# 8. Regression

*Albulene Grajcevci*

edutech

SAARLAND UNIVERSITY

# Agenda

- An introduction to regression
- Bias in regression models
  - Regression with one predictor
  - Multiple regression
  - Regression with several predictors
- Interpreting multiple regression

# An introduction to regression

- Regression predicts the `future`
- The main idea: A specific outcome can be predicted by a model

**Outcome variables**

**Predictor variables**

# Bias in Regression models

- Regression predicts the `future`
- The main idea: A specific outcome can be predicted by a model

**Outcome variables**

**Predictor variables**

- 1 outcome variable  is generally referred to as simple regression
- More than one variable is generally referred to as multiple regression

# Bias in Regression models

Violation of the Generalization assumption

1. Unusual cases that impact the model
    1. The formula for calculating a z-score is $z = (x-\mu)/\sigma$, where x is the raw score, $\mu$ is the population mean, and $\sigma$ is the population standard deviation. As the formula shows, the z-score is simply the raw score minus the population mean, divided by the population standard deviation
2. Influential cases
    1. Adjusted predicted value

# Assumptions of the linear model

Assumptions of the model:

1. Additivity and linearity: the outcome variable is linearly related to any of the predictors
2. Independent errors: independence assumptions (Durbin-Watson tests)
3. Homoscedasiticity: violation leads to significant tests- variance of the residual does not change along with the value of the predictor variable
4. Normally distributed errors: difference between the model and observed data is zero or close to zero

# Assumptions of the linear model

Assumptions of the model:

4. Variable types: Predictors must be quantitative or categorical (two categories)

5. No perfect multicollinearity: bad because it limits the size of R

**can conduct a variance inflation factor VIF: Bad news if it is greater than 10 but a value of 1 already points to bias in regression.**

6. Non-zero variance: predictors should have variation

# Multiple regression

Methods of regression

1. Hierarchical (block wise entry)
    1. Based on past research
2. Forced entry
3. Stepwise  methods: non very well liked by researchers
    1. Forward method:
    2. Stepwise method: each time a new predictor is added one is omitted that is considered unusual
    3. Backward method: add all predictors then calculating the contribution of each of them no removal
    4. Removal

# Reporting multiple regression

Multiple regression analysis was used to test if the personality traits significantly predicted participants' ratings of aggression. The results of the regression indicated the two predictors explained 35.8% of the variance ($R = .38$, $F(2,55) = 5.56$, $p < .01$).