TU DORTMUND
Fakultät Statistik
Prof. Dr. Jörg Rahnenführer
Dr. Franziska Kappenberg
M.Sc. Marieke Stolte

Winter semester 2022/23

Project III
Regression Analysis

The dataset `bodymeasurements.csv` contains body measurements of 424 persons aged between 18 to 40 years old, which were collected in a study. The aim of the study was to determine a model that can be used to explain the body height. The following variables have been collected:

Height Body height (in cm)

Chest Chest circumference (in cm)

Belly Belly circumference (in cm)

Biceps Biceps circumference (in cm)

Knee Knee circumference (in cm)

Ankle Ankle circumference (in cm)

Wrist Wrist circumference (in cm)

Thigh Thigh circumference (in cm)

Calf Calf circumference (in cm)

Age Age at the time of the survey (in years)

Weight Body weight (in kg)

Sex Sex ('m' for males, 'f' for females)

Tasks:

1. Determine a linear regression model that explains the body height based on all other given variables.
2. Find the 'best' set of explanatory variables for the body height using Best Subset Selection. Use the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) as the selection criteria. Compare the included variables of the two resulting 'best' models.
3. Estimate the 'best' linear model for the dependent variable w.r.t. the BIC from Task 2. Interpret the coefficients of the model and their statistical significance, provide confidence intervals for the regression parameters and evaluate the goodness of fit.

Submission

Submission of the report and the corresponding (executable and commented) program code until *Friday, January 27, 2023, 08:30 am*, in Moodle.