

Multiple Linear Regression

Your task for this project is to apply the tools we've learned in chapters 6, 7, 9, & 10 (Multiple Linear Regression) to answer questions about the relationship between multiple explanatory variables and one response variable. The report should include:

- Introduction (Briefly refresh the reader's mind as to the variables of interest)
- The regression model you'll be fitting. Use at least 4 other variables. If a factor variable has more than one level, it will need to be made into $g-1$ indicator variables, where g is the number of levels.
- Before running the analysis, create a pairs plot on the explanatory and response variables. Comment on any interesting relationships you see. Are any of the explanatory variables highly correlated? Is there any reason to fit a quadratic term? Or do a log transformation?
- Comment on whether or not you are using interaction variables. (You can use interactions with non-factor variables, it's just that they are slightly more difficult to interpret.) If you think interaction variables are necessary, comment on why the slope of the equation would change based on the level of one of the other variables.
- Fit your model. Include quadratic, log, or interaction terms as you see fit.
- Interpret your β coefficients to the best of your ability. Are your coefficients significant? You can perform a test of significance $H_o : \beta_i \geq 0$ **OR** $H_o : \beta_i \geq c$ if you think there is a reason that the slope would increase by a certain factor greater than 0 (or that the intercept would increase by a certain factor if the variable of interest is an **indicator** variable.)
- Use the F test to compare two nested models. (That means that the larger model contains all the variables in the smaller model.) Make the smaller model have at least 2 fewer variables than the larger model. Comment on the soundness of the model. Which one would you report to your boss if you could only give her one model? (You can use Occam's Razor as a guideline.)
- Report the R^2 and Adjusted- R^2 values. Comment on the fit of the model as determined by how much variability is explained. Is this a guarantee that the model will accurately describe the population? Why or Why not?
- A complete analysis of the residuals and influence points. Use plots to get an idea of which points may be contributing to the fit. Consider re-fitting a model with and without certain data that have both high leverage and large residuals. Do not include every plot, but consider including plots that give the reader an idea of your analysis.

- Use either manual, stepwise, forward, or backward selection procedures to create the best model for your data. Use residual plots, significance tests, and criteria (C_p , R_a^2 , R^2 , AIC, SBC,...) to justify your model.
- Try to give an interpretation of the model that makes sense. Why do you think some variables stayed significant and others dropped out? Are any of your variables highly correlated (could one have taken the place of another?)
- Give CIs for a mean predicted value and a future predicted value for at least one combination of X's (from your final generalized linear model).
- Summarize your report.

Notes:

- Summarize any output from R. Do not turn in the print outs, but make new tables and summarize so that it flows nicely in the text. I don't need to see the technical calculations.
- I've asked you to do a series of things above, make sure the sections flow nicely into one another. This is a report on the data not a homework assignment. (Try to tell a good story.)
- Do not be tempted to turn in everything you do. Only turn in the interesting parts of the analysis. One of the hardest parts of being a consultant is figuring out what to tell the researcher.
- Computer output that is attached and not described will be ignored.
- Please double space your work.
- Turn in the previous projects with this project. (But let this assignment stand alone, that is, don't expect me to remember your variables.)
- Remember to label all graphs, email me if you are having trouble in R with labels (or really, any troubles in R!)