

Korelace

# Korelace a kovariance

- Korelace:

- Míra asociace dvou proměnných
- Ukazatel síly efektu (*effect size*)
- Statistická korelace neznamená nutně kauzalitu

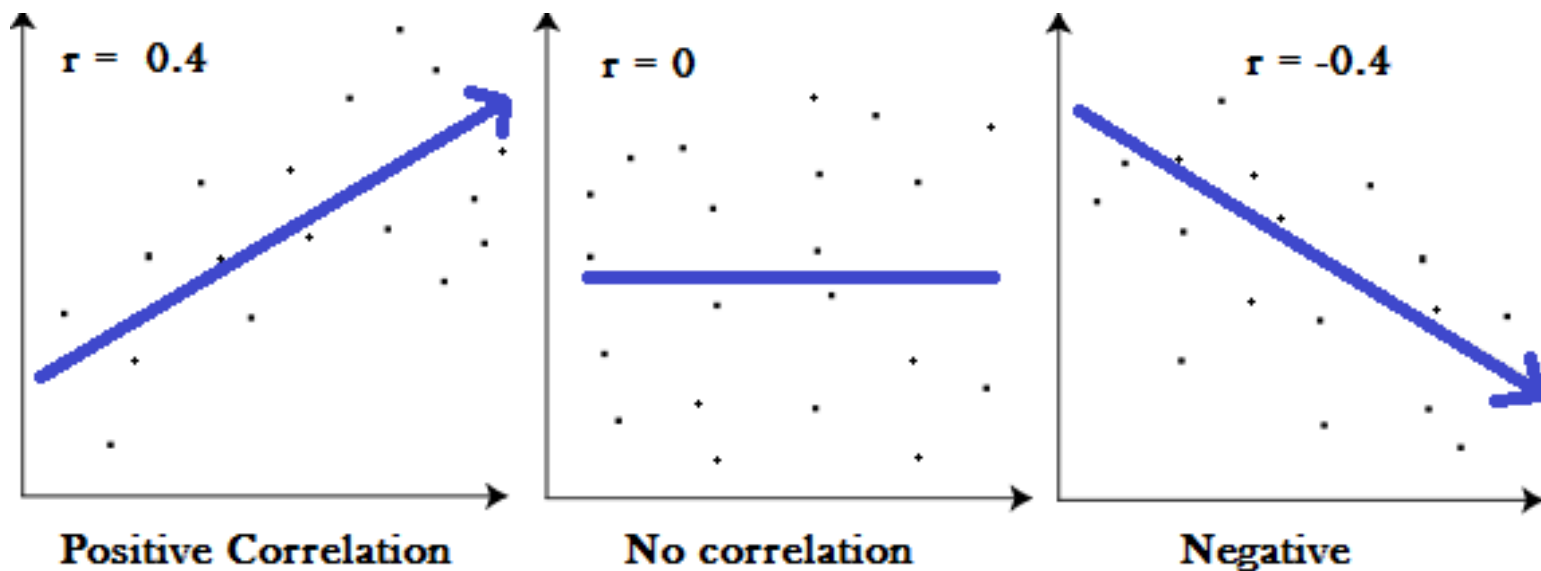
- Kovariance:

- Míra společného rozptylu dvou kardinálních proměnných
- $$cov = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$$
- Vyšší číslo = větší těsnost vztahu
- Ale výsledná velikost záleží na jednotkách (tzn. není standardizovaná), nemá horní hranici
- Korelační koeficienty pro ordinální proměnné později (např. gamma, Kendalovo tau)

# Pearsonův a Spearmanův korelační koeficient

- Standardizované míry těsnosti vztahu mezi dvěma kardinálními proměnnými
- Standardizovaná kovariance, resp. kovariance v jednotkách směrodatné odchylky
- Vychází v intervalu  $\langle -1;1 \rangle$
- $-1$  = dokonale negativní vztah
- $0$  = žádný vztah
- $1$  = dokonale pozitivní vztah

# Pearsonův a Spearmanův korelační koeficient



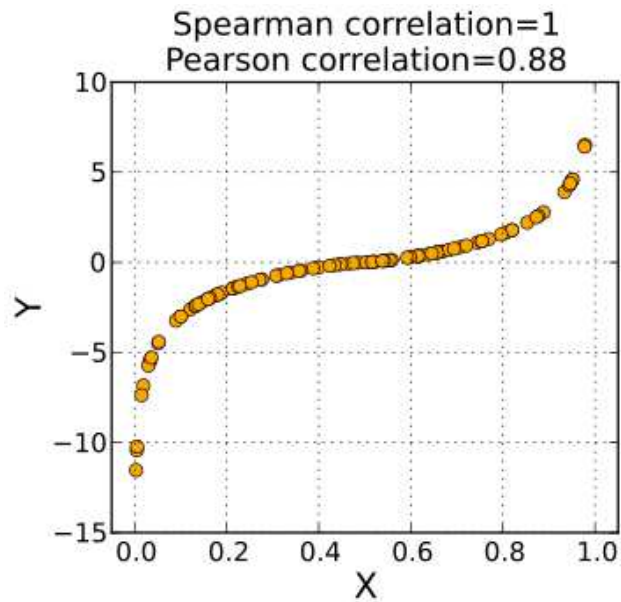
# Pearsonův a Spearmanův korelační koeficient

- Pearsonův koeficient ( $r$ ):
  - Standardizovaná míra lineárního vztahu
  - $r = \frac{\textit{kovariance}}{\textit{Směr. odchylka}_{XY}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)SD_xSD_y}$
  - V intervalu  $\langle -1;1 \rangle$
  - $r^2$  = procento sdíleného rozptylu
    - Např.  $r = 0,3$  -> proměnné sdílejí 9% rozptylu
- Předpoklady:
  - Koeficient samotný vyžaduje pouze kardinalitu proměnných
  - ALE silně ovlivněn odlehlými pozorováními
  - Pro statistické testování (a maximum informací?) vyžadována normalita rozdělení a stejnost rozptylu

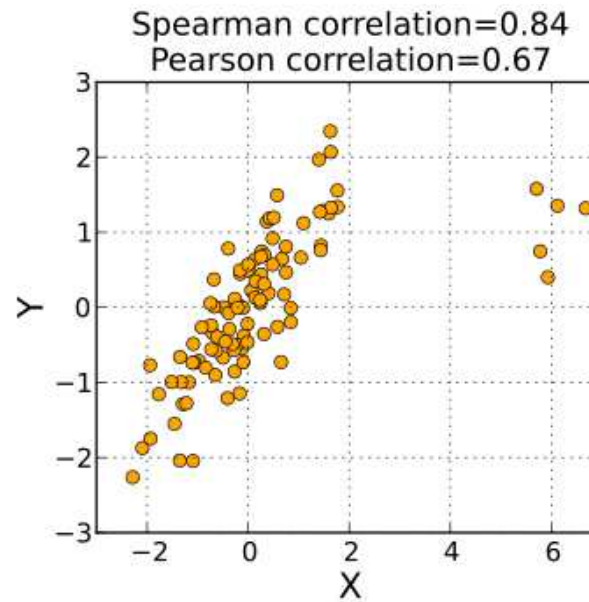
# Pearsonův a Spearmanův korelační koeficient

- Spearmanův korelační koeficient
  - Standardizovaná míra monotónního vztahu (tzn. ne pouze lineárního)
  - Počítá se s pořadím, jinak vzorec stejný jako pro Pearsonův koeficient
  - V intervalu  $\langle -1; 1 \rangle$
  - Méně citlivý na odlehlá pozorování než Pearson
  - Ideálně kardinální data (teoreticky zvládne i ordinální, ale existují lepší alternativy)
  - Pro výpočet statistické signifikance stále nutná normalita rozdělení a stejnost rozptylů

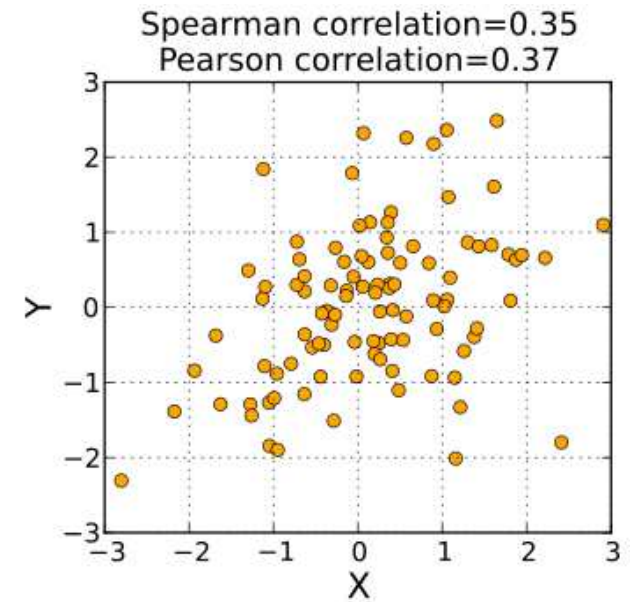
# Pearson vs. Spearman



Lineární  
vs  
Monotónní vztah



Odlehlá pozorování



# Parciální korelace

- Standardizovaná míra těsnosti dvou proměnných, pokud je třetí kardinální proměnná konstantní
- Vztah dvou proměnných, pokud kontrolujeme třetí
- $$r_{xy.z} = \frac{r_{xy} - (r_{xz} * r_{yz})}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$
- Např. Jak silný je vztah mezi věkem a příjmem, pokud je rok narození konstantní?



# Test nekorelovanosti

- Testuje, zda je koeficient statisticky významně odlišný od nuly
- $H_0$ : Rozdíl koeficientu od nuly je nulový
- $t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$
- Předpoklady:
  - Normalita rozdělení
  - Stejnost rozptylů
- V praxi chceme  $H_0$  (většinou) zamítnout

# Test shody koeficientů

- Testuje, zda jsou dva korelační koeficienty jsou stejné
- $H_0$ : Rozdíl mezi prvním a druhým koeficientem je nulový

- Obdoba T testu založená na z skórech koeficientů

$$Z = \frac{z(r_1) - z(r_2)}{SE_{r_1 r_2}} = \frac{z(r_1) - z(r_2)}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

$$z(r) = \frac{1}{2} \log_e \left( \frac{r + 1}{r - 1} \right)$$

- Výsledek porovnat s hladinou kritického oboru
- Např. Liší se korelace mezi věkem a příjmem mezi muži a ženami?

# Korelace v SPSS

The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar indicates the file name is \*V1706a\_F1.sav [DataSet1]. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, and Window. The Analyze menu is open, showing various statistical options. The 'Correlate' option is highlighted in yellow. A sub-menu is open for 'Correlate', showing 'Bivariate...', 'Partial...', and 'Distances...'. The main data editor window is visible in the background, showing a list of variables with their names and types.

	Name	Type
1	CD	Numeric
2	NUTS3	Numeric
3	VSO	Numeric
4	OCT	Numeric
5	KRAJ	Numeric
6	IDE_9	Numeric
7	IDE_8	Numeric
8	IDE_2	Numeric
9	t_VEK_5	Numeric
10	t_VZD	Numeric
11	IDE_6B	Numeric

Analyze -> Correlate

Bivariate = klasická korelace

Partial = parciální korelace

# Korelace v SPSS

Vztah mezi věkem a osobním příjmem

Correlations

		IDE.2 Věk	IDE.10a Osobní čistý měsíční příjem
IDE.2 Věk	Pearson Correlation	1	-,175**
	Sig. (2-tailed)		,000
	N	982	602
IDE.10a Osobní čistý měsíční příjem	Pearson Correlation	-,175**	1
	Sig. (2-tailed)	,000	
	N	602	603

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Correlations

			IDE.2 Věk	IDE.10a Osobní čistý měsíční příjem
Spearman's rho	IDE.2 Věk	Correlation Coefficient	1,000	-,250**
		Sig. (2-tailed)	.	,000
		N	982	602
	IDE.10a Osobní čistý měsíční příjem	Correlation Coefficient	-,250**	1,000
		Sig. (2-tailed)	,000	.
		N	602	603

\*\* . Correlation is significant at the 0.01 level (2-tailed).

