

MÍRY ASOCIACE

Jde o míry, které měří směr a sílu vztahu mezi dvěma proměnnými. Existuje několik různých typů mír asociace a každá má své specifické vlastnosti a využití v závislosti na charakteru zkoumaných dat.

Dále budou popsány míry asociace, které jsou k dispozici v proceduře *Analyza-Descriptive Statistics – Crosstabs* v záložce *Statistics*.

Crosstabs: Statistics

☒ **Chi-square**

Nominal

☐ Contingency coefficient

☐ Phi and Cramer's V

☐ Lambda

☐ Uncertainty coefficient

Ordinal

☐ Gamma

☐ Somers' d

☐ Kendall's tau-b

☐ Kendall's tau-c

Nominal by Interval

☐ Eta

☒ **Cochran's and Mantel-Haenszel statistics**

Test common odds ratio equals: 1

Continue **Cancel** **Help**

MÍRY PRO NOMINÁLNÍ PROMĚNNÉ

Míry pro nominální proměnné lze rozdělit na základě principu, z něhož vycházejí, do dvou skupin: na míry asociace založené na Pearsonově chí-kvadrát statistice a na míry vycházející z rozkladu variability. Do první skupiny patří koeficient kontingence, ϕ a Cramérovo V , do druhé skupiny se řadí koeficient λ , Goodman-Kruskalovo τ a koeficient nejistoty.

Samotnou Pearsonovu chí-kvadrát statistiku není vhodné používat pro porovnání síly vztahu u tabulek, které se liší počtem řádků či sloupců nebo celkovým počtem případů, protože její hodnota roste se zvětšujícími se rozměry tabulky a s rostoucím počtem případů. Koeficient kontingence, ϕ a Cramérovo V penalizují hodnotu Pearsonovy chí-kvadrát statistiky počtem případů, případně také počtem kategorií řádkové či sloupcové proměnné.

KOEFICIENT KONTINGENCE CC

Penalizuje Pearsonovu chí-kvadrát statistiku pouze počtem případů, a to následujícím způsobem:

$$CC = \sqrt{\frac{\chi^2}{\chi^2 + n}}.$$

Koeficient kontingence nabývá hodnot z intervalu $[0, \sqrt{(q-1)/q}]$, kde q je minimum z počtu řádků a sloupců tabulky – nula indikuje nezávislost proměnných, hodnoty blízké horní mezi intervalu ukazují na silnou závislost. **Maximální možná hodnota tohoto koeficientu však závisí na rozměrech tabulky, což znesnadňuje jeho použití pro porovnání síly vztahů v tabulkách odlišných rozměrů.**

KOEFICIENT FÍ

Upravuje Pearsonovu chí-kvadrát statistiku pouze počtem případů a počítá se následovně:

$$\varphi = \sqrt{\frac{\chi^2}{n}}.$$

Koeficient φ může nabýt nezáporných hodnot, menších nebo rovných výrazu $\sqrt{(q-1)}$, kde q je minimum z počtu řádků a sloupců tabulky. Čím vyšší je hodnota φ , tím silnější je vztah mezi proměnnými, nula odpovídá nezávislosti proměnných. **Závislost maximální možné hodnoty tohoto koeficientu na rozměrech tabulky, znesnadňuje jeho použití pro porovnání síly vztahů v tabulkách odlišných rozměrů.**

Pro čtyřpolní tabulky, tj. tabulky s rozměry 2x2, je φ rovno Pearsonovu korelačnímu koeficientu. V tomto případě nabývá hodnot z intervalu $[-1,1]$, znaménko přebírá od Pearsonova korelačního koeficientu. Druhá mocnina koeficientu φ se vyskytuje v korespondenční analýze pod názvem inercie.

CRAMEROVO V

Penalizuje Pearsonovu chí-kvadrát statistiku nejen počtem případů, ale i rozměry tabulky. Získá se následovně:

$$V = \sqrt{\frac{\chi^2}{n(q-1)}}.$$

Q je minimum z počtu řádků a sloupců tabulky. Cramérovo V nabývá hodnot z intervalu $[0,1]$ – nula ukazuje na absenci asociace, hodnoty blízké jedné na silnou asociaci mezi řádkovou a sloupcovou proměnnou. Horní mez intervalu již není (na rozdíl od dvou předešlých měr) závislá na rozměrech tabulky. Tuto mez koeficient dosahuje v případě úplné asociace, což je například

lineární závislost řádkové a sloupcové proměnné a jiné typy závislosti. **Jestliže zkoumáme symetrický vztah dvou nominálních proměnných, lze doporučit použití Cramérova V.**

MÍRY ZALOŽENÉ NA ROZKLADU VARIABILITY

Druhá skupina měr pro nominální proměnné vychází z rozkladu variability jedné proměnné na variabilitu vysvětlenou znalostí druhé proměnné a zbylou variabilitu. Jedná se tedy o míry asymetrického vztahu, kdy je jedna proměnná považovaná za závislou a druhá za nezávislou. Míry lambda, Goodman-Kruskalovo tau a koeficient nejistoty vyjadřují podíl variability závislé proměnné vysvětlené znalostí nezávislé proměnné. Jedná se tedy o koeficienty determinace a liší se pouze volbou míry variability. Koeficient lambda je založen na variačním poměru, Goodman-Kruskalovo tau na Giniho míře variability a koeficient nejistoty na entropii.

Při výpočtu těchto měr lze jednou uvažovat rozklad variability sloupcové proměnné a podruhé rozklad variability řádkové proměnné. Tímto způsobem získáme dvě varianty asymetrického koeficientu. Koeficient lambda a koeficient nejistoty navíc existují i v symetrické podobě, kdy se hodnoty dvou asymetrických měr kombinují (přesněji viz níže). Všechny tyto míry nabývají hodnoty z intervalu $[0,1]$, kde nula indikuje absenci vztahu mezi proměnnými a hodnoty blízké jedné silný vzájemný vztah proměnných. Hodnotu 1 nabývají pouze v případě, pokud je v každém řádku tabulky maximálně jedna buňka s nenulovou četností.

MÍRY PRO ORDINÁLNÍ PROMĚNNÉ

Koeficienty gama, Somersovo d, Kendallovo tau-b a tau-c jsou míry určené pro dvě ordinální proměnné. Kromě Somersova d jsou to míry symetrické, Somersovo d je primárně asymetrická míra, ale existuje i jeho symetrická varianta.

Všechny tyto koeficienty mají vlastnosti koeficientů pořadové korelace: mohou nabývat hodnot od -1 do 1, v případě nezávislosti proměnných jsou rovny nule.

Všechny dále vycházejí z počtu konkordantních a diskordantních dvojic. Za konkordantní dvojici se považuje taková dvojice případů, kde má případ s vyšší hodnotou proměnné A i vyšší hodnotu proměnné B. Diskordantní dvojicí rozumíme takovou dvojici případů, kde případ s vyšší hodnotou proměnné A má nižší hodnotu proměnné B. Za shodu označíme situaci, kdy se hodnoty u proměnné A nebo B shodují.

Dále zavedme toto značení:

P ... počet konkordantních dvojic,

Q ... počet diskordantních dvojic,

T_A ... počet dvojic, které mají stejnou hodnotu proměnné A, ale odlišnou hodnotu proměnné B,

T_B ... počet dvojic, které mají stejnou hodnotu proměnné B, ale různou hodnotu proměnné A.

GAMA

Je symetrickou mírou založenou pouze na počtu konkordantních a diskordantních dvojic, shody nebere v úvahu. Značíme ji γ a spočítáme takto:

$$\gamma = \frac{P-Q}{P+Q}.$$

Jedná se tedy o rozdíl podílu konkordantních a diskordantních dvojic ze všech dvojic kromě shod. Krajních hodnot -1 a 1 nabývá nejen v případě perfektní lineární závislosti, tj. kdy se nenulové četnosti vyskytují pouze na hlavní či vedlejší diagonále tabulky, ale i když jsou buňky s nenulovou četností rozmístěny podél hlavní či vedlejší diagonály tabulky (všechny však musí být na stejné straně diagonály).

KENDALLOVO TAU-B

je symetrická míra vycházející z počtu konkordantních a diskordantních dvojic a z počtu shod. Spočítá se následovně:

$$\tau_b = \frac{P-Q}{\sqrt{(P+Q+T_A)(P+Q+T_B)}}.$$

Hraniční hodnoty 1, resp. -1 může tato míra dosáhnout pouze pro čtvercové tabulky, a to v případě perfektní lineární závislosti (nenulové četnosti pouze na hlavní, resp. vedlejší diagonále tabulky). Hodnota Kendallova tau-b je v absolutní hodnotě vždy menší než hodnota koeficientu gama, ale větší než hodnota symetrické varianty Somersova d. Platí tedy:

$$|\text{symetrické Somersovo } d| < |\text{Kendallovo tau-b}| < |\text{gama}|.$$

Kendallovo tau-b lze spočítat také pomocí procedury *Bivariate Correlations (Analyze, Correlate, Bivariate)*.

KENDALLOVO TAU-C

Spočítá se následovně:

$$\tau_c = \frac{2q(P-Q)}{n^2(q-1)},$$

kde q je minimum z počtu řádků a sloupců tabulky. Hraniční hodnoty 1, resp. -1 může dosáhnout pouze pro čtvercové tabulky, a to v případě perfektní lineární závislosti, nicméně pro obdélníkové tabulky se hraniční hodnotě 1, resp. -1 blíží více než Kendallovo tau-b.

Pokud zkoumáme symetrický vztah dvou ordinálních proměnných lze doporučit Kendallovo tau-b pro čtvercové tabulky a Kendallovo tau-c pro obdélníkové tabulky, pro asymetrické vztahy použijeme Somersovo d.

SOMERSOVO D

je míra vycházející nejen z počtu konkordantních a diskordantních dvojic, ale i z počtu shod. Pro sloupcovou proměnnou B při znalosti řádkové proměnné A se Somersovo $d_{B|A}$ počítá následovně:

$$d_{B|A} = \frac{P-Q}{P+Q+T_A}.$$

NOMINÁLNÍ A KARDINÁLNÍ PROMĚNNÁ

Koeficient eta je totožný s koeficientem eta počítaným v proceduře *Means (Analyze, Compare Means, Means)*. Jedná se o míru vycházející z rozkladu celkové variability kardinální proměnné na variabilitu uvnitř skupin daných nominální proměnnou a na variabilitu mezi skupinami. Variabilita je měřená pomocí rozptylu počítaného z hodnot uložených v datové matici (kódů jednotlivých kategorií). Koeficient je v proceduře *Crosstabs* počítán oběma možnými způsoby, tj. za kardinální proměnnou se jednou považuje sloupcová proměnná a podruhé řádková proměnná kontingenční tabulky.