**HOMEWORK 9**

*2-sample Inference, 1-way ANOVA, and Multiple Comparisons*

Reading: This assignment focuses on content from your textbook, *STAT2: Building Models for a World of Data*, Chapter 5, Sections 1-4, 7. Read these sections of your textbook.

Notes:

* For questions requiring you to use JMP, you must provide a copy of your output at the end of your assignment or embedded within your assignment. No credit will be given if you do not include your output, even if your answer is correct.
* Recall that you can download JMP to your personal computer for free. See the JMP information posted on Canvas. Problems due to not getting JMP working will not allow you to submit your assignment late. Please plan to work ahead and email your instructor questions if needed.
* You must use your own words to answer all homework questions. You cannot copy information from the book or other sources.
* Round all numbers to 2 decimal places unless otherwise specified.
* For all questions requiring calculations, show your work to receive credit.

Answer the following questions from your textbook, *STAT2: Building Models for a World of Data*.

For this assignment, you will use a dataset described in exercise 4.13 – North Carolina births. Use the exercise described in the book on page 191, along with the dataset posted on Canvas, to answer the following questions. Note that the sample size in the dataset provided to you is smaller than the book because rows with missing data have been excluded. You do not need to answer the questions listed in the book.

For all questions, the response variable will be the birthweight in ounces (BirthWeightOz).

1. Suppose we use mothers’ race (MomRace; white, black, Hispanic, and other) as the explanatory variable. Given the MSGroup = 4744.48, SSTotal = 689271.30, and the sample size is 1409, fill in the rest of the 1-way ANOVA table. Show **ALL** calculations, regardless of how simple they are, to get credit.
   1. Fill in the table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F |
| MomRace | 4-1 = 3 | MSGroup \* DFGroup = 4744.48\*3 = 14233.44 | 4744.48 | MSGroup / MSE = 4744.48 / 480.45 = 9.88 |
| Error | 1409-4 = 1405 | 689271.30 – 14233.44 = 675037.86 | SSE/DFE = 675037.86 / 1405 = 480.45 | X |
| Total | 1409-1 = 1408 | 689271.30 | X | X |

* 1. You do not have to find the exact value of the *p*-value, but do you think the *p*-value will be small or large? Why?

P<0.001. The p-value will be small because it is inversely proportional to the F value. Considering that if F>3 is already a big value which results in a small p-value, because F=9.88 for the model, a very small p-value would result.

* 1. Verify your answers to parts a and b using JMP. Make sure to include a copy of your JMP output.

Table

Description automatically generated

* 1. Are the conditions met to conduct the 1-way ANOVA? Include appropriate graphs and explain.

Samples are randomly selected and are independently from the K groups. All 4 populations have distributions that are normal and the sample size is 1409 which is >30 meeting the sample size condition. The four population variances are not equal as the smax/smin = 2.5830432 / 0.7444265 = 3.46984316114 which is not ≤ 2. Considering that only 2 of the 3 assumptions are met, it may not be appropriate for the ANOVA to be used.

Chart

Description automatically generated with low confidence

Chart, histogram

Description automatically generated

* 1. What are the hypothesis test statements for the 1-way ANOVA?

𝐻0: 𝜇1 = 𝜇2 = 𝜇3 = 𝜇4 where 𝜇1 = the mean of black race, 𝜇2 = the mean of Hispanic race, 𝜇3 = mean of other race, 𝜇4 = the mean of white race

𝐻𝐴: at least one equality doesn’t hold

* 1. What is your conclusion for the 1-way ANOVA?

F ratio: 9.8750

p-value: <0.0001

Conclusion: There is overwhelming evidence to suggest that the population mean mother’s race differs depending on the field.

* 1. Which groups have evidence of a difference? Use Tukey comparisons with 95% confidence. Answer by commenting on the Ordered Differences Report from JMP. Round your *p*-values to 4 decimal places.

The groups white and black have a significant difference based on the ordered differences report with a p-value of <0.0001. The groups Hispanic and black have a strong difference based on the ordered differences report with a p-value of 0.0021. The other relationships between the groups do not have significant p-values, so there is no significant difference for any other pairing.

We are 95% confident that the population mean birthweight in ounces for white and black is between 6.79 and 8.31 ounces. We are 95% confident that the population mean birthweight in ounces for hispanic and black is between 6.54 and 8.86 ounces.

Chart

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

1. Suppose we use whether or not the mom is a smoker (Smoke) as the explanatory variable. For this question, you will use JMP for all computations; you do not have to calculate values by hand.
2. Are the conditions met to conduct a two-sample t-test and confidence interval for a difference in means using the *t*-distribution? Explain.

Samples are randomly selected and are independently from the K groups. All 2 populations have distributions that are normal and the sample size is 1409 which is >30 meeting the sample size condition. The two population variances are equal as the smax/smin = 21.970023/21.723102 = 1.011 which is ≤ 2.

Chart, box and whisker chart

Description automatically generated

1. Is there evidence of a difference in mean birthweight for babies whose mother smokes versus not for the population? Show the hypothesis statements, test statistic, *p*-value, and conclusion. Assume the variances are roughly equal (you will need to choose the “Pooled t Test” in JMP).

Ho: μ1=μ2 where μ1=no smoking, μ2 = yes smoking

Ha: μ1≠μ2

*T* test statistic = (-5.06644)^2 = 25.6688142736

*p*-value = <0.0001

Conclusion: There is overwhelming evidence to suggest that the population mean birth weight in ounces is different when the mother is not a smoker and is a smoker.

A picture containing chart

Description automatically generated

1. Interpret the 95% confidence interval for the difference in means.

CI=[-11.624, -5.135]

We are 95% confident the population mean birth weight in ounces for mothers who do not smoke is between 11.624 lower and 5.135 higher than mothers who smoke.