STP 531 Final project (take home)

- Please do NOT discuss this final project with anyone, except for the instructor, and your team member(s).

- By 3:00PM on Tuesday, 05/01/2018, at my office WXLR 525, please return in:

  1. individual effort report signed by all the team members
  2. final project report
  3. individual survey

- Your must type your solutions and print them out; pages must be numbered. For each problem, your solution should have two parts:

  - part I (summary section, up to two pages for each question, supported by tables and/or figures), directly address the question of interest and clearly state your results. You will present a summary that can be used by an intelligent person without referring to other documents or conducting additional calculations.

  - part II (analysis section, up to five pages for each question) show your work here; this should consist of a logical progression of steps; each step should be summarized in words, and with supporting computer codes, computer outputs, figures, and/or mathematical formulae. Please include only relevant computer outputs, and provide your interpretations of these outputs for answering the questions. You do NOT need to present the data set.

1. An experiment was designed to study the association between the wheat variety and its growth (Y). There are four wheat varieties of interest. And, five greenhouse benches were set up as blocks. Within each block, the four varieties of wheat were planted; all the other conditions were kept the same as possible. The data on measurements of plant heights (in inches) after a period of time are given below.

Block	Variety			
	1	2	3	4
1	9.7	11.8	6.3	4.6
2	6.6	9.7	5.3	3.4
3	7.6	10.9	4.7	2.3
4	8.1	11.3	5.5	3.6
5	6.4	10.7	4.5	2.8

   (a) State an appropriate ANOVA model for this study. State and check the model assumptions.

   (b) Test whether or not the main effect of variety is present. State the hypotheses, test statistic, p-value and conclusion. Use $\alpha = 0.05$.

   (c) Obtain confidence intervals for all pairwise comparisons between the treatment (variety) means; use the most efficient multiple comparison procedure with a 90% family confidence coefficient. Interpret your results.

   (d) Estimate the difference in the mean plant height for the first two groups of variety $(\mu_{\cdot 1} - \mu_{\cdot 2})$ with a 95% confidence interval. Interpret your findings.

   (e) Test for $H_0 : \mu_{\cdot 1} = \mu_{\cdot 2}$ vs. $H_a : \mu_{\cdot 1} \neq \mu_{\cdot 2}$ with $\alpha = 0.05$. State the test statistic, p-value and your conclusion. Does your conclusion agree with the result for (d)? Explain.

   (f) Test for $H_0 : \mu_{\cdot 1} \geq \mu_{\cdot 2}$ vs. $H_a : \mu_{\cdot 1} < \mu_{\cdot 2}$ with $\alpha = 0.025$.

   (g) Comment on the efficiency of the blocking variable.

2. Consider the same data set as in the previous problem. But now, suppose the four varieties are randomly selected from a population of 20 varieties of wheat.

   (a) State an mixed effects model for this setting, and write down the model assumptions.

   (b) Provide the ANOVA table, and discuss how this table can help to make inference about the difference in the plant heights among the 20 wheat varieties.

   (c) Following (b), is there a statistically significant difference in the plant heights across different varieties of wheat? State the hypotheses, test statistic, p-value and conclusion. Use $\alpha = 0.05$.

3. An experimenter would like to compare the effects of the four diets on the weight gain, Y, (pounds per day) of pigs. The data can be found in Diet.txt. The four columns in the data set are, respectively, diets, initial age (1: ≤ 77 days, and 2: > 77 days), initial weight (pounds) and the weight gain. For this problem, use only the initial weight as the only concomitant variable, and ignore the initial age.

   (a) Write down a statistical model that you would use for analyzing the data. State and check the model assumptions.

   (b) Obtain a point estimate and a 95% confidence interval for the effect of the initial weight on the weight gain. Interpret your results.

   (c) Obtain a 95% confidence interval to estimate the mean weight gain for pigs in the second treatment group (diet of the second type) that have an initial weight of 56 pounds.

   (d) Obtain confidence intervals for all pairwise comparisons between the treatment means; use the most efficient multiple comparison procedure with a 90% family confidence coefficient. State your findings.

4. As in the previous problem, now analyze the data by including both the initial weight (as concomitant variable) and initial age (as blocking variable) in the model.

   (a) State an appropriate statistics model. Test whether or not the main effects of diet are present by assuming that all level combinations of diet and initial age are equally important. Use $\alpha = 0.1$.

   (b) Suppose that, in the target population, 75% of the pigs are 77 days or younger. Test whether or not the main effects of the diets are present in this target population. Use $\alpha = 0.1$.

   (c) Consider the same population as in (b). Find a 95% confidence interval for the mean weight gain for pigs that are in the second treatment group (diet of the second type) with an initial weight of 56 pounds.

   (d) Test whether or not the mean weight gain obtained in (c) is positive. State the hypotheses, test statistic, p-value and conclusion.