# LINGI2262 : Assignment 1 - Introduction to R

Group 8: Maxime Dimidschstein, Doriane Olewicki

February 12, 2018

## 1 Linear Model

We established two different models: one using 2 covariates (Height and Weight), and one with 3 covariates (Weight, Abdomen, Biceps).

Using the `lm` function, we obtained as coefficients for the first model $-0.7025959$ and $0.2013835$ respectively for the Height and the Weight, with an intercept of $32.4046404$. This shows that one's BFI increases as the Weight does too, while it decreases as the Height grows.

For the second model, we got $-0.1741240$, $0.9990568$ and $0.2796236$ respectively for the Weight, the Abdomen and the Biceps. Here we observe that an increase in Weight induces a decrease in BFI. In fact, if the Abdomen and Biceps size keep the same value while one's Weight increases, it means the bodyfat has been replaced by muscle, which has a bigger density.
The results obtained by this model can be seen on figure 1.

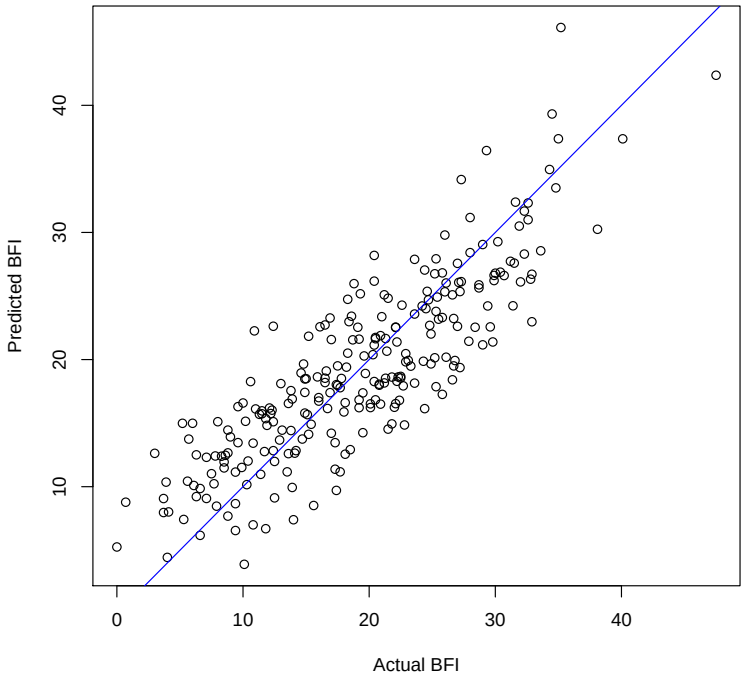


Figure 1: Results from the second model

## 2 Optimistic Bias

When performing both the training and the testing on the whole set, we obtain a square error of 4879.435936. This corresponds to a mean square error of 19.36284.

When performing training on half of the set and testing on the other half, with 100 repetitions, we get a square error of 2607.387, thus a mean square error of 20.69355. Multiple iterations give similar results.

This shows that the error in the second experiment is higher than the error in the first one. The optimistic bias is thus very well observable.
The results obtained by one of the 100 experiments are shown on figure 2.

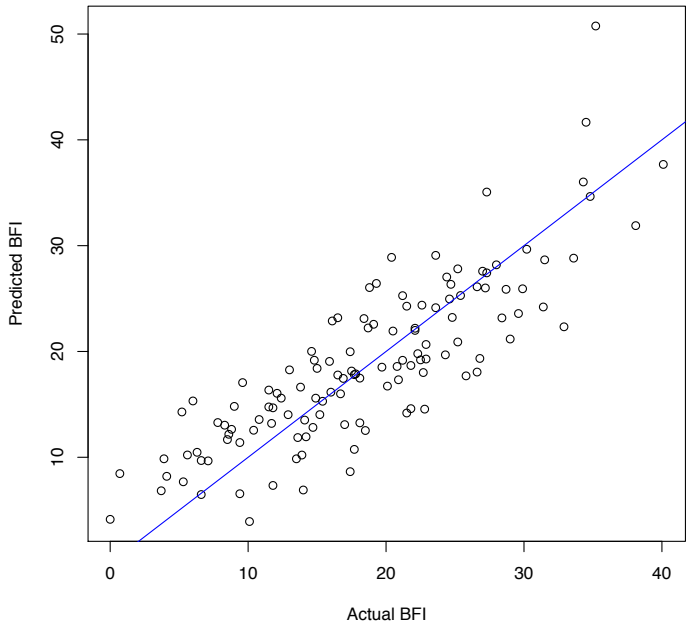


Figure 2: Results from one experiment