STA 100 Fall 2018
Homework 6 - Due Wednesday, Nov 14^{th}

Book Homework

These prompts correspond to "Book Homework" portion of the homework on Canvas. You turn in the answers to these questions online.

1. Note, this is revisiting a problem in Homework 5 (the sorority GPA problem). Group means and sample sizes are found below:

Sorority	A	B	C	D
$\bar{y}_i$	3.22	3.57	2.87	2.98
n_i	10	10	10	10

The value of $MS(within)$ is 0.1278. Let A = group 1, B = group 2, C = group 3, and D = group 4.

(a) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 99% confidence intervals for $(\mu_2 - \mu_1)$, assuming you will make $k = 3$ total confidence intervals.

(b) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 99% confidence intervals for $(\mu_2 - \mu_1)$, assuming you will make $k = 3$ total confidence intervals.

(c) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 99% confidence intervals for $(\mu_2 - \mu_3)$, assuming you will make $k = 3$ total confidence intervals.

(d) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 99% confidence intervals for $(\mu_2 - \mu_3)$, assuming you will make $k = 3$ total confidence intervals.

(e) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 99% confidence intervals for $(\mu_2 - \mu_4)$, assuming you will make $k = 3$ total confidence intervals.

(f) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 99% confidence intervals for $(\mu_2 - \mu_4)$, assuming you will make $k = 3$ total confidence intervals.

(g) Which confidence intervals suggest a significant difference in the means?

(h) Which confidence interval suggests the smallest significant difference between two means?

(i) Interpret the family-wise confidence interval for $(\mu_2 - \mu_3)$ in terms of the problem.

2. A particular type of tree was put into three groups, a control group (with 10 trees), a vitamin deficient group (with 10 trees), and a fertilizer group (with 9 trees). The goal is to see if their average growth is affected. Summary statistics are below:

	Control	Deficient	Fertilizer
$\bar{y}_i$	17.31	11.70	17.84
n_i	10	10	9

The value of $MS(within)$ is 16.8889. Let Control = group 1, Deficient = group 2, and Fertilizer = group 3.

(a) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 90% confidence intervals for $(\mu_1 - \mu_2)$, assuming you will make $k = 3$ total confidence intervals.

(b) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 90% confidence intervals for $(\mu_1 - \mu_2)$, assuming you will make $k = 3$ total confidence intervals.

(c) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 90% confidence intervals for $(\mu_1 - \mu_3)$, assuming you will make $k = 3$ total confidence intervals.

(d) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 90% confidence intervals for $(\mu_1 - \mu_3)$, assuming you will make $k = 3$ total confidence intervals.

(e) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 90% confidence intervals for $(\mu_2 - \mu_3)$, assuming you will make $k = 3$ total confidence intervals.

(f) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 90% confidence intervals for $(\mu_2 - \mu_3)$, assuming you will make $k = 3$ total confidence intervals.

(g) Which confidence intervals suggest a significant difference in the means?

(h) Which confidence interval suggests the largest significant difference between two means?

(i) Interpret the family-wise confidence interval for $(\mu_1 - \mu_2)$ in terms of the problem.

3. Assume a research wants to test if a new drug is better at reducing the average pain score, and they have four groups : a placebo (P), a control (C), a standard drug group (S), and the new drug group (D).

(a) State the null hypothesis in mathematical symbols.

(b) State the alternative hypothesis.

(c) Interpret a Type I error in terms of the problem.

(d) Interpret a Type II error in terms of the problem.

(e) If we rejected the null hypothesis in this case, would we immediately be able to say that all group have significantly different averages from each other?

4. To evaluate the policy of routine vaccination of infants for whooping cough, adverse reactions were monitored in 340 infants who received their first injection of the vaccine. Reactions were noted in 68 of the infants.

(a) Find the lower bound for the 95% confidence interval for the true probability of an adverse reaction to the vaccine.

(b) Find the upper bound for the 95% confidence interval for the true probability of an adverse reaction to the vaccine.

(c) Interpret the confidence interval from (a/b) in terms of the problem.

(d) Does your interval suggest that under 25% of infants had an adverse reaction?

(e) If we made many, many 95% confidence intervals, what percentage would we expect to cover the true proportion?

5. In a study of non-human primates, a sample of 71 orangutans were tested, and 14 were found to be blood type B.

(a) Find the lower bound for the 99% confidence interval for the true proportion of type B orangutans.

(b) Find the upper bound for the 99% confidence interval for the true proportion of type B orangutans.

(c) Interpret your confidence interval from (a/b) in terms of the problem.

(d) Does your interval support the claim that approximately 25% of orangutans have type B blood?

(e) If we made many, many 99% confidence intervals, what percentage would we expect does not cover the true proportion?

(f) What assumptions are required for a confidence interval for a single proportion?

6. Answer the following questions with TRUE or FALSE. It is good practice to explain your answer.

(a) The calculated upper bound of a Wilson-Adjusted confidence interval for p can be above 1.

(b) If we create 4 confidence intervals for pairwise difference in means, we should correct $t_{\alpha/2}$ to $t_{\alpha/4}$ when we use the familywise/simultaneous/Bonferroni correction.

(c) If we fail to reject the null hypothesis of ANOVA, all pairwise (familywise/simultaneous/Bonferroni) confidence intervals comparing group means will contain the value 0.

(d) For a confidence interval for a proportion, we must assume our sample data is normally distributed.

R Homework

These prompts correspond to "R Portion" of the homeworks on Canvas. You use R to find the answers to the following questions, and submit your answers online.

I. On Canvas under Files you will find a dataset `IQ.txt` (or `IQ.csv`). It has two columns, the first of which denotes what major a student is from (A, B, or C). The second is the IQ measured by the Stanford–Binet Intelligence Scales. The goal is to determine if this IQ measure differs between majors.

(a) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 95% confidence intervals for $(\mu_A - \mu_B)$, assuming you will make $k = 3$ total confidence intervals.

(b) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 95% confidence intervals for $(\mu_A - \mu_B)$, assuming you will make $k = 3$ total confidence intervals.

(c) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 95% confidence intervals for $(\mu_A - \mu_C)$, assuming you will make $k = 3$ total confidence intervals.

(d) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 95% confidence intervals for $(\mu_A - \mu_C)$, assuming you will make $k = 3$ total confidence intervals.

(e) Calculate the lower bound for the family-wise (simultaneous, Bonferroni) 95% confidence intervals for $(\mu_B - \mu_C)$, assuming you will make $k = 3$ total confidence intervals.

(f) Calculate the upper bound for the family-wise (simultaneous, Bonferroni) 95% confidence intervals for $(\mu_B - \mu_C)$, assuming you will make $k = 3$ total confidence intervals.

(g) What is the largest significant difference in averages suggested by all of the confidence intervals? (This should be one number).

(h) What is the smallest significant difference in averages suggested by all of the confidence intervals? (This should be one number).

(i) What two majors were the most significantly different according to the confidence intervals?

II. On Canvas under Files you will find a dataset `IQ.txt` (or `IQ.csv`). It has two columns, the first of which denotes what major a student is from (A, B, or C). The second is the IQ measured by the Stanford–Binet Intelligence Scales. The goal is to determine if this IQ measure differs between majors.

(a) Plot the normal probability plot of the errors. Does this data look approximately normally distributed?

(b) Conduct the Shapiro-Wilk's test for normality. Do we fail to reject, or reject the null hypothesis at $\alpha = 0.05$?

(c) Based on the above, were the main assumptions for ANOVA met in this case, or not?

III. In a study of hand hygiene compliance in hospital surgical wards, researchers observed random instances for which hand washing would have been appropriate. In the file `hand.csv` (or `hand.txt`) we have one column, called `result`, where the values are either `washed` (they washed their hands when it was appropriate) or not (they did not wash their hands when it was appropriate). Use this data and R to answer the following questions:

(a) Find the lower bound for 99% confidence interval for the true proportion of people who washed their hands when it was appropriate.

(b) Find the upper bound for 99% confidence interval for the true proportion of people who washed their hands when it was appropriate.

(c) Does the interval in (a/b) suggest that the majority of people was their hands when it is appropriate?

(d) Are the assumptions for a confidence interval for a proportion met in this case?