# Cvičení MV011 Statistika I 11. Lineární regresní model

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

jaro 2019



Pro řešení lineárního regresního modelu  $Y_i = m(x_i) + \varepsilon_i$ , i = 1, ..., n, v R slouží příkaz lm (linear model):

```
model <-lm (formule, data = DatovaTabulka), příp.
model <-lm (formule, data = DatovaTabulka, weights = VektorVah).
```

Pro tzv. *formuli* se používá speciální syntaxe, kde Y je název sloupce závisle proměnné, x je název sloupce nezávisle proměnné:

m(x)	formule
$\beta_0 + \beta_1 x$	Y ~ x nebo Y ~ 1 + x, člen $eta_0$ je totiž vkládán implicitně
$\beta_1 x$	Y ~ 0 + x, odstranění členu $eta_0$ nutno zapsat explicitně
$\beta_0 + \beta_1 x + \beta_2 x^2$	$Y \sim x + I(x^2)$
$\beta_2 x^2$	$Y \sim 0 + I(x^2)$
$\beta_1 x $	Y~0+I(abs(x))
$\beta_0 + \beta_1 e^x$	Y ~ I(exp(x))
$\beta_0 + \beta_1 \ln x$	Y~I(log(x))
$\beta_0 + \beta_1 \sqrt{x}$	Y ~ I(sqrt(x))

prehled <-summary (model),
příp. prehled <-summary (model, correlation=TRUE) pro výběrovou korelační matici parametrů.</pre>

Detailní výsledky a další číselné charakteristiky získáme příkazem

MV011 Statistika I. 2019 Cvičení 11: Lineární regresní model 2 / 9

$\widehat{oldsymbol{eta}}$	MNČ-odhady parametrů	<pre>model\$coefficients coef(model)</pre>
$\left(\widehat{\beta}_{j}, SD(\widehat{\beta}_{j}), T_{j}, p_{j}\right)$	odhady, směrodatné odchylky, testy významnosti, p-hodnoty	<pre>prehled\$coefficients coef(prehled)</pre>
Ŷ	aproximované hodnoty	<pre>model\$fitted.values fitted.values(model)</pre>
r	rezidua	<pre>model\$residuals residuals(model)</pre>
n-k	stupně volnosti modelu	model\$df.residual
X	matice plánu	<pre>model.matrix(model)</pre>
w	váhy	model\$weights
S	odhad sm. odchylky chyb $arepsilon_i$	prehled\$sigma
$R^2$	index determinace	prehled\$r.squared
$\overline{R}^2$	korigovaný index determinace	prehled\$adj.r.squared
(F,k-1,n-k)	celkový F-test	prehled\$fstatistic
(k,n-k,k)	stupně volnosti	prehled\$df
$R(\widehat{oldsymbol{eta}})$	korelační matice odhadů $\widehat{oldsymbol{eta}}$	prehled\$correlation

# Úkoly v příkladech:

- lacksquare MNČ-odhady parametrů  $\widehat{oldsymbol{eta}}$  regresní funkce m(x), sledujte i jejich významnost,
- lacksquare zapište matematický tvar regresní funkce m(x),
- $\blacksquare$  reziduální součet čtverců  $S_{\ell}$  a odhad směrodatné odchylky s náhodných chyb,
- index determinace R<sup>2</sup>, proveď te celkový F-test,
- vykreslete data a grafy regresních funkcí (predict), příp. s pásy spolehlivosti,
- vykreslete boxploty reziduí,
- modely porovnejte (mj. anova), zvolte z nich nejvhodnější.

#### Příklad 1

Datový soubor KysMlecna.csv: zkoumejte závislost množství kyseliny mléčné u novorozence na množství stejné látky u matky-prvorodičky (v mg ve 100 ml krve) pomocí regresní přímky a paraboly.

# Příklad 2

Datový soubor prodlouzeni .csv: zkoumejte závislost prodloužení měděné trubky v závislosti teplotním rozdílu  $\Delta t$  od referenční hodnoty  $t_0=20\,^\circ$ C pomocí vhodné regresní přímky a paraboly. Dle fyzikálních zákonů by při  $\Delta t=0$  prodloužení mělo být nulové.

#### Příklad 3

Datový soubor spotreba2.csv: zkoumejte závislost spotřeby paliva motorového vozidla (v l/100 km) na rychlosti (v km/h) pomocí regresní přímky a paraboly.

#### Příklad 4

Datový soubor  ${\tt C02.csv}$ : zkoumejte závislost koncentrace  ${\tt CO_2}$  (v ppm) v atmosféře v letech 1764–1995 pomocí několika polynomických regresních funkcí.

#### Příklad 5

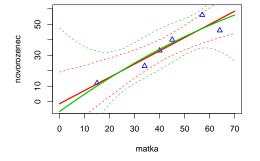
Datový soubor <code>EmiseUhliku.csv</code>: zkoumejte závislost uhlíkových emisí (v milionech tun) v letech 1950–1995 pomocí několika polynomických regresních funkcí.

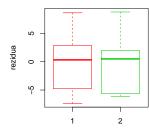
# Příklad 6

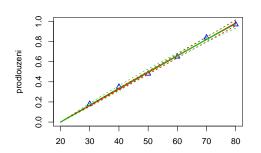
Datový soubor teplota.csv: zkoumejte závislost průměrné teploty (ve  $^{\circ}$ C) v letech 1866–1996 pomocí několika polynomických regresních funkcí.

# Příklad 7

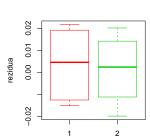
Datový soubor ropa.csv: zkoumejte závislost logaritmu objemu vytěžené ropy (v tisících barelů) v letech 1880–1988 pomocí několika polynomických regresních funkcí, grafy vykreslete i pro nelogaritmované hodnoty.







tenlota



MV011 Statistika I, 2019

