STA 2201/442 Assignment 11

1. Steel is made by heating iron and adding some carbon. A steel company conducted an experiment in which knife blades were manufactured using two different amounts of carbon (Low and High), and three different temperatures (Low, Medium and High). Of course even the Low temperature was very hot. A sample of knife blades was manufactured at each combination of carbon and temperature levels, and then the breaking strength of each blade was measured by a specially designed machine. The response variable is breaking strength.

   (a) In a table with one row for each treatment combination, please make columns giving the coefficients of the contrast or contrasts you would use to test for main effects of Temperature.

   (b) In another table with one row for each treatment combination, please make columns giving the coefficients of the contrast or contrasts you would use to test the Temperature by Carbon Level interaction.

   (c) In one last table with one row for each treatment combination, please make columns showing how you would set up dummy variables for both independent variables, using *effect coding* (that's the scheme with the -1s).

   (d) Write $E(Y|\mathbf{X} = \mathbf{x})$ for the regression model, using the names from your table above. Include the interactions!

   (e) Using the β values from your answer to the preceding question, state the null hypothesis you'd use to test whether the effect of carbon level on breaking strength depends on the temperature.

2. Consider a two-factor analysis of variance in which each factor has two levels. Use this regression model for the problem:

$$Y_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + \beta_3 d_{i,1} d_{i,2} + \epsilon_i,$$

   where $d_{i,1}$ and $d_{i,2}$ are dummy variables.

   (a) Make a two-by-two table showing the four treatment means in terms of β values. Use *effect coding.* In terms of the β values, state the null hypothesis you would use to test for

      i. Main effect of the first factor
      ii. Main effect of the second factor
      iii. Interaction

   (b) Make a two-by-two table showing the four treatment means in terms of β values. Use *indicator dummy variables* (zeros and ones). In terms of the β values, state the null hypothesis you would use to test for

      i. Main effect of the first factor

     ii. Main effect of the second factor

   iii. Interaction

  (c) Which dummy variable scheme do you like more?

3. In the slide show on power for the F test, I obtained a special case formula for comparing two means. It was $\phi = nf(1 - f)d^2$, where $f = \frac{n_1}{n}$ and $d = \frac{|\mu_1 - \mu_2|}{\sigma}$. I did this the hard way, using a regression model with an intercept and a single indicator dummy variable. I believe it would be a lot easier with cell means coding. Try it. You should get the same formula for ϕ.

   The formula sheet has the general expression for the non-centrality parameter ϕ. The formula sheet will be provided with the final exam, and also with Quiz 11 if necessary. There is a link to the formula sheet on the course home page unless the one in this document does not work.

4. Remember the rotten potato example of factorial ANOVA. It was a 2×3 design with equal sample sizes. Buried in all those tests of special contrasts was a test of equality of three Bacteria means at cool temperature. Suppose the means for Bacteria types 1 and 2 were equal, but the mean for Bacteria type 3 was one-half standard deviation lower.

   (a) What was the power to detect this effect? The answer is a number between zero and one.

   (b) What *total sample size* would have been required for a power of 0.80? Remember, all 6 sample sizes are equal.

   Please use R and bring your printout to the quiz. You may use my function fpow2 if you wish. The lecture slides show how to get it. After you use the source command, type fpow2 to see the definition of the function.

5. I know this is pretty gruesome, but the data are real. An experiment in dentistry seeks to test the effectiveness of a drug (HEBP) that is supposed to help dental implants become more firmly attached to the jaw bone. This is an initial test on animals. False teeth were implanted into the leg bones of rabbits, and the rabbits were randomly assigned to receive either the drug or a saline solution (placebo). Technicians administering the drug were blind to experimental condition.

   Rabbits were also randomly assigned to be "sacrificed" after either 3, 6, 9 or 12 days. At that time, the implants were pulled out of the bone by a machine that measures force in newtons and stiffness in newtons/mm. For both of these measurements, higher values indicate more healing. A measure of "pre-load stiffness" in newtons/mm is also available for each animal. This may be another indicator of how firmly the false tooth was implanted into the bone, but it might even be a covariate. Nobody can seem to remember what "preload" means, so we'll ignore this variable for now.

The data are available in the file `bunnies.data`. The variables are

- Identification code
- Time (3,6,9,12 days of healing)
- Drug (1=HEBP, 0=saline solution)
- Stiffness in newtons/mm
- Force in newtons
- Preload stiffness in newtons/mm

Classify the factors as within cases or between cases. Then please do the following using SAS. Remember that in the `input` statement, character-valued variables should be followed by a dollar sign.

(a) Use `proc freq` to find out how many rabbits are in each experimental condition.

(b) Using `proc glm`, conduct a two-way ANOVA, with force as the dependent variable. Use the means statement to get cell means and marginal means. Be prepared to answer the following questions about each of the significance tests that SAS produces by default (I count 4 default tests).

   i. What is the value of the test statistic? The answer is a number from your printout.

   ii. What proportion of the remaining variation is explained? You can use `proc iml`, or just be ready to do it with a calculator. The formula will be supplied with the quiz if this question is asked.

   iii. What is the p-value? The answer is a number from your printout.

   iv. Do you reject the null hypothesis at the 0.05 level? Yes or No.

   v. What, if anything, do you conclude? This is not the place for statistical jargon. "What do you conclude" means say something about the drug, healing, time – something like that.

(c) I know this is a bit redundant with the preceding question, but *did the drug work?* If the results justify an answer, then answer Yes or No.

(d) Now, make a table with a row for each treatment combination. Give the coefficients of the constrast or contrasts that would be used to test for

   i. The main effect of Drug

   ii. The main effect(s) of Time

   iii. The Drug by Time interaction.

(e) Make another table with a row for each treatment combination. Make columns showing the dummy variables for effect coding.

(f) Give $E[Y|\mathbf{X} = \mathbf{x}]$ for a regression model with both main effects and the interaction. Use your variable names from the preceding question.

(g) In terms of the β values of your regression model, give the null hypothesis you would test in order to answer each of the following questions.

    i. Averaging across time periods, is there a differnece between the drug and placebo in mean force required to extract the tooth?

    ii. Averaging across drug and placebo, is does elapsed time affect the mean force required to extract the tooth?

    iii. Does the effect of the drug depend upon elapsed time?

(h) Now please return to SAS. Using `proc reg` and cell means coding, conduct tests to answer the following questions. Just do regular one-at-a-time (custom) tests. Don't bother with any Bonferroni correction this time. Just consider one dependent variable: Force. As usual, we are guided by the $\alpha = 0.05$ significance level.

    i. Are the marginal means different at 3 and 6 days?

    ii. Are the marginal means different at 6 and 9 days?

    iii. Are the marginal means different at 9 and 12 days?

    iv. Is there a difference between Drug and Placebo just at 3 days?

    v. Is there a difference between Drug and Placebo just at 6 days?

    vi. Is there a difference between Drug and Placebo just at 9 days?

    vii. Is there a difference between Drug and Placebo just at 12 days?

    viii. Be able to answer questions like these for each test:

        A. What is the value of the test statistic? The answer is a number from your printout.

        B. What proportion of the remaining variation is explained? You can use `proc iml`, or just be ready to do it with a calculator. The formula will be supplied with the quiz if this question is asked.

        C. What is the p-value? The answer is a number from your printout.

        D. Do you reject the null hypothesis at the 0.05 level? Yes or No.

        E. What, if anything, do you conclude? This is not the place for statistical jargon. "What do you conclude" means say something about the drug, healing, time – something like that.

Please bring your `SAS` log file and list file to the quiz. Remember, the log file is not the same as a plain listing of the program. Please be sure that the log file and list file come from the same SAS run. *Do not write anything on your printouts in advance except maybe your name* (and student number, if you wish).