Homework 2

Due in class or by email before 9am on Thursday January 27

[Data analysis problems: use R.]

1. (10 pts) KNNL 19.12. Notes:
   - In (c), substitute the dot plot with a plot of residuals versus predicted values
   - In (d), a Normal qqplot is enough. Skip the coefficient of correlation.

2. (10 pts) KNNL 19.13. Notes:
   - In (a), the problem refers to an interaction plot

3. (10 pts) KNNL 19.31

4. (10 pts) Suppose that we’d like to repeat the 2-way factorial experiment in KNNL 19.12. In this future experiment, the primary interest will be in two comparisons, $L_1 = \mu_1 - \mu_2$, and $L_2 = \mu_1 - \mu_2$. What is the minimal number of replicates that will be necessary to collect for each combination of eye contact and gender to detect a difference in success rating of 2 units? Control the family-wise error rate at the level of $\alpha = 0.05$ using Bonferroni approach, set power to 0.8, and use $\hat{\sigma}^2$ from problem 19.12 as a planning value of the variance.

5. (10 pts) KNNL 20.2 (a) and (b). Notes:
   - In (b), it is enough to conduct the tests. Skip the upper bound for the family level of significance.

6. (10 pts) KNNL 20.4

7. (20 pts) [Methods qualifying exam, January 2011: use paper and pencil.] A market researcher is interested in assessing the variability of caffeine levels in different types of caffeinated drinks and coffee brands.

   (a) At the first stage of the study, she selects 3 types of drinks and 4 coffee brands, and measures the caffeine content in 2 servings of each combination of drink type and brand.

   i. Write out the ANOVA model, specify any distributional assumptions, and interpret the parameters.

   ii. The researcher would like to make the following two comparisons: (1) whether the drink types 1 and 2 have the same expected caffeine concentration, and (2) whether drink type 1 have the same expected caffeine concentration as the
mean of expected concentrations of drink types 2 and 3. State the null and the alternative hypotheses, the test statistic and the decision rule. Control the family-wise error rate of the comparisons at 5%, while selecting the most appropriate procedure among Bonferroni, Scheffé or Tukey.

iii. State the coefficients of model parameters corresponding to the first of the two comparisons above. Clearly label the coefficients. (Hint: the coefficients can be used, e.g. as input to the ‘contrast’ statement in SAS.)

(b) The second experiment focuses on the drink type 1, and the researcher is interested in the variability of caffeine levels in this drink across all possible coffee brands. She randomly selected 8 coffee brands, and measured the caffeine content in 4 replicate servings of the drink type 1 from each brand.

i. Specify the ANOVA model, and the associated assumptions, that can be used to analyze these data.

ii. Using words that a non-statistician can understand, explain to the market researcher the interpretation of the ANOVA model above.

iii. Suppose the sum of squares of the error in the model in (i) is 0.72 and the sum of squares of brand is 3.92. What percent of the overall variation in the caffeine level of drink type 1 is due to the different coffee brands?

iv. The overall average caffeine level across all servings of drink type 1 sampled in this experiment was 3.6 mg/oz. Suppose that Starbucks coffee boasts a caffeine content of 20 mg/oz. Does the examined drink have the expected caffeine level less than 1/5 of this amount? Specify the null and the alternative hypotheses, test statistic, and decision using $\alpha = .05$. 