Math 158, Spring 2009
Jo Hardin
semester project (part 5)
Seniors: due Thursday, April 30
Non-Seniors: due Thursday, May 14
(earlier would be wonderful!)

Summary Plus

Your task for this project is to summarize the semester project and apply the tools from a chapter / section which we have not covered.

Possible Topics

- Simultaneous Inference:
  - Inference about 2 regression lines (section 4.2)
  - Inference at multiple explanatory variables (section 4.3)
  - Inverse predictions (section 4.6)
- 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)
  - Coefficient of partial determination (chp 7.4)
  - Weighted least squares (for non-constant variance) (11.1: regression, 18.4: ANOVA)
- Multicollinearity
  - Ridge Regressions (section 11.2)
  - Variance Influence Factor (section 10.5)
- Extended regression models
  - Polynomial models (sections 6.1, 8.1, 8.2)
  - Lowess (in R: `loess`) models (sections 3.10, 11.4)
  - Randomization tests (section 16.9, note: you'll have to write R code)
- Logistic regression, when the response variable is binary (chp 14)
- 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)

Plus

- Why did you pick the method you did? That is, why does it work for your data?
- Give some background/theory to the method (convince me that you understand what is going on).
- What assumptions does the model make? How sensitive are the results to the assumptions? Have you violated them?
- Perform at least 4 tasks (tests, plots, CI, etc.) that we haven't covered in class.

Summary

- Report on the most interesting or significant findings in your data analysis this semester.
- If one (or more) of the methods we 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.)
- 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?
- Give any final/concluding thoughts on the project and analysis. (Not whether you liked doing it... you can give me that feedback on the evaluation forms!)

Same notes as on all the previous assignments. And:

- Remember that this should be your final report. Think of it as an analysis that you are turning into your supervisor after collecting the data. You are trying to answer legitimate questions that will give insight into the data and population of interest.