Summary Plus

In this project, you have 3 separate tasks. The first is to apply the topics from the end of the semester (ANOVA). The second is to apply something new (see below). The final task is to summarize the semester project in a meaningful way for a client. Feel free to re-do anything from previous projects to make your final report even better.

1 ANOVA

The first section includes applications of the tools from chapters 16 - 19 (Analysis of Variance) to answer questions about the differences in means across categorical groups. The report should include:

- Introduction (Briefly refresh the reader’s mind as to the variables of interest)
- The hypotheses that you’ll be addressing. The original null hypothesis will be that the means across 3 or more groups are equal.
- You’ll also be hypothesizing about a subset of the large number of groups. (I.e., you will combine some of the groups like we did with the mice data: you should be running a nested F-test.) The hypotheses need to be in the paper, but you can write them down in a natural place (they don’t both have to come immediately after the introduction – the key to a good paper is telling a good story.)
- The ANOVA table(s) and a description of the pertinent items.
- Any graphs (boxplots?) that you think give evidence to the point you’re making (story you’re telling with the data.)
- After running your hypothesis test, compare or contrast means across groups. Depending on your comparisons (planned? unplanned?) adjust for multiple comparisons. Justify your choice of adjustment for multiple comparisons. (Give at least one linear combination that is not a pairwise comparison.)
- Diagnostics: check your model assumptions using residual plots and consider possible transformations (see section 18.5)
- Run a 2-way ANOVA using 2 categorical variables. Test for interaction (and main effects). Include a plot of treatment means (see figure 19.7, page 829); try the R command `interaction.plot`. Suggest a strategy for analysis given your interaction results (see section 19.7, pgs 847-848).
- A Conclusion (Summarize your results. Comment on anything of interest that occurred in doing the project. Were the data approximately what you expected or did some of the results surprise you? What other questions would you like to ask about the data?)
2 Something New

- Perform at least 2 different tasks (tests, plots, CI, etc.) that we haven’t covered in class.

- Why did you pick the method(s) you did? That is, why does it work for your data? Explain why this method was important for understanding the complete analysis of your data.

- Give some background/theory to the method (demonstrate that you understand what is going on). This is crucial! For example, describe the derivation and intuition behind a new test statistic. Give as much information as possible about what you understand of the new idea.

- What assumptions does the model make? How sensitive are the results to the assumptions? Have you violated them?

- If you have any questions about the new topic, please come talk to me. I am happy to walk through the new idea with you to make sure that you are describing the main parts in sufficient detail.

Possible Topics

- More diagnostics
  - Normal probability plots (aka qq plots) (section 3.2)
  - Tests of assumptions (sections 3.5, 3.6, 6.8, 18.2)
  - Lack of fit (sections 3.7, 6.8)
  - Added variable plots (section 10.1)
  - Inverse predictions (section 4.6)
  - Weighted least squares (for non-constant variance) (11.1: regression, 18.4: ANOVA)

- Multicollinearity
  - Ridge Regressions (section 11.2; 6.2.1 in ISLR). Go beyond what we did in class: e.g., choice of ridge trace with VIF
  - Principal Components Regression (6.3.1 in ISLR)
  - Partial Least Squares (6.3.2 in ISLR)

- Extended regression models
  - Regression Spline / Local Regression (sections 3.10, 11.4; 7.4, 7.6 in ISLR). Go beyond what we did in class: e.g., using cross validation to set the smoothing parameters (either knots or span)
  - Smoothing Splines (section 7.5 in ISLR)
  - Generalized Additive Models (section 7.7 in ISLR)
  - Randomization tests (section 16.9, note: you’ll probably have to write R code)
  - Tree-Based Methods (chp 8 in ISLR)
  - Robust Regression (section 11.3)

- Logistic regression, when the response variable is binary (chp 14; section 4.3 in ISLR)

- Sample size calculations for ANOVA (sections 16.10, 17.8)

- Two-way ANOVA with one observation per treatment (chp 20)
• Randomized Complete Block Design (chp 21)
• Analysis of Covariance (ANCOVA), some categorical, some continuous variables (chp 22)
• Two-way ANOVA with unequal sample sizes (chp 23)
• Three-way+ ANOVA (chp 24)
• Random, Fixed, Mixed effects (chp 25)
• Nested designs, when second variable is a subset of first variable (chp 26)
• Repeated measures, not independent observations (chp 27)

3 Summary

• Report on the most interesting or significant findings in your data analysis this semester. Report as if to a client. If you give a model, report the entire model (variables, coefficient estimates, and p-values). Also, include the residual plot which is convincing evidence that the model is appropriate.

• If one (or more) of the methods you used didn’t give any interesting or applicable results, leave it out.

• Feel free to repeat the parts of the analysis that were particularly interesting or insightful.

• Give some justification for why the method(s) you chose worked well (for example, if you used ANOVA, comment on the fact that your 6 groups were very natural and led to an obvious choice of an ANOVA analysis as opposed to a linear model with factor variables.)

• Make some conclusions about the data overall. Did you see anything that should be further investigated? Do you think maybe the results are fascinating, but the sampling was poorly done and so the analysis should be re-done on a better sample?

• Be sure to communicate about the inference part: to who/what can you infer your result

• Give any final/concluding thoughts on the project and analysis. (Not whether or not you liked doing it... you can give that feedback on the course evaluation forms!)

Same notes as on all the previous assignments. And:

• There is no page limit. But you will be graded down for things that don’t belong in the report (e.g., warnings / errors of R code, lists of numbers, tables that aren’t readable, etc.).

• Remember that this should be your final report. Think of it as an analysis that you are submitting to your supervisor after collecting the data. You are trying to answer legitimate questions that will give insight into the data and population of interest. Your supervisor will expect it to be simultaneously concise and informative.