Quantitative Methoden 1

Tutorium 16/04/2021

- Regression with multiple predictors
- Model selection
- Chapter exercise

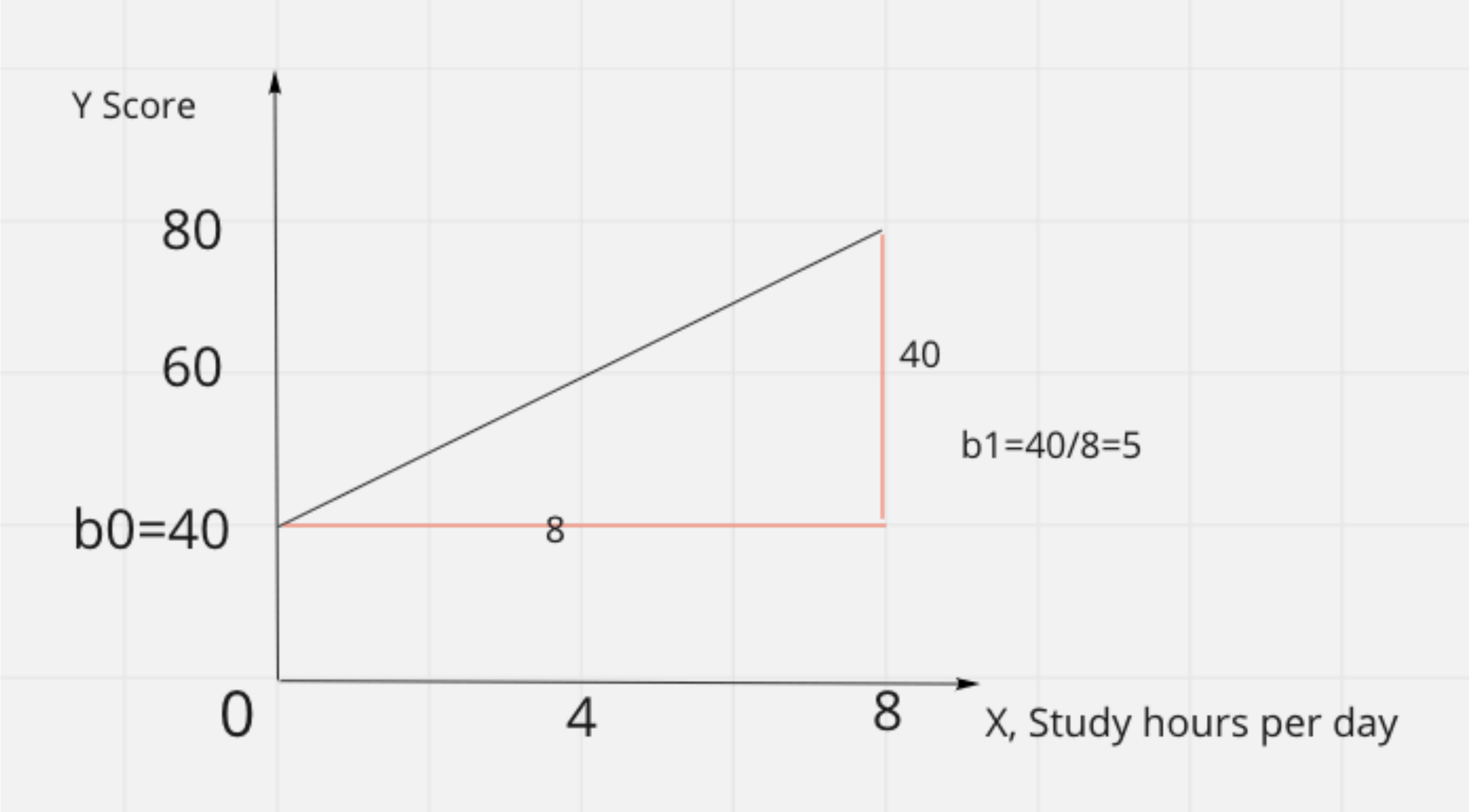
Review Linear model one predictor

- Linear regression is the statistical method for fitting a line to data where the relationship between two variables x and y, can be modeled by a straight line with some error,
- When we use x to predict y, we usually call x the **predictor** variable and we call y the **outcome**

$Y = b0 + b1\ x$
b0: intercept, b1:slope

$Y = 40 + 5\ x$

The steepness of a hill is called a **slope**



https://openintro-ims.netlify.app/summarizing-visualizing-data.html#chp2-review-exercises

Model with multiple predictors

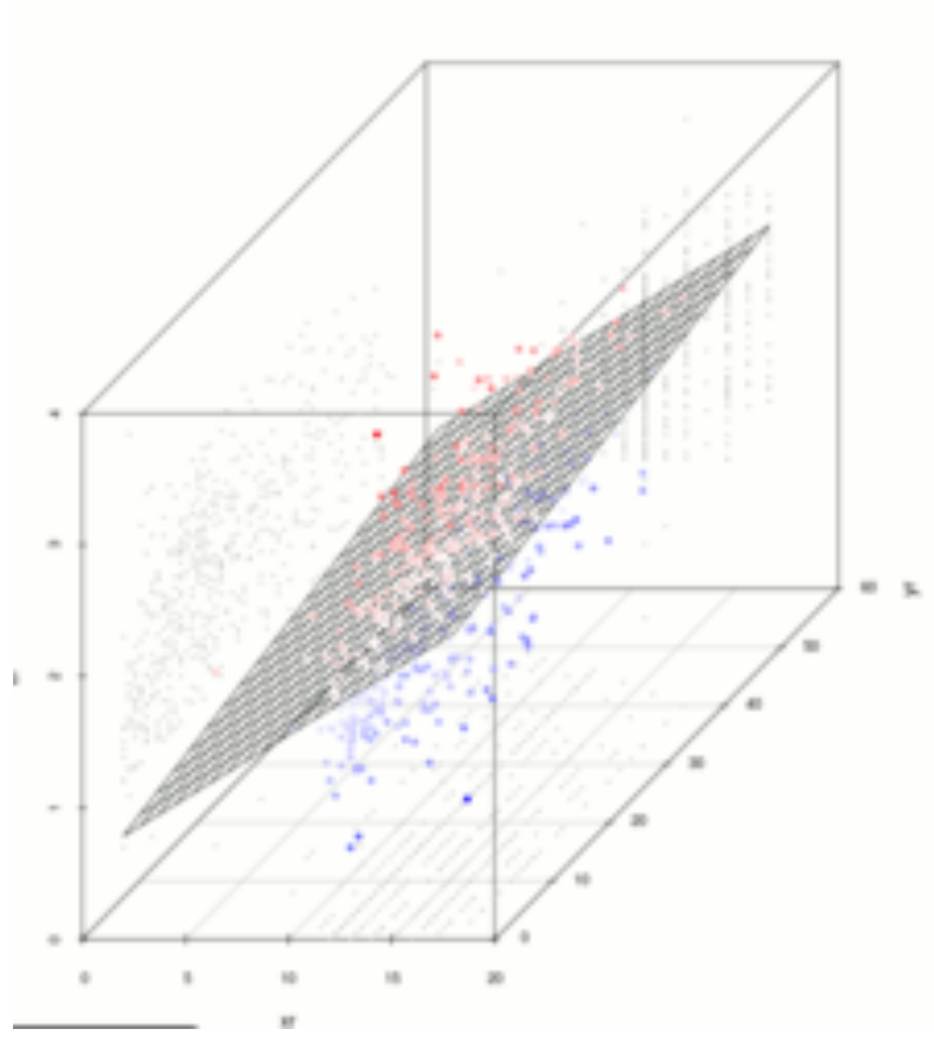
Multiple dimensional model

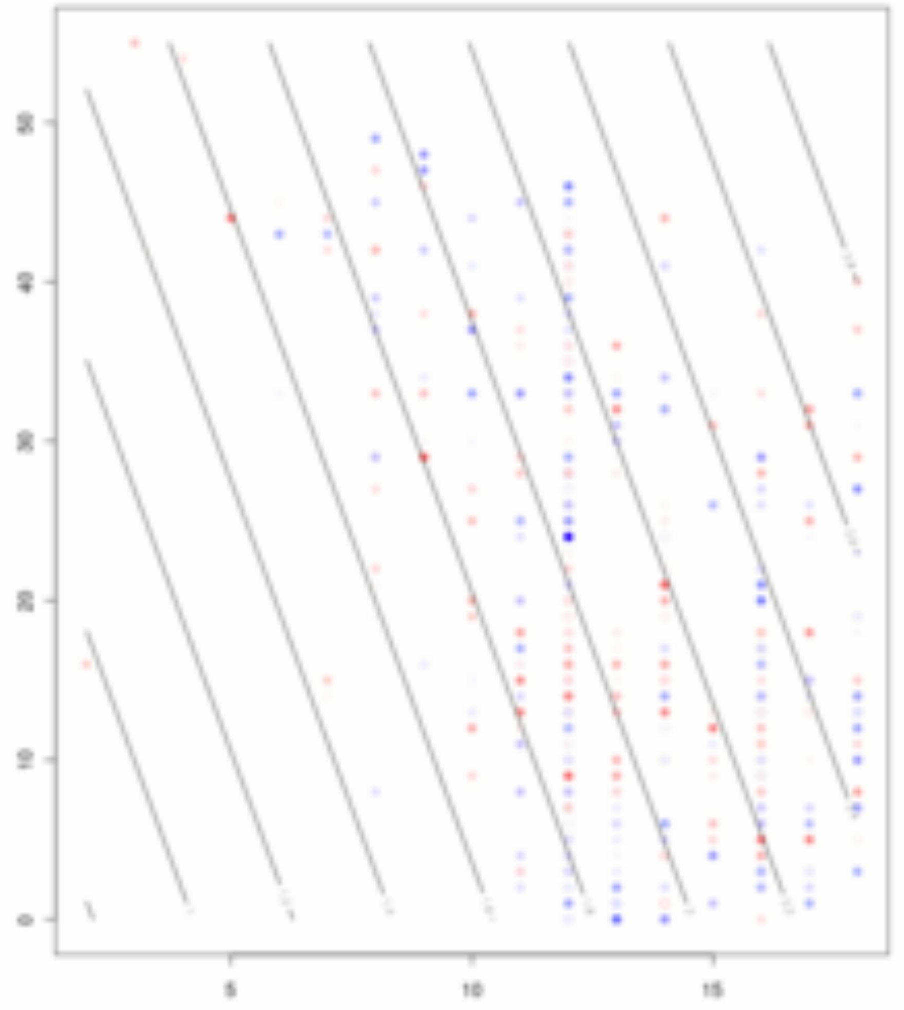
Multiple regression model.

A multiple regression model is a linear model with many predictors. In general, we write the model as

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

when there are k predictors. We always calculate b_i using statistical software.





https://openintro-ims.netlify.app/multi-logistic-models.html#ind-and-cat-predictors
https://de.wikipedia.org/wiki/Regressionsanalyse

Adjusted R-squared, R-squared ?

- **When single predictor: R-squared is ok!**
- **When multiple predictors: Adjusted R-squared is better!**

Adjusted R-squared as a tool for model assessment

The **adjusted R-squared** is computed as

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

where n is the number of cases used to fit the model and k is the number of predictor variables in the model. Remember that a categorical predictor with p levels will contribute $p-1$ to the number of variables in the model.

https://openintro-ims.netlify.app/multi-logistic-models.html#many-predictors-in-a-model

Exercise time

IMS book 4.1.4.2, 4.1.4.3, 4.1.4.4

Model selection

- Backward elimination: from full to step by step elimination
- Forward selection: reverse

Common choose standard:
Larger adjusted R squared, smaller P value.

https://openintro-ims.netlify.app/multi-logistic-models.html#ind-and-cat-predictors

Exercise time

IMS book 4.2.3.1, 4.2.3.2, 4.2.3.3, 4.2.3.4