

# Project 9 - One-way ANOVA

Statistics 401: Spring 2006

Due Monday, November 20

Turn in your answers in a type-written report. Number the problems appropriately.

1. Explain why it is not appropriate to perform an ANOVA for Problem 15.4 on page 676 of your textbook.
2. Do problem 15.6 on page 677.
  - (a) Show your work for the degrees of freedom calculations!
  - (b) Perform the test by answering the questions below, which you will need to do by hand since we do not have the data. Assume that the assumptions for the ANOVA are met. You will perform the follow-up test in #3.
    - i. Hypotheses:
    - ii. Test statistic value:
    - iii. Distribution of the test statistic:
    - iv. p-value:
    - v. Decision at  $\alpha = .05$ :
    - vi. Conclusion:
  - (c) When computing  $MSE$ , show your work!
  - (d) Summarize all of your previous calculations by filling in an ANOVA table as in the Chapter 15 handout. Label and reference this table from the body of your report.
  - (e) Fill in the following table, label it, and include it in your report.

parameter	estimate	Explanation in English words of the parameter
Estimate for $\mu_1$		
Estimate of $\sigma$		

3. Do problem 15.22 on page 685.
  - (a) Why is it appropriate to perform a follow-up analysis to #2: problem 15.6 on page 677?
  - (b) Perform a Tukey's follow-up test, which you'll need to do by hand since we do not have the data. Show your work!
  - (c) Indicate all significant differences between the means. What is your conclusion?
4. The Kenton Food Company wants to determine the effect of four different package designs (Design 1, Design 2, Design 3, and Design 4) on monthly sales (in hundreds of dollars) for a new breakfast cereal. Nineteen stores are selected as the experimental units. Each store is randomly assigned to sell the cereal with only one of the package designs. Each package design is assigned to five stores, except for Design 3, which, due to spilled hot coffee, is only assigned to four stores. The stores are chosen to be comparable in location and sales volume. Other relevant conditions that could affect sales, such as price, amount and location of shelf space, and special promotional efforts, were kept the same for all of the stores in the experiment.

The "stores.txt" data file can be found on the STAT401 website. Use R to complete this problem. Attach all R commands and R output used in an appendix. Label all necessary Figures and Tables and refer to these figures and tables from the text of your report.

- (a) Give the value of  $x_{2,3}$  and explain what this value is in terms of the problem.
  - (b) Construct side-by-side boxplots for a visual comparison of the groups. Include a figure of the plot in your report.
  - (c) Give the null and alternative hypotheses to be tested by a one-way ANOVA.
  - (d) Fit the ANOVA model. Include a table that displays the one-way ANOVA table as in the Chapter 15 notes.
  - (e) Check the assumptions.
    - i. Does the evidence suggest that the data for each group are not normal? Include the a normal probability plot, a smoothed histogram of the studentized residuals, and the correlation test of the studentized residuals to justify your answer. Include the plots in a figure in your report.
    - ii. Does the assumption of constant variance hold? Why or why not?
  - (f) Give the distribution of the  $F$  statistic assuming the  $H_0$  is true.
  - (g) What is your decision regarding  $H_0$ ?
  - (h) Give a conclusion in terms of the problem. Is it appropriate to make a cause and effect conclusion? Why? If your answer is yes, then address cause-and-effect in your conclusion.
5. (a) Is it appropriate to conduct a follow-up test for the Cereal Sales versus Box Design ANOVA? Why or why not?
- (b) If it is appropriate, compute 95% Tukey confidence intervals for the pairwise differences between means. Provide a table of the results of the simultaneous confidence intervals as in the Chapter 15 notes.
- (c) Summarize the results of this study (i.e. give a conclusion). In particular, specify which package design is best for maximizing mean sales.
- (d) Give estimates for  $\mu_4$  and  $\sigma$ , and interpret each of these values in terms of the problem. Give your answers in a table as in #2(e).