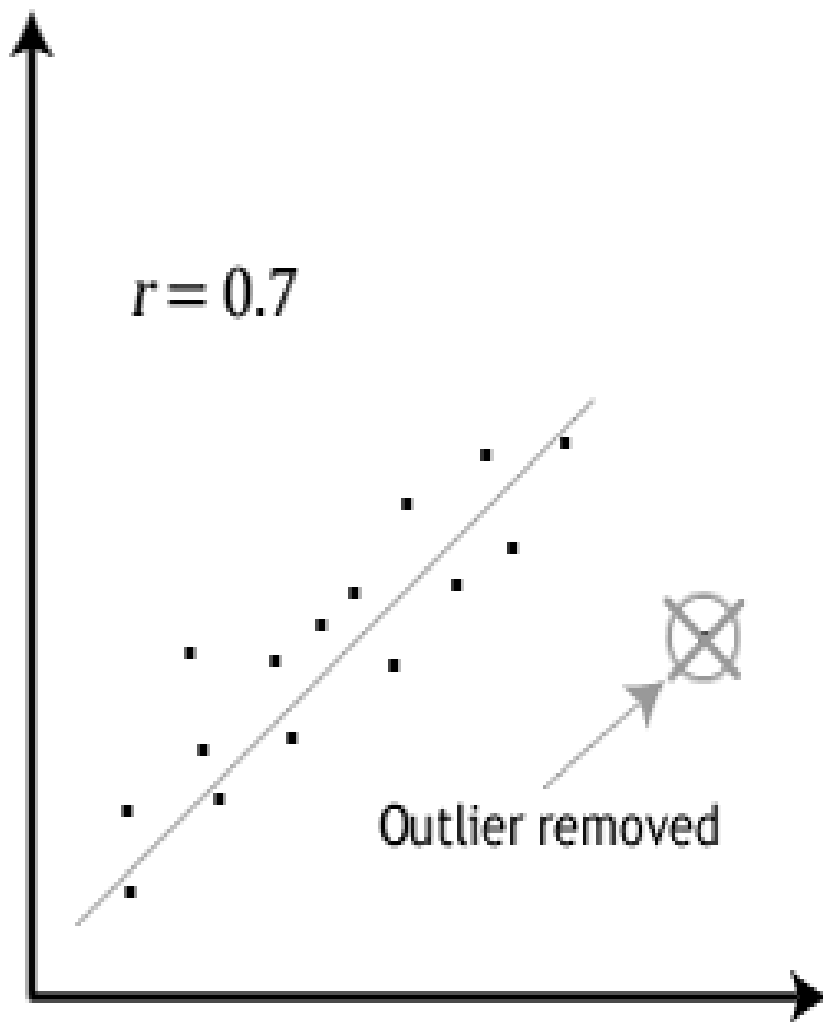
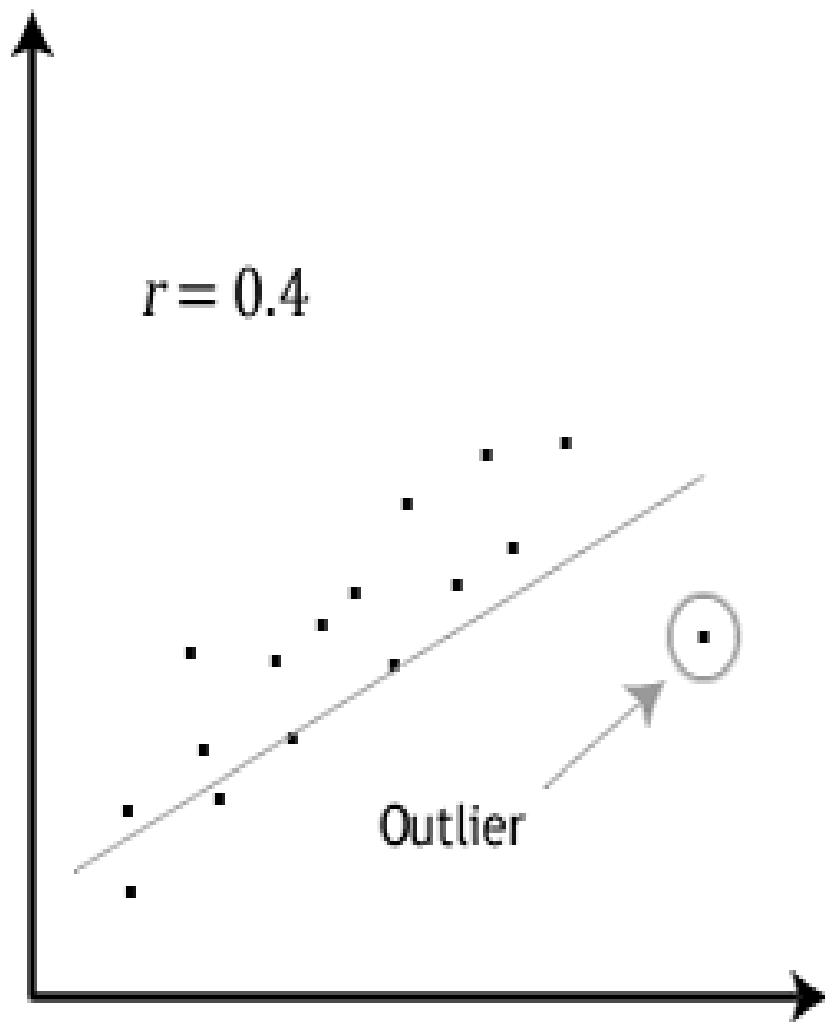
relig_prot x
obama2012

Kvantilový graf

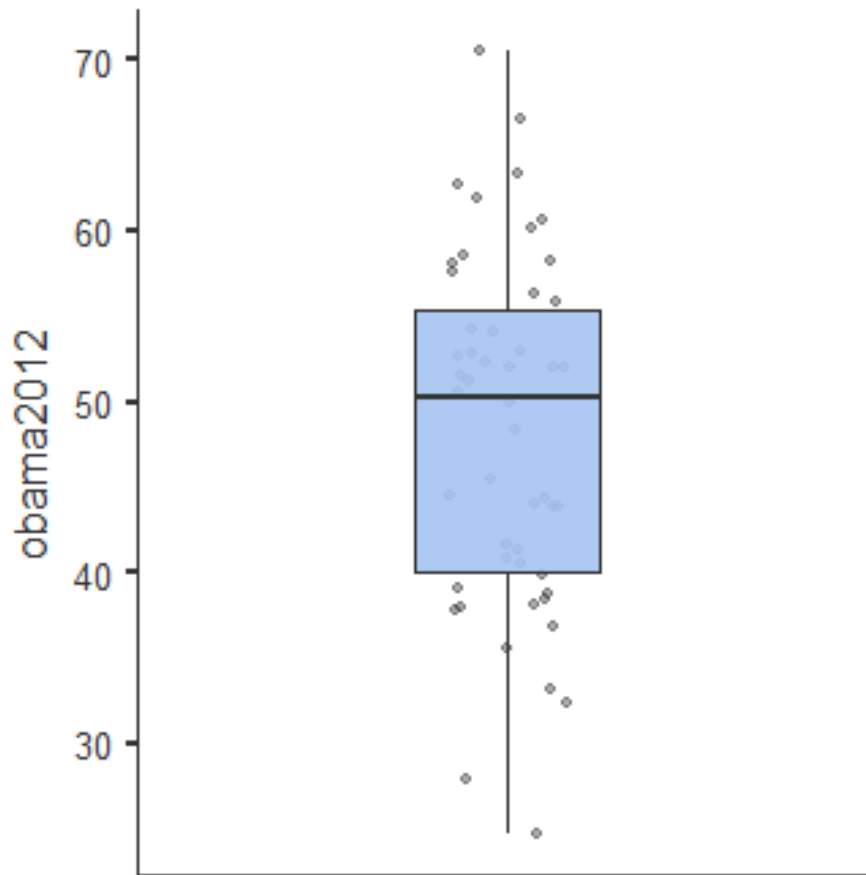


Pearsonovo r: předpoklady a omezení

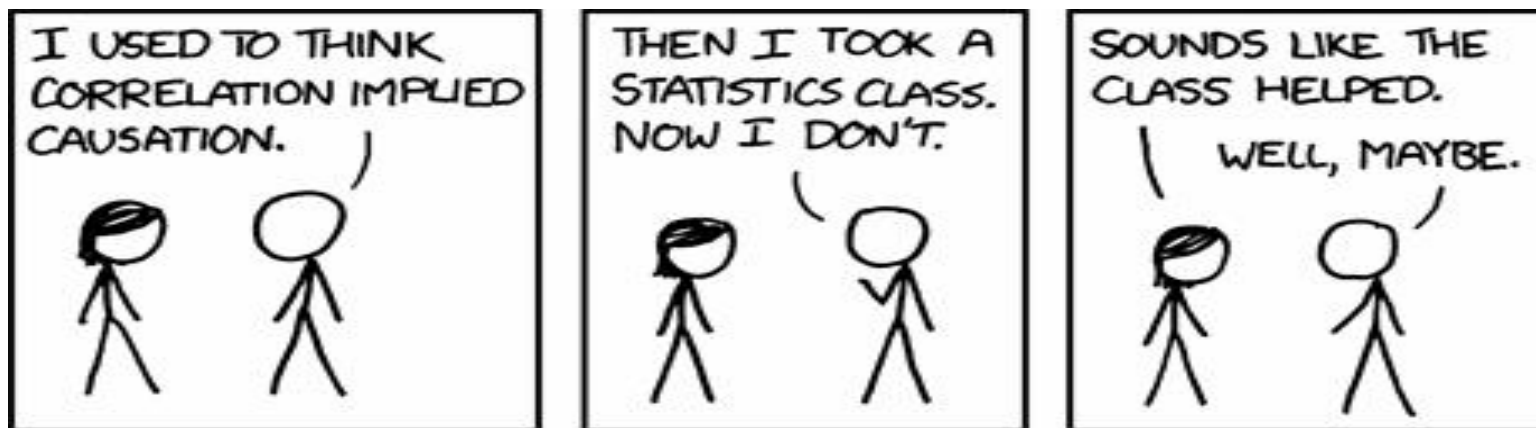
- Metrická úroveň (intervalová nebo poměrová) měření
- Lineární vztah mezi proměnnými X a Y
- Homoskedasticita (nezávislost rozptylu)
- **Citlivost vůči odlehlým hodnotám (outliers)**



Krabicový graf (box plot)



Korelace a kauzalita



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

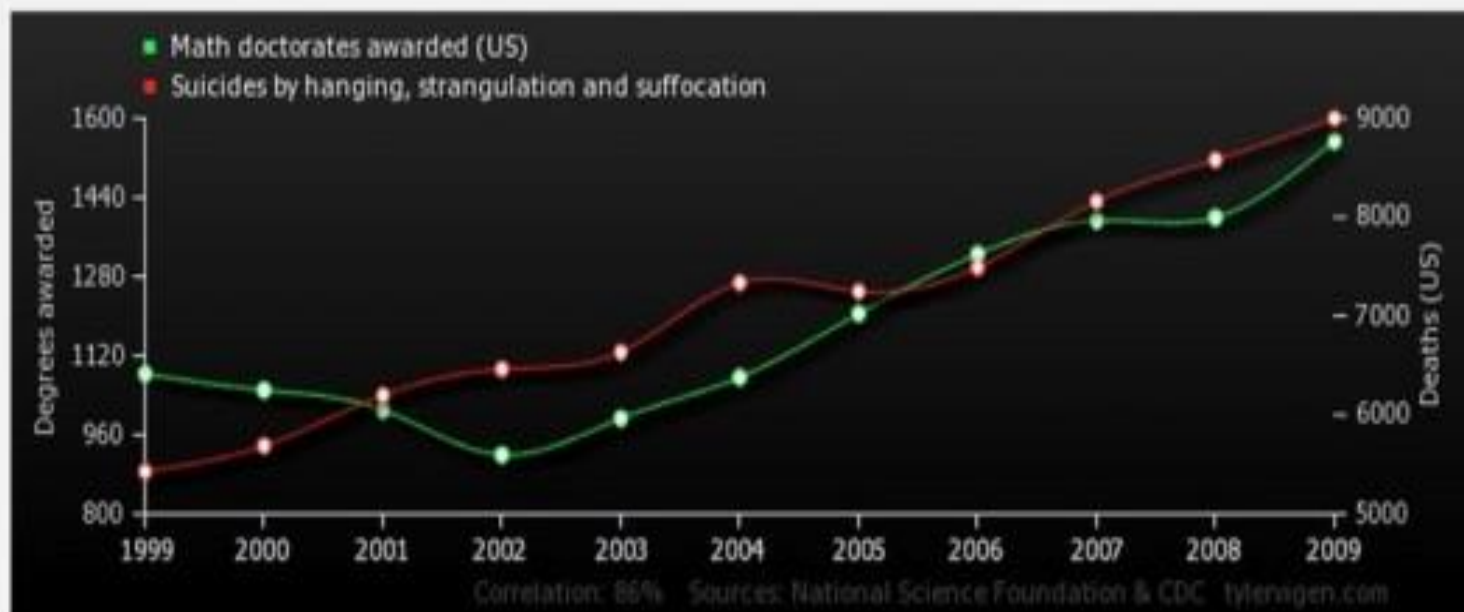
Obecné vzorce

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> Degrees awarded (National Science Foundation)	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> Deaths (US) (CDC)	5,427	5,688	6,198	6,462	6,635	7,336	7,248	7,491	8,161	8,578	9,000

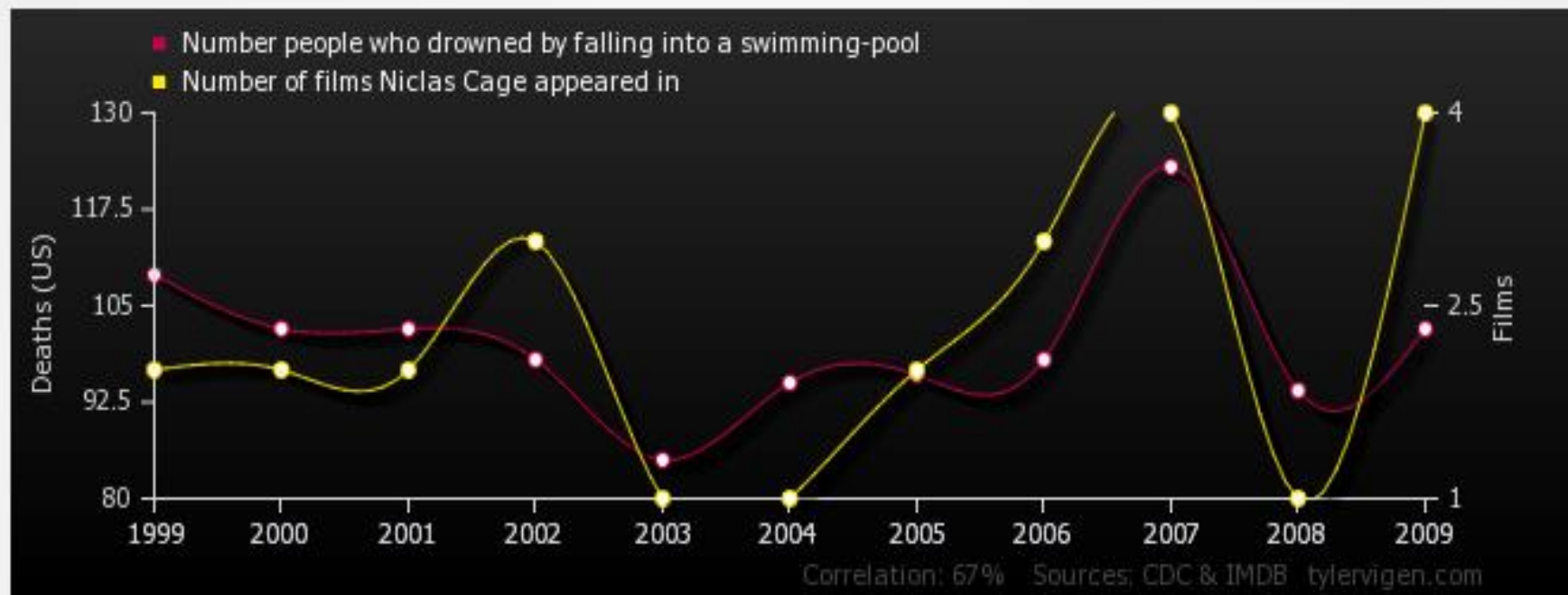
Correlation: 0.860176

Permalink - Not interesting

Number people who drowned by falling into a swimming-pool

correlates with

Number of films Nicolas Cage appeared in



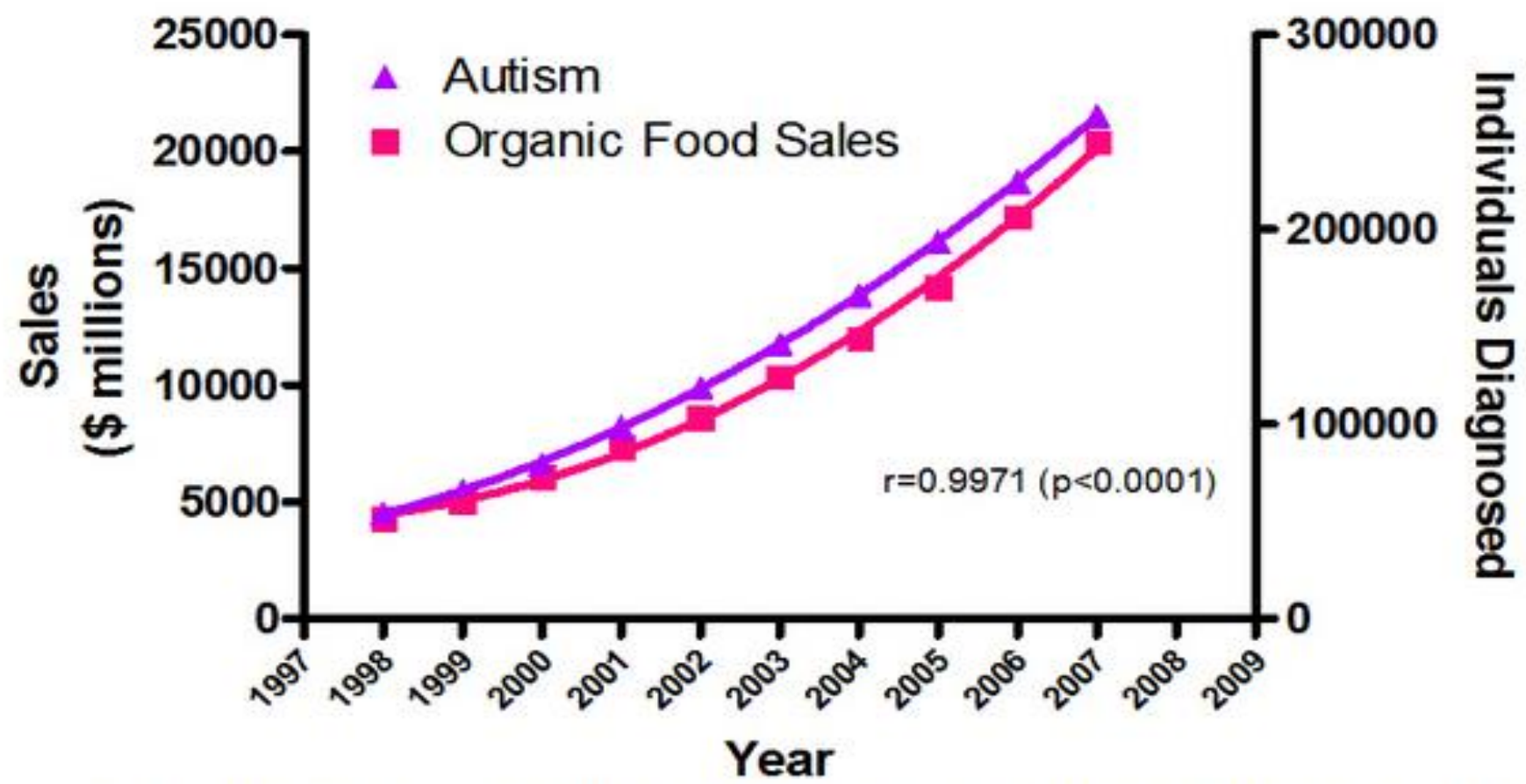
[Upload this image to imgur](#)

	<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

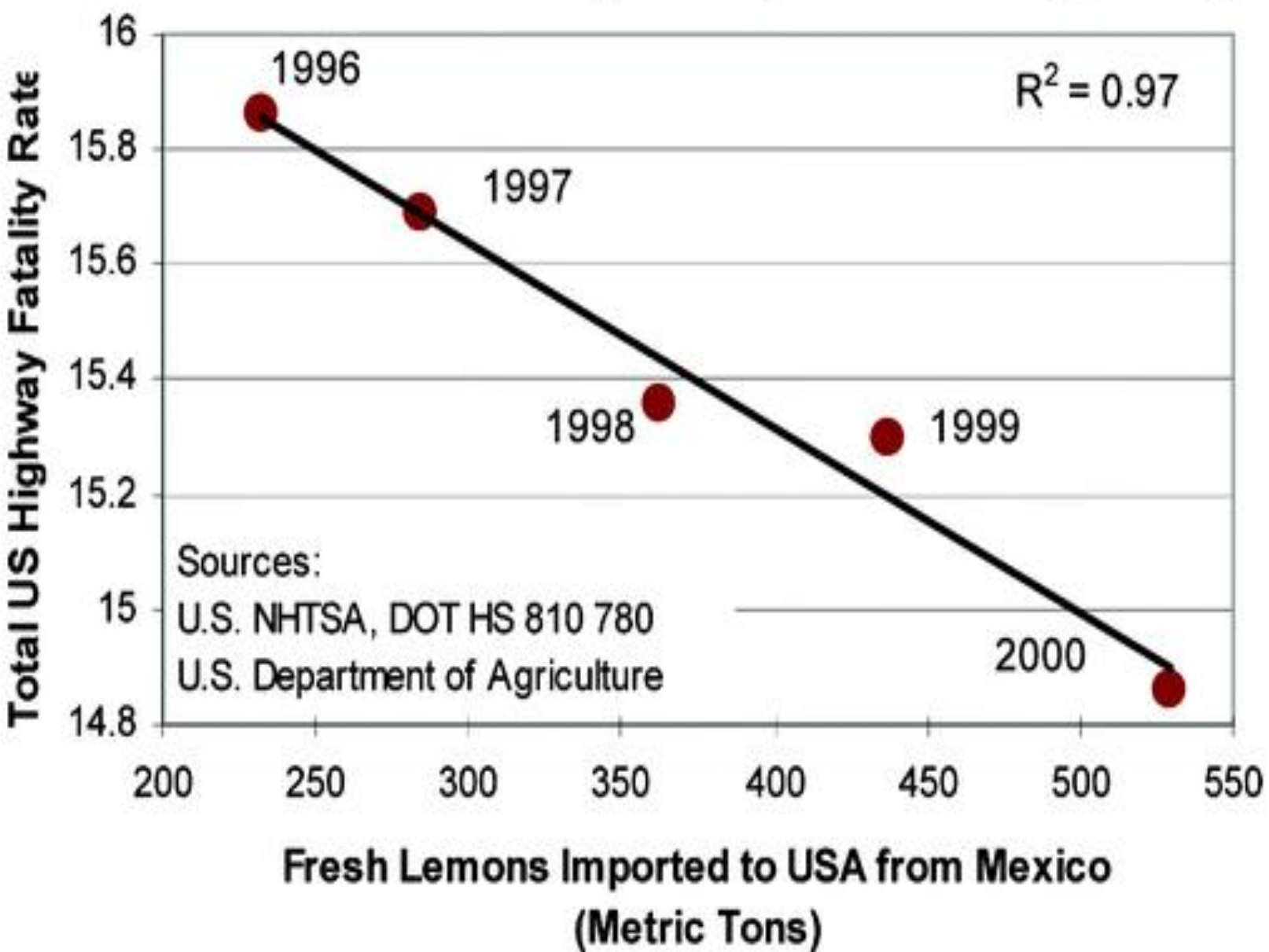
4. Eating organic food causes autism.

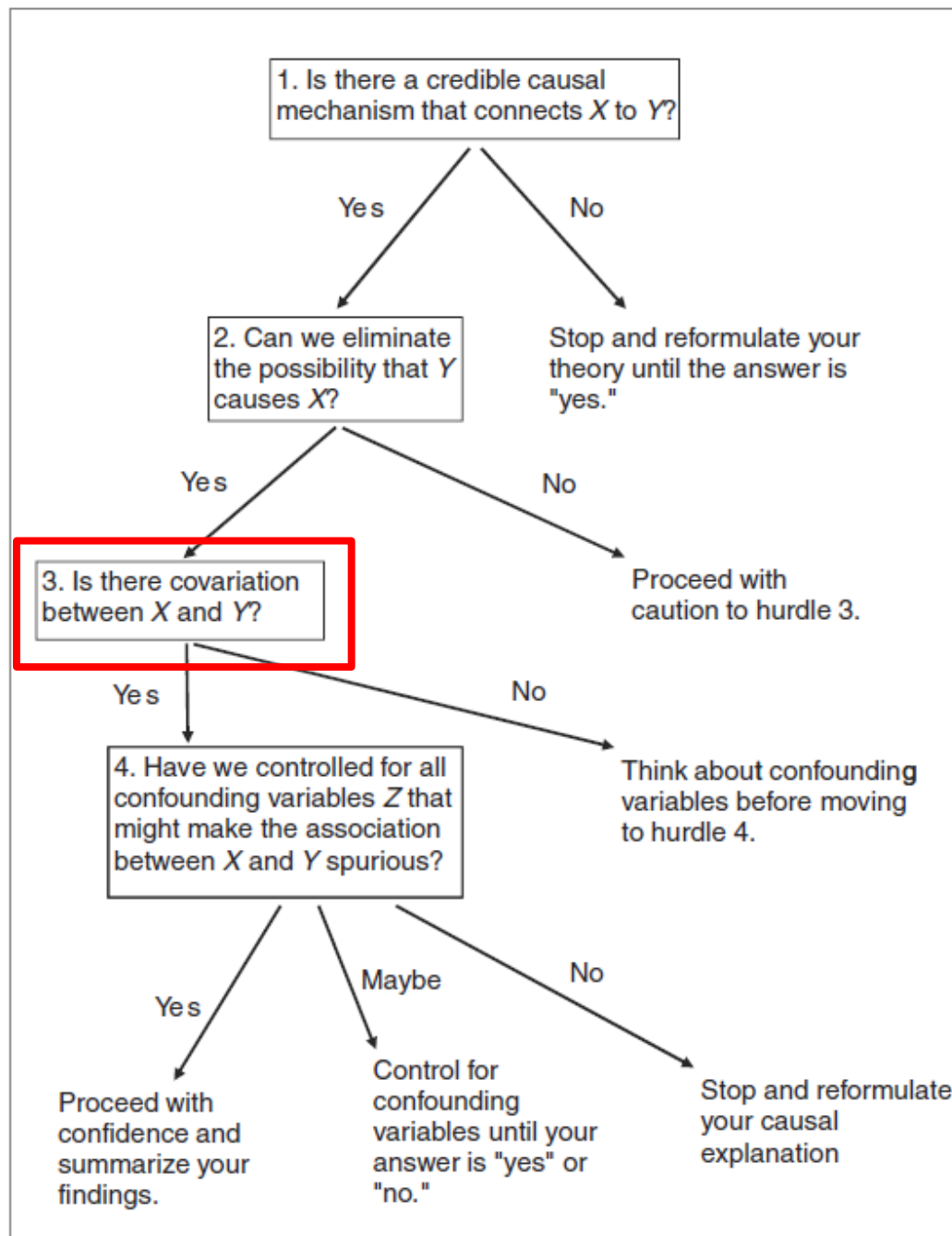
The real cause of increasing autism prevalence?



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

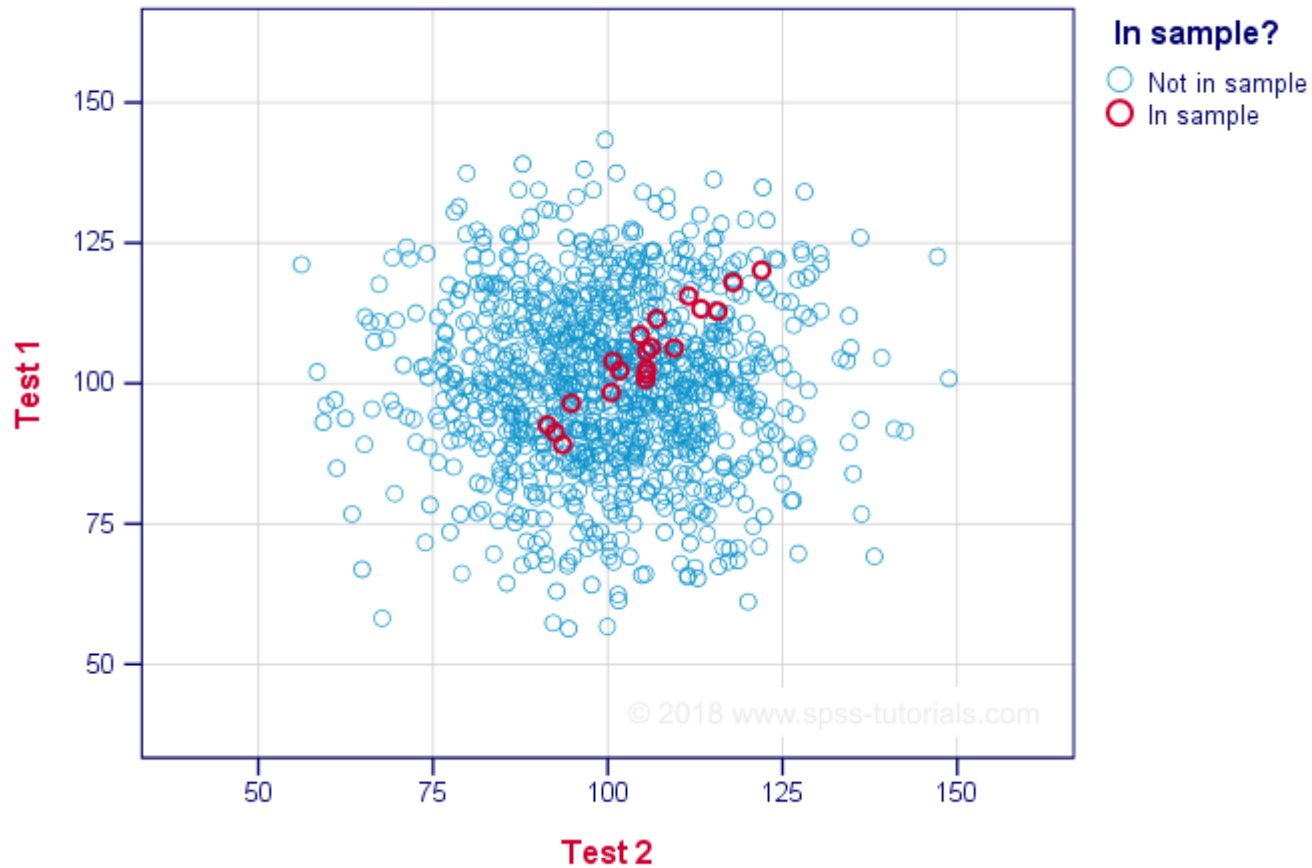
- H_0 : Mezi proměnnými X a Y *neexistuje* vztah lineární závislosti (korelace r)
- H_A : Mezi proměnnými X a Y *existuje* vztah lineární závislosti (korelace r)
- $H_0: r = 0 ; r \geq 0 ; r \leq 0$
- $H_A: r \neq 0$ (oboustranná hypotéza)
- $H_A: r > 0$ (jednostranná hypotéza, pozitivní korelace)
-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

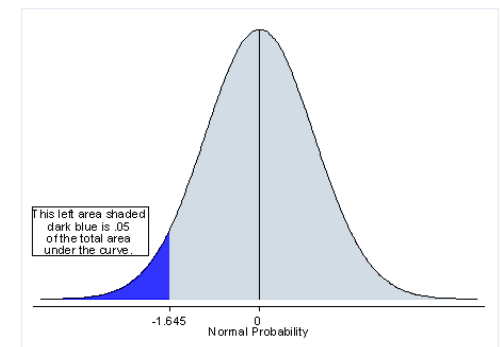
Sample 2 | N = 20

Sample Correlation = 0.95



Test hypotézy: Pearsonovo r

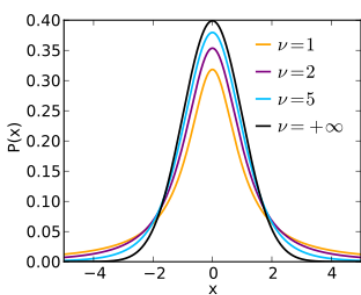
- **Teoretický rámec:** volební chování je do určité míry určeno sociokulturními štěpnými liniemi (Norris & Inglehart 2019; Lipset & Rokkan 1967)
- **H₀:** Mezi volebním výsledkem Obamy v roce 2012 (X) a podílem protestantů (Y) *není* korelace; $r(x, y) \geq 0$
- **H_A:** Mezi volebním výsledkem Obamy v roce 2012 (X) a podílem protestantů (Y) *je negativní* korelace; $r(x, y) < 0$



Test hypotézy: Pearsonovo r

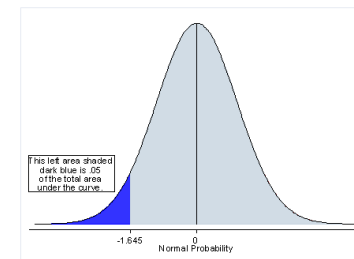
- **H₀**: Mezi volebním výsledkem Obamy v roce 2012 (X) a podílem protestantů (Y) *není* korelace; $r(x, y) \geq 0$
- **H_A**: Mezi volebním výsledkem Obamy v roce 2012 (X) a podílem protestantů (Y) *je negativní* korelace; $r(x, y) < 0$
- **Data**: 50 pozorování (státy USA)
- Stanovení **stupňů volnosti** 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

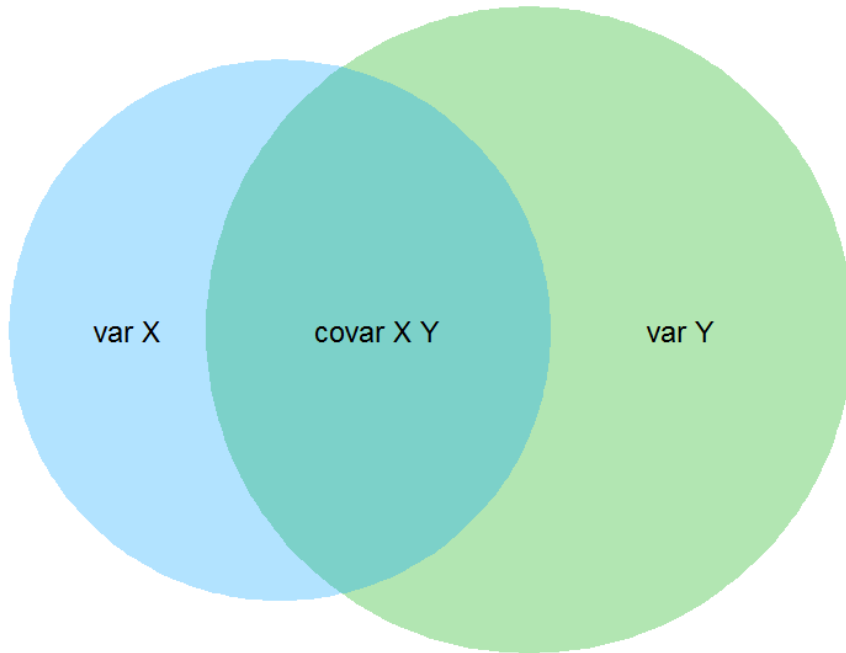
90% hodnota pro oboustranný test, tj. 95% pro jednostranný



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 Obama2012 a relig_prot: $r = -0.413$

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 (H0: $r \geq 0$)?*

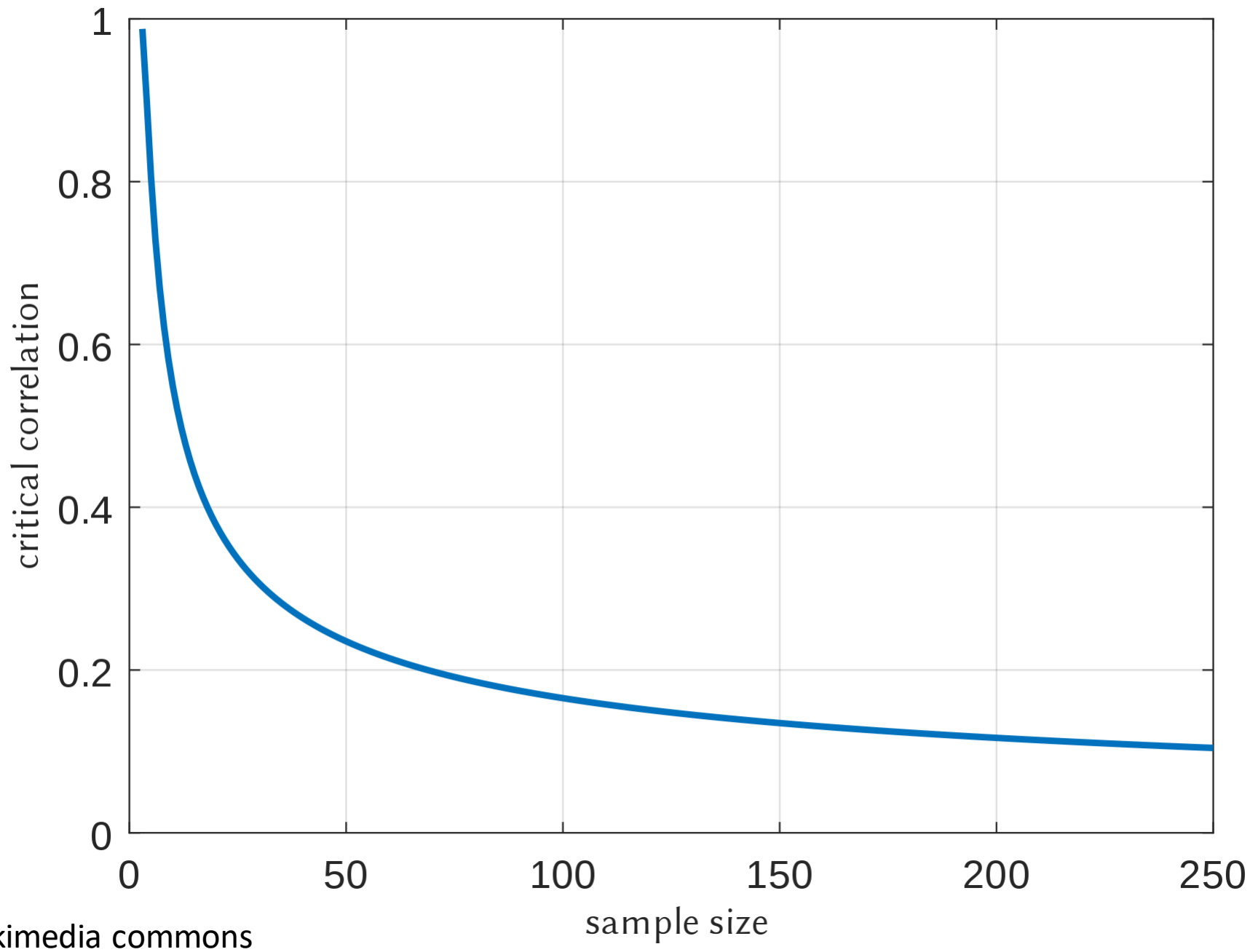
- $t = \frac{\text{signal}}{\text{noise}} ; t = \frac{r * \sqrt{n - 2}}{\sqrt{1 - r^2}} ; n = \text{velikost vzorku}$

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

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

Rozhodnutí o H_0

- **H_0** : Mezi volebním výsledkem Obamy v roce 2012 (X) a podílem protestantů (Y) *není* korelace; $r(x, y) \geq 0$
- **H_A** : Mezi volebním výsledkem Obamy v roce 2012 (X) a podílem protestantů (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 $t = -3.14 < (\alpha = 0.05, 48) = -1.677$, **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 volebním výsledkem Obamy (X) a podílem protestantů (Y) je, na 5% hladině st. významnosti, lineární závislost (korelace r).



wikimedia commons

sample size

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/>.



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R



Modules

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

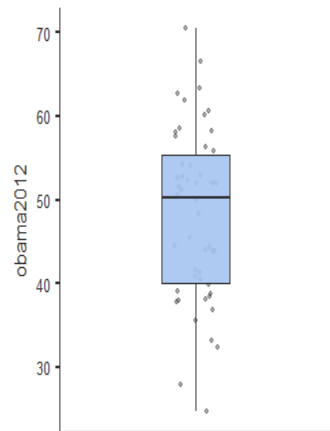
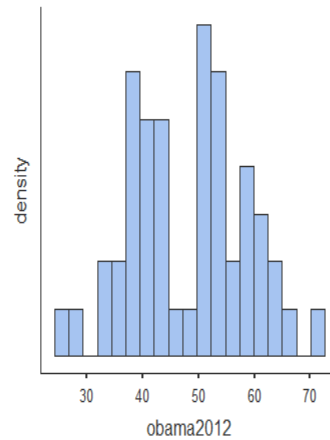
Plots

Histograms 
 Histogram
 Density
Box Plots 
 Box plot
 Violin
 Data

Bar Plots 
 Bar plot
Q-Q Plots 
 Q-Q

Plots

obama2012



relig_prot

Descriptives



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

- Variables
- obama2012
 - relig_prot

Split by

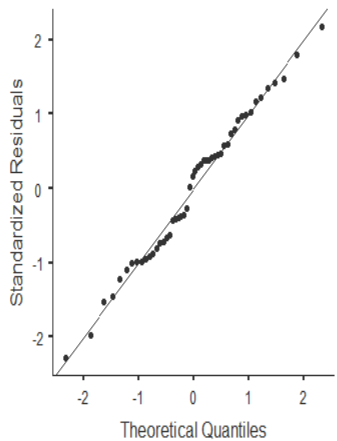
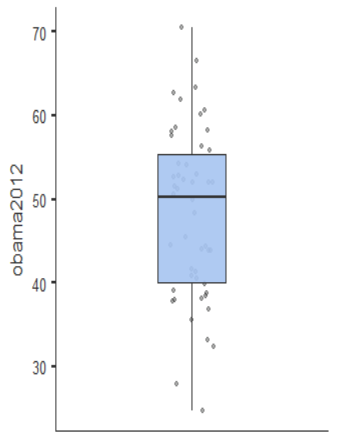
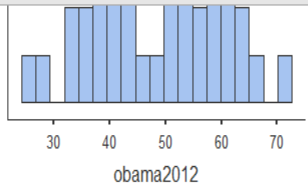
Frequency tables

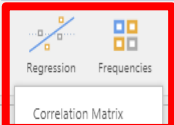
Statistics

- 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
- Bar Plots**
- Bar plot
- Q-Q Plots**
- Q-Q
- Jittered





Descriptives

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

Frequency tables

Statistics

Sample Size

N Missing

Percentile Values

Quartiles
 Cut points for 4 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

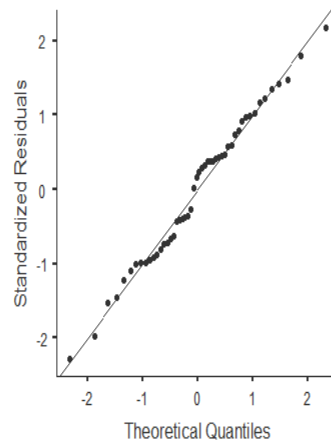
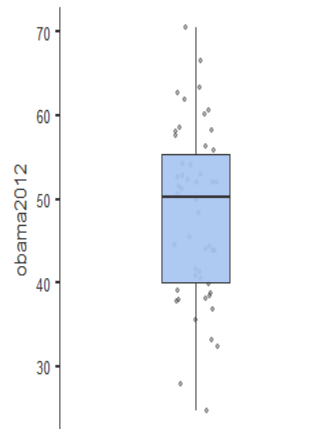
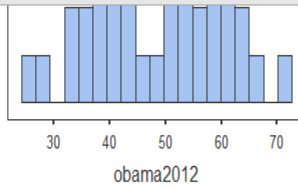
Box plot
 Violin
 Data

Bar Plots

Bar plot

Q-Q Plots

Q-Q Jittered



Correlation Matrix



ns_or_more
 pop2000
 pop2010
 pop2010_hun_thou
 popchg0010
 popchgpc
 relig_cath
 relig_high
 relig_low

→

obama2012
 relig_prot

Correlation Coefficients

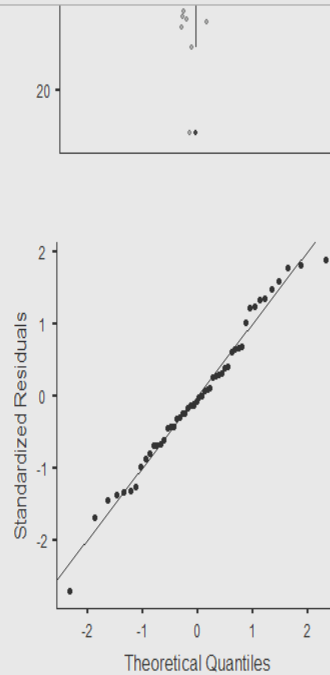
Additional Options

- 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

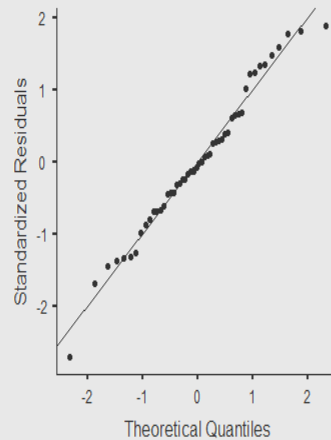
Correlated negatively

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

Hypothesis

- Correlated
 Correlated positively
 Correlated negatively

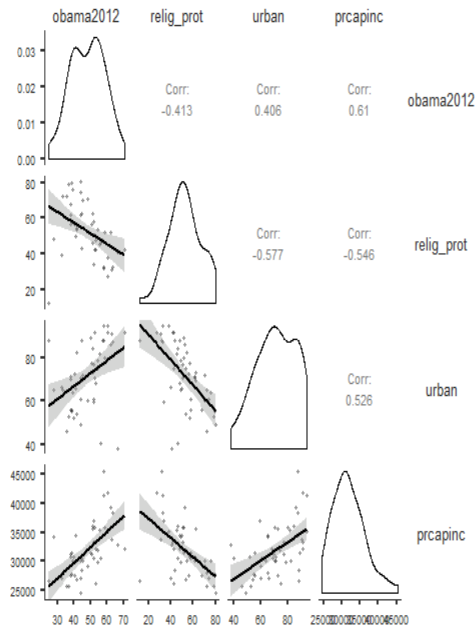
Plot

- 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 $n-1$

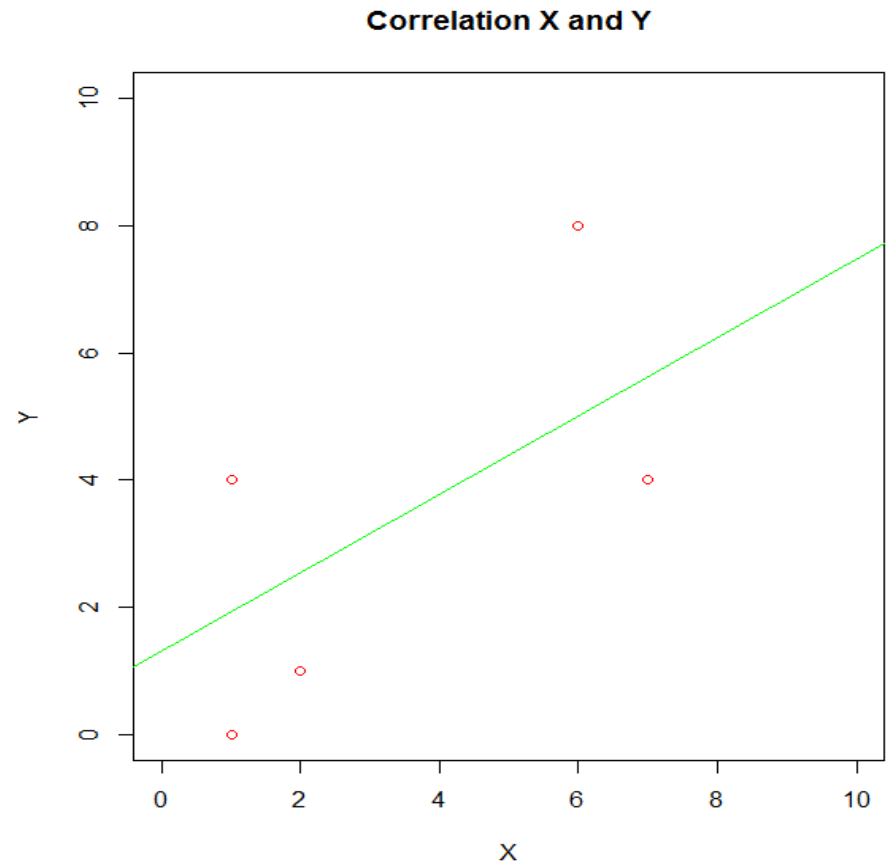
(x - m)	(y - m)	cross-prod.
(1 - 3.4)	(0 - 3.4)	8.16
(2 - 3.4)	(1 - 3.4)	3.36
(1 - 3.4)	(4 - 3.4)	-1.44
(6 - 3.4)	(8 - 3.4)	11.96
(7 - 3.4)	(4 - 3.4)	2.16
0	0	24.2

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

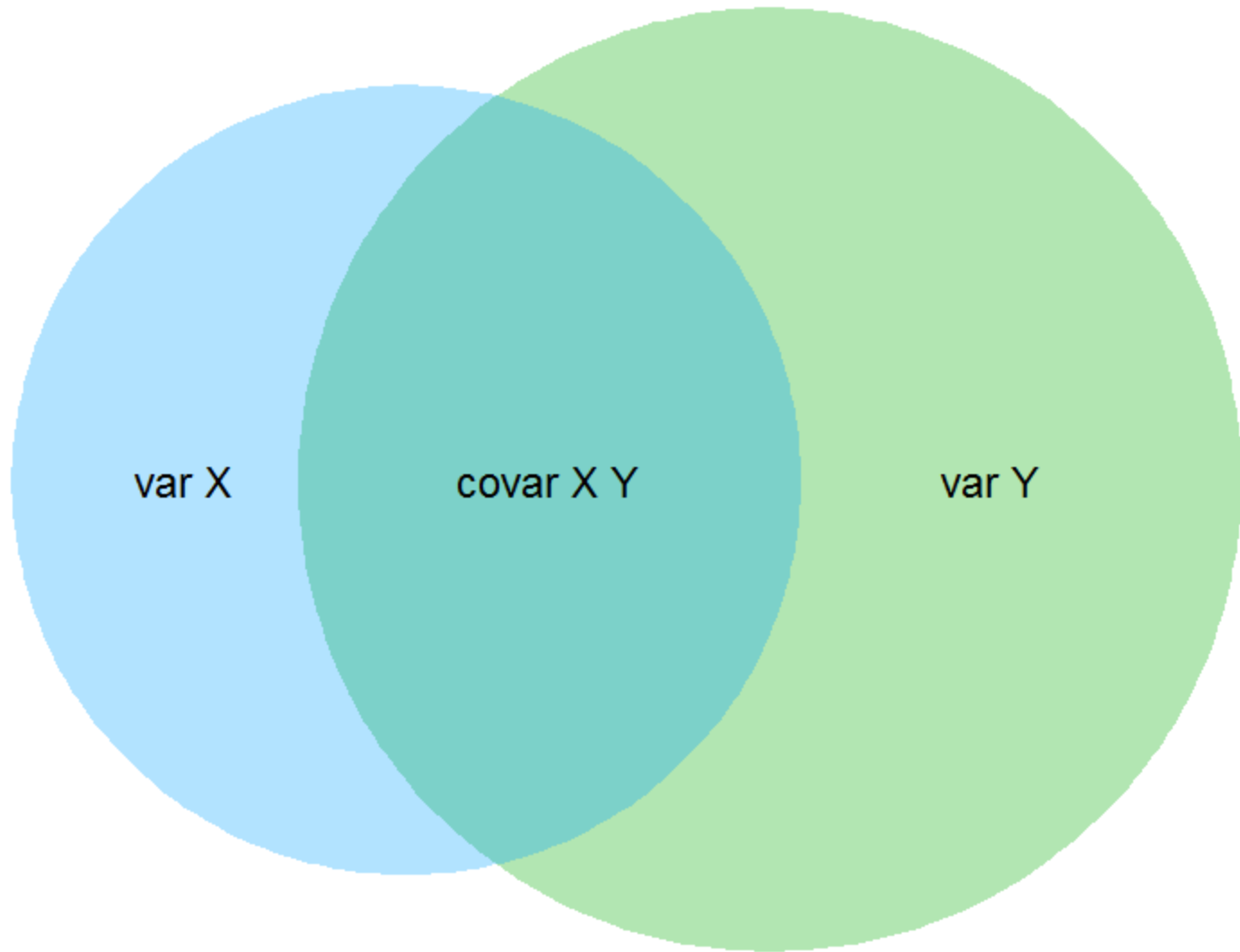
$$\text{COV}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- (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