

**Stat 401, Section F Homework 3**  
**Due Date:** Wednesday, September 12

**Please do not include any computer output besides graphs. All other information should be copied from R into your written answer.**

1. Download and study the data set `steers.txt` available on the course website. The data come from a study designed to compare the effectiveness of two diets for increasing the weight of beef steers. Nine pairs of animals were chosen from a herd. The members of each pair were matched closely as possible with respect to hereditary factors and randomly allocated to diets such that one member of each pair received Diet 1 and the other Diet 2. The data are weight gains (in lbs) of the animals over a 140-day test period.
  - (a) Estimate the mean difference  $\mu$  in weight gain (Diet 1 - Diet 2).
  - (b) Provide a 95% confidence interval for the mean difference  $\mu$  in weight gain.
  - (c) A representative from a company that makes the feed Diet 1 claims that Diet 1 is superior to Diet 2 and that the difference estimated in part(a) is large enough to warrant charging those who raise steers a higher price for the feed used in Diet 1 compared to the feed used in Diet 2. On the basis of the confidence interval computed in part(b), write a response to the company representative explaining whether you agree or disagree with this claim.
  
2. Do Computational Exercise 12 from Chapter 2 on marijuana use during pregnancy.  
*(If the value for degrees of freedom does not appear in Table A.2, you should take the closest value for degrees of freedom which does appear in the table. For example, if  $df = 113$ , then 113 is not in the Table A.2, but the closest value which does appear is  $df = 100$ , use 100 instead of 113.)*
  
3. This problem is basically Data Problem 21 in Chapter 2, but I decided to rewrite the question and guide you through the various issues you need to address, rather than simply asking you to “write a short summary report”. Read the description of the data set found in Display 2.15. This data set can be found on the course website as `sparrow.txt`.
  - (a) The data contains two columns representing two variables. The second column consists of 1s and 2s. What do these numbers represent?
  - (b) Read the data set into **R** and use the function `t.test` to compute a 95% confidence interval for the mean weights of male sparrows who would have perished in the storm had they been exposed.
  - (c) Visually compare the data in the two groups. Draw side by side boxplots of the two groups of male sparrows, choosing meaningful labels for the axes of the graph and for labeling your boxplots (e.g. you could use labels like “Survived” and “Perished”).
    - i. Find the interquartile range for the male sparrows that perished.
    - ii. Find  $Q1 - 1.5(Q3 - Q1)$  and  $Q3 + 1.5(Q3 - Q1)$  for the male sparrows that perished.
    - iii. Verify that the largest weight of a male sparrow that perished is greater than  $Q3 + 1.5 * (Q3 - Q1)$ .
    - iv. Verify that no other weight for a male sparrow that perished falls outside the limits computed in part (ii).

- (d) The function `t.test` will conduct a ttest to compare the means of the groups specified when you call the function. State the degrees of freedom, test statistic, and  $p$ -value for the test of  $H_0 : \mu_1 = \mu_2$  vs.  $H_a : \mu_1 \neq \mu_2$  assuming the standard deviations in the two groups are equal. (Here  $\mu_1$  is the mean weight of male sparrows who would have died had they been exposed to the storm, and  $\mu_2$  is the mean weight of male sparrows who would have survived the storm.) What do you conclude? To use `t.test` to perform a two-sample t-test with equal variances (assuming the two samples are called `j1` and `j2`) type `t.test(j1, j2, var.equal=TRUE)`
- (e) Provide a 99% confidence interval for  $\mu_2 - \mu_1$ . How does this interval agree with your conclusion above?  
*Note:* By default, the confidence level is set to 95% in **R**. To change the confidence level, simply specify `conf.level=0.99` when you call the function `t.test` as above.
4. Suppose 6 pairs of hogs are used in an experiment to determine if a dietary supplement can increase weight gain. Each pair of hogs are full siblings. One sibling is randomly assigned to receive the dietary supplement. The other sibling receives a supplement that does not contain the active ingredient believed to affect weight gain. The gains in pounds are provided below.

Sib Pair	Treatment Weight Gain	Placebo Weight Gain
1	70.4	69.3
2	38.2	37.2
3	59.7	58.7
4	40.3	39.3
5	93.5	92.5
6	50.7	49.6

Note: You don't have to use **R** to answer the next two problems, but I think you will save yourself some time and tedious work if you do use **R**. When running a matched pairs analysis in **R** you could either compute the sequence of differences first and run a one-sample t-test on those, or you can simply input the two groups directly into the function `t.test` and specify that you want a paired t-test by adding `paired=TRUE` in the function call. For example, for two data sets named `j1` and `j2`, I would type `t.test(j1, j2, paired=TRUE)`.

- (a) Is there significant evidence that the supplement increases weight gain? First analyze the data by examining the mean of the differences in weight gain for each sib pair. Compute the test statistic,  $p$ -value, and state a conclusion.
- (b) Now analyze the data by assuming that the two columns of data represent independent samples from two populations (hogs treated with the supplement and hogs treated with the placebo). Compute the test statistic,  $p$ -value, and state a conclusion.
- (c) Which analysis of the data is correct? Explain briefly.
- (d) Can you explain what features of the data cause the results of the two analyses to differ?

*How funny, I had just realized that all the examples used in this assignment deal with weights, what a coincidence!!!*