DCM: Assignment 1

Richard Öhrvall

2023-01-25

Assignment 1

Your assignment is to **individually** carry out the tasks described below. You should submit 2 files in Lisam: one R Markdown file with your code and text and one file that is the knitted pdf version of the same R Markdown file.

If you want more information about the dataset and its variables, you can look here: https://www.kaggle.com/c/titanic/data.

The assignment is due on Thursday, February 2, at 09.00. You submit in Lisam, under assignments. Late submissions are not accepted, so do not wait until the last minute to submit. If Lisam is not working for you, e-mail your files to Ben, benjamin.jarvis@liu.se, before the deadline.

Please note that LiU takes plagiarism very seriously. Plagiarism can both be to hand in assignments that in parts are identical to others student's assignment or other work, to not refer to other peoples' texts and ideas properly, and more. Please consult LiU's guidelines at https://liu.se/en/article/plagiering-upphovsratt.

Tasks

Note: all tables should appear in the pdf and have relevant titles, labels, captions, and so on. The pdf file should not include your code, just the tables and your written answers. Remember that you can control the output from the code chunks by setting echo, messages, etc. to FALSE, see https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf for more information or consult the resources mentioned in the study guide.

A (max. 300 words)

The passengers on Titanic embarked from different ports. Create a new variable, Southampton, that refers to if they embarked from Southampton (=1) or not (=0). Make a table with descriptive statistics for the variables: Sex, Age, Survived, Pclass, and your new variable Southampton. The continuous variables and the dummy variables (with 0 or 1) should have information on min, max, mean, SD, and number of observations (one could question the value of SD of a dummy variable, but include it anyway). The categorical variables should have information on number of observations and the distribution over the categories in percent.

Describe the table as you would do in a paper. Use clear and concise English.

B (300-500 words)

Let us say that we want to study if the old phrase "women and children first" accurately describes the event on Titanic, or rather if women and younger persons were more likely to survive.

Create a new dummy variable, female, that indicates if a passenger is female (=1) or not (=0). Estimate a linear probability model with survival as dependent and age and your new variable female as independent variables. Estimate a second model where you also include passenger class and the variable Southampton as an independent variables.

Produce one table that includes both of the two estimated models.

Describe the estimated models and your findings, including interpretation of the results (including at least one measurement of the model fit), in a way that is understandable for someone who is not familiar with the data, i.e. as you would do in an article. Use clear and concise English.