

MÍRY ASOCIACE, KORELACE

ZUR357 Statistická analýza dat --

30. listopadu 2017

KORELACE – MÍRY ASOCIACE

- Úkolem dvourozměrné analýzy nebo vícerozměrné analýzy (více proměnných) je zjistit, je-li mezi nimi prokazatelná souvislost (matematická – kterou nelze zaměnit za vztah příčiny a následku).

„Osoby s vysokoškolským vzděláním mají vyšší příjmy“.

- Prokázání ASOCIACE není důkazem její KAUZALITY.
- Souvislost či vztah je vlastně jen pravděpodobnosti uhádnout správně stav (hodnotu) jedné proměnné na základě stavu (hodnoty) druhé proměnné.

3 PODMÍNKY KAUZALITY

1. EXISTUJÍ SOUBĚŽNÉ ZMĚNY V OBOU PROMĚNNÝCH.
2. JE VYLOUČENA EXISTENCE NĚJAKÉ DALŠÍ, VNĚJŠÍ PŘÍČINY.
3. ZMĚNY V OBOU PROMĚNNÝCH SE OBJEVUJÍ V LOGICKÉM POŘADÍ (ROZLIŠENÍ PŘÍČINY A DŮSLEDKU).

V PŘÍPADĚ ASOCIACE ŘEŠÍME

- Zda-li vůbec *existuje*, či nikoliv.
 - Zda je *symetrická* či *nesymetrická*
 - Jaký má *směr*: je-li pozitivní, či negativní
 - Jakou má *sílu (těsnost)*: Do jaké míry rozložení variant jedné proměnné určuje rozložení variant druhé proměnné (rozdíly párového výskytu jednotlivých variant proměnných).
 - Jakou má *povahu*: je-li monotónní (lineární) či jiného druhu.
- + Určujeme i významnost (když máme reprezentativní

SÍLA ASOCIACE

Měří statistické koeficienty asociace.

Použití konkrétních koeficientů je určeno:

- Úrovní měření.
- Velikostí či tvarem kontingenční tabulky.
- Linearitou či nelinearitou vztahu.
- Symetrií či asymetrií vztahu.

Pro každou úroveň měření (nominální, ordinální a kardinální) jsou určeny zvláštní koeficienty.

Hodnoty koeficientů se většinou pohybují v intervalech:

$\langle 0; 1 \rangle$ Koeficient vypovídá o síle tohoto vztahu, ne o jeho směru (u nominálního znaku nemá směr žádný smysl).

$\langle -1; +1 \rangle$ Koeficient vypovídá o síle tohoto vztahu, znaménko o jeho směru (ordinální a kardinální proměnné)

ukrývá i Spearmanův koeficient

Crosstabs: Statistics

Chi-square

Nominal

- Contingency coefficient
- Phi and Cramér's V
- Lambda
- Uncertainty coefficient

Nominal by Interval

- Eta

Correlations

Ordinal

- Gamma
- Somers' d
- Kendall's tau-b
- Kendall's tau-c
- Kappa
- Risk
- McNemar

Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals:

Continue
Cancel
Help

HODNOTA KOEFICIENTU ASOCIACE

- ◆ Nulová hodnota obvykle znamená, že vztah neexistuje.
- ◆ Někdy ovšem je to jen výraz toho, že vztah není lineární.
- ◆ Hodnota 1,00 znamená perfektní vztah.

KOEFICIENTY PRO NOMINÁLNÍ PROMĚNNÉ ZALOŽENÉ NA χ^2

- Phi: Pro tabulky 2x2, u větších může nabývat hodnot nad 1,00.
- C: Pearsonův koeficient kontingence. Nabývá hodnot $\langle 0; 1 \rangle$.
- Cramer's V: Pro větší tabulky (u tabulky 2x2 je identický s Phi).

KOEFICIENTY PRO NOMINÁLNÍ PROMĚNNÉ ZALOŽENÉ NA PROPORCIÁLNÍ REDUKCI CHYBY

- Symetrické LAMBDA.
- Asymetrické LAMBDA. Hodnota závisí na tom, která proměnná predikuje druhou proměnnou.

KOEFICIENTY ASOCIACE ORDINÁLNÍCH PROMĚNNÝCH

Nabývají hodnot $\langle -1; 1 \rangle$

Pro lineární vztahy. U nelineárních vztahů se použije míra asociace pro nominální proměnné.

Pro symetrické vztahy

- Goodman-Kruskalovo Gamma.
- Kendallovo τ_b (čtvercová tabulka).
- Kendallovo τ_c (obdélníkové tabulky).
- Spearmanův koeficient pořadové korelace

Pro asymetrické vztahy

- Somersovo D – jedna z proměnných brána jako závislá.



PŘÍKLAD:

Předpokládejme, že míra religiozity pozitivně koresponduje s mírou odporu vůči potratům (obě ordinální!).

Je-li tedy osoba A religióznější než osoba B, lze předpokládat, že i míra odporu proti potratům u ní bude větší než u osoby B.

Porovnáváme všechny páry a GAMA je podílem párových srovnání, která tomuto předpokladu vyhovují.

KOEFICIENTY ZALOŽENÉ NA POŘADOVÉ KORELACI

SPEARMANŮV koeficient pořadové korelace je neparametrickou metodou a nabývá hodnot $\langle -1; +1 \rangle$.

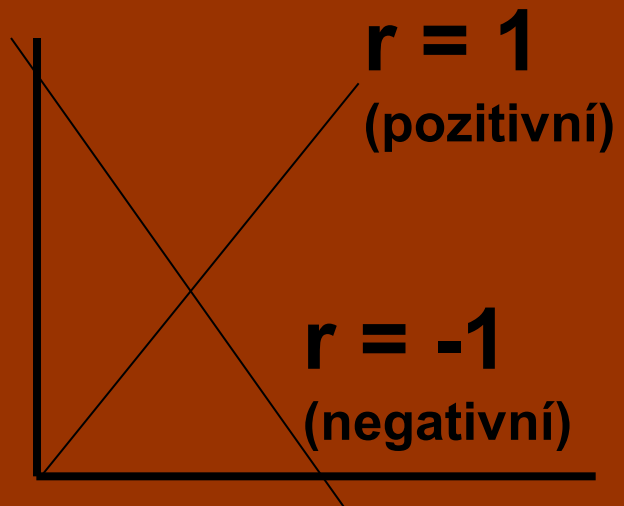
Je vhodný pro ordinální proměnné s větším počtem hodnot.

KOEFICIENTY KOREALCE PRO KARDINÁLNÍ PROMĚNNÉ

PEARSONŮV KORELAČNÍ KOEFICIENT r

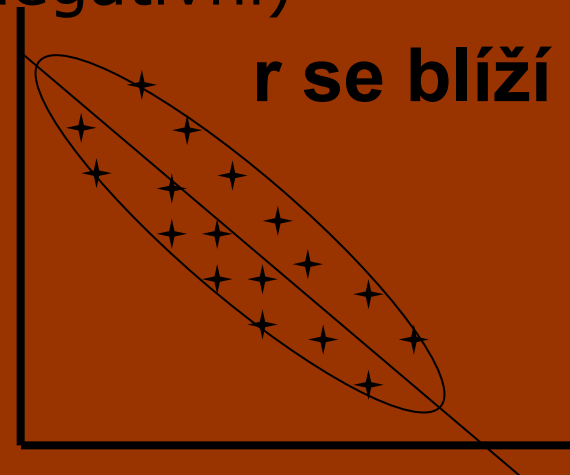
Koeficient nerozlišuje co je příčina a co důsledek (nezávislá a závislá proměnná).

PERFEKTNÍ KORELACE



SILNÁ KORELACE (negativní)

r se blíží -1



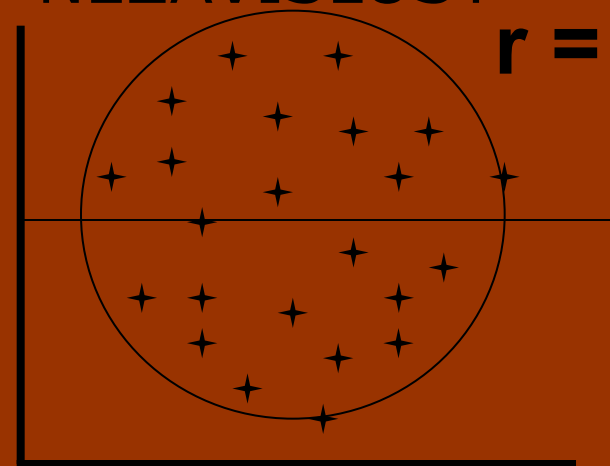
SILNÁ KORELACE (pozitivní)

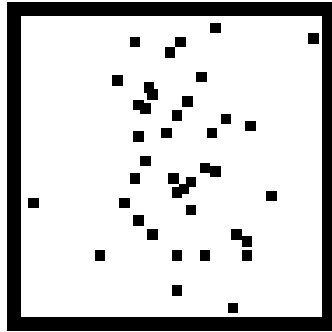
r se blíží 1



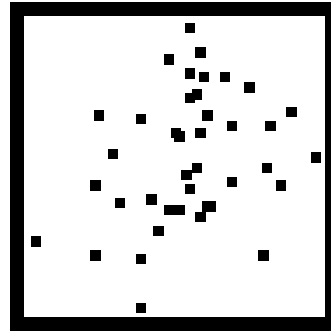
PERFEKTNÍ NEZÁVISLOST

$r = 0$

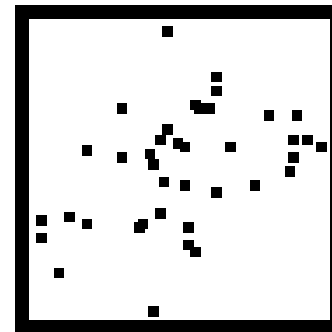




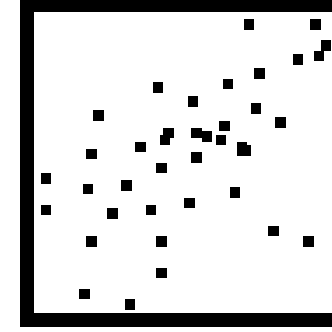
$r=0$



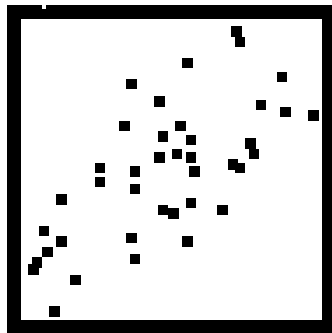
$r=.28$



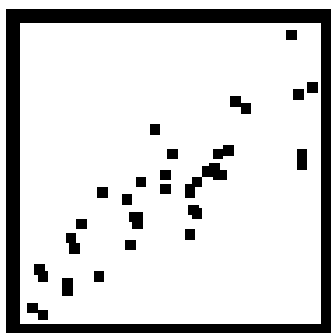
$r=.42$



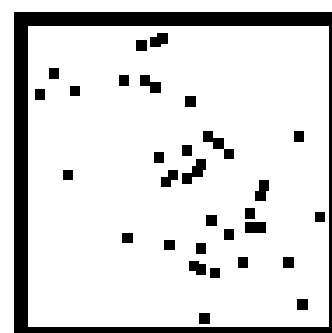
$r=.55$



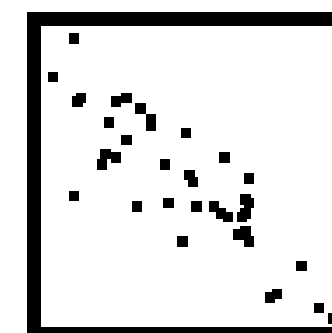
$r=.67$



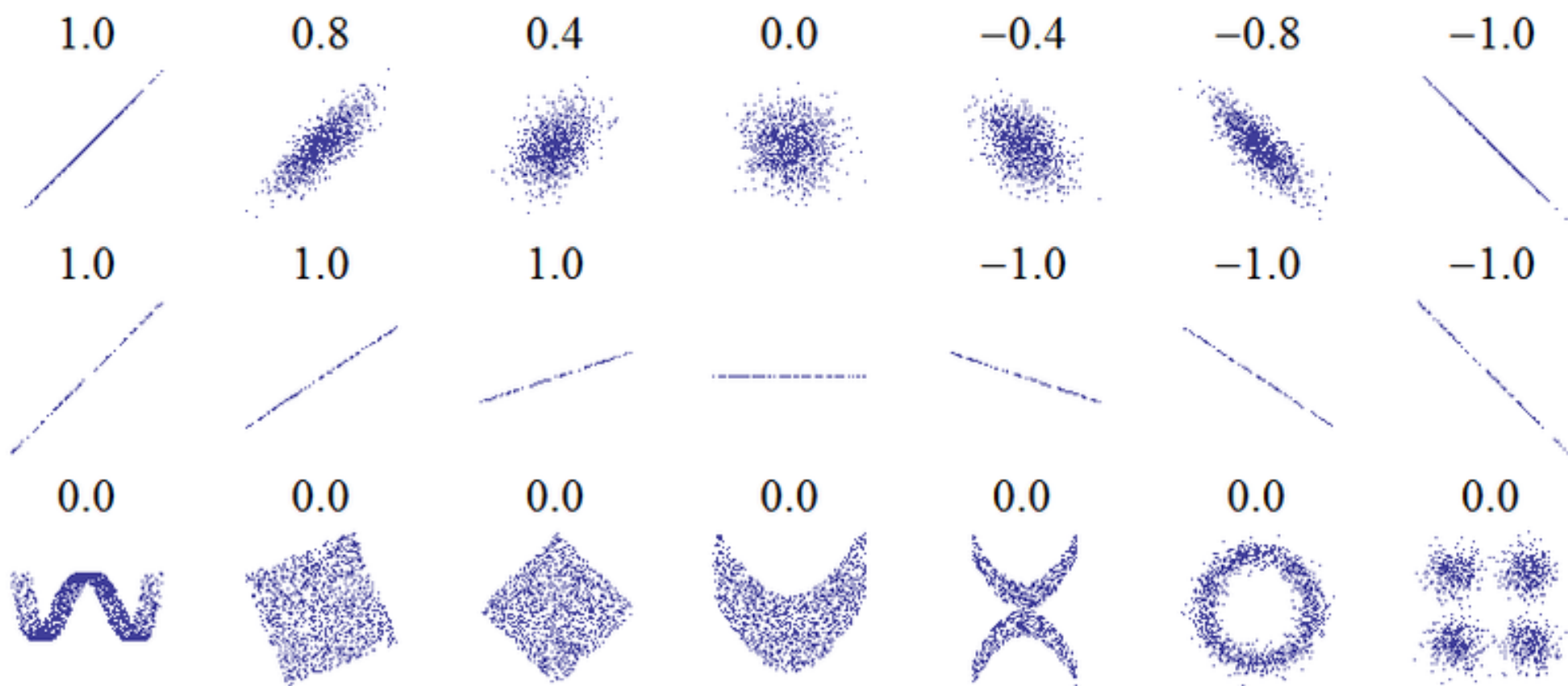
$r=.86$



$r=-.55$



$r=-.85$





ETA koeficient

- když závisle proměnná je kardinální (měřená na intervalové škále) a
- nezávisle proměnná je nominální nebo ordinální (měřená na nominální nebo ordinální škále).

TEST LINEARITY ASOCIACE

Použít kontingenční tabulku.

Porovnat skupinové průměry.

Porovnat lineární a nelineární koeficienty.

Analyzovat residuály.

Correlations

		smůla, osud	šude jsou chudí	lenivost, pasivita chudých	daň pokrok	nespravedl. společnost	rozmarilost chudých
smůla, osud	Kendall's Sig. (2-tail N	1,000 . 999	,194* ,000 999	,143* ,000 998	,145* ,000 999	,027 ,310 999	,136* ,000 998
vsude jsou	Kendall's Sig. (2-tail N	,194* ,000 999	1,000 . 999	,280* ,000 998	,376* ,000 999	-,110* ,000 999	,194* ,000 998
lenivost, pa	Kendall's Sig. (2-tail N	,143* ,000 998	,280* ,000 998	1,000 . 998	,187* ,000 998	-,203* ,000 998	,381* ,000 997
dan pokrok	Kendall's Sig. (2-tail N	,145* ,000 999	,376* ,000 999	,187* ,000 998	1,000 . 999	-,052* ,046 999	,193* ,000 998
nespravedl. společnost	Kendall's Sig. (2-tail N	,027 ,310 999	-,110* ,000 999	-,203* ,000 998	-,052* ,046 999	1,000 1000	-,169* ,000 998
rozmarilost chudých	Kendall's Sig. (2-tail N	,136* ,000 998	,194* ,000 998	,381* ,000 997	,193* ,000 998	-,169* ,000 998	1,000 . 998

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

*Correlation is significant at the 0.05 level (2-tailed).



$R^2 =$ KOEFICIENT
DETERMINANCE