

Exploring Statistics with R

Matthias Bannert

8 Nov 2015

Task A: Load the Data

Load EITHER the mtcars dataset (easy), the mlb dataset (medium) or the KOF Business situation (hard) into R.

1. mtcars: `data(mtcars)`
2. mlb: `http://whatsgood.io/data/mlb.RData`
3. KOF: `https://datenservice.kof.ethz.ch/api/sets/ds_bs_monthly?mime=csv`

Task A C'd: Basic Report

- ▶ Which variables are in the dataset?
- ▶ What's the size (rows,cols) of your dataset?
- ▶ Which is the most/least economic car (1), who is the tallest/shortest player (2), which is the best/worth month in total (3)?
- ▶ Provide some summary statistics such as means quantiles etc?
- ▶ What about the correlation structure of your data set? Inspect visually. What can you see? (hint: see `pairs()`, `cor()` and `image()`)
- ▶ Form an interesting sub groups and compare means in these subgroups. (hint: factor variables are good starting point)

Math Refresher: OLS coefficients

$$e = Y - X\beta$$