

2. cvičení

$\def\mc{#1}{\mathcal{#1}} \def\sc{#1#2}{\langle #1, #2 \rangle} \def\N{\mathbb{N}}$
 $\def\R{\mathbb{R}} \def\Q{\mathbb{Q}} \def\Z{\mathbb{Z}} \def\D{\mathbb{D}}$
 $\def\bm{#1}{\boldsymbol{#1}} \def\vv{#1}{\mathbf{#1}} \def\vp{#1}{\mathbf{#1}}$
 $\def\fl{#1}{\lfloor #1 \rfloor} \def\ce{#1}{\lceil #1 \rceil} \def\grad{#1}{\mathrm{grad} \, #1}$
 $\def\ve{\varepsilon}$

Lineární model

Obecně má tvar $Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k} + \varepsilon_i$, kde $i = 1, \dots, n$. Y_i je naše "cílová proměnná" (regresand). Proměnné $x_{i,1}, \dots, x_{i,k}$ jsou kovariáty (regresor, prediktor) a jsou pevně dané. Dále máme regresní koeficienty β_0, \dots, β_k a ε_i je náhodná proměnná chyby.

Také platí $\varepsilon_i \sim^{iid} (0, \sigma^2)$ $E(\varepsilon_i) = 0$
 $\text{var}(\varepsilon_i) = \sigma^2$ $\text{cov}(\varepsilon_i) = 0$

Celkem máme
$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,k} \\ 1 & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
$$\tag{LSM}$$
$$A \text{ vektorově } \mathbf{Y} = \underbrace{\mathbf{X}}_{\text{matice plánu}} \cdot \mathbf{v}(\beta) + \mathbf{v}(\varepsilon)$$

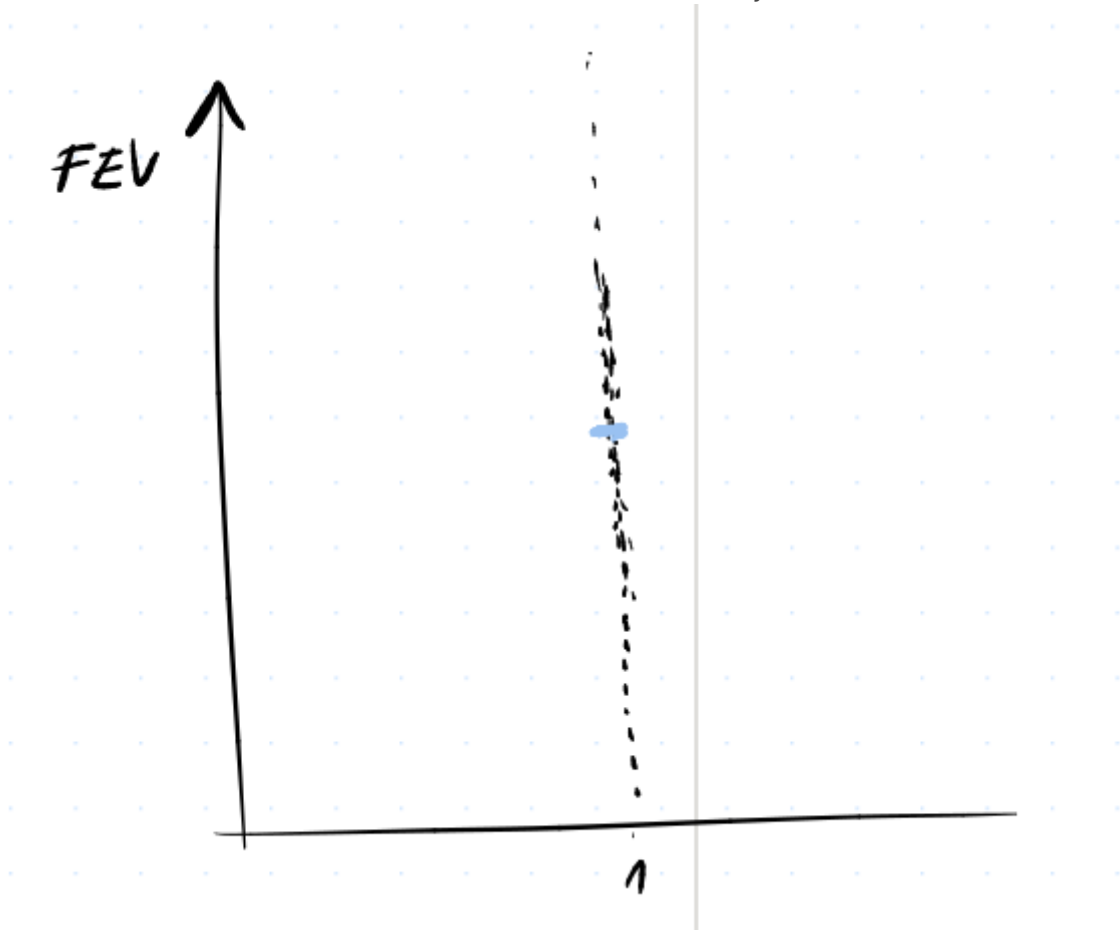
Tedy pro 2. cvičení

02 / a)

$$FEV_i = \beta_0 + \varepsilon_i$$

Můžeme si představit jako funkci $y = \beta_0$

A matice plánu bude $\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}$ Tedy kapacita plic je podle tohoto modelu konstantní a vizuálně znázorněné jako

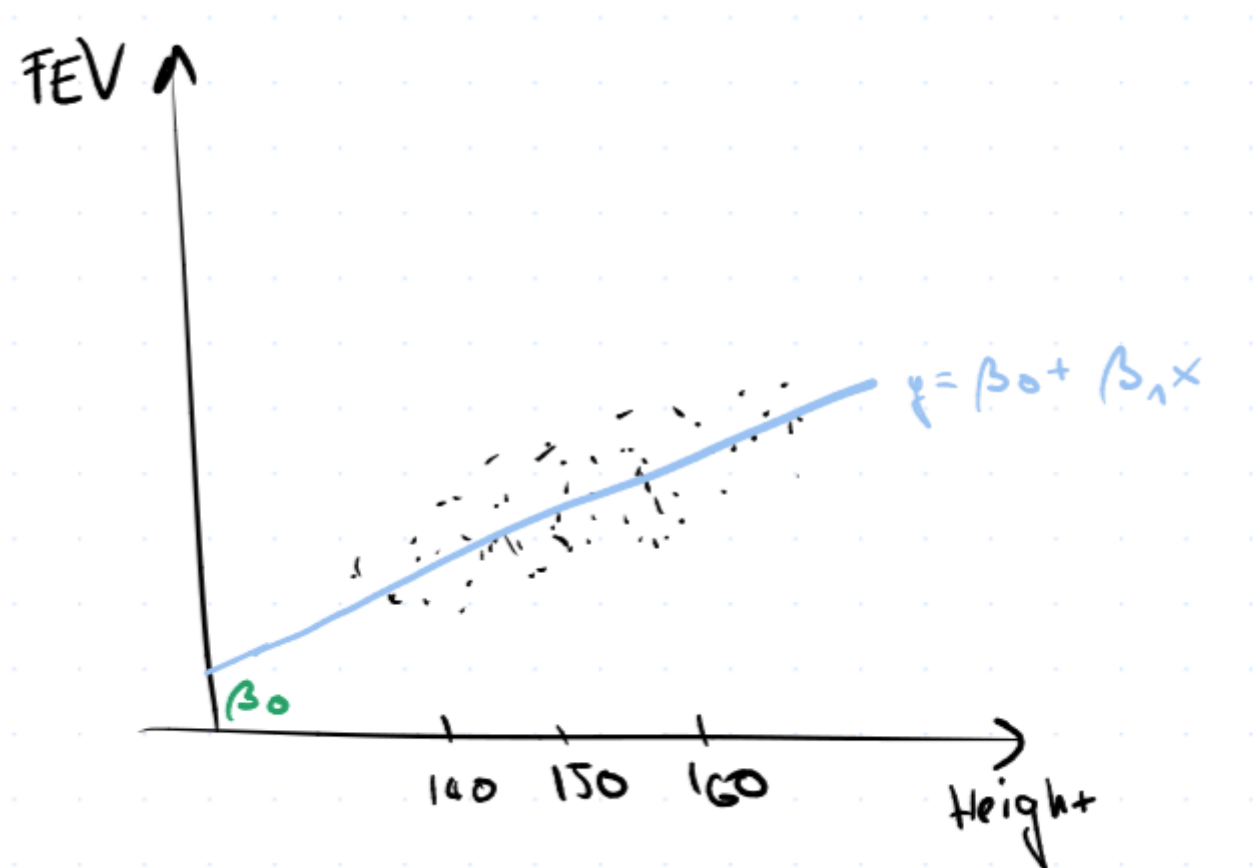


02 / c)

$$FEV_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \epsilon_i$$

Kapacitu plic modelujeme pomocí výšky. Tedy v tomto případě chceme funkci $y = \beta_0 + \beta_1 x$.

A graficky



A matice plánu tentokrát bude $\begin{pmatrix} 1 & \text{Height}_1 & 1 & \text{Height}_2 & \vdots & 1 & \text{Height}_n \end{pmatrix}$, což dosazujeme do (LSM).

Tedy máme model hledáme $y = \beta_0 + \beta_1 x$ a zde

- β_0 ... střední hodnota predikce při nulových hodnotách ostatních prediktorů
- β_1 ... nárůst střední hodnoty predikce při nárůstu výšky (prediktor Height) o 1 cm

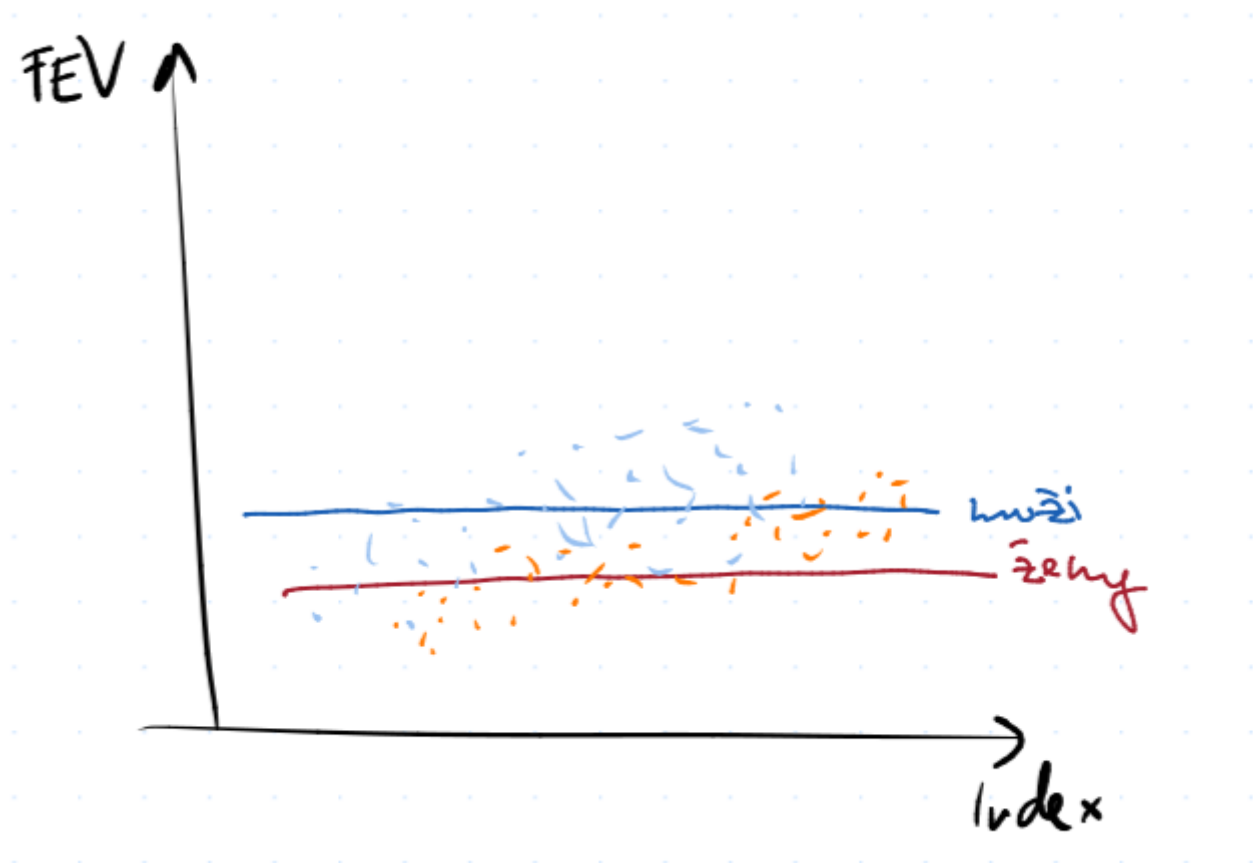
02 / b)

$$FEV_i = \beta_0 + \beta_1 \cdot \text{Sex}_i + \epsilon_i$$

a β_1 zde reprezentuje rozdíl predikce mezi muži a ženami s maticí plánu $\begin{pmatrix} 1 & 1 & 1 & 0 & \vdots & 1 & 1 \end{pmatrix}$,

kde 1 reprezentuje muže.

A graficky



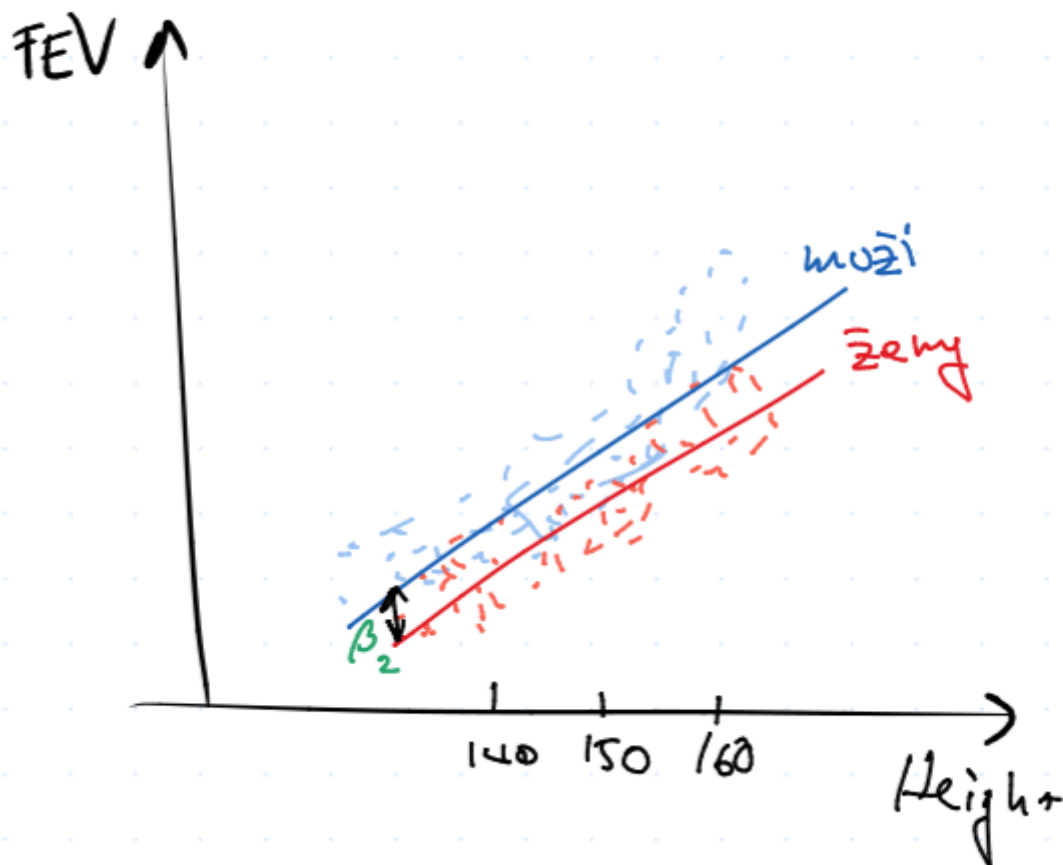
02/ d)

$FEV_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \beta_2 \cdot \text{Height}_i^2 + \epsilon_i$ V tomto případě je β_1, β_2 jsou složité na interpretaci

A matice plánu by v tomto případě byla
$$\begin{pmatrix} 1 & \text{Height}_1 & \text{Height}_1^2 \\ 1 & \text{Height}_2 & \text{Height}_2^2 \\ \vdots & \vdots & \vdots \\ 1 & \text{Height}_n & \text{Height}_n^2 \end{pmatrix},$$

02 / e)

$FEV_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \beta_2 \cdot \text{Sex}_i + \epsilon_i$ s maticí plánu
$$\begin{pmatrix} 1 & \text{Height}_1 & 0 \\ 1 & \text{Height}_2 & 1 \\ \vdots & \vdots & \vdots \\ 1 & \text{Height}_n & 1 \end{pmatrix},$$
 přičemž ve 3. sloupci jsou 1\$ značí muže.



A hledáme přímku $y = \beta_0 + \beta_1 x + \beta_2 \text{ I } \set{\text{Sex} = \text{"male"}}$

s významem koeficientů

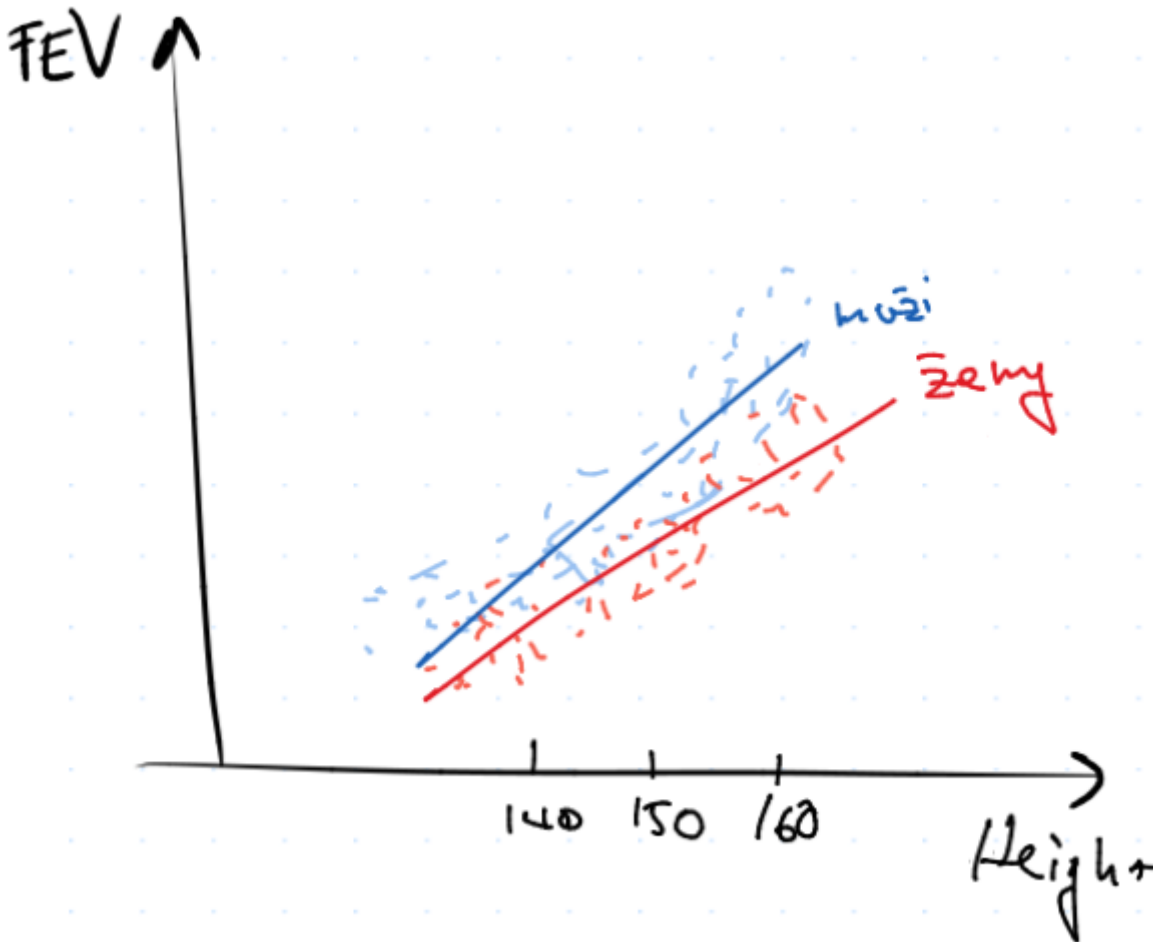
- β_0 ... střední hodnota predikce při nulových hodnotách ostatních prediktorů (nulová výška a ženské pohlaví)
- β_1 ... změna střední hodnoty predikce při nárůstu výšky (prediktor Height) o 1 cm pro ženy
- β_2 ... rozdíl střední hodnoty predikce mezi muži a ženami

02 / g)

$$\text{FEV}_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \beta_2 \text{Sex}_i + \beta_3 (\text{Sex}_i \times \text{Height}_i) + \epsilon_i$$
kde členu $\text{Sex}_i \times \text{Height}_i$ **interakce** a matice plánu bude
$$\begin{pmatrix} 1 & \text{Height}_1 & 0 & 0 \\ 1 & \text{Height}_2 & 1 & \text{Height}_2 \\ 1 & \text{Height}_3 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{Height}_n & 1 & \text{Height}_n \end{pmatrix}$$
, přičemž ve 3. sloupci jsou 1 značí muže.

A hledáme přímku $y = \beta_0 + \beta_1 x + \beta_2 \text{ I } \set{\text{Sex} = \text{"male"}} + \beta_3 x \text{ I } \set{\text{Sex} = \text{"male"}}$, tj.

- žena ... $y = \beta_0 + \beta_1 x$
- muž ... $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x$



Zde β_3 značí rozdíl střední hodnoty predikce mezi muži a ženami při nárůstu výšky o 1 cm

“rozdíl rychlosti růstu FEV mezi muži a ženami

Interpretation

- “keeping the values of all the other covariates fixed, a unit increase in x_i is associated with a $\hat{\beta}_i$ increase in $E Y$ ”
 - ▶ suitably adapted for categorical predictors and potentially interactions, and depends on the choice of the identifiability conditions
 - ▶ polynomials need a more complex interpretation
- is it meaningful to imagine that a covariate changes while all the other remain fixed?

Revision #4

Created 12 January 2023 11:05:32 by Sceptri

Updated 14 January 2023 11:30:00 by Sceptri