Multiple regression

E. Pastucha

November 2024

Multiple regression

- Linear
- Multiple explanatory variables (Predictors)
- One Response variable

Multiple regression

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Multiple regression

$$\hat{y} = 3 + 0.5x + 2.73z$$

This would be a plane in 3D

Multiple regression

Conditions:

- linearity - each variable is linearly related to the outcome,
- nearly normal residuals,
- constant variability,
- independent residuals - residuals in order of their data collection/residuals against each predictor variable.

Multiple regression

Adjusted R^2

$$R^2 = 1 - \frac{\text{variability in residuals}}{\text{variability in the outcome}} = 1 - \frac{s^2_{residuals}}{s^2_{outcome}}$$

$$R^2_{adj} = 1 - \frac{s^2_{\text{residuals}} \,/(n-k-1)}{s^2_{\text{outcome}} \,/(n-1)} = 1 - \frac{s^2_{\text{residuals}}}{s^2_{\text{outcome}}} \times \frac{n-1}{n-k-1}$$

n - number of cases

k - number of predictors

Multiple regression

Model selection - model cleaning.

- Forward selection - R^2_{adj} or $p - value$ approach
- Backwards elimination - R^2_{adj} or $p - value$ approach