Lingvistický seminář I

ZS 2024

Proměnná

Proměnná

- vysvětlete

Proměnná

- vysvětlete

- abstraktní jednotka

Proměnná

- vysvětlete

- abstraktní jednotka
- nabývá různých hodnot

Proměnná

- vysvětlete

- abstraktní jednotka
- nabývá různých hodnot
  - jakých?

Proměnná

- vysvětlete

- abstraktní jednotka
- nabývá různých hodnot
  - jakých?
    - pojmenování -> nominální p.
    - pojmenování, které je možné seřadit -> ordinální p.
    - číselné vyjádření -> numerické p.

Vztah mezi proměnnými

- nezávislá proměnná

- závislá proměnná

Vztah mezi proměnnými

- nezávislá proměnná
  - předpokládaná příčina
- závislá proměnná
  - předpokládaný následek
- cíl analýzy
  - zjistit, zda změna hodnoty nezávislé proměnné vede ke změně závislé proměnné

Vztah mezi proměnnými

- analýza jedné proměnné
    - má to smysl?

Korelace

- pouze vztah
- nepředpokládá, že by jedna proměnná byla příčinou a druhá následkem

Korelace

- „Velmi důležitou vlastností korelačního vztahu je to, že nám umožňuje s určitou pravděpodobností predikovat vlastnosti dalších, předtím nezkoumaných jednotek“
  (Chromý, 2023, s. 107)

Korelace

- vztah mezi proměnnými

- „Velmi důležitou vlastností korelačního vztahu je to, že nám umožňuje s určitou pravděpodobností predikovat vlastnosti dalších, předtím nezkoumaných jednotek“
  (Chromý, 2023, s. 107)

  - co to znamená? uveďte příklady (prodiskutujte nejdříve ve dvojicích)

Korelace

- nepředpokládá, že by jedna proměnná byla příčinou a druhá následkem

- smysluplná analýza korelace předpokládá její teoretické důvody

Korelace

- nepředpokládá, že by jedna proměnná byla příčinou a druhá následkem

- smysluplná analýza korelace předpokládá její teoretické důvody

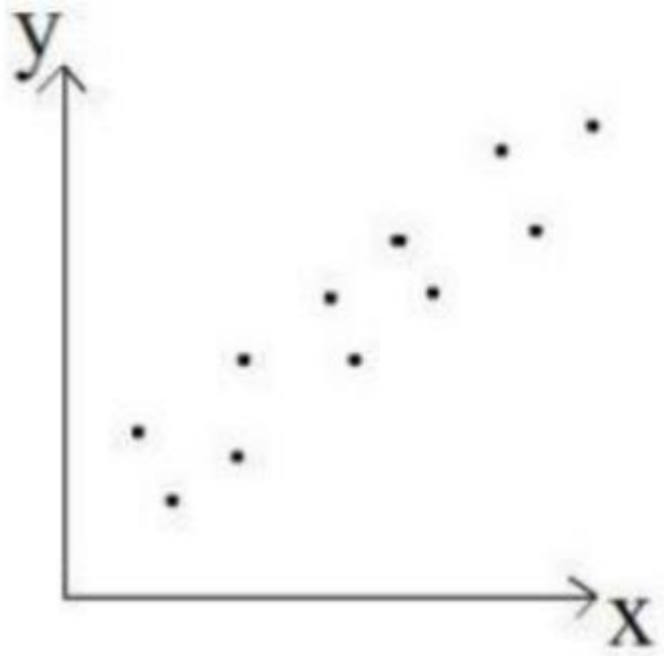
- nesmysluplné korelace
  - počet přezutých pneumatik a počet vran v Ostravě
  - délka palce a velikost slovní zásoby
  - spotřeba nealkoholických nápojů a počet utopených lidí
  - množstvím alkoholu v krvi u řidičů a počet nehod, které lidé s daným promile způsobí

Korelace

- korelační koeficient
  - <-1, +1>

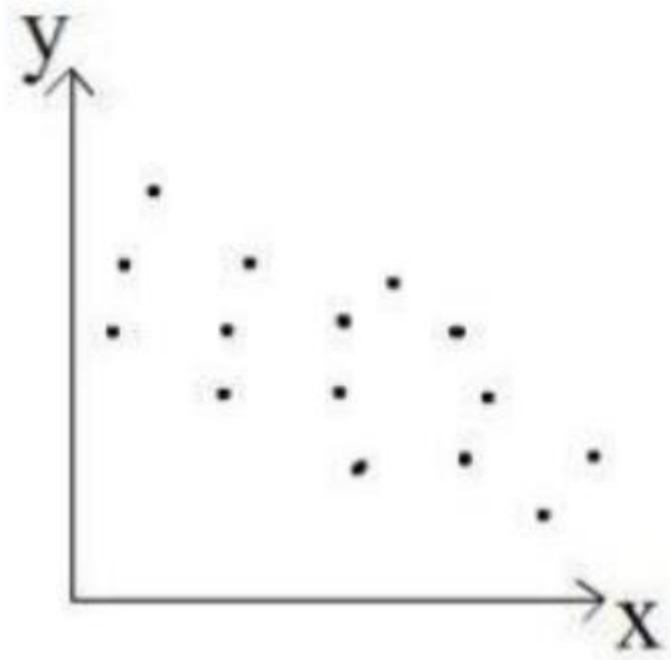
Korelace – typy

- kladná
  - pokud roste první proměnná, roste i druhá



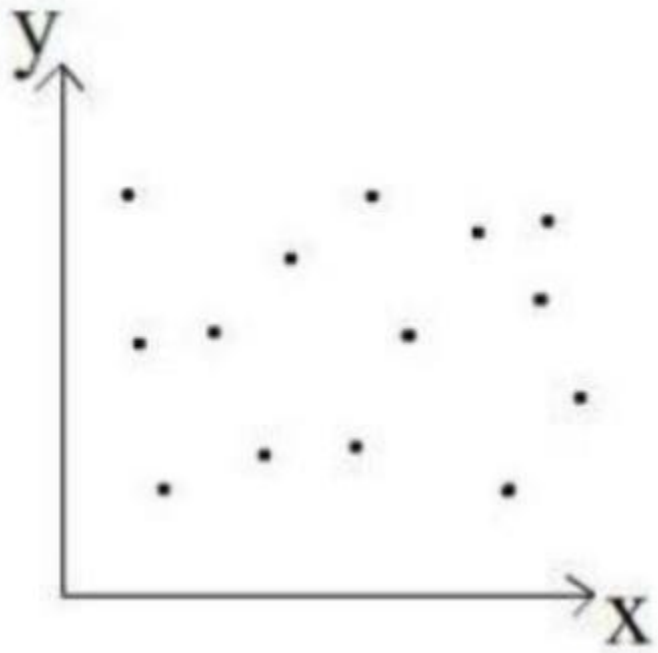
Korelace – typy

- záporná
  - pokud roste první proměnná, druhá klesá



Korelace – typy

- nulová
  - pokud roste první proměnná, druhá nemá tendenci ani stoupat, ani klesat

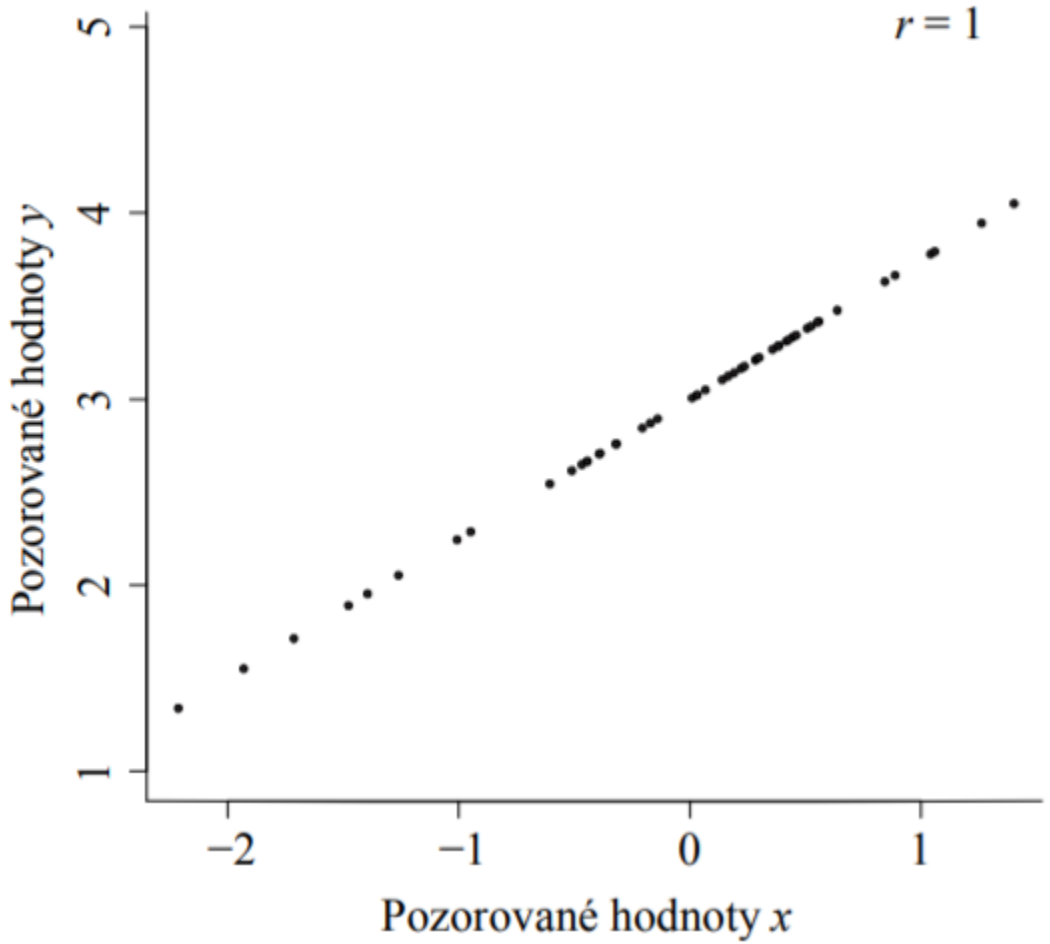


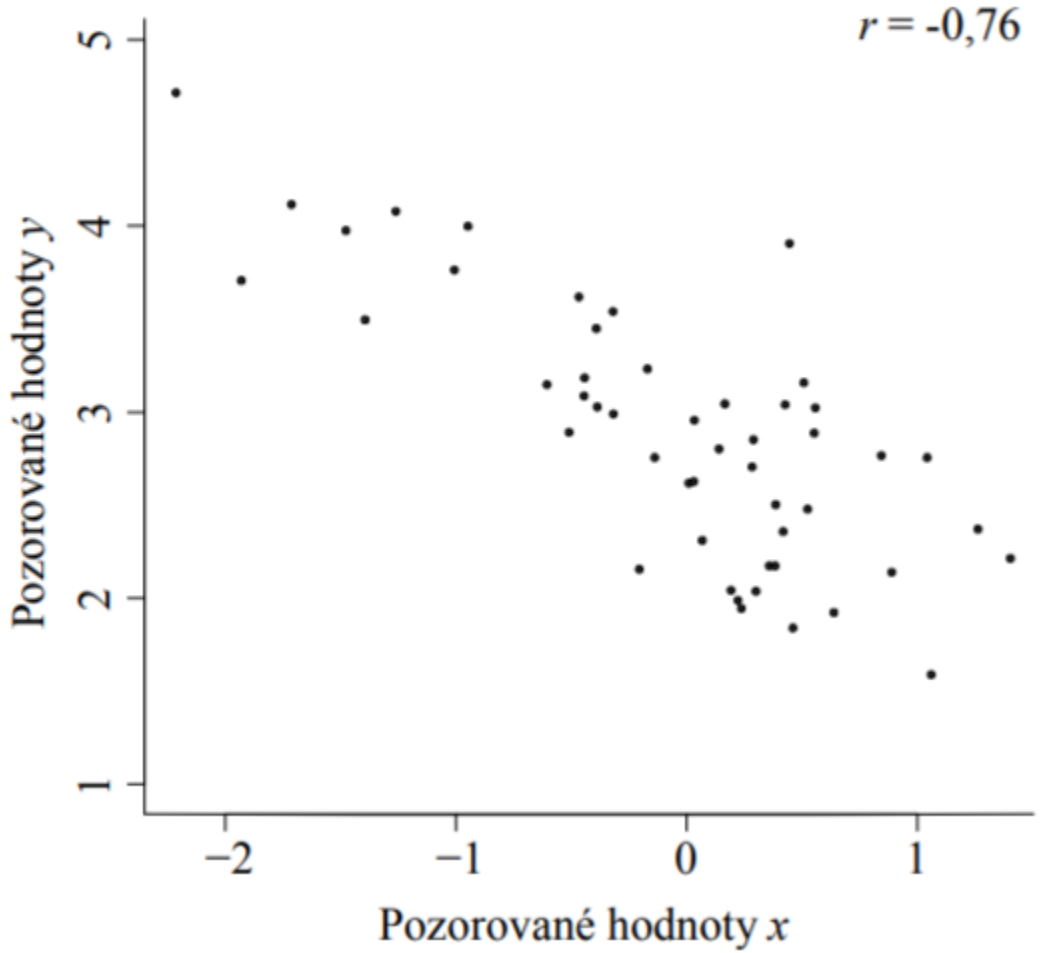
Korelační koeficient
(Pearsnův, Spearmanův, Kendallův)

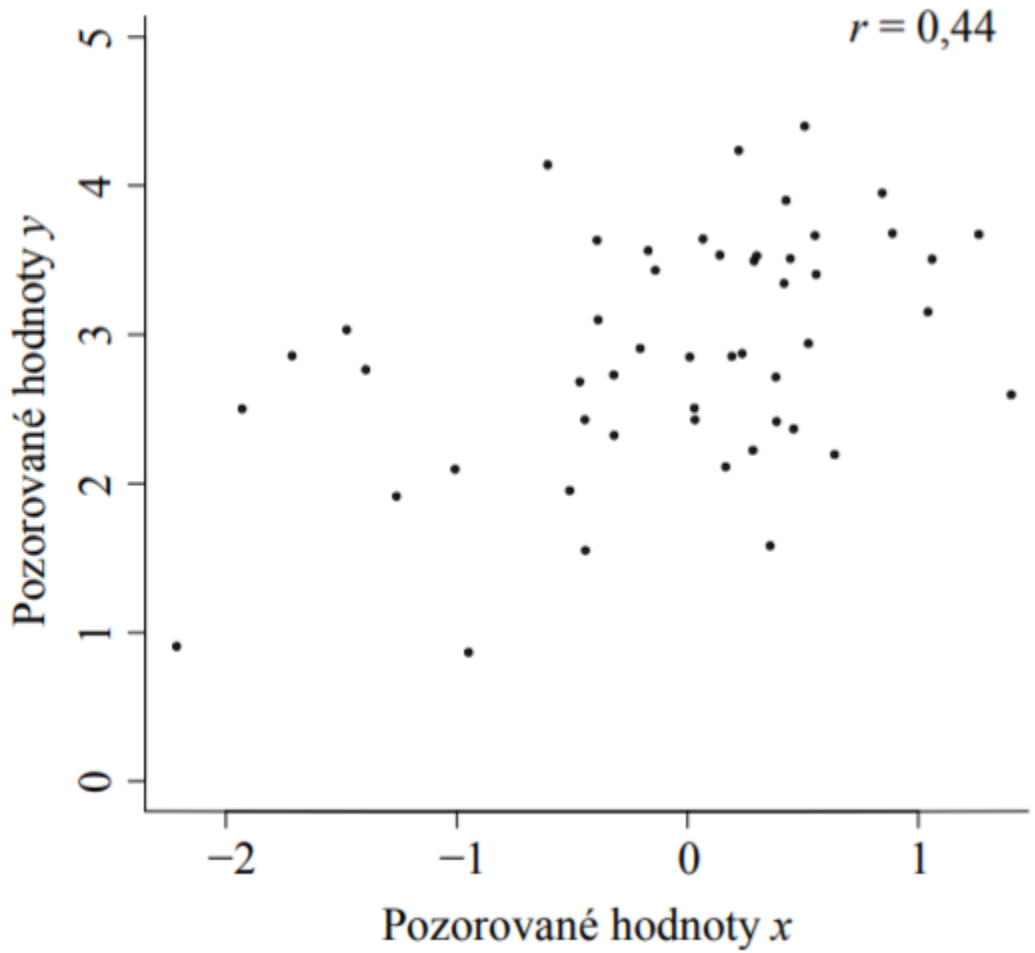
- statistický ukazatel síly lineárního vztahu

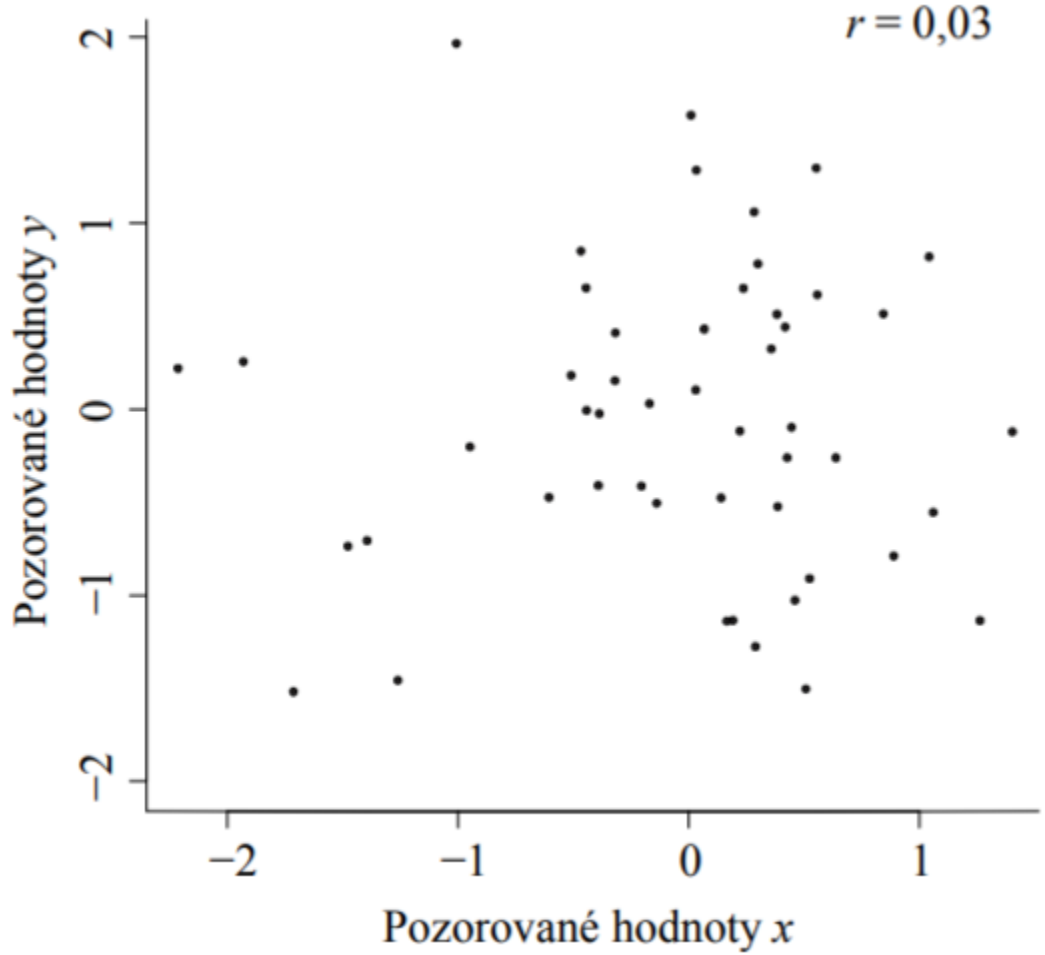
- $-1 \leq r \leq 1$

  - kladná hodnota *r* → kladná lineární korelace
  - záporná hodnota *r* → záporná lineární korelace
  - *r* = 0 → mezi proměnnými není lineární korelace
  - čím je hodnota *r* blíže 1 nebo -1, tím je korelace silnější


r = 1
Pozorované hodnoty y
Pozorované hodnoty x
1
2
3
4
5
−2
−1
0
1


r = -0,76
Pozorované hodnoty y
Pozorované hodnoty x
1
2
3
4
5
−2
−1
0
1


r = 0,44
Pozorované hodnoty y
Pozorované hodnoty x
0
1
2
3
4
5
−2
−1
0
1


r = 0,03
Pozorované hodnoty y
Pozorované hodnoty x
−2
−1
0
1
2
−2
−1
0
1

Klasifikace korelace

- 0,00 – 0,19 velmi slabá
- 0,20 – 0,39 slabá
- 0,40 – 0,59 střední
- 0,60 – 0,79 silná
- 0,80 – 1,00 velmi silná

Koeficient determinace

- síla korelace

- umocnění korelačního koeficientu
- podíl případů daná korelace „vysvětluje“

- r = 0.5
- R^2 = 25 %

- zjistěte koeficienty determinace u nižší hodnoty z intervalu klasifikace korelace (viz předchozí slide)

Korelační koeficienty

- parametrické
  - Pearsonův
- neparametrické
  - Spearmanův
  - Kendallův

Korelační koeficienty

- test statistické významnosti korelace
- p-hodnota

- může nastat
  - silná korelace, ALE statisticky nevýznamná
  - slabá korelace, ALE statisticky významná

Korelace – statistická významnost

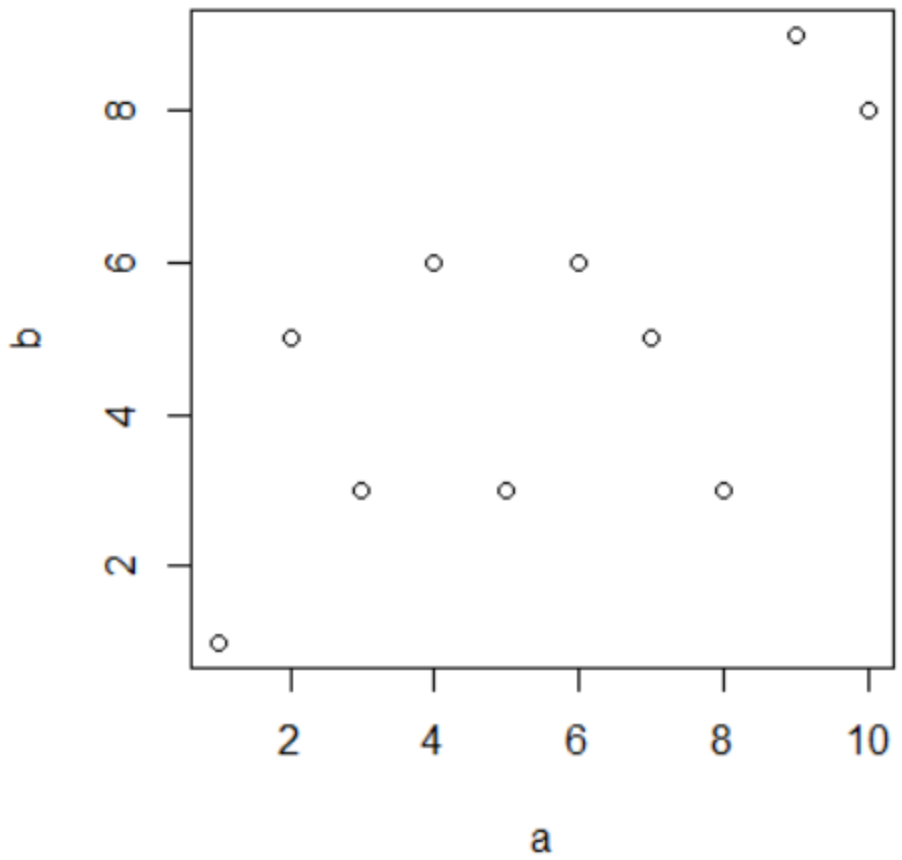
a = (1,2,3,4,5,6,7,8,9,10)

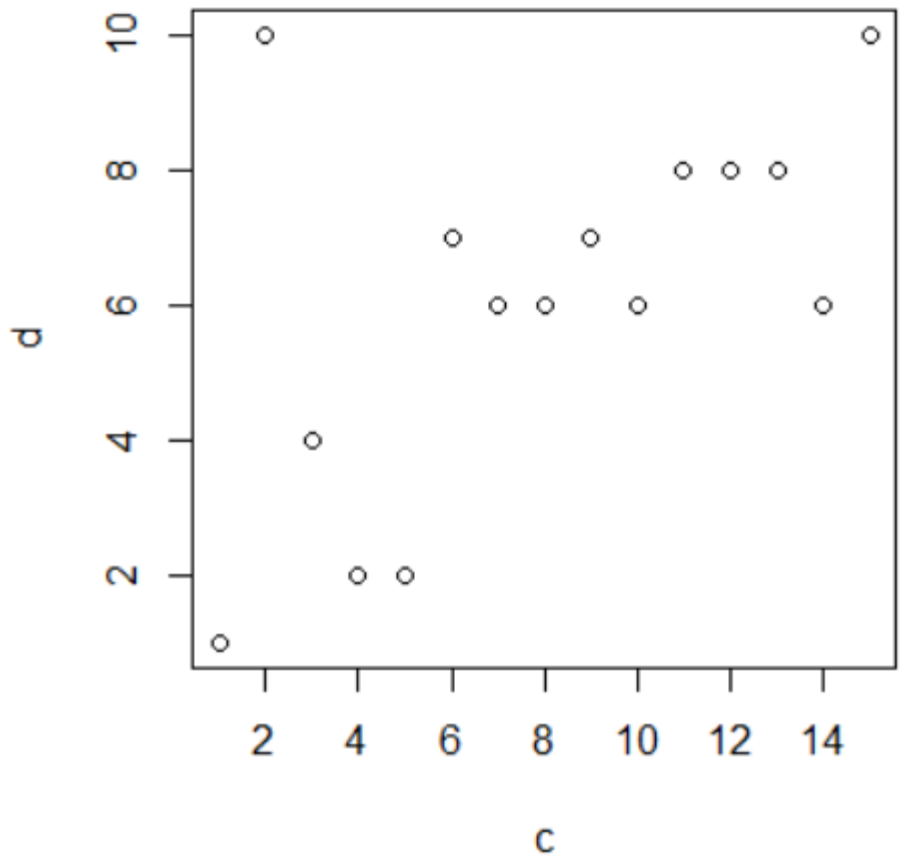
b = (1,5,3,6,3,6,5,3,9,8)

c = (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)

d = (1,10,4,2,2,7,6,6,7,6,8,8,8,6,10)

Korelace – statistická významnost





Korelace – statistická významnost

a = (1,2,3,4,5,6,7,8,9,10)

b = (1,5,3,6,3,6,5,3,9,8)

Kendallův koeficent

tau = 0,471

c = (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)

d = (1,10,4,2,2,7,6,6,7,6,8,8,8,6,10)

Kendallův koeficent

tau = 0,475

Korelace – statistická významnost

a = (1,2,3,4,5,6,7,8,9,10)

b = (1,5,3,6,3,6,5,3,9,8)

Kendallův koeficent

tau = 0,471

p-value = 0,067

c = (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)

d = (1,10,4,2,2,7,6,6,7,6,8,8,8,6,10)

Kendallův koeficent

tau = 0,475

p-value = 0,017

Korelace – výběr testu

- pokud data normálně rozdělena:
  - Pearsonův korelační koeficient
    - R
    - https://www.socscistatistics.com/tests/pearson/default.aspx

Korelace – výběr testu

- pokud data normálně rozdělena:
  - Pearsonův korelační koeficient
    - R
    - https://www.statskingdom.com/correlation-calculator.html
    - https://www.socscistatistics.com/tests/pearson/default.aspx

- pokud data neodpovídají normálnímu rozdělení
  - Spearmanův korelační koeficient
    - R
    - https://www.statskingdom.com/correlation-calculator.html
    - Spearman's Rho (Correlation) Calculator
  - Kendallův test
    - R
    - https://www.wessa.net/rwasp_kendall.wasp

Kauzalita

Kauzalita

- co to je?

Kauzalita

- předpokládá, že jedna proměnná je příčinou a druhá následkem
- silnější interpretační síla
- kauzalita – příklady?

Kauzalita

- předpokládá, že jedna proměnná je příčinou a druhá následkem
- silnější interpretační síla
- kauzalita – příklady?
  - vzdělání – míra porozumění textu

Kauzalita - podmínky

- Dané proměnné musí být ve vztahu korelace
- Příčina musí stát před následkem, tj. vztah proměnných má jasnou časovou posloupnost (když se změní příčina, změní se následek, ale ne naopak).
- Vztah mezi danými proměnnými přetrává i tehdy, pokud ve výzkumu zohledníme další, do té doby neanalyzované proměnné.

Korelace & kauzalita

- korelace ≠ kauzalita
  - více viz heslo „Korelace neimplikuje kauzalitu“
    - https://cs.wikipedia.org/wiki/Korelace_neimplikuje_kauzalitu