Korelace

POLb 1139

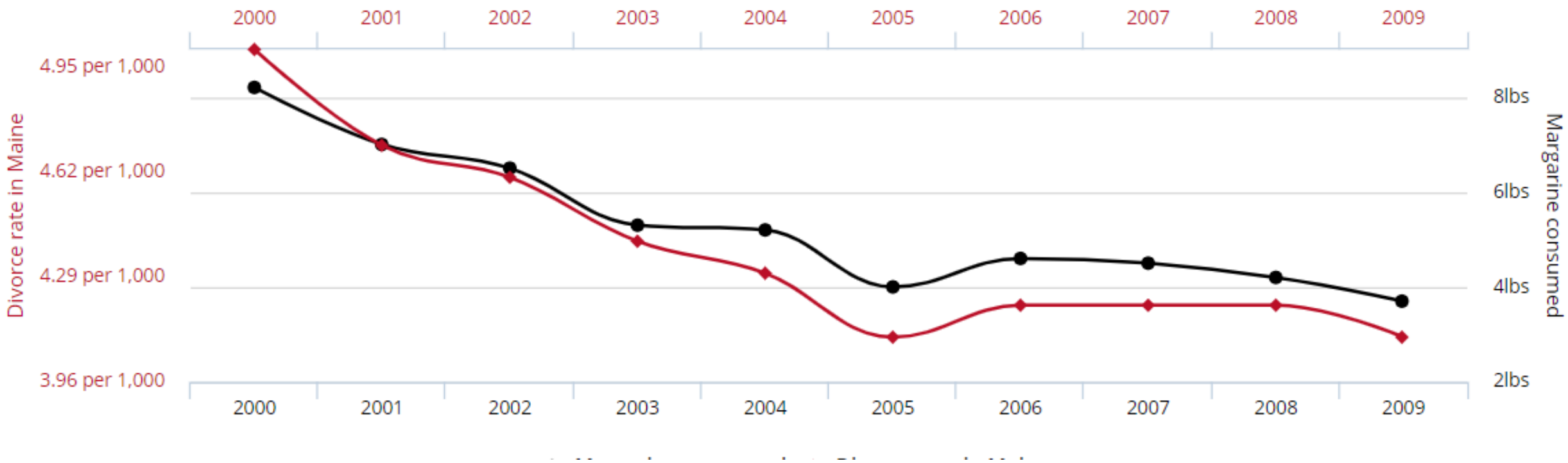
Peter Spáč

Korelace - základy

- Vzájemná souvislost mezi proměnnými
- Nárůst hodnot jedné proměnné je spojený s nárůstem (anebo poklesem) hodnot druhé proměnné
- Korelace neimplikuje kauzalitu

Divorce rate in Maine
correlates with
Per capita consumption of margarine

Correlation: 99.26% (r=0.992558)



Korelace - druhy

- Bivariační – souvislost mezi dvěma proměnnými
- Parciální (partial) – souvislost mezi dvěma proměnnými za určité kontroly vlivu jiných proměnných
- JASP umožňuje pouze bivariační korelaci
- Pro kontrolu vlivu třetích proměnných typicky využíváme pokročilejší techniky (regrese)

Korelace a proměnné

- Podoba analýzy závisí na typu proměnných

- Kontingenční tabulky (crosstabs):
  - Dvě kategorické proměnné – nominální, ordinální
  - Nižší počet kategorií v proměnných (podmínka jsou minimálně dvě)

- Korelace (correlation):
  - Dvě kardinální proměnné, kardinální a ordinální, dvě ordinální
  - Specifický případ – kardinální a dichotomická proměnná

Bivariační korelace

- Posuzuje souvislost mezi dvěma proměnnými bez dalšího

- Tři základní postupy:
  - Pearsonův korelační koeficient
  - Spearmanovo rho
  - Kendallovo tau

- Odlišné podmínky a předpoklady

Pearsonův korelační koeficient

- Parametrický postup

- Předpoklady:
  - Kardinální data (možná výjimka – kardinální a binární proměnná)
  - Pokud zjišťujeme i statistickou signifikanci, tak i normální rozložení (nebo dostatečná velikost vzorku)

- Citlivost na odlehlé případy (outliers)

Pearsonův korelační koeficient

- Vychází z výpočtu rozptylu, pro srovnatelnost se standardizuje

- Hodnoty koeficientu:
  - Rozsah od -1 po 1
  - +1 = perfektní kladná souvislost
  - -1 = perfektní záporná souvislost
  - 0 = žádná souvislost

- Čím více je hodnota vzdálena od nuly, tím je souvislost silnější

Pearsonův korelační koeficient

- Vychází z výpočtu rozptylu, pro srovnatelnost se standardizuje

- Síla vztahu:
  - ± 0,1 – slabý
  - ± 0,3 – střední
  - ± 0,5 – silný

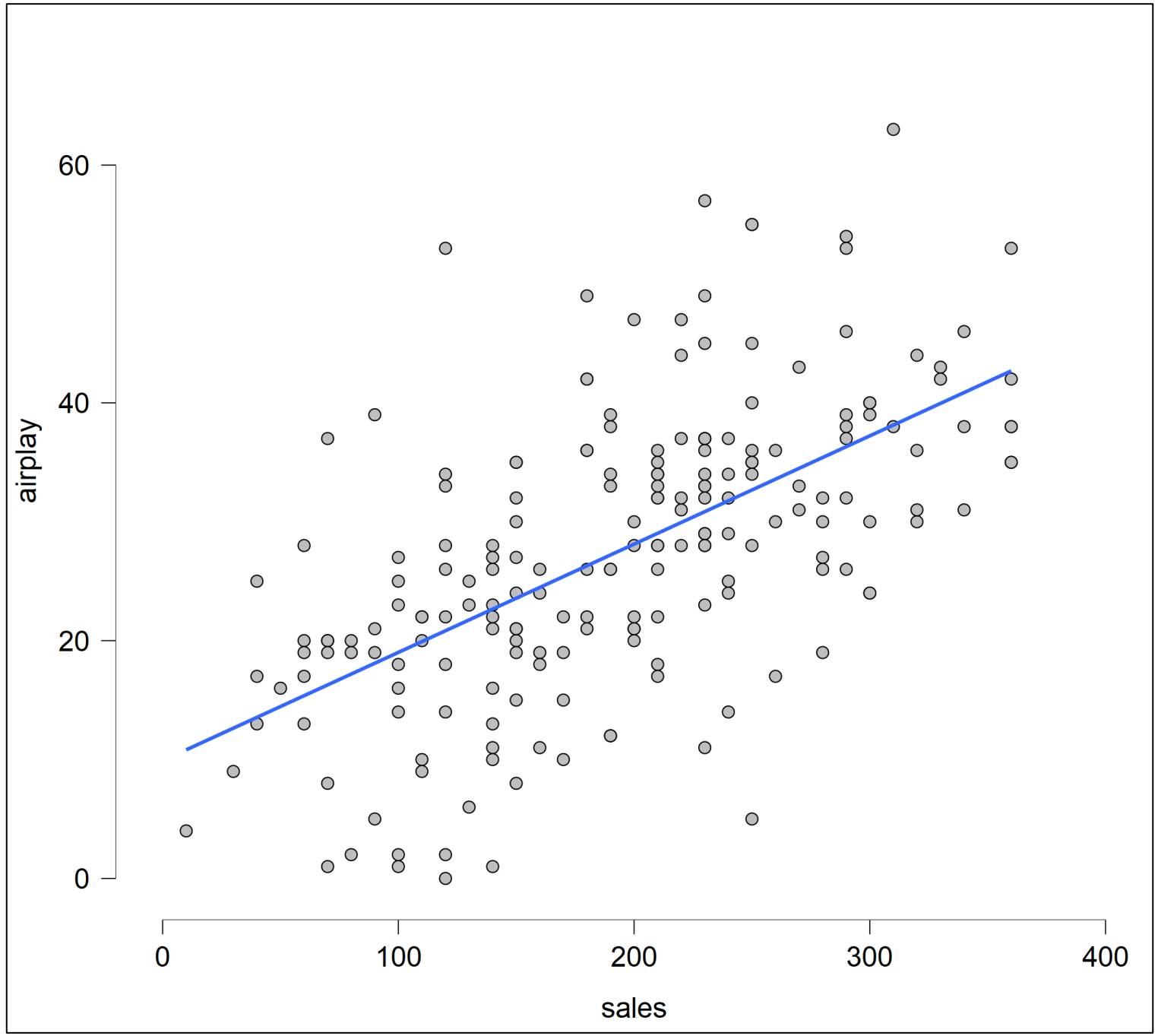
- Spíše orientační, podstatnější je věcný význam zjištění

Postup

- 1. Zhodnocení a kontrola dat
  - Podmínky pro jednotlivé koeficienty
  - Vizualizace dat

- 2. Výběr vhodného korelačního koeficientu

- 3. Spočítání výsledků a jejich (především věcná) interpretace
  - Existence vztahu, směr, síla
  - Možnost zobecnění na populaci


60
40
20
0
airplay
0
100
200
300
400
sales

Pearson's Correlations

Variable		sales	airplay
1. sales	Pearson's r	—	
	p-value	—	
2. airplay	Pearson's r	0.599	—
	p-value	< .001	—

Pearson's Correlations

			Pearson	
			r	p
sales	-	airplay	0.599***	< .001

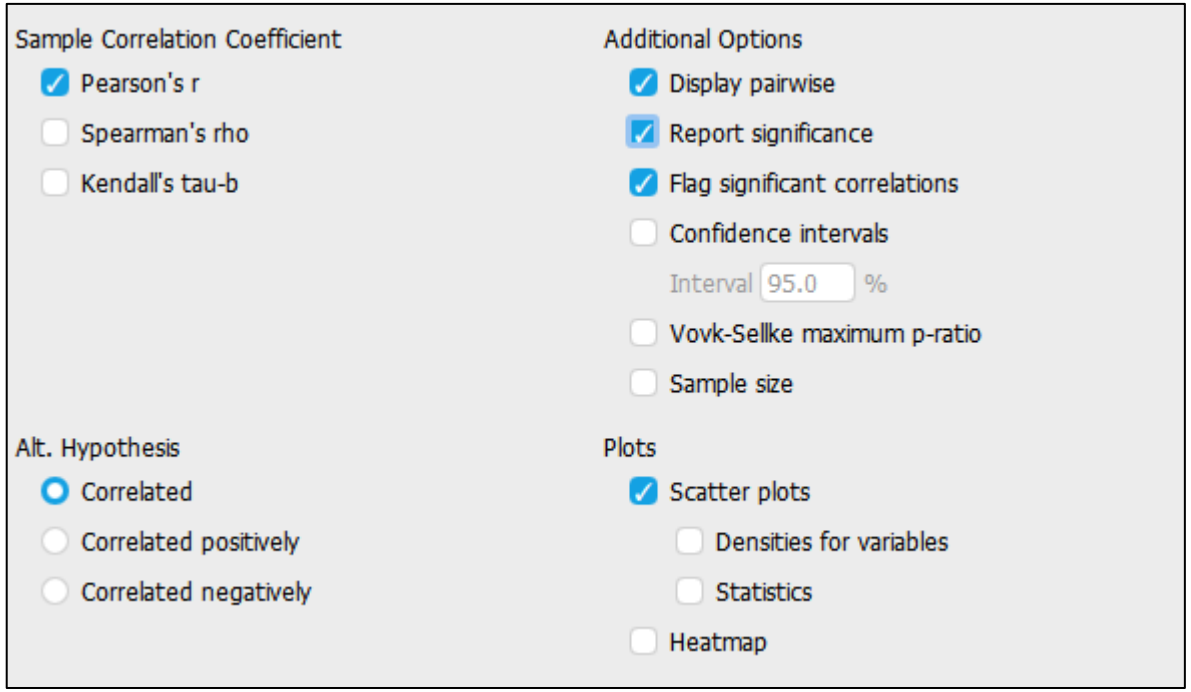
* p < 0.5, ** p < 0.01, *** p < 0.001

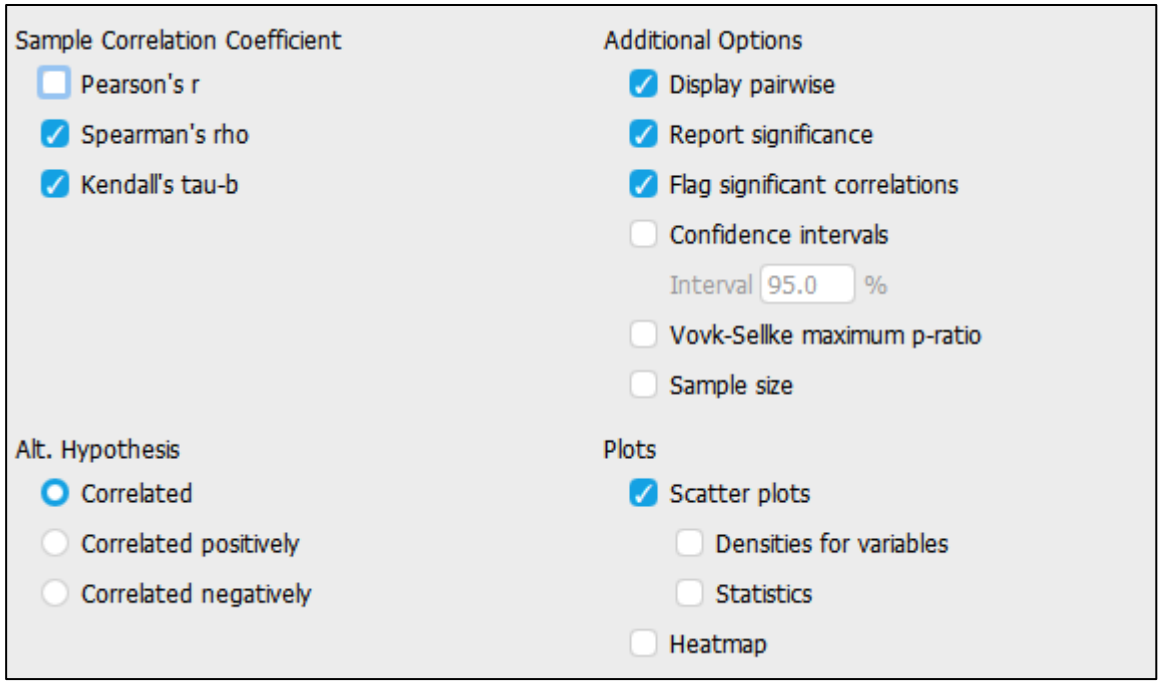
Pearsonův korelační koeficient

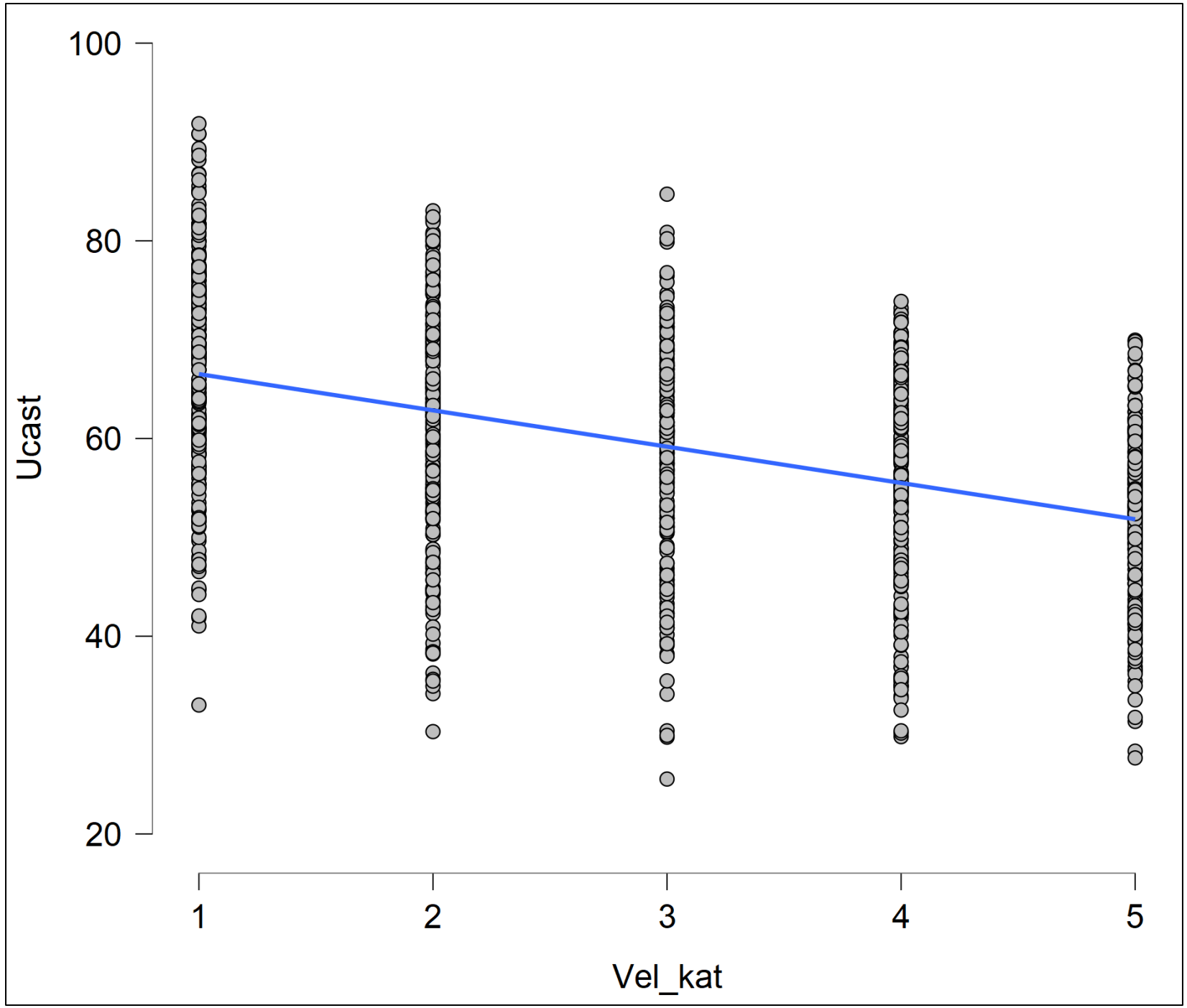
- Po umocnění r získáváme tzv. Index determinace (R^2)
- R^2 vymezuje, jaký podíl variability jedné proměnné je sdílený s druhou proměnnou
- Pro názornost se R^2 násobí číslem 100 a vyjadřuje v procentech
- Nadále však daná hodnota neříká nic o kauzalitě

Spearmanovo rho a Kendallovo tau

- Neparametrické postupy
- Např. nominální a ordinální proměnná, dvě ordinální proměnné
- Data nejdřív seřadí a následně toto pořadí využívá pro výpočet korelačního koeficientu
- Výsledky a interpretace stejné jako u Pearsonova korelačního koeficientu
- Kendallovo tau:
  - Vhodnější při menším počtu dat a opakujících se hodnotách
  - Výsledné hodnoty o něco nižší než u Spearmanova rho


Sample Correlation Coefficient
Pearson's r
Spearman's rho
Kendall's tau-b
Additional Options
Display pairwise
Report significance
Flag significant correlations
Confidence intervals
Interval 95.0 %
Vovk-Sellke maximum p-ratio
Sample size
Alt. Hypothesis
Correlated
Correlated positively
Correlated negatively
Plots
Scatter plots
Densities for variables
Statistics
Heatmap


Sample Correlation Coefficient
Pearson's r
Spearman's rho
Kendall's tau-b
Additional Options
Display pairwise
Report significance
Flag significant correlations
Confidence intervals
Interval 95.0 %
Vovk-Sellke maximum p-ratio
Sample size
Alt. Hypothesis
Correlated
Correlated positively
Correlated negatively
Plots
Scatter plots
Densities for variables
Statistics
Heatmap


100
80
60
40
20
Ucast
1
2
3
4
5
Vel_kat

Spearman's Correlations

Variable		Vel_kat	Ucast
1. Vel_kat	Spearman’s rho	—	
	p-value	—	
2. Ucast	Spearman’s rho	-0.445	—
	p-value	< .001	—

Spearman's Correlations

			Spearman	
			rho	p
Vel_kat	-	Ucast	-0.445***	< .001

* p < 0.5, ** p < 0.01, *** p < 0.001

Limity

- Základní pravidlo – **korelace ≠ kauzalita**
- Korelace vyjadřuje pouze souvislost mezi proměnnými, ne příčinu a následek
  - Vliv třetích proměnných
  - Korelace neuvádí směr působení proměnných
- Nemožnost konstatovat kauzalitu trvá i pokud se příčinný vztah jeví jako „logický“
- Statistické zjištění nemá automaticky věcný význam
- Uživatelský limit – pozor na kódování kategorických proměnných