

Working lecture assignments

Statistics

Working lecture 13 & 14: Submission task.

Introduction

This final hand-in assignment involves analyzing a real dataset. This is a somewhat freer assignment that requires turning in a report of the analysis and conclusions. This assignment counts twice as much as previous hand-in assignments. This assignment may be made alone or in pairs.

When writing the report keep in mind that a client has approached you with the text below.

01-02-1987

Dear statistician,

Attached you will find a dataset with data on 52 US companies from the Forbes500 list of best companies in the US for the year 1986. In it you will find the revenue, capital and market value of each of these companies. In addition, each company is indicated to which of the following four sectors it belongs: energy, financial, manufacturing or retail.

I would like to ask you to investigate how sales are affected by these factors so that I can take them into account in my business plan. I await your urgent response.

Sincerely,

B. Gates

Suggestions

Below are some hints and suggestions for analysis. The idea is that you should also consider for yourself what makes sense to do, give arguments for choices made, and explain what you did. Your report is a coherent narrative in which you present your findings. In particular, it is not recommended that you limit yourself to completing the list below point by point. You may use functions available in R, but your client may not have specific knowledge of R. However, he did take the Statistics course during his studies, using a book by Van der Vaart, Bijma and Jonker. So you can use sentences like "We fit a linear model with . . . as independent variables and calculate the least squares estimator for "

For this assignment, you are expected to explain your methods and conclusions, because we also want to be able to verify exactly what you did. In real life, this might not be necessary, or perhaps not even intended. Even conclusions that you would not normally report to your client may be important in this case.

1. The variables in the dataset are called sales, assets, market trap and sector.
2. Start by making relevant plots to get an idea of the data.
3. Set yourself the goal of fitting a good linear regression model.
4. Consider not including the financial variables directly in the model, but taking the logarithm first. You can also do this for the dependent variable.

5. The most basic model has 6 regression parameters and a parameter σ . Add interaction terms with the categorical variable **sector**.
6. Consider whether it makes sense to estimate variance by category. If so, what does your model look like? Try to write it down compactly. To test that $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$ you can use that

$$\frac{SS_{\text{res}}}{\sigma^2} \sim \chi^2 (n - k)$$

in een model met k regressieparameters. De volgende toetsingsgrootte is geschikt:

$$T = (n - k) \log \frac{SS_{\text{res}}}{n - k} - \sum_{i=1}^4 (n_i - k_i) \log \frac{SS_{\text{res}}^{(i)}}{n_i - k_i}.$$

Here n is the sample size, k is the total number of regression parameters, and k_i is the number for category i . You can simulate the distribution of T .

7. One can ask whether the continuous variables are of significant influence, and whether this influence depends on the industry. An F -test is appropriate for this type of question.