MA012 Statistika II

5. Korelační analýza: korelační koeficienty, mnohonásobná lineární regrese

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.

Teorie pravděpodobnosti: korelační koeficient

Z teorie pravděpodobnosti si zopakujme číselné charakteristiky:

$$\text{rozptyl:}\;\; \mathrm{D}X = \mathrm{E}(X^2) - (\mathrm{E}X)^2,$$

$$\text{kovariance:}\;\; \mathrm{C}(X,Y) = \mathrm{E}\left[(X-\mathrm{E}X)(Y-\mathrm{E}Y)\right] = \mathrm{E}(XY) - \mathrm{E}X\,\mathrm{E}Y,$$

$$\textbf{korelační koeficient}:\;\; \rho_{XY} = \frac{\mathrm{C}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}} \in [-1; 1].$$

Uvědomme si, že všechny uvedené číselné charakteristiky jsou teoretické povahy, neboť k výpočtu středních hodnot potřebujeme znát hustoty pravděpodobnosti, resp. pravděpodobnostní funkce, náhodných veličin X, Y:

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x, \quad \mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x, \quad \mathrm{E}(XY) = \iint_{\mathbb{R}^2} xy f(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

$$\mathrm{E}X = \sum_x x\,p(x)\,\mathrm{d}x, \quad \mathrm{E}(X^2) = \sum_x x^2\,p(x)\,\mathrm{d}x, \quad \mathrm{E}(XY) = \sum_x \sum_y x\,y\,p(x,y)\,\mathrm{d}x.$$

V praxi však máme pouze náhodné výběry, ne hustoty/pravděpodobnostní funkce.

Výběrový korelační koeficient

Definice 1 (výběrový korelační koeficient)

Předpokládáme, že máme dvoudimenzionální náhodný výběr rozsahu n,

$$\big((X_1, Y_1), \ldots, (X_n, Y_n)\big)'.$$

Míru lineární závislosti (asociovanosti) náhodných veličin X a Y odhadujeme pomocí **Pearsonova výběrového korelačního koeficientu**

$$r_{XY} = \frac{S_{XY}}{\sqrt{S_X^2}\sqrt{S_Y^2}} = \frac{\text{výběrová kovariance}}{\text{součin výběrových směrodatných odchylek}}.$$

Pro praktické výpočty se používá tvar

$$r_{XY} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2}} = \frac{\sum_{i=1}^{n}(X_i Y_i) - n\overline{X}\,\overline{Y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}\sqrt{\sum_{j=1}^{n} Y_j^2 - n\overline{Y}^2}}.$$

Příklad 1: výpočet r_{CV}

$$r_{CV} = \frac{\sum_{i=1}^{n}(C_i\, V_i) - n\,\overline{C}\,\overline{V}}{\sqrt{\sum_{i=1}^{n} C_i^2 - n\overline{C}^2}\sqrt{\sum_{j=1}^{n} V_j^2 - n\overline{V}^2}}.$$

								průměry
V	40	30	40	10	60	40	50	38,571
C	4	2	4	1	5	3	4	3,286
P	100	80	120	30	150	120	130	104,286
								součty
C·V	160	60	160	10	300	120	200	1010
C·P	400	60	160	10	300	120	200	2700
P·V	4000	2400	4800	300	9000	4800	6500	31800
V^2	1600	900	1600	100	3600	1600	2500	11900
C^2	16	4	16	1	25	9	16	87
P^2	10000	6400	14400	900	22500	14400	16900	85500

$$r_{CV} = \frac{1010 - 7 \cdot 3{,}286 \cdot 38{,}571}{\sqrt{87 - 7 \cdot 3{,}286^2}\sqrt{11900 - 7 \cdot 38{,}571^2}} = \frac{122{,}790}{\sqrt{11{,}415 \cdot 1485{,}846}} = 0{,}942$$

Výběrové číselné statistiky v R

výběrový průměr	$\overline{X}$	`mean (X)`
výběrový rozptyl	S_X^2	`var (X)`
výběrová směrodatná odchylka	$\sqrt{S_X^2}$	`sd (X)`
výběrová kovariance	S_{XY}	`cov (X, Y)`
výběrový korelační koeficient	r_{XY}	`cor (X, Y)`

Test významnosti korelačního koeficientu

Kromě vlastního odhadu korelačního koeficientu v praxi potřebujeme testovat hypotézu **nekorelovanosti** veličin X a Y,

$$H_0 : \rho_{XY} = 0, \qquad H_1 : \rho_{XY} \neq 0.$$

K tomu slouží statistika T, která má za platnosti H_0 Studentovo t-rozdělení,

$$T = r\sqrt{\frac{n-2}{1-r^2}} \sim t(n-2).$$

Věta 2 (Test významnosti Pearsonova korelačního koeficientu)

H_0 *zamítneme na hladině významnosti* α*, pokud*

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

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

Obecný test korelačního koeficientu

Pro testování hypotézy, že korelační koeficient je rovný zvolenému ρ_0,

$$H_{A0} : \rho_{XY} = \rho_0, \qquad H_{A1} : \rho_{XY} \neq \rho_0,$$

navrhnul R. A. Fisher statistku

Definice 3 (Fisherova Z-transformace)

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r},$$

která má za platnosti H_{A0} normální rozdělení s

$$\mathrm{E}Z = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}, \qquad \mathrm{D}Z = \frac{1}{n-3}.$$

Věta 4 (Obecný test Pearsonova korelačního koeficientu)

H_{A0} zamítneme na hladině významnosti α, pokud

$$\sqrt{n-3} \left| Z - \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} \right| \geq u_{1-\alpha/2}.$$

Příklad 1: test významnosti r_{CV}

$$T = 0{,}942\sqrt{\frac{5}{1-0{,}942^2}} = 6{,}326 > t_{0{,}975}(5) = 2{,}571 \implies \rho_{CV} \text{ je významný}$$

$$\sqrt{4}\left|\frac{1}{2}\ln\frac{1+0{,}942}{1-0{,}942} - 0\right| = 3{,}526 > u_{1-\alpha/2} = 1{,}96 \implies \rho_{CV} \text{ je významný}$$

```
cor.test (C, V)

data:  C and V
t = 6.3263, df = 5, p-value = 0.001455
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6544368 0.9917452
sample estimates:
      cor
0.9428374
```

Praktický význam korelačního koeficientu

- H_0 nezamítneme, $r \approx 0$:
  $\Longrightarrow$ není prokázaná lineární závislost veličin X a Y
- $r \approx 1$, H_0 zamítneme:
  $\Longrightarrow$ je prokázaná lineární závislost veličin X a Y, čím větší X tím větší Y
- $r \approx -1$, H_0 zamítneme:
  $\Longrightarrow$ je prokázaná lineární závislost veličin X a Y, čím větší X tím menší Y

Příklad 1

Spočítali jsme $r_{CV} = 0{,}942$.
Lze tedy tvrdit, že čím větší počet členů v domácnosti (C), tím větší celkový čistý příjem (P) domácnosti? Opravdu toto naše data potvrzují?

Kovarianční matice

Definice 5 (Výběrová kovarianční matice)

Předpokládejme, že sledujeme celkem p náhodných veličin $X_1, \ldots, X_p$ a máme p-rozměrný náhodný výběr rozsahu n,

$$\begin{pmatrix} (X_{11}, & \cdots, & X_{1p}) \\ & \vdots & \\ (X_{n1}, & \cdots, & X_{np}) \end{pmatrix}.$$

Výběrová kovarianční matice je matice výběrových kovariancí $S_{ij} = S_{X_i X_j}$:

$$\mathbf{S} = \left\{S_{ij}\right\}_{i,j=1}^{p}.$$

$$S_{X_i X_j} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ik} - \overline{X_i})(Y_{jk} - \overline{Y_j}), \qquad i, j = 1, \ldots, p$$

$\mathbf{S}$ je symetrická čtvercová matice rozměru $p \times p$, hodnoty na hlavní diagonále jsou rovny výběrovým rozptylům $S^2_{X_1}, \ldots, S^2_{X_p}$. Matice je vždy pozitivně semidefinitní.

Korelační matice

Definice 6 (Výběrová korelační matice)

Výběrová korelační matice je matice výběrových korel. koeficientů $r_{ij} = r_{X_i X_j}$:

$$\mathbf{R} = \left\{r_{ij}\right\}_{i,j=1}^{p}.$$

$$r_{X_i X_j} = \frac{S_{X_i X_j}}{\sqrt{S^2_{X_i}}\sqrt{S^2_{X_i}}}$$

$\mathbf{R}$ je symetrická čtvercová matice rozměru $p \times p$, na hlavní diagonále jsou jedničky.

Kovarianční a korelační matice v R

`cov (X)`, `cor (X)`,
kde `X` je matice, jejíž sloupce tvoří náhodné výběry z jednotlivých veličin.
Pro výpočty p-hodnot testů významnosti korelačních koeficientů je však vhodnější pracovat s funkcí `rcorr` z `library (Hmisc)`.

Příklad 1: průměry, kovarianční a korelační matice

```
apply (X, 2, mean)
cov (X)
cor (X)
```

```
       C           P          V
3.285714 104.285714  38.571429

       C         P       V
C  1.904    50.238  20.476
P 50.238  1561.904 607.142
V 20.476   607.142 247.619

      C     P     V
C 1.000 0.921 0.942
P 0.921 1.000 0.976
V 0.942 0.976 1.000
```

Grafická zobrazení

Scatter plot

Tzv. *scatter plot* je zobrazení dvourozměrných závislostí mezi všemi dvojicemi náhodných veličin. Kreslí se ve tvaru matice grafů, kdy na pozici (i, j) je vykreslena závislost X_i na X_j pomocí bodového grafu z hodnot odpovídajících náhodných výběrů.

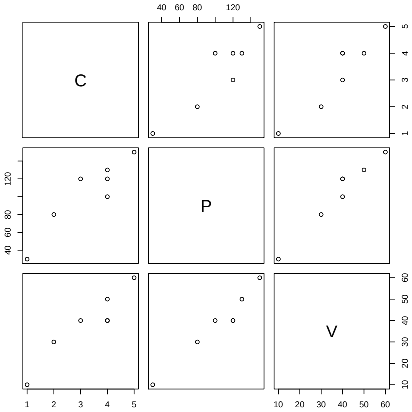
V *R*: `pairs (X)`

Korelogram

Vizuální reprezentaci korelační matice $\mathbf{R}$ dává tzv. *korelogram*. Kreslí se ve tvaru matice grafů, kdy na pozici (i, j) je graficky zobrazena hodnota $r_{X_i X_j}$ spolu s informací o významnosti daného korelačního koeficientu.

V *R*: pomocí funkce `corrplot` z `library (corrplot)`

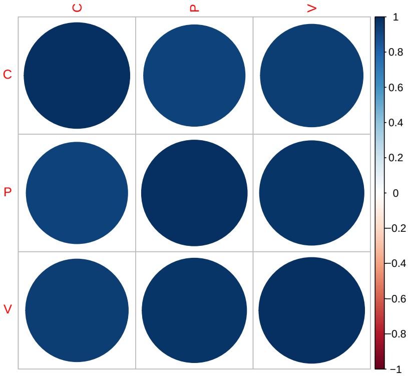
Příklad 1: scatter-plot



Příklad 1: korelační marice, p-hodnoty, korelogram

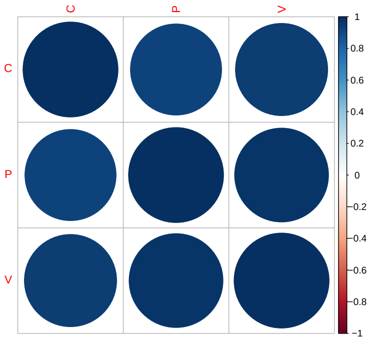
```
R <- rcorr (X)
R$r
R$P
corrplot (R$r, p.mat=R$P)
```

```
      C     P     V
C 1.000 0.921 0.942
P 0.921 1.000 0.976
V 0.942 0.976 1.000

      C     P     V
C    NA 0.003 0.001
P 0.003    NA 0.000
V 0.001 0.000    NA
```



Příklad 1

```
      C     P     V
C 1.000 0.921 0.942
P 0.921 1.000 0.976
V 0.942 0.976 1.000

      C     P     V
C    NA 0.003 0.001
P 0.003    NA 0.000
V 0.001 0.000    NA
```



Všechny tři dvojice korelačních koeficientů jsou velmi blízké hodnotě 1, a zároveň jsou všechny významné. To ukazuje na lineární závislost v souhlasném smyslu mezi všemi třemi dvojicemi náhodných veličin.

Např. čím větší počet členů v domácnosti (C), tím větší celkový čistý příjem (P) domácnosti? Opravdu toto naše data potvrzují?

Opakování: lineární regresní model

V lineárním regresním modelu jsme dosud pracovali jen s regresní funkcí $Y = f(x)$, která byla lineární vzhledem ke svým parametrům $\boldsymbol{\beta}$. V případě znalosti dvoudimenzionálního náhodného výběru rozsahu n,

$$\big((X_1, Y_1), \ldots, (X_n, Y_n)\big)',$$

lze lineární model s k parametry zapsat v maticovém tvaru

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta},$$

$$\text{kde} \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{k-1} \end{pmatrix}$$

a sloupce matice plánu $\mathbf{X}$ rozměru $(n \times k)$ jsou tvořeny odpovídajícími funkcemi hodnot $x_1, \ldots, x_n$.

Připomeňte si např. model s regresní funkcí tvaru polynomu $Y = \sum_{i=0}^{k-1} \beta_i x^i$, kde parametry $\beta_0, \ldots, \beta_{k-1}$ odpovídají koeficientům mocnin x^i.

Mnohonásobná lineární regrese

Obecně však lze pracovat i s regresním modelem, kde jednorozměrnou máhodnou veličinu Y modelujeme skupinou p náhodných veličin $X_1, \ldots, X_p$. Vycházíme ze znalosti $(p+1)$-dimenzionálního náhodného výběru rozsahu n,

$$\begin{pmatrix} (X_{1\,1}, & \cdots, & X_{1\,p}) \\ & \vdots & \\ (X_{n\,1}, & \cdots, & X_{n\,p}) \end{pmatrix}.$$

a model **mnohonásobné lineární regrese** zapíšeme v analogickém tvaru

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}.$$

První sloupec matice $\mathbf{X}$ je tvořen n jedničkami, další sloupce jsou postupně tvoří náhodné výběry rozsahu n jednotlivých náhodných veličin,

$$\mathbf{X} = \begin{pmatrix} 1 & X_{1\,1}, & \cdots, & X_{1\,p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n\,1}, & \cdots, & X_{n\,p} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Mnohonásobná lineární regrese

Pracujeme tedy s modelem s maticovým zápisem

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$$

s maticí plánu $\mathbf{X}$ rozměru $n \times (p+1)$ a s $(p+1)$ parametry, pro jehož řešení používáme klasický postup odhadu parametrů metodou nejmenších čtverců, tzn.

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}, \qquad \widehat{\mathbf{Y}} = \underbrace{\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'}_{\mathbf{H}}\mathbf{Y}.$$

Parametry $\beta_1, \ldots, \beta_p$ (ne β_0) jsou přitom lineární koeficienty jednotlivých náhodných veličin $X_1, \ldots, X_p$, pomocí nichž modelujeme výslednou veličinu Y, tj.

Geometrický význam

$$\widehat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_1 + \cdots + \widehat{\beta_p}X_p.$$

Grafem je **p-dimenzionální (nad)rovina v prostoru dimenze $(p+1)$**.

Příklad 1: mnohonásobná lineární regrese

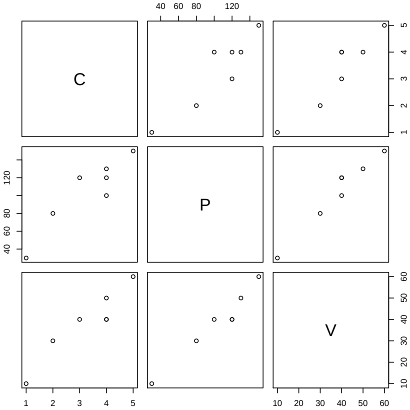
model mnohonásobné lineární regrese: $V = \beta_0 + \beta_1 P + \beta_2 C$

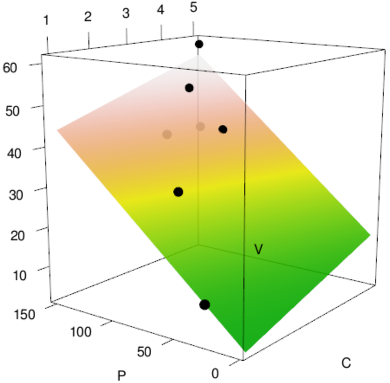
```
model <- lm (V ~ P + C, data = tabulka)
summary (model)
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.74142    4.08119  -0.427   0.6916
P            0.28320    0.09473   2.990   0.0404 *
C            3.28056    2.71254   1.209   0.2931

Residual standard error: 3.571 on 4 degrees of freedom
Multiple R-squared:  0.9657, Adjusted R-squared:  0.9485
F-statistic: 56.25 on 2 and 4 DF,  p-value: 0.001179
```

MNČ-odhady: $\beta_1 = 0{,}283{*} > 0, \quad \beta_2 = 3{,}281 > 0, \quad R^2 = 0{,}966, \quad F{*}$

Příklad 1: scatter-plot a regresní rovina





rovnice regresní roviny: $V = -1{,}741 + 0{,}283\,P + 3{,}281\,C$

Mnohonásobná korelace (multiple correlation)

Definice 7 (Koeficient mnohonásobné korelace)

Celkovou míru závislosti mezi náhodnou veličinou Y a náhodnými veličinami $\mathbf{X} = (X_1, \ldots, X_p)$ popisujeme **koeficientem mnohonásobné korelace**, což je korelační koeficient mezi Y a **nejlepší lineární aproximací** $\widehat{Y}$ pomocí veličin $\mathbf{X}$,

$$\rho_{Y \cdot \mathbf{X}} = \rho(Y, \widehat{Y}).$$

Jedná se tedy o největší ze všech absolutních hodnot korelačních koeficientů mezi Y a libovolnou lineární kombinací Y^* veličin $\mathbf{X}$,

$$\rho_{Y \cdot \mathbf{X}} = \max_{\text{lin. } Y^*} |\rho(Y, Y^*)|.$$

Výběrový koeficient mnohonásobné korelace
mezi náhodnou veličinou Y a náhodnými veličinami $\mathbf{X} = (X_1, \ldots, X_p)$ je

$$r_{Y \cdot \mathbf{X}} = r(Y, \widehat{Y}),$$

kde $\widehat{Y}$ je odhad v LRM $\quad Y = \beta_0 + \mathbf{X}\boldsymbol{\beta}_X$.

```
Y ~ X1 + ... + Xp
```

Parciální korelace (partial correlation)

Definice 8 (Koeficient parciální korelace)

Míru *ryzí* závislosti mezi náhodnými veličinami Y a Z při eliminaci vlivu náhodných veličin $\mathbf{X} = (X_1, \ldots, X_p)$ popisujeme **koeficientem parciální korelace**, což je korelační koeficient mezi rezidui $Y - \widehat{Y}$ a $Z - \widehat{Z}$ při nejlepších lineárních aproximacích $\widehat{Y}$ a $\widehat{Z}$ pomocí veličin $\mathbf{X}$,

$$\rho_{YZ \cdot \mathbf{X}} = \rho(Y - \widehat{Y}, Z - \widehat{Z}).$$

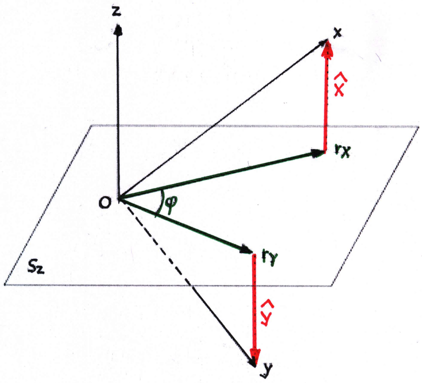
Výběrový koeficient parciální korelace mezi náhodnými veličinami Y a Z při eliminaci vlivu náhodných veličin $\mathbf{X} = (X_1, \ldots, X_p)$ je

$$r_{YZ \cdot \mathbf{X}} = r(\underbrace{Y - \widehat{Y}}_{\mathbf{r}_Y}, \underbrace{Z - \widehat{Z}}_{\mathbf{r}_Z}),$$

kde $\widehat{Y}$ a $\widehat{Z}$ jsou odhady v LRM $\quad Y = \beta_0 + \mathbf{X}\boldsymbol{\beta}$ a $Z = \alpha_0 + \mathbf{X}\boldsymbol{\alpha}$.

`Y ~ X1 + ... +Xp`, `Z ~ X1 + ... +Xp`

Geometrický význam parciální korelace $r_{XY\cdot Z}$



Předloha: wikipedia.org

- S_z = (nad)rovina kolmá na z
- $\widehat{x}$ = nejlepší lineární odhad veličiny x pomocí z v modelu

$$X = (\mathbf{1}, Z)\boldsymbol{\beta}$$

- $r_x = x - \widehat{x}$ = rezidua veličiny x, r_x = kolmý průmět x do S_z
- $\widehat{y}$ = nejlepší lineární odhad veličiny y pomocí z v modelu

$$Y = (\mathbf{1}, Z)\boldsymbol{\beta}$$

- $r_y = y - \widehat{y}$ = rezidua veličiny y, r_y = kolmý průmět y do S_z
- parciální korelační koeficient

$$r_{XY\cdot Z} = \cos\varphi = \cos|\sphericalangle\, r_X, r_Y|$$

Semiparciální korelace (semipartial/part correlation)

Definice 9 (Koeficient semiparciální korelace)

Míru *ryzí* závislosti mezi náhodnými veličinami Y a Z při eliminaci vlivu náhodných veličin $\mathbf{X} = (X_1, \ldots, X_p)$ na veličinu Z popisujeme **koeficientem semiparciální korelace**, což je korelační koeficient mezi Y a reziduem $Z - \widehat{Z}$ při nejlepší lineární aproximaci $\widehat{Z}$ pomocí veličin $\mathbf{X}$,

$$\rho_{Y\,(Z\cdot\mathbf{X})} = \rho(Y, Z - \widehat{Z}).$$

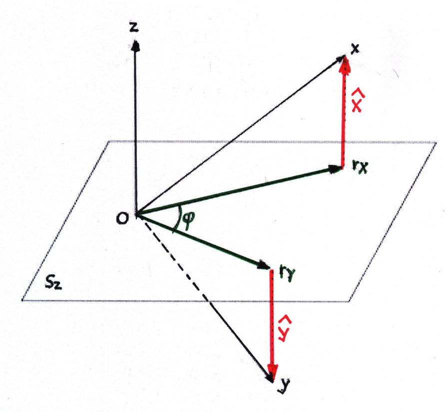
Výběrový koeficient parciální korelace mezi náhodnými veličinami Y a Z při eliminaci vlivu náhodných veličin $\mathbf{X} = (X_1, \ldots, X_p)$ na veličinu Z je

$$r_{Y\,(Z\cdot\mathbf{X})} = r(Y, \underbrace{Z - \widehat{Z}}_{\mathbf{r}_Z}),$$

kde $\widehat{Z}$ je odhad v LRM $\quad Z = \alpha_0 + \mathbf{X}\boldsymbol{\alpha}$.

`Z ~ X1 + ... + Xp`

Geometrický význam semiparciální korelace



- semiparciální korelační koeficient

$$r_{X\,(Y\cdot Z)} = \cos|\sphericalangle\, x, r_Y|$$

- semiparciální korelační koeficient

$$r_{Y\,(X\cdot Z)} = \cos|\sphericalangle\, y, r_x|$$

Předloha: wikipedia.org

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)`

Výpočty pomocí korelační matice $\mathbf{R}$

V softwarech se obvykle výběrové korelační koeficienty a p-hodnoty testů významnosti počítají pomocí maticových operací s korelační maticí $\mathbf{R}$.

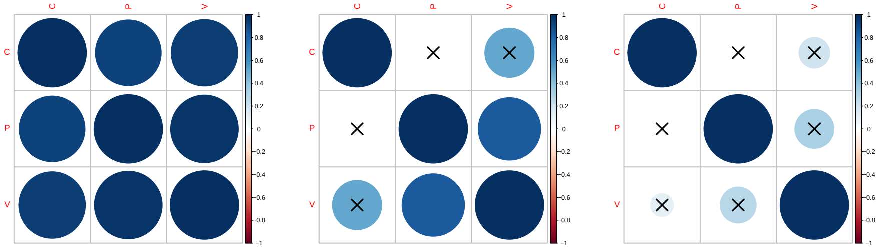
Např. pro skupinu 3 náhodných veličin X_1, X_2, X_3 s výběrovou korelační maticí $\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}$ platí:

$$r_{3\cdot 12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2\, r_{31}\, r_{32}^2\, r_{12}}{1 - r_{12}^2}},$$

$$r_{12\cdot 3} = r_{21\cdot 3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}},$$

$$r_{1\,(2\cdot 3)} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{23}^2}}, \qquad r_{2\,(1\cdot 3)} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}}$$

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

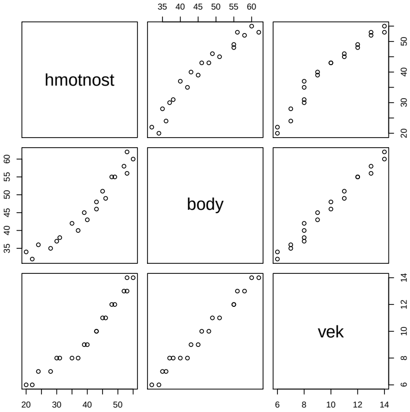
model mnohonásobné lineární regrese: $Body = \beta_0 + \beta_1 Hmotnost + \beta_2 Vek$

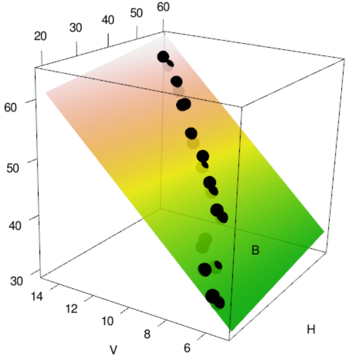
```
model <- lm (body ~ hmotnost + vek, data = tabulka)
summary (model)
```

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

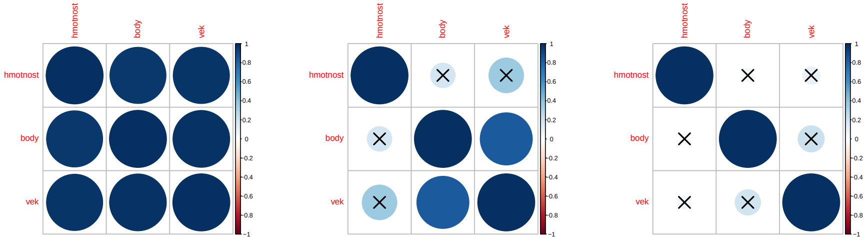
Příklad 2: scatter-plot a regresní rovina





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

Příklad 2: korelogramy



	hmotnost	body	vek
hmotnost	1	0.97	0.97
body	0.97	1	0.99
vek	0.97	0.99	1

	hmotnost	body	vek
hmotnost	1	0.19	0.37
body	0.19	1	0.83
vek	0.37	0.83	1

	hmotnost	body	vek
hmotnost	1	0.04	0.09
body	0.03	1	0.21
vek	0.05	0.2	1