Statistics 500 - Homework #7, Fall 2020
Due: noon Friday, 10/16/2020

Reading Assignment: Statistical Sleuth, Chapters 13, 14 (Analysis of factorial experiments), and 24.

1. Standing from a seated position generates stress on the knee joint. Finding ways to minimize this stress is important in individuals who have had knees replaced (Total Knee Arthroplasty, TKA). An experiment is conducted to test whether different feet placement (neutral, back, staggered) affect the amount of torque on the knee (measured in Newton meters, Nm) of 14 old men with TKA. Each of the men will experience all three treatments, in randomly selected order, during the experiment. For all trials, the height of the chair will be the same and participants will all wear comfortable clothing and tennis shoes. The data are located in the file **knee.txt** and appear in the table below.

Subject	Neutral (N)	Back (B)	Staggered (S)
1	26.3	23.0	21.3
2	22.7	19.7	19.8
3	21.1	20.0	19.5
4	25.9	23.3	22.4
5	25.6	21.1	18.5
6	22.5	19.9	19.6
7	26.1	23.4	22.5
8	21.4	18.8	19.0
9	17.4	17.7	16.6
10	22.8	21.9	20.0
11	23.0	22.2	22.1
12	18.4	17.3	18.6
13	29.6	25.5	21.8
14	27.4	23.2	22.3

   a. Using SAS, construct an ANOVA table for the model with treatment and subject effects.
   b. Are there significant differences in the torque between the three feet placements? Apply an F-test and state your conclusion.
   c. Apply the Tukey's method with an experiment-wise Type I error level of $\alpha = 0.05$. State your conclusions about the pairwise comparison of the three replacement methods.
   d. Specify and test two orthogonal contrasts that compares (i) the mean of the feet neutral group to the mean of the other two groups, and (ii) the mean of the back group to the mean of the staggered group. How do these results support your conclusions from part (c)?
   e. Examine the normal probability plot of the residuals. What does this plot suggest?
   f. Plot the residuals versus the estimated means (predicted values). What does this plot suggest?
   g. Estimate and interpret the relative efficiency of blocking for this experiment relative to a completely randomized experiment.

2. A group of investigators are interested in the effect of planting density on the yield of millet, a small grain. The investigators are comparing five planting densities, labeled 2, 4, 6, 8, and 10. A field of land was divided into 25 plots (5 rows and 5 columns) to account for known differences in soil quality in the field. In their first study, the investigators randomly assigned density treatments to plots in a 5 x 5 Latin Square. The data are in **millet.txt**.
   a. Use SAS to obtain an ANOVA table for the data.

b. Test the hypothesis of no effect of treatment (i.e., no differences between planting densities). Report the F statistic and p-value and draw your conclusion on whether there is any effect of planting density on the yield of millet.
c. Conduct an analysis of the pairwise comparisons for the mean yields of the five planting densities using Tukey's method. State your conclusions about the relationships between the mean yields of five planting densities.
d. If the study is repeated next year, would you recommend that it still be a Latin Square? Or, should the investigators just use row blocks, just use column blocks, or just use a CRD? Justify your answer with appropriate calculation and explanation.

3. (The study we looked at in Problem 3 of Midterm Exam) In a study of crop losses due to air pollution at a farm in North Carolina, plots of green beans were grown in open-top field chambers and continuously exposed to various conditions of air pollution. Four plots were randomly assigned to each of four treatment groups. Treatment A is air without pollution (control); treatment B is air with sulfur dioxide pollution; treatment C is air with nitrous oxides pollution, and treatment D is air with both sulfur dioxide and nitrous oxides pollution. After a month, the plants were harvested and the yield (kg) of bean pods was recorded for each of 16 plots.

a. What are the two treatment factors included in this experiment? What are their levels?
b. Is this a full factorial treatment design? Justify your answer.
c. Construct an appropriate ANOVA table by listing the source of variation and the degrees of freedom.
d. Suppose that the investigators are interested in the following questions:
(i) Is there any difference between treatments with air pollution and the control treatment (air without pollution)?
(ii) Averaged over sulfur dioxide pollution status (yes or no), is there any effect of nitrous oxides pollution?
(iii) Does the effect of nitrous oxides pollution depend on the inclusion of sulfur dioxide pollution or not?
For each question in parts (i)-(iii), give a corresponding contrast to answer the question.
e. For each contrast in (d), indicate whether it is a main effect, a simple effect, or an interaction effect.
f. For this experiment, how many contrasts in a set of orthogonal contrasts can you get at most?
g. Are the two contrasts (ii) and (iii) orthogonal? Give a Yes or No answer, and support your answer with appropriate calculations.

4. A completely randomized two factor experiment consisted of burning fuel with levels of two additives in a laboratory setting and determining the CO emissions released. Eighteen batches of a standard fuel were available for this study. Two of the batches were randomly assigned to each of nine combinations of two additives corresponding to three levels of added ethanol (0.1, 0.2, or 0.3) and three air/fuel ratio settings (14, 15, or 16). Units for the ethanol levels were not reported. Carbon monoxide (CO) emissions concentrations (g/meter3) were determined for each burning the same amount of fuel from each of the 18 batches. The data are shown below and are located in the file **emissions.txt**.

	Air/Fuel Ratio		
Added Ethanol	**14**	**15**	**16**
0.1	66 60	72 65	68 64
0.2	78 81	80 81	66 69
0.3	90 94	75 78	60 58

a. Construct an appropriate ANOVA table. Which factors or interactions have significant effects on CO concentrations in emissions?
b. Study the interaction plot from the SAS output. Explain why, based on this plot, the interaction is significant in the model.
c. Partition the sum of squares for the ethanol effects, averaging across air/fuel ratio levels, into sums of squares for linear and quadratic components. The coefficients for these contrasts are (-1, 0, 1) and (-1, 2, -1). Is there a significant linear or quadratic effect in the model for the ethanol effects?
d. Partition the sum of squares for the air/fuel ratio effects, averaging across levels of ethanol, into sums of squares for linear and quadratic components. The coefficients for these contrasts are (-1, 0, 1) and (-1, 2, -1). Is there a significant linear or quadratic effect in the model for the air/fuel ratio effects?
e. Use Tukey's HSD method to make pairwise comparisons of the marginal means for the three ethanol values. State your conclusions based on this analysis.
f. Use Tukey's HSD method to make pairwise comparisons of the marginal means for the air/fuel ratio values. State your conclusions based on this analysis.
g. Study the residuals versus predicted value plot. Is there anything of concern in this plot?
h. Study the residuals by ethanol values plot and the residuals by air/fuel ratio levels plot. Is there anything of concern in these plots?
i. Study the normal probability plot for the residuals. Is there anything of concern in this plot?