

**Stat 401 A XM Homework 8**

**On-Campus Due Date:** Wednesday, October 29

**Off-Campus Due Date:** November, November 5

1. Suppose a completely randomized experiment with 4 treatments yielded the following data.

Treatment	Data
1	3, 4, 5, 8
2	5, 7, 10
3	13, 17, 18
4	7, 10, 13

Let  $\mu_i$  denote the mean of the  $i$ th treatment. Conduct one  $F$ -test of

$$H_0 : \mu_1 = \mu_2 \text{ and } \mu_3 = \mu_4 \quad \text{vs.} \quad H_A : \text{Either } \mu_1 \neq \mu_2 \text{ or } \mu_3 \neq \mu_4 \text{ or both.}$$

State your test statistic, its degrees of freedom, an approximate  $p$ -value, and a brief conclusion. *Hint: Consider this test as a comparison of full and reduced models. You may wish to review the handout "The ANOVA Table and Comparison of Full and Reduced Models." This handout doesn't tell you exactly how to do this problem, but you can do this problem if you understand the ideas on this handout.*

2. It can be difficult to make a line plot showing significant differences among means when there is great disparity among sample sizes. To illustrate the problem, consider the following hypothetical summary statistics.

Group ( $i$ )	$\bar{Y}_i$	$s_i$	$n_i$
1	7.0	2.7	3
2	7.3	3.0	50
3	9.0	3.3	50

- (a) Estimate  $\sigma$  the population standard deviation that is assumed to be the same for all three groups.
  - (b) Obtain a two-sided  $p$ -value for testing  $H_0 : \mu_1 = \mu_2$ .
  - (c) Obtain a two-sided  $p$ -value for testing  $H_0 : \mu_1 = \mu_3$ .
  - (d) Obtain a two-sided  $p$ -value for testing  $H_0 : \mu_2 = \mu_3$ .
  - (e) Suppose you consider any difference to be significant if the  $p$ -value corresponding to the test is 0.05 or less. (Don't worry about any adjustment for multiple testing for this problem.) What difficulty do you encounter if you try to make a line plot that illustrates the significant and nonsignificant differences among the three group means?
3. Twenty plots were randomly selected in a large field of corn. The plant density (number of plants per plot) and the mean cob weight (grams of grain per cob) were computed for each plot. The average plant density across the 20 plots was 128 plants per plot. The average of the mean cob weights computed across the 20 plots was 224 grams of grain per cob. The standard deviation of the 20 plant densities was 32.6 plants per plot. The standard deviation of the 20 mean cob weights was 25.0 grams of grain per cob. The correlation between plant density of cob weight was -0.94.
- (a) We will use linear regression to describe the relationship between plant density and mean cob weight. Which of the two variables would most naturally be considered the explanatory variable and which would most naturally be considered the response variable?
  - (b) Compute the equation of the least-squares regression line.

- (c) Compute a 95% confidence interval for the slope of the regression line.
  - (d) Estimate the mean cob weight for plots with 135 plants.
  - (e) Do you think your estimate in part (d) involved interpolation or extrapolation? Explain.
  - (f) Estimate the standard deviation of mean cob weights for plots with 135 plants.
4. Manatees (a.k.a. sea cows) live off the coast of Florida. Many manatees are killed or injured by powerboats. The program *manatee.sas* contains data on  $X$  =the number of Florida powerboat registrations (in 1000s) and  $Y$  =number of manatees killed near Florida. There is one point for each year from 1977 to 1990. Examine and run *manatee.sas*. Use the output to answer the following questions. (Most of your work will be in deciphering the SAS output. SAS will compute almost everything for you.)
- (a) Examine the scatter plot produced by SAS. Is the relationship between  $Y$  and  $X$  roughly linear?
  - (b) Is the correlation negative, zero, or positive?
  - (c) Find the exact sample linear correlation between  $X$  and  $Y$  in the SAS output under the section for *The CORR Procedure*. Only one of the numbers in this section could possibly be the correlation between  $X$  and  $Y$ . You should be able to figure out which number is the correct one on your own.
  - (d) Give the equation of the least-squares regression line for predicting the number of manatee deaths as a function of powerboat registrations.
  - (e) Predict the number of manatee deaths for a year in which 600,000 powerboats are registered in Florida ( $X = 600$ ). You can do this using your answer above or you can get SAS to do the computation for you. To have SAS do the computation:
    - (i) Add another observation to the data set that has a 600 value for  $X$  and a period in place of a  $Y$  value. (Periods indicate missing data in SAS.)
    - (ii) Replace *model*  $Y=X$ ; with *model*  $Y=X/p$ ; This will give you a predicted value for each observation in the data set, including the one you added with  $X = 600$ .