

STAT2120: Linear models 2020/21

Project

Important:

- Work in groups of two or three people.
- Submit a written report by **Saturday, January 2, 2021, 12h**, recto-verso, in the box of the assistant, as well as by email (pdf only) to the assistant and the professor.
- Validate the dataset and the composition of your group by email to the assistant before **Thursday, November 12, 2020**.
- The title page of the report contains the names of the group members, a title, the course number and the academic year.
- Include your programs as an appendix.
- Put graphics and tables either in the text, or appendix.
- Page limit: 25 pages altogether. Text: maximum 15 pages, Appendix: maximum 10 pages.

Data:

Every group uses its own dataset, to be validated by the assistant. Try to find a meaningful empirical study, where you determine a response variable Y and the explanatory variables X . You must have at least seven quantitative variables, and at least two qualitative variables.

Tasks:

1. Separate arbitrarily a small subset of observations from the dataset. These observations will be used for prediction, but not for model estimation.

2. Start your report by formulating an objective. What is the idea and the goal of your project? At the end of the report, explain why or why not the objective has been attained.
3. Do a descriptive analysis of the variables of the model. Provide a table with mean, standard deviation, skewness and kurtosis. Show boxplots of the variables, and the correlation matrix.
4. Select an adequate model for the response variable by considering all quantitative and qualitative variables and by using our model selection strategies. Consider also possible interactions of the qualitative variables with one or several quantitative variables. Verify the underlying hypotheses and, if necessary, take remedial actions. For example, check for
 - (a) nonlinearity
 - (b) outliers and/or influential observations,
 - (c) multicollinearity,
 - (d) heteroskedasticity, and
 - (e) autocorrelation.

If necessary, try to improve the model by using the methods seen in class. Check also for normality of the model residuals. If you reject, say why your results might be an approximation.

5. Test for significance of the estimated coefficients of the obtained model, and interpret their signs. Give a more detailed interpretation of the coefficients of the qualitative variables.
6. Test a linear combination of at least two coefficients, which makes sense in the context of your project.
7. Test a subset of coefficients (at least two), for example corresponding to the qualitative variables, to be equal to zero.
8. Calculate a prediction interval for the observations that were excluded in the beginning. Do your intervals cover the excluded observations?
9. Give a general conclusion about the adequateness and usefulness of your model.