MA012 Statistika II

6. Korelační analýza: korelace a koeficient determinace, pořadové korelační koeficienty

Ondřej Pokora (pokora@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

(podzim 2015)

Motivační příklady

Příklad 1

Byly sledovány výdaje (V) 7 domácností (v tisících Kč za 3 měsíce) za potraviny a nápoje v závislosti na počtu členů domácnosti (C) a na čistém příjmu (P) domácnosti (v tisících Kč za 3 měsíce).

V	40	30	40	10	60	40	50
C	4	2	4	1	5	3	4
P	100	80	120	30	150	120	130

Zkoumejte závislosti (asociovanost) veličin.

Příklad 2

20 dětí různého věku se podrobilo pedagogicko-psychologickému výzkumu, v rámci něhož mj. odpovídaly na tytéž otázky testu a byly váženy. Překvapivý výsledek přinesl korelační koeficient mezi hmotností dětí a počtem bodů dosažených v testu, jehož hodnota vyšla 0,968. Znamená to, že obezita má pozitivní vliv na schopnost učení? Prozkoumejte závislosti (asociovanost) veličin.

Funkce pro výběrové korelační koeficienty v R

Pearsonův	r_{YZ}	`rcorr (X)` $*$
		`cor (X, Y)`
		`cor.test (X, Y)`
		`cor (X)`

parciální	$r_{YZ\cdot\mathbf{X}}$	`pcor (X)` $\star$ `pcor.test (X, Y)` $\star$
semiparciální	$r_{Y(Z\cdot\mathbf{X})}$	`spcor (X)` $\star$ `spcor.test (X, Y)` $\star$
mnohonásobný	$r_{Y\cdot\mathbf{X}}$	R^2 ve výsledku LRM funkcí `lm`

$*$ `library (Hmisc)`, $\star$ `library (ppcor)`

Testy významnosti korelačních koeficientů

Věta 1

Za platnosti $\rho_{YZ\cdot\mathbf{X}} = 0$ *je*

$$T = r_{YZ\cdot\mathbf{X}}\sqrt{\frac{n-p-2}{1-r^2_{YZ\cdot\mathbf{X}}}} \ \sim\ t(n-p-2);$$

koeficient parciální korelace je tedy na hladině α *významný, pokud*

$$|T| \geq t_{1-\alpha/2}(n-p-2).$$

Věta 2 (analogie celkového F-testu v lineárním regresním modelu)

Za platnosti $\rho_{Y\cdot\mathbf{X}} = 0$ *je*

$$F = \frac{n-p-1}{p}\cdot\frac{r^2_{Y\cdot\mathbf{X}}}{1-r^2_{Y\cdot\mathbf{X}}} \ \sim\ F(p, n-p-1);$$

koeficient mnohonásobné korelace je tedy na hladině α *významný, pokud*

$$F \geq F_{1-\alpha}(p, n-p-1).$$

Testování významnosti semiparciálních korelačních koeficientů se provádí podobně, pomocí statistiky s F-rozdělením, avšak s jinými stupni volnosti.

Výpočty korelačních koeficientů podle definice

Pearsonův	$r_{YZ} = r_{ZY} = r(Y,Z) = r(Z,Y)$	`cor (Y, Z)`
parciální	$r_{YZ\cdot\mathbf{X}} = r_{ZY\cdot\mathbf{X}} = r(Y-\widehat{Y}, Z-\widehat{Z})$	`cor (rY, rZ)`
semiparciální	$r_{Y(Z\cdot\mathbf{X})} = r(Y, Z-\widehat{Z})$	`cor (Y, rZ)`
semiparciální	$r_{Z(Y\cdot\mathbf{X})} = r(Z, Y-\widehat{Y})$	`cor (Z, rY)`
mnohonásobný	$r_{Y\cdot\mathbf{X}} = r(Y, \widehat{Y})$	`cor (Y, Yhat)`

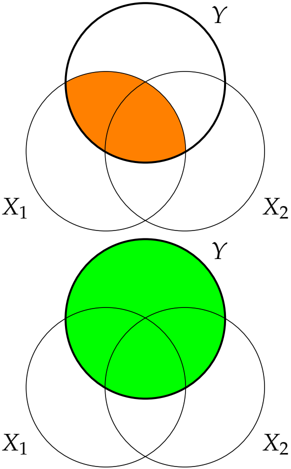
Odhady a rezidua přitom získáme vyřešením lineárních regresních modelů:

$$Y = \beta_0 + \mathbf{X}\boldsymbol{\beta} \implies \widehat{Y}, \qquad Z = \alpha_0 + \mathbf{X}\boldsymbol{\alpha} \implies \widehat{Z}$$

Symbolický zápis v *R*:

```
modelY <- lm (Y ~ X1 + ... + Xp)
modelZ <- lm (Z ~ X1 + ... + Xp)
rY <- modelY$residuals
rZ <- modelZ$residuals
Yhat <- modelY$fitted.values
```

Koeficient determinace a Pearsonova korelace



- veličinu Y modelujeme veličou X_1
- Pearsonův korelační koeficient $r_{Y\,X_1}$ popisuje míru závislosti mezi veličinami Y a X_1
- koeficient determinace $R^2_{Y\cdot X_1}$ v LRM

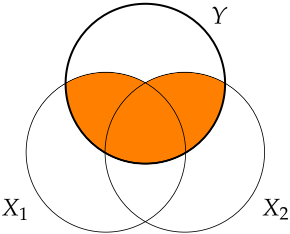
```
Y ~ X1
```

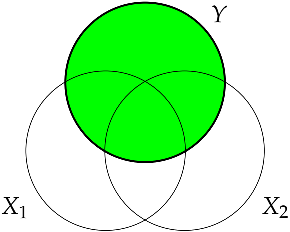
popisuje, jakou část celkové variability veličiny Y lze vysvětlit veličinou X_1

⇒ kvadrát Pearsonova korelačního koeficientu popisuje, jakou část celkové variability veličiny Y lze vysvětlit veličinou X_1,

$$r^2_{Y\,X_1} = \frac{R^2_{Y\,X_1}}{1} = R^2_{Y\,X_1}$$

Koeficient determinace a mnohonásobná korelace





- veličinu Y modelujeme veličinami X_1, X_2
- koeficient mnohonásobné korelace $r_{Y \cdot X_1 X_2}$ popisuje míru závislosti mezi Y a nejlepší lineární kombinací $\widehat{Y}$ veličin X_1, X_2
- koeficient determinace $R^2_{Y \cdot X_1 X_2}$ v LRM

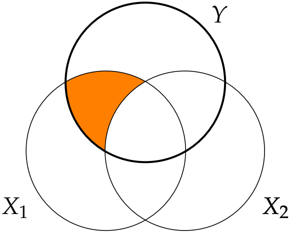
```
Y ~ X1 + X2
```

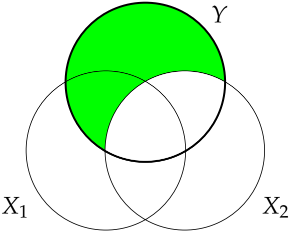
popisuje, jakou část celkové variability veličiny Y lze vysvětlit veličinami X_1, X_2

$\Rightarrow$ kvadrát koeficientu mnohonásobné korelace popisuje, jakou část celkové variability veličiny Y lze vysvětlit veličinami X_1, X_2,

$$r^2_{Y \cdot X_1 X_2} = \frac{R^2_{Y \cdot X_1 X_2}}{1} = R^2_{Y \cdot X_1 X_2}$$

Koeficient determinace a parciální korelace





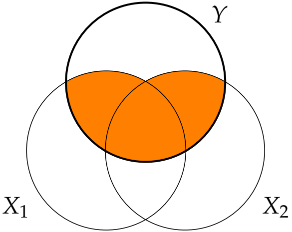
- veličinu Y modelujeme veličinou X_1, přičemž vylučujeme vliv veličiny X_2 na obě tyto veličiny zároveň
- koeficient parciální korelace $r_{Y X_1 \cdot X_2}$ popisuje míru závislosti mezi Y a X_1 při vyloučení vlivu X_2 na obě tyto veličiny zároveň

$\Rightarrow$ kvadrát koeficientu parciální korelace popisuje, jakou část variability veličiny Y nezávislé na veličině X_2 lze vysvětlit samotnou veličinou X_1,

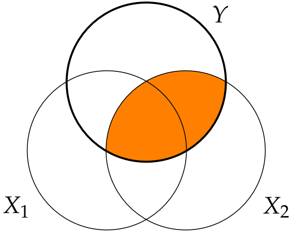
$$r^2_{Y X_1 \cdot X_2} = \frac{\text{variabilita } Y \text{ v oranžové oblasti}}{\text{variabilita } Y \text{ v zelené oblasti}}$$

Koeficient determinace a parciální korelace

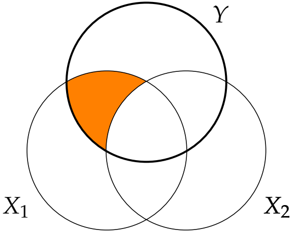
$R^2_{Y\cdot X_1 X_2}$



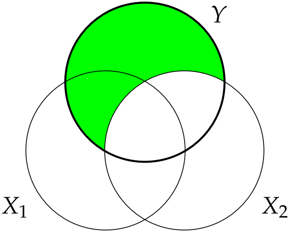
`Y ~ X2`: $R^2_{Y\cdot X_2}$



$R^2_{Y\cdot X_1 X_2} - R^2_{Y\cdot X_2}$



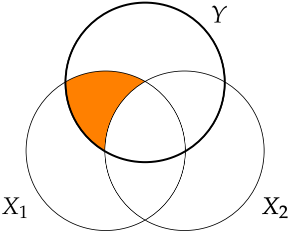
$$\boxed{\Rightarrow} \qquad r^2_{Y X_1 \cdot X_2} = \frac{R^2_{Y\cdot X_1 X_2} - R^2_{Y\cdot X_2}}{1 - R^2_{Y\cdot X_2}}$$

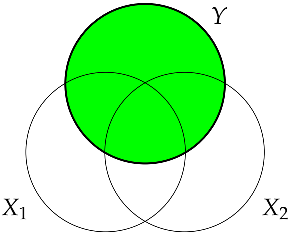


$1 - R^2_{Y\cdot X_2}$

Všimněte si, že `Y ~ X2` je podmodel `Y ~ X1 + X2`.
Koeficient determinace $R^2_{Y\cdot X_2}$ v `Y ~ X2` popisuje, jakou část celkové variability Y lze vysvětlit pomocí X_2.

Koeficient determinace a semiparciální korelace





- veličinu Y modelujeme veličinou X_1, přičemž vylučujeme vliv veličiny X_2 na veličinu X_1
- koeficient parciální korelace $r_{Y\,X_1\,\cdot\,X_2}$ popisuje míru závislosti mezi Y a X_1 při vyloučení vlivu X_2 na veličinu X_1

⇒ kvadrát koeficientu parciální korelace popisuje, jakou část celkové variability veličiny Y lze vysvětlit samotnou veličinou X_1,

$$r^2_{Y\,\cdot\,(X_1\,X_2)} =$$

$$= \frac{R^2_{Y\,\cdot\,X_1\,X_2} - R^2_{Y\,\cdot\,X_2}}{1} = R^2_{Y\,\cdot\,X_1\,X_2} - R^2_{Y\,\cdot\,X_2}$$

Souvislost koeficientů korelace a determinace

Věta 3

Kvadrát koeficientu mnohonásobné korelace je rovný koeficientu determinace,

$$r^2_{Y\cdot\mathbf{X}} = R^2_{Y\cdot\mathbf{X}}\ , \qquad r^2_{Y\cdot Z\mathbf{X}} = R^2_{Y\cdot Z\mathbf{X}}\ .$$

Věta 4

Pro kvadrát výběrového parciálního korelačního koeficientu platí

$$r^2_{YZ\cdot\mathbf{X}} = r^2_{ZY\cdot\mathbf{X}} = \frac{R^2_{Y\cdot Z\mathbf{X}} - R^2_{Y\cdot\mathbf{X}}}{1 - R^2_{Y\cdot\mathbf{X}}} = \frac{R^2_{Z\cdot Y\mathbf{X}} - R^2_{Z\cdot\mathbf{X}}}{1 - R^2_{Z\cdot\mathbf{X}}}.$$

Věta 5

Pro kvadrát výběrového semiparciálního korelačního koeficientu platí

$$r^2_{Y(Z\cdot\mathbf{X})} = R^2_{Y\cdot Z\mathbf{X}} - R^2_{Y\cdot\mathbf{X}}\ , \qquad r^2_{Z(Y\cdot\mathbf{X})} = R^2_{Z\cdot Y\mathbf{X}} - R^2_{Z\cdot\mathbf{X}}\ .$$

$R^2_{Y\cdot\mathbf{X}}$, resp. $R^2_{Y\cdot Z\mathbf{X}}$, je koeficient determinace R^2 v lineárním regresním modelu `Y ~ X`, resp. `Y <- Z + X`. Přitom $\mathbf{X} = (X_1, \ldots, X_p)$ může být vektor veličin.

Vlastnosti korelačních koeficientů

Korelační koeficienty nabývají hodnot z intervalu $[-1; 1]$:

$$-1 \leq r_{YZ} \leq 1, \qquad -1 \leq r_{YZ\cdot\mathbf{X}} \leq 1, \qquad -1 \leq r_{Y(Z\cdot\mathbf{X})} \leq 1.$$

Výjimkou je koeficient mnohonásobné korelace, který je navíc nezáporný:

$$0 \leq r_{Y\cdot\mathbf{X}} \leq 1.$$

Pro libovolnou lineární kombinaci Y^* veličin $X_1, \dots, X_p$ platí:

$$|r_{YY^*}| \leq r_{Y\cdot\mathbf{X}}\ , \qquad \text{spec.} \quad |r_{YX_j}| \leq r_{Y\cdot\mathbf{X}}\ .$$

Pro Pearsonův, parciální a semiparciální korelační koeficient platí:

$$|r_{Y(X_1\cdot X2)}| \leq |r_{YX_1\cdot X2}|\ , \qquad |r_{Y(X_1\cdot X2)}| \leq |r_{YX_1}|\ .$$

Přitom $r_{Y(X_1\cdot X2)} = r_{YX_1}$, pokud Y lze modelovat pomocí X_1 nezávisle na X_2.

Platnost uvedených nerovností mezi korelačními koeficienty ověřte porovnáním odpovídajících schémat variability a vyjádření pomocé koeficientů determinace.

Příklad 1: mnohonásobná korelace $r_{V\cdot CP}$

Vycházíme z nejbohatšího modelu `V ~ C + P`.

```
# 1. jako korelace s nejlepsim linearnim odhadem
m <- lm (V ~ C + P, data = tabulka)
v <- summary (m)
cor (V, m$fitted.values)

# 2. jako r.squared v linearnim regresnim modelu
sqrt (v$r.squared)

0.9826827
```

$r_{V\cdot CP} = 0{,}983, \qquad r^2_{V\cdot CP} = 0{,}966$

Koeficient mnohonásobné korelace mezi veličinou V a skupinou veličin C, P je 0,983. Jedná se o (v absolutní hodnotě) největší korelaci, jaké lze na základě lineárního modelu pro V s veličinami C, P na daných datech dosáhnout.

Pomocí modelu mnohnásobné linární regrese jsme tedy na základě našich dat schopni vysvětlit 96,6 % variability veličiny V (výdaje domácností) pomocí veličin C (počet členů domácnosti) a P (čisté příjmy domácnosti).

Příklad 1: parciální korelace $r_{PV \cdot C}$, $r_{VP \cdot C}$

Výběrový parciální korelační koeficint počítáme jako Pearsonovu korelaci mezi rezidui dvojice lineárních regresních modelů `P ~ C` a `V ~ C`.

```
# 1. jako korelace mezi dvema rezidui
m1 <- lm (P ~ C, data = tabulka)
m2 <- lm (V ~ C, data = tabulka)
v1 <- summary (m1)
v2 <- summary (m2)
cor (m1$residuals , m2$residuals)

0.8311677
```

Příklad 1: parciální korelace $r_{PV \cdot C}$, $r_{VP \cdot C}$

Alternativně lze výpočet (až na znaménko) provést pomocí koeficientů determinace dvojice lineárních regresních modelů `P ~ C + V` a `P ~ C` nebo dvojice modelů `V ~ C + P` a `V ~ C`.

```
# 2. pomoci r.squared v linearnich regresnich modelech
m01 <- lm (P ~ C + V, data = tabulka)
v01 <- summary (m01)
sqrt ( (v01$r.squared - v1$r.squared) / (1 - v1$r.squared) )

m02 <- lm (V ~ C + P, data = tabulka)
v02 <- summary (m02)
sqrt ( (v02$r.squared - v2$r.squared) / (1 - v2$r.squared) )

0.8311677
```

Příklad 1: semiparciální korelace $r_{P(V \cdot C)}$

Semiparciální korelaci lze spočítat jako Pearsonovu korelaci mezi `P` a rezidui modelu `V ~ C` nebo pomocí koeficientů determinace dvojice lineárních regresních modelů `P ~ V + C` a `P ~ C`.

```
# 1. jako korelace mezi velicinou a reziduem druhe veliciny
m2 <- lm (V ~ C, data = tabulka)
cor (P, m2$residuals)

# 2. pomoci r.squared v linearnich regresnich modelech
m01 <- lm (P ~ V + C, data = tabulka)
v01 <- summary (m01)
m1 <- lm (P ~ C, data = tabulka)
v1 <- summary (m1)
sqrt (v01$r.squared - v1$r.squared)

0.3236838
```

Příklad 1: semiparciální korelace $r_{V(P \cdot C)}$

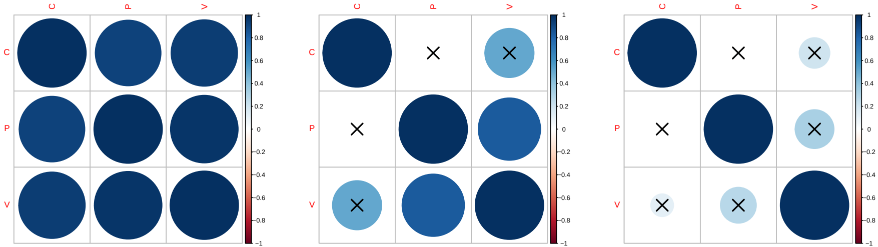
Podobně, jako Pearsonovu korelaci `V` a rezidui modelu `P ~ C` nebo pomocí koeficientů determinace dvojice lineárních regresních modelů `V ~ P + C` a `V ~ C`.

```
# 1. jako korelace mezi velicinou a reziduem druhe veliciny
m1 <- lm (P ~ C, data = tabulka)
cor (V, m1$residuals)

# 2. pomoci r.squared v linearnich regresnich modelech
m02 <- lm (V ~ P + C, data = tabulka)
v02 <- summary (m02)
m2 <- lm (V ~ C, data = tabulka)
v2 <- summary (m2)
sqrt (v02$r.squared - v2$r.squared)

0.2769893
```

Příklad 1: Korelogramy



	C	P	V
C	1	0.92	0.94
P	0.92	1	0.98
V	0.94	0.98	1

	C	P	V
C	1	0.01	0.52
P	0.01	1	0.83
V	0.52	0.83	1

	C	P	V
C	1	0	0.2
P	0	1	0.32
V	0.11	0.28	1

Příklad 2: mnohonásobná lineární regrese

model mnohonásobné lineární regrese: $Body = \beta_0 + \beta_1 Hmotnost + \beta_2 Vek$

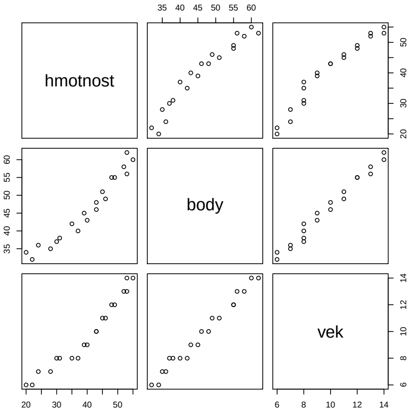
```
model <- lm (body ~ hmotnost + vek, data = tabulka)
summary (model)
```

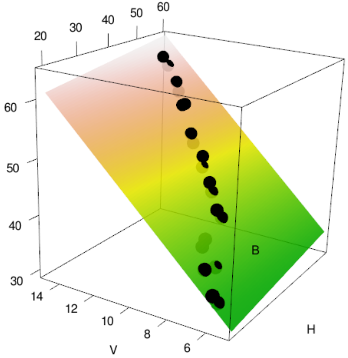
```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.06490    1.23693   8.945 7.72e-08 ***
hmotnost     0.09466    0.12090   0.783    0.444
vek          3.19203    0.51058   6.252 8.77e-06 ***

Residual standard error: 1.377 on 17 degrees of freedom
Multiple R-squared:  0.9806,  Adjusted R-squared:  0.9784
F-statistic: 430.4 on 2 and 17 DF,  p-value: 2.753e-15
```

MNČ-odhady: $\beta_1 = 0{,}095 > 0, \quad \beta_2 = 3{,}192 > 0 *{*}{*}, \quad R^2 = 0{,}981, \quad F*$

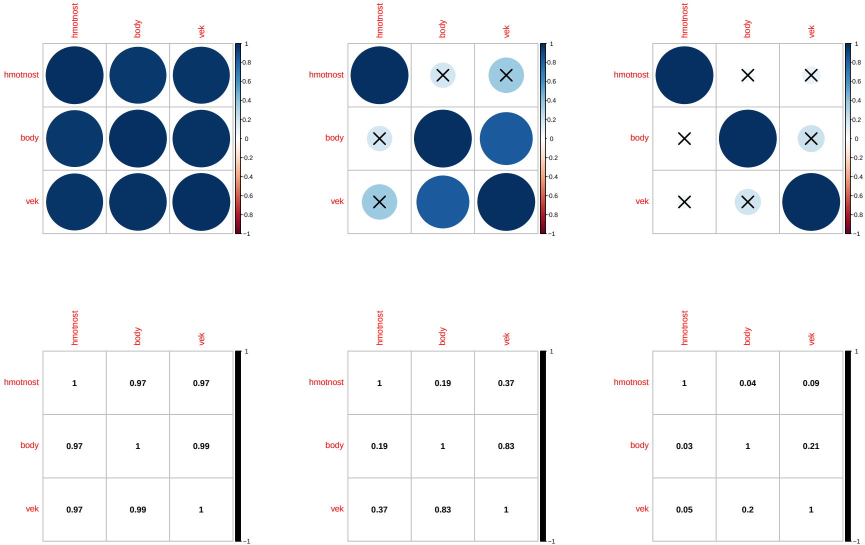
Příklad 2: scatter-plot a regresní rovina





rovnice regresní roviny: $Body = 11{,}065 + 0{,}095\,Hmotnost + 3{,}192\,Vek$

Příklad 2: korelogramy



Spearmanův pořadový korelační koeficient

Neparametrickou analogií korelačního koeficientu r je Spearmanův pořadový korelační koeficient r_S. Je definován jako Pearsonův korelační koeficient mezi (průměrnými) pořadími v uspořádaných náhodných výběrech. Používáme jej zejména v situacích, kdy náhodné výběry mají výrazně nenormální rozdělení pravděpodobnosti.

Označme $R_1, \ldots, R_n$, resp. $S_1, \ldots, S_n$ pořadí X_i a Y_i v uspořádaných výběrech

$$X_{(1)} \leq \cdots \leq X_{(n)}, \quad \text{resp.} \quad Y_{(1)} \leq \cdots \leq Y_{(n)}.$$

Definice 6 (Spearmanův korelační koeficient)

$$r_S = r(R, S) = 1 - 6 \, \frac{\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \in [-1; 1],$$

kde $d_i = R_i - S_i$ jsou rozdíly pořadí v X-ovém a Y-ovém náhodném výběru.

V *R* je výpočet r_S implementován ve funkci `cor (X, Y, method="spearman")`

Test významnosti Spearmanova r_S

Test významnosti r_S, tedy test hypotézy o nulovosti r_S, lze provádět pomocí některé z následujících testovacích statistik.

Věta 7 (Test významnosti Spearmanova korelačního koeficientu)

Hypotézu $H_0 : r_S = 0$ na hladině významnosti α,

$$\textit{pokud} \qquad |T| \geq t_{1-\alpha/2}(n-2)$$

$$\textit{pro} \qquad T = r_S \sqrt{\frac{n-2}{1-r_S^2}}$$

$$\textit{nebo pokud} \qquad |Z| \geq u_{1-\alpha/2}$$

$$\textit{pro} \qquad Z = \sqrt{\frac{n-3}{1{,}06}} \cdot \frac{1}{2} \ln \frac{1+r_S}{1-r_S}.$$

V *R* je test implementován ve funkci `cor.test (X, Y, method="spearman")`

Kendallův korelační koeficient

Na principu pořadí, konkrétně souhlasného či nesouhlasného pořadí párů, je založen i další pořadový korelační koeficient, tzv. Kendallovo τ,

Definice 8 (Kendallův korelační koeficient)

$$\tau = \frac{n_+ - n_-}{\sqrt{n_0 - n_X}\sqrt{n_0 - n_Y}} \in [-1;1],$$

- $n_0 = \frac{1}{2}n(n-1)$ = počet všech párů,
- n_+ = počet konkordantních párů,
- n_- = počet diskordantních párů,
- $n_X = \sum_i \frac{1}{2}u_i(u_i - 1)$, resp. $n_Y = \sum_j \frac{1}{2}v_j(v_j - 1)$,

  kde u_i, resp. v_j, jsou počty opakování hodnot v X-ovém, resp. Y-ovém výběru.

Páry (X_i, Y_i) a (X_j, Y_j) nazýváme

- **konkordantní**, pokud jsou pořadí jejich elementů souhlasná, tzn. $X_i < X_j$ & $Y_i < Y_j$, anebo $X_i > X_j$ & $Y_i > Y_j$
- **diskordantní**, pokud jsou pořadí jejich elementů nesouhlasná, tzn. $X_i < X_j$ & $Y_i > Y_j$, anebo $X_i > X_j$ & $Y_i < Y_j$

Test významnosti Kendallova τ

Výpočet Kendallova τ v *R*: `cor (X, Y, method="kendall")`

Asymptotický test významnosti τ je v případě neopakovaných hodnot v náhodných výběrech založen na asymptotické normalitě τ, s $\mathrm{E}\tau = 0$, za platnosti nulové hypotézy $H_0 : \tau = 0$.

Věta 9 (Asymptotický test významnosti Kendallova τ)

Hypotézu $H_0 : \tau = 0$ zamítáme na asymptotické hladině α, pokud

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}}\,|\tau| \geq u_{1-\alpha/2}.$$

Pro malé rozsahy n a v případě výskytu opakovaných hodnot v některém náhodném výběru se používají korigované τ-statistiky.

Test významnosti Kendallova τ v *R*: `cor.test (X, Y, method="kendall")`

Příklad 1: výpočet Spearmanova $r_S(C,V)$

V	40	30	40	10	60	40	50	
C	4	2	4	1	5	3	4	
P	100	80	120	30	150	120	130	
								průměry
R = pořadí V	4	2	4	1	7	4	6	4
S = pořadí C	5	2	5	1	7	3	5	4
T = pořadí P	3	2	4,5	1	7	4,5	6	4
								součty
R·S	20	4	20	1	49	12	30	136
S·T	15	4	22,5	1	49	13,5	30	135
R·T	12	4	18	1	49	18	36	138
R^2	16	4	16	1	49	16	36	138
S^2	25	4	25	1	49	9	25	138
T^2	9	4	20,25	1	49	20,25	36	139,5

$$r_S(C,V) = r(R,S) = \frac{\sum_{i=1}^n (R_i\, S_i) - n\,\overline{R}\,\overline{S}}{\sqrt{\sum_{i=1}^n R_i^2 - n\overline{R}^2}\sqrt{\sum_{j=1}^n S_j^2 - n\overline{S}^2}} = \frac{136 - 7\cdot 4^2}{138 - 7\cdot 4^2} = 0{,}923$$

Příklad 1: Spearmanův $r_S(C, V)$ v R

```
R <- rank (V)
S <- rank (C)
cor (R, S)

cor (C, V, method = "spearman")
cor.test (C, V, method = "spearman")

  Spearmans rank correlation rho

data:  C and V
S = 4.3077, p-value = 0.003023
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9230769
```

Příklad 1: Kendallovo $\tau(C, V)$ v R

```
cor (C, V, method = "kendall")
cor.test (C, V, method = "kendall")

  Kendalls rank correlation tau

data:  C and V
z = 2.6146, p-value = 0.008933
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.8888889
```

Příklad 1: výpočet Kendallova $\tau(C, V)$

i,j	1	2	3	4	5	6	7
V	40	30	40	10	60	40	50
C	4	2	4	1	5	3	4
P	100	80	120	30	150	120	130

Je celkem 21 párů, z toho 16 je konkordantních, žádný diskordantní (konkordantní jsou všechny páry kromě párů na pozicích i–j: 1–3, 1–6, 1–7, 3–6, 3–7).

- $n_0 = \frac{1}{2} 7 \cdot 6 = 21$,
- $n_+ = 16$
- $n_- = 0$
- $n_C = \frac{1}{2} 3 \cdot 2 = 3$, $n_V = \frac{1}{2} 3 \cdot 2 = 3$

$$\tau(C, V) = \frac{n_+ - n_-}{\sqrt{n_0 - n_C}\sqrt{n_0 - n_V}} = \frac{16 - 0}{\sqrt{21 - 3}\sqrt{21 - 3}} = 0{,}889$$

Korelační analýza: shrnutí

- Pearsonův korelační koeficient: definice, výpočet, vlastnosti, interpretace
- Mnohonásobná lineární regrese: zápis, řešení modelu, geometrický význam
- Koeficienty mnohonásobné, parciální a semiparciální korelace: definice, interpretace (vysvětlování závislostí mezi sledovanými náhodnými veličinami)
- Struktura korelační matice, korelogram, scatter-plot
- Souvislost korelačních koeficientů a koeficientů determinace v LRM (vysvětlení variability)
- Pořadové korelační koeficienty – Spearmanův a Kendallův: definice, konkordantní a diskordantní páry
- Význam a interpretace výsledků testů významnosti korelačních koeficientů