Enostavna linearna regresija, povzetek 1

Odvisnost proučevane spremenljivke Y od pojasnjevalne spremenljivke X iščemo v obliki zveze (modela):

$$Y = \alpha + \beta \cdot X + \varepsilon.$$

Tu je ε slučajna spremenljivka, katere vloga je lahko:

- meritvena (zaokrožitvena) napaka,
- drugo slučajno odstopanje od linearne zveze $Y = \alpha + \beta X$.

Predpostavljamo $E(\varepsilon) = 0$. V povprečju velja torej linearna zveza. Včasih predpostavljamo, da $\varepsilon \sim N(0, \sigma)$ za neki neznani odklon σ.

Vzorec velikosti n je n parov meritev (x_i, y_i). Dobimo jih z n replikacijami slučajnega eksperimenta.

Parametra α in β ocenimo tako, da se premica $\widehat{y} = \alpha + \beta x$ čim bolj prilega podatkom. Z drugimi besedami, minimiziramo vsoto kvadratov residualov $\mathrm{VKR}(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$.

Enostavna linearna regresija, povzetek 2

Oceni za α in β

- $\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})/(n-1)}{\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})/(n-1)} = \frac{\text{vzorčna kovarianca}}{S_X^2}$,
- $\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$.

Regresijska premica ...

- ... je premica z enačbo $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$.
- V obliki napovedi vrednosti spremenljivke Y pišemo $\widehat{Y} = \hat{\alpha} + \hat{\beta} \cdot X$.

Tesnost prileganja

Tesnost prileganja vzorca $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ regresijski premici $\widehat{y} = \hat{\alpha} + \hat{\beta} \cdot x$ merimo s koeficientom determinacije $R^2 = \frac{(\text{vzorčna kovarianca})^2}{S_X^2 \cdot S_Y^2}$.