Case: $CaCO_3$ in cement

Introduction

Calcium carbonate is important for the properties of cement, and one intends to measure the value of $CaCO_3$ in the raw meal.

40 samples were taken from a cement mixer at regular intervals. The content of $CaCO_3$ were determined by duplicate titrations.

In this exercise we will predict the value of the second titration using the value of the first titration. We will apply linear regression, $y_i \approx \alpha + \beta x_i$.

Variables

variable name	description
`sample`	Sample no.
`tit1`	Kalkindhold ved første titrering i %
`tit2`	Kalkindhold ved første titrering i %

Exercise

- Make a scatter plot af data. Does it seem reasonable to predict the value of `tit2` from the value of `tit1`?
- Split the data into training and test data.
- "Train" the model with linear regression. Which unknown parameters are in the the model?
- Repeat using different sizes of training and test data. Make a graph of training and test errors as functions of training size. Comment.