

Základné štatistické metódy

Portál: edu.ukf.sk - Vzdelávací portál - Univerzita
Konštantína Filozofa, Nitra

Kurz: Počítačová analýza dát 21/22 (KI/PAD/21)

Kniha: Základné štatistické metódy

Vytlačil(a): Zuzana Pavlendová

Dátum: Streda, 1 december 2021, 17:49

Obsah

1. Popisná štatistika

- 1.1. Početnosti
- 1.2. Popisné charakteristiky

2. Testy rozdelenia

3. Testy o rozptyle a ich neparametrické alternatívy

- 3.1. Jedna vzorka
- 3.2. Dve nezávislé vzorky
- 3.3. Dve a viac nezávislých vzoriek

4. Testy o strednej hodnote a ich neparametrické alternatívy

- 4.1. Jedna vzorka
- 4.2. Závislé vzorky
- 4.3. Nezávislé vzorky

5. Vzťahy medzi premennými

- 5.1. Analýza kontingencie
- 5.2. Korelačná analýza

1. Popisná štatistika

Pre lepšie porozumenie získaným údajom je potrebné utriediť dáta a vypočítať základné charakteristiky.

1.1. Početnosti

Ak sa hodnoty premennej opakujú, používame na zistenie početností **jednoduché triedenie**. V prípade, že hodnoty sú rôznorodé, používame **intervalové triedenie**, t.j. zisťuje sa, koľko hodnôt sa nachádza vo vytvorených intervaloch.

K vizualizácii početností sa používa **polygón, histogram, listový graf a kruhový diagram**.

K vyjadreniu interakčných početností medzi premennými slúži **kontingenčná tabuľka**. K vizualizácii kontingenčnej tabuľky sa používa **interakčný graf**.

1.2. Popisné charakteristiky

Ak premenná má normálne rozdelenie, tak súbor charakterizujeme najčastejšie **priemerom, smerodajnou (štandardnou) chybou a smerodajnou (štandardnou) odchýlkou**. V prípade normálneho rozdelenia priemer, medián a modus sú približne rovnaké hodnoty a koeficient šikmosti a špicatosti je približne rovný nule.

Ak hodnoty premennej nepochádzajú z normálneho rozdelenia, resp. sa v súbore vyskytujú extrémne hodnoty, tak súbor charakterizujeme najčastejšie **mediánom, kvartilovým rozpätím a variačným rozpätím**.

K vizualizácii uvedených trojíc popisných štatistík sa používa **krabicový graf**.

2. Testy rozdelenia

K testovaniu, či premenná pochádza z predpokladaného rozdelenia používame testy z kategórie „dobrej zhody“. Pomocou týchto testov skúmame priebeh celej distribučnej funkcie. Tieto testy používame hlavne k overovaniu predpokladu normality, t.j. overujeme, či sledovaná premenná má normálne rozdelenie. **Kolmogorov-Smirnovov test** môžeme použiť pre akékoľvek rozdelenie. **Lillieforsov test** je modifikáciou predchádzajúceho a týka sa iba normálneho rozdelenia. Ak máme namerané údaje v triednych intervaloch s príslušnou frekvenciou, preverujeme normalitu **chí-kvadrát testom dobrej zhody**. Najväčšiu silu zo všetkých testov normality má vo väčšine situácií **Shapiro-Wilkov W test**. Ak je štatistika W významná, zamietame nulovú hypotézu, ktorá tvrdí, že dáta pochádzajú z normálneho rozdelenia.

Ďalšou možnosťou je použitie **normálneho pravdepodobnostného grafu**, pomocou ktorého kontrolujeme, či dáta pochádzajú z normálneho rozdelenia. Na rozdiel od testovania umožňuje identifikovať prípadné odchýlky od normality (negatívnu a kladnú šikmosť), bimodalitu a extrémne hodnoty.

3. Testy o rozptyle a ich neparametrické alternatívy

Testy o rozptyle a ich neparametrické alternatívy

3.1. Jedna vzorka

Test o rozptyle normálneho rozdelenia testuje hypotézu o rozdiel rozptylu a konštanty (Tabuľka 19). Zisťuje, či sú namerané rozdiely medzi rozptylom a konštantou iba náhodné, alebo štatisticky významné.

Na použitie testu je nutné splniť nasledujúci predpoklad:

- Normálne rozdelenie premennej.

Procedúra testovania - test o rozptyle normálneho rozdelenia

	Test o rozptyle normálneho rozdelenia
1.	$H_0: \sigma^2 = \sigma_0^2$ proti obojstrannej alternatíve
2.	$\alpha = 0,05 \vee \alpha = 0,01 \vee \dots$
3.	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
4.	Zamietame H_0 ak nebude platiť $\chi_{n-1}^2(1-\alpha/2) < \chi^2 < \chi_{n-1}^2(\alpha/2) \vee \frac{1}{n-1}\sigma_0^2\chi_{n-1}^2(1-\alpha/2) < s^2 < \frac{1}{n-1}\sigma_0^2\chi_{n-1}^2(\alpha/2)$

Z rozdelenia štatistiky chí-kvadrát a z definície kritickej hodnoty vyplýva, že interval $\left(\frac{(n-1)s^2}{\chi_{n-1}^2(\alpha/2)}, \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha/2)} \right)$ je obojstranný 100(1 - α)% interval spoľahlivosti pre σ^2 . Interval pokrýva skutočnú hodnotu σ^2 s pravdepodobnosťou 1 - α .

3.2. Dve nezávislé vzorky

F-test testuje hypotézu o rozdiel rozptylov dvoch nezávislých skupín (Tabuľka 20). Zisťuje, či sú namerané rozdiely medzi rozptylmi iba náhodné, alebo štatisticky významné. Posúdenie rozptýlenosti dát je dôležité pri overení predpokladu rovnosti rozptylov, napr. pri dvojvýberovom t-teste pre nezávislé vzorky. Skúmanie variability dát môže byť i vlastným cieľom výskumu, aj keď metódy pre hodnotenie stredných hodnôt sa používajú v podstatne väčšej miere. Napríklad môžeme porovnať stabilitu dvoch výrobných procesov, ak porovnáme nielen priemerné hodnoty indikátora kvality, ale i hodnoty jeho premenlivosti. Na použitie testu je nutné splniť nasledujúci predpoklad:

- Normálne rozdelenie premennej v obidvoch výberoch.

Procedúra testovania - F-test

	F-test
1.	$H_0: \sigma_1^2 = \sigma_2^2$ proti obojstrannej alternatíve
2.	$\alpha = 0,05 \vee \alpha = 0,01 \vee \dots$
3.	$F = \frac{s_1^2}{s_2^2}, s_1^2 \geq s_2^2$
4.	$F \geq F_{n_1-1, n_2-1}(\alpha/2) \vee p \leq \alpha$

3.3. Dve a viac nezávislých vzoriek

Táto skupina testov testuje hypotézu o rozdiely rozptylov dvoch a viac nezávislých skupín. Zisťuje, či sú namerané rozdiely medzi rozptylmi iba náhodné, alebo štatisticky významné. Posúdenie rozptýlenosti dát je dôležité pri overení predpokladu rovnosti rozptylov, napr. pri analýze rozptylu, kovariancie. Testy môžeme rozlišovať podľa toho, či sú určené pre situácie, kde vystupuje jedna závislá premenná – univariačné testy, alebo sú určené pre situácie, kde vystupujú viaceré závislé premenné – multivariačné testy.

Tieto testy testujú nasledovnú nulovú hypotézu

$$H_0: \sigma_1^2 = \dots = \sigma_I^2$$

proti alternatíve H_1 , že H_0 neplatí.

Na použitie testov je nutné splniť nasledujúci predpoklad:

- Normálne rozdelenie premennej v každom výbere.

Medzi univariačné testy patrí **Cochranov, Hartleyov, Bartlettov test a Leveneov test**. Nevýhodou Bartlettovho testu je to, že je veľmi citlivý na porušenie predpokladu normality. Cochranov a Hartleyov test sa používa na testovanie nulovej hypotézy iba vtedy, keď je početnosť všetkých skupín rovnaká. Neparametrický Leveneov test na testovanie rovnosti rozptylov predstavuje jednocestnú analýzu rozptylu na absolútnych odchýlkach pozorovaní od mediánov skupín.

Boxov M test je multivariačný test rovnosti rozptylov pre závislé premenné alebo kovariáty. Umožňuje špecifikovať, či chceme zahrnúť kovariáty do testu. M test je veľmi citlivý na porušenie predpokladu normality. Ak je tento test významný, variančno-kovariančné matice sú významne odlišné od seba. **Senov a Puriov neparametrický test** je neparametrickou alternatívou k Boxovmu M testu. Nevyžaduje predpoklad normality ako predchádzajúci. Taktiež umožňuje špecifikovať, či chceme zahrnúť kovariáty do testu.

4. Testy o strednej hodnote a ich neparametrické alternatívy

Testy o strednej hodnote a ich neparametrické alternatívy

4.1. Jedna vzorka

Jednovýberový z-test a t-test testujú hypotézu o rozdieli priemeru a konštanty (Tabuľka 21). Zisťujú, či sú namerané rozdiely medzi priemerom a konštantou iba náhodné, alebo štatisticky významné. Z-test sa používa, ak je rozptyl známy, t.j. pracujeme so skutočným parametrom.

Na použitie testov je nutné splniť nasledujúci predpoklad:

- Normálne rozdelenie premennej.

Procedúra testovania - jednovýberový z-test/t-test

	Z-test	Jednovýberový t-test
1.	H0: $\mu = \mu_0$ proti obojstrannej alternatíve	
2.	$\alpha = 0,05 \vee \alpha = 0,01 \vee \dots$	
3.	$Z = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$	$T = \frac{\bar{x} - \mu_0}{s} \sqrt{n}$
4.	$ Z \geq z(\alpha/2) \vee p \leq \alpha$	$ T \geq t_{n-1}(\alpha) \vee p \leq \alpha$

Z rozdelenia štatistiky Z a z definície kritickej hodnoty vyplýva, že interval s koncovými bodmi $\bar{x} \mp \frac{\sigma^* z(\alpha/2)}{\sqrt{n}}$ prekrýva skutočnú hodnotu μ s pravdepodobnosťou $1 - \alpha$. Nazývame ho obojstranný $100(1 - \alpha)\%$ interval spoľahlivosti pre μ pri známom σ^2 . Z rozdelenia štatistiky T a z definície kritickej hodnoty vyplýva, že interval s koncovými bodmi $\bar{x} \mp \frac{s^* t_{n-1}(\alpha)}{\sqrt{n}}$ je $100(1 - \alpha)\%$ interval spoľahlivosti pre μ pri neznámom σ^2 .

Ak je vzorka malá, menšia ako 30, a premenná nepochádza z normálneho rozdelenia, treba použiť neparametrické alternatívy: **znamienkový test a Wilcoxonov test**, ktoré testujú hypotézu o zhode mediánu s konštantou.

4.2. Závislé vzorky

Párový t-test/t-test pre závislé vzorky testuje hypotézu o priemernom rozdiel hodnôt párov. Zisťuje, či je priemerný rozdiel párov iba náhodný, alebo štatisticky významný. Realizácia t-testu v tejto situácii je rovnaká ako pri jednovýberovom t-teste pre rozdiely .

Na použitie testu je nutné splniť nasledujúci predpoklad:

- Normálne rozdelenie rozdielov párov .

Procedúra testovania - párový t-test

	Párový t-test
1.	$H_0: \mu_1 = \mu_2$ proti obojstrannej alternatíve
2.	$\alpha = 0,05 \vee \alpha = 0,01 \vee \dots$
3.	$T = \frac{ \bar{d} }{s_d} \sqrt{n}$
4.	$ T \geq t_{n-1}(\alpha) \vee p \leq \alpha$

Ak je vzorka malá, menšia ako 30, a premenná nepochádza z normálneho rozdelenia, treba použiť neparametrické alternatívy: **znamienkový test a Wilcoxonov test**, ktoré testujú medián rozdielov párov. Za predpokladu, že závislá premenná je binárna môžeme použiť **McNemarov test**.

V prípade, že by sme chceli porovnať viac ako dve závislé vzorky použili by sme **opakované merania analýzy rozptylu**. Jej neparametrickou alternatívou je **Friedmanov test a Kendallov koeficient zhody** alebo **Cochranov Q test** v prípade, že závislá premenná je binárna. Cochranov Q test predstavuje rozšírenie McNemarovho testu na viac ako dve závislé vzorky.

Kendallov koeficient zhody predstavuje mieru zhody medzi závislými vzorkami. Hodnota koeficientu je z intervalu $<0,1>$, pričom 1 znamená dokonalú zhodu a 0 znamená nezhodu, t.j. nezávislosť vzoriek.

4.3. Nezávislé vzorky

Dvojvýberový t-test testuje hypotézu o rozdiel priemerov dvoch nezávislých skupín (Tabuľka 23). Zisťuje, či sú namerané rozdiely medzi priermi iba náhodné, alebo štatisticky významné.

Na použitie testu je nutné splniť nasledujúce predpoklady:

- Normálne rozdelenie závislej premennej v oboch výberoch.
- Rovnosť rozptylov.

V prípade nesplnenia predpokladu o rovnosti rozptylov môžeme použiť **Cochran-Coxov t-test**.

Procedúra testovania - dvojvýberový t-test/Cochran-Coxov t-test

	Dvojvýberový t-test	Cochran-Coxov t-test
1.	$H_0: \mu_1 = \mu_2$ proti obojstrannej alternatíve	
2.	$\alpha = 0,05 \vee \alpha = 0,01 \vee \dots$	
3.	$T = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$	$T = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
4.	$ T \geq t_{n_1+n_2-2}(\alpha) \vee p \leq \alpha$	$ T \geq \frac{\frac{s_1^2}{n_1} t_{n_1-1}(\alpha) + \frac{s_2^2}{n_2} t_{n_2-1}(\alpha)}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \vee p \leq \alpha$

Ak je vzorka malá (menšia ako 30 v každej skupine) a premenná nemá normálne rozdelenie v skupinách, treba použiť neparametrické alternatívy: **Mann-Whitneyov U test/dvojvýberový Wilcoxonov test, Kolmogorov-Smirnovov dvojvýberový test a Wald-Wolfowitzov test**.

V prípade, že by sme chceli porovnať viac ako dve nezávislé vzorky použili by sme **analýzu rozptylu jednoduchého triedenia**. Jej neparametrickou alternatívou je **Kruskal-Wallisova analýza rozptylu**, ktorá pracuje s poradovými číslami alebo **mediánový test**, ktorý pracuje s početnosťami. Kruskal-Wallisova analýza rozptylu je rozšírením Mann-Whitneyho U testu na viac ako dve nezávislé vzorky.

5. Vztáhy medzi premennými

Vztáhy medzi premennými

5.1. Analýza kontingencie

Tieto testy zisťujú vzťahy medzi dvoma nominálnymi premennými, t.j. používajú sa k analýze závislosti nominálnych premenných. Patrí sem skupina neparametrických testov, ktoré vychádzajú z kontingenčnej tabuľky (Tabuľka 24). Tieto testy overujú nulovú hypotézu, ktorá tvrdí, že premenné sú nezávislé. Jedna skupina testov je určená iba pre štvorpoľné kontingenčné tabuľky (2x2), v ktorých vystupujú dve dichotomické premenné.

Kontingenčná tabuľka pozorovaných početností $R \times S$

$X \backslash Y$	y_1	y_2	...	y_s	Σ
x_1	a_{11}	a_{12}	...	a_{1s}	r_1
x_2	a_{21}	a_{22}	...	a_{2s}	r_2
...
x_R	a_{R1}	a_{R2}	...	a_{Rs}	r_R
Σ	s_1	s_2	...	s_s	n

Chí-kvadrát test nezávislosti môžeme použiť k analýze viacpoľných tabuliek. Chí-kvadrát test nezávislosti predstavuje rozšírenie chí-kvadrát testu dobrej zhody a vychádza z kontingenčnej tabuľky pozorovaných početností, kde pozorovaná početnosť a_{ij} je početnosť (frekvencia) kombinácie $x_i \wedge y_j$.

Očakávané početnosti e_{ij} sú také, ktoré zodpovedajú nulovej hypotéze o nezávislosti dvoch premenných.

Očakávaná početnosť príslušnej bunky sa rovná podielu súčinu príslušnej pozorovanej početnosti riadku a stĺpca a celkovému počtu pozorovaní.

$$e_{ij} = \frac{r_i s_j}{n}$$

Chí-kvadrát test overuje, či môžu byť rozdiely skutočných a očakávaných početností iba náhodné (premenné sú nezávislé) alebo štatisticky významné (premenné sú závislé).

Chí-kvadrát test môžeme použiť iba v prípade, že očakávané početnosti sú dostatočne veľké.

$$e_{ij} = \frac{r_i s_j}{n} \geq 5$$

Taktiež výsledky z kontingenčných tabuliek, ktoré obsahujú nulové početnosti, treba brať s rezervou.

V prípade, že očakávané početnosti nebudú dostatočne veľké, môžeme použiť **Fisherov test**, ktorý je však použiteľný iba pre štvorpoľné tabuľky.

Procedúra testovania - chí-kvadrát test nezávislosti/Fisherov test

	Chí-kvadrát test nezávislosti	Fisherov test
1.	$H_0: P(X \wedge Y) = P(X)P(Y)$ proti alternatíve X, Y sú závislé	
2.	$\alpha = 0,05 \vee \alpha = 0,01 \vee \dots$	
3.	$\chi^2 = \sum_{i=1}^R \sum_{j=1}^S \frac{(a_{ij} - e_{ij})^2}{e_{ij}}$	$P = \sum_{i=0}^{\min(r_1, s_1) - a_{11}} \frac{r_1! r_2! s_1! s_2!}{n! (a_{11} + i)! (a_{12} - i)! (a_{21} - i)! (a_{22} + i)!}$
4.	$\chi^2 \geq \chi^2_{(R-1)(S-1)}(\alpha) \vee p \leq \alpha$	$P \leq \alpha$

Kontingenčné koeficienty (Pearsonov, Cramerov V) predstavujú mieru vzťahu medzi dvoma nominálnymi premennými. Nadobúdajú hodnoty z intervalu 0 (žiadny vzťah) až 1 (dokonalý vzťah). Chí-kvadrát testom nezávislosti môžeme testovať **významnosť kontingenčných koeficientov**.

Ďalej ako mieru vzťahu dvoch nominálnych premenných môžeme použiť symetrické korelačné koeficienty **Spearmanov, Kendallov, koeficient gama** a asymetrický koeficient **Sommerovo D**, ktorý je rozšírením gama koeficientu. Asymetrický koeficient rozlišuje závislú a nezávislú premennú na rozdiel od predchádzajúcich koeficientov. Štatistický softvér väčšinou ponúkne k výpočtu obidve možnosti: $D(X/Y)$, $D(Y/X)$.

Fí-kvadrát môžeme použiť ako mieru vzťahu iba v štvorpoľných kontingenčných tabuľkách (2x2), t.j. medzi dvoma dichotomickými premennými.

Všetky miery môžeme interpretovať rovnako ako korelačné koeficienty.

5.2. Korelačná analýza

Korelačná analýza skúma tesnosť štatistickej závislosti medzi kvantitatívnymi premennými. Korelačná analýza na rozdiel od regresie nevyjadruje príčinnno-následný vzťah. Premenná Y nezávisí na premennej X , ale dve náhodné premenné X a Y sa spoločne menia. Regresná analýza predpokladá, že premenná Y je náhodná a premenná X fixná. Korelačný koeficient je mierou lineárnej závislosti dvoch premenných.

Korelačný koeficient:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Výberový/Pearsonov korelačný koeficient:

$$r_{XY} = \frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}},$$

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$r_{XY} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{\left(n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \cdot \left(n \cdot \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}.$$

Výberový korelačný koeficient r definujeme pomocou analogického vzorca, v ktorom sa neznáme rozptyly a neznáma kovariancia nahradí ich nestrannými odhadmi. Čitateľ sa nazýva kovariancia a vyjadruje ako sa súčasne menia hodnoty dvoch premenných. Kladná hodnota znamená, že sa menia spoločne jedným smerom, záporná hodnota znamená, že sa menia opačným smerom a nula, že sa menia nezávisle. Vydelením kovariancie smerodajnými odchýlkami sa vypočíta korelačný koeficient, ktorého hodnota sa nachádza v intervale od -1 do 1. Pričom -1 znamená nepriamu úmernosť, 1 znamená priamu úmernosť a 0 nezávislosť X a Y .

Interpretácia korelačného koeficientu

Interpretácia korelačného koeficientu podľa Cohena	
ρ_{XY}	interpretácia
0,0 - 0,1	triviálna korelácia
0,1 - 0,3	malá
0,3 - 0,5	stredná
0,5 - 0,7	veľká
0,7 - 0,9	veľmi veľká
0,9 - 1,0	takmer dokonalá

Viacnásobný koeficient korelácie je mierou lineárnej závislosti premennej Y a náhodného vektora. Jeho štvorec sa nazýva **koeficient determinácie** a vyjadruje podiel variability premennej Y vysvetlenej regresným modelom k celkovej variabilite premennej Y . Pre dvojrozmerný náhodný vektor (X, Z) sa viacnásobný koeficient korelácie $r_{Y,(X,Z)}$ vypočíta podľa vzorca:

$$r_{Y,(X,Z)}^2 = \frac{r_{YX}^2 + r_{YZ}^2 - 2r_{YX}r_{YZ}r_{XZ}}{1 - r_{XZ}^2}.$$

Ktorá z nezávislých premenných má najväčší vplyv na premennú Y sa zisťuje z **parciálnych korelačných koeficientov**. Nasledujúci vzorec parciálnej korelácie určuje mieru závislosti premenných Y a Z s vylúčením vplyvu premennej X (Y je závislá premenná, Z a X sú nezávislé premenné):

$$r_{Y,Z.X} = \frac{r_{YZ} - r_{YX}r_{ZX}}{\sqrt{(1-r_{YX}^2)(1-r_{ZX}^2)}}$$

Testy významnosti korelačných koeficientov testujú, či je koeficient korelácie štatisticky významne rôzny od nuly.

Na použitie testov je nutné splniť nasledujúci predpoklad:

- Normálne rozdelenie premenných.

Procedúra testovania - test významnosti výberového korelačného koeficientu

	Test významnosti výberového korelačného koeficientu
1.	$H_0: \rho = 0$ proti obojstrannej alternatíve
2.	$\alpha = 0,05$ v $\alpha = 0,01$ v ...
3.	$T = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$
4.	$ T \geq t_{n-2}(\alpha) \vee p \leq \alpha$

Testy významnosti viacnásobného korelačného koeficientu a parciálneho korelačného koeficientu sú ekvivalentné s testovaním významnosti regresných koeficientov.

Korelačný koeficient môže byť ovplyvnený extrémnymi hodnotami, a to v oboch smeroch. Jediný extrémny prípad vo veľkom súbore môže významne znížiť silnú závislosť, ale aj vyrobiť silnú závislosť tam, kde žiadna nie je. Dôležité závery sa nesmú robiť iba na základe hodnoty koeficientu. Vždy je nutné preskúmať graf zobrazujúci objekty definované súradnicami X a Y v rovine. Z grafu možno zistiť aj nelineárny, ale silný vzťah medzi premennými. V takom prípade treba linearizovať vzťah tak, že sa jedna premenná transformuje a následne sa použije na výpočet korelácie.

Na prítomnosť extrémnych hodnôt netrpia neparametrické korelačné koeficienty. Ako mieru vzťahu v prípade výskytu extrémnych hodnôt medzi dvoma intenzívnymi premennými (ordinálne, metrické) môžeme použiť:

Spearmanov korelačný koeficient, Kendallov koeficient tau a Koeficient gama.