**Fall 2023 Midterm Exam**

**for DS 6371**

IMPORTANT: By taking and submitting this exam, you are promising on your honor that you did not communicate about this test with anyone that could have helped on this test (except for Bivin Sadler) from 11am CST Saturday, October 14, 2023 to 11:59pm CST Sunday October 15, 2023. Please simply put your initials here to verify that this statement was both read and agreed to: \_ML\_

Thank you! Now rock this test!

Please make your responses organized and only include output that is used to address the question. If you use ChatGPT, do not copy and paste responses directly from a generative AI or any other source. This is plagiarism and will result in significant points off. Sorry to get serious there for a second…. Just have to be clear. 😊 Ace the test and have fun!

Give results that have been log transformed. Maybe ChatGPT output.

1. (6 points) What is the symbol for:
   1. Population Standard Deviation: δ (sigma)
   2. Sample Standard Deviation: *s*
   3. Population Mean: μ (mu)
   4. Sample Mean: X̅ (x-bar)
   5. Use some of the symbols above to fill in the blanks:
      1. \_\_\_X̅\_\_\_ estimates \_\_μ\_\_\_ and
      2. \_\_\_*s*\_\_\_\_ estimates \_δ\_\_\_\_
2. (9 points) Consider the scenario below:

**Scenario: Effect of a Memory-Boosting Supplement on Students**

**Objective:** To test the effectiveness of a new memory-boosting supplement on the memory retention abilities of students.

**Procedure:**

1. **Selection of Participants:**
   * Randomly select 30 SMU students from a list of student ID numbers.
2. **Baseline Test:**
   * Administer a memory test to all the students to assess their baseline memory capabilities. This test involves memorizing a list of 50 words in 5 minutes and then recalling as many words as possible after a 30-minute break.
   * Record the number of words each student recalls.
3. **Intervention:**
   * Provide the selected students with the memory-boosting supplement. Instruct them to take the supplement daily for a month.
4. **Post-Intervention Test:**
   * After a month of taking the supplement, administer the same memory test to the students.
   * Record the number of words each student recalls.
5. Which test would be appropriate to obtain the objective? (More than one test may be appropriate.)
   * **Paired Sample t-test** because I will assume all the paired instances are normally distributed, I want to compare the means of the same group but before and after they were given the supplement (the intervention), if the p-value is less than α, I will reject the null hypothesis to believe that the supplement does affect the memory.
   * **Wilcoxon Signed Rank Test:** I will assume normality was not met, if the p-value is below the significance level, I will reject the null hypothesis.
6. Write the Ho and Ha of this test for this scenario.

* **Ho**(Null Hypothesis): The memory-boosting supplement has no effect on memory retention.
* **Ha**(Alternative Hypothesis): The memory-boosting supplement has an effect on memory retention

1. Indicate if the result could be generalized to the larger population.

* The study was a relatively small sample size (30 students) just from SMU, no other universities were involved in the study. It is challenging to state if the findings would apply to everyone else. There were no details as to how this group of students were selected or if the students granted permission properly, which adds further uncertainty to the study’s generalizability.
* The Wilcoxon Signed-Rank Test gives us a hint about the supplement's potential benefits for our group of students. However, because our study only involves a small group and there might be other factors playing a role, we can't be totally sure these findings would apply to everyone. Even though the results seem promising, it's best to be cautious and not jump to conclusions until we've tested this on a bigger and more diverse group.

1. (12 points) Consider the scenario below:

**Scenario: Effect of Two Different Diets on Weight Loss**

**Objective:** To compare the effectiveness of Diet A (a low-carb diet) and Diet B (a low-fat diet) on weight loss over a period of three months. Specifically, to test if the median weight loss of Diet A is different than the median weight loss of Diet B.

**Procedure:**

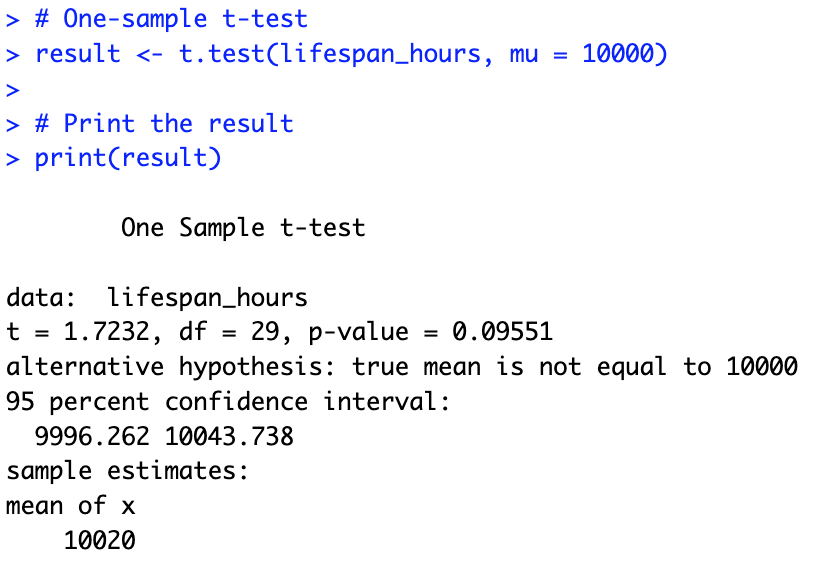
1. **Random Selection of Participants:**
   * Run a commercial and recruit 50 volunteers to participate in the study.
   * Ensure that these participants do not have any medical conditions or are on medications that might significantly affect weight.
2. **Random Assignment to Diets:**
   * Once the 50 participants are selected, for each participant, flip a coin and if it is heads, assign the subject to Diet A and if it is tails assign the subject to Diet B.
3. **Intervention:**
   * The Diet A group will receive a low carb diet for three months.
   * The Diet B group will receive a low fat diet for three months.
   * Both groups receive guidance on maintaining a similar level of physical activity.
4. **Measurement:**
   * Record the weight of each participant at the start and end of the three months.
   * Calculate the weight loss for each participant.
5. Which test would be appropriate to obtain the objective? (More than one test may be appropriate.)
6. Write the Ho and Ha of this test for this scenario.
7. Indicate if the result could be generalized to the larger population.
8. Assume the pvalue of the test was less than alpha. Could we conclude that the Diets caused the difference?
9. (15 points) Consider the scenario below as well as the results of the test:

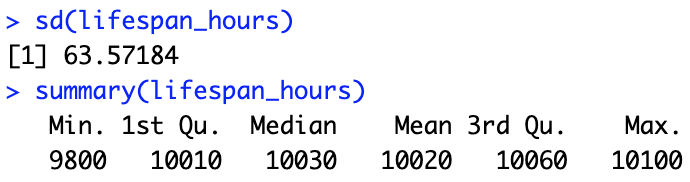
**Scenario: Testing the Efficiency of a New Light Bulb**

**Objective:** A manufacturer claims that their new energy-saving light bulb lasts an average of 10,000 hours before burning out. A consumer protection agency wants to verify this claim.

**Procedure:**

1. **Selection of Light Bulbs:**
   * Randomly select a sample of 30 light bulbs from the manufacturer's production line.
2. **Testing:**
   * Use the light bulbs continuously and record the number of hours each bulb lasts before burning out. The data are stored in the R array *lifespan\_hours*.
3. **One-Sample t-test:**
   * The consumer protection agency will perform a one-sample t-test to compare the average lifespan of the sample to the manufacturer's claimed average of 10,000 hours.





1. Write all 6 steps of this test (you may assume the assumptions are met). Assume alpha is .05.
   * **Step 1 – Hypothesis:**

**Ho:** The average lifespan of the light bulbs is that of manufacturer’s claim of 10,000 hours.

μ = 10,000 hours

**Ha:** The average lifespan of the light bulbs is that of manufacturer’s claim not to be that of 10,000 hours.

μ ≠ 10,000 hours

* + **Step 2 – Identification of Critical Values:**

With a t=1.7232, df=29, α=0.05, for a two-tailed test (with each tail critical values {0.05/2 = 0.025} of ±2.045 for a 95% confidence level), T-value with a greater than 2.045 magnitude will lead me to reject the null hypothesis at the 0.05 significance level.

* + **Step 3 – Value of Test Statistic:**

t = 1.7232

* + **Step 4 – Give p-value:**

p-value = 0.09551

* + **Step 5 – Decision:**

With a p-value (0.09551) is a greater than the significance **level** α=0.05, I will fail to reject **Ho.**

* + **Step 6 – Conclusion:**

Using the t-test with a 5% significance level, I cannot confidently state the light bulbs last vastly different than the manufacturer’s claim of 10,000 hours. The data shows the light bulbs last between 9,996.262 and 10,043.738 hours on average, which includes the realm of the manufacturer’s claim of 10,000 hours. So, my test could not dispute the manufacturer's claim of the light bulbs’ lifespan of 10,000.

1. Write a scope of inference.

**Scope of Inference:**

Given the results of my one-sample t-test, my conclusion is that the mean of the light bulbs is restricted to this obtained batch. My test results did not provide compelling evidence to contradict the manufacturer’s claim of 10,000 hours for this obtained batch. However, the caveat is to tread lightly when applying the findings to other batches during a distinct timeframe. My assumptions were that the light bulbs tested gave me a good idea of the all the light bulbs and that I was thinking every light bulb was tested the same way.

1. Now assume alpha is .1. Rewrite you step 5 and 6 below.

**Step 5 – Decision:**

The same p-value is 0.09551, which is less than the new α=0.1, I would reject the Ho.

**Step 6 – Conclusion:**

With α=0.1, I will be inclined to think the light bulbs do not last 10,000 hours; however, my results did state they could last anywhere between 9996.262 to 10,043.738 hours, while there is a small difference, the lifespan of the bulbs do come close to the manufacturer’s claim of the 10,000 hours.

1. We know a 90% confidence interval will always “agree” with the alpha = .1 two-sided test. Calculate and interpret this 90% confidence interval for . Show you work.

|  |
| --- |
| A close-up of a math problem |
|  |
| Based on the 90% interval, I estimate the true average lifespan (μ) of the light bulbs’ lifespan, from this sample will be within the range of 9998.2 hours and 10041.8 hours. With a confidence level that 90% of the time, mu (μ) will be within this range. |

1. Your client would like to be sure and detect a difference if the true mean lifespan of the bulbs is 9900? If the true lifespan was 9900 hours, what would be the power of the this test?

|  |
| --- |
|  |

The power is the probability the t-statistic under μa falls within the critical region calculated in the tcritical, which is approximately 8.515. Thus, the average lifespan of 9,900 hours, the power of this test would be very close to 100%, which I will then be inclined to favor the alternative.

1. (20 points) The dataset *airline\_delay1* has a random sample of flights from July 2023 from two different airlines (Delta and Spirit) and how long the flight’s departure was delayed.
2. Conduct a two-sample t-test for the difference of means if the assumptions are met or are at least reasonably robust to an assumption.

The assumptions:

* **Independence of observations**: It was a given assumption.
* **Normality**: If the p-value is greater than 0.05, the assumption of the normality is not violated.
  + Shapiro-Wilk Test p-value for Delta: 7.24 x 10-13
  + Shapiro-Wilk Test p-value for Spirit: 2.34 x 10-11

Normality, per these p-values, was violated by both airlines. Relatively per the large size of the data set, the t-test can be robust to this violation. On the

On the contrary, if the size is small, the violation of the normality assumption will have to a consideration. As an alternative, the Wilcoxon Rank-Sum Test might be suggested to be used as a non-parametric test.

* **Equality of variances**: A p-value greater than 0.05 may be an indication of the variances are equal across the two groups.
  + Levene's Test p-value: 2.42 x 10-05
  + With this p-value, the assumption of equal variances is also violated.
  + As an alternative, the Welch’s t-test might be suggested to be used as a variant of the independent two-sample t-test that does not assume equal variances the two airlines.
* Both assumptions of normality and equality of variances were violated, therefore, the standard two-sample t-test for the difference of means is not appropriate for this data set.

The client would like an inference on the mean if the t-test is feasible.

Address each assumption of the two-sample t-test and conduct the test if you believe, for each assumption, the test either does not violate that assumptions or the test is robust to that assumptions.

Addressing the assumptions in the Welch’s test:

* + **Independence of observations:** It was a given assumption as both airlines’ data are set to be independent.
  + **Normality:** The Welch’s t-test is robust enough to meet this assumption with a p-value of 1.065 x 10-7, and with confidence in my inference about the difference on the means of the two airlines.
  + **Equality:** The Welch’s t-test is specifically designed to handle unequal variances, which makes appropriate to conduct the t-test with this data set.

Welch Two Sample t-test

data: airline\_data$Delta and airline\_data$Spirit

t = -5.4964, df = 220.86, p-value = 1.065e-07

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-3.885658 -1.834622

sample estimates:

mean of x mean of y

2.818182 5.678322

Given the violations of the two-sample t-test, that t-test is not feasible. As an alternative t-test, I conducted a Welch’s t-test, which allow me to make an inference about the differences on the means of the two airlines. There is a statistically significant difference on the means of the two airlines, despite the violation of normality assumption.

If you do not believe an assumption is met and the test is not robust to that assumption, provide succinct and clear evidence as to back up your thoughts. With respect to the last assumption, you may assume the data are independent.

Assumptions for the two-sample t-test:

**Independence of observation:** I assumed the data set points were independent.

**Normality:**

* + **Evidence –** The airlines did not follow the normal distribution, with

p-values for Delta: 7.24 x 10-13 and Spirit: 2.34 x 10-11 from the Shapiro Wilk test, proved they did not meet this assumption. The Q-Q plots further proof

this violation. The data points are not aligning with the red line, which it is an indication separation from normality.

* + - **Q-Q plot for Delta:**

For the most part, a good set of points are aligned to the redline;

however, the left tail end demonstrates the deviation from the line.

* **QQ plot for Spirit:**

Also, for the most part the set of points are aligned to the redline;

however, also on the left tail end points to skewness, deviating from

the normality.

A screenshot of a graph

Description automatically generated

**Robustness –** With the Welch’s t-test, the normality of the data set was handled.

**Equality of Variances:**

* + **Evidence** – The Levene’s test demonstrated the spread of variances of the two airlines, with a p-value: 2.42 x 10-05 less than 0.05, which makes a violation

of the assumption. The box plots further proof this violation.

* Here in the boxplot, the box for Delta is shorter, which indicates there is less variability in the middle for 50% of the delays.
* The box for Spirit is bigger, meaning there is more variability in the middle 50% of the delays.
* The whickers for Delta are shorter, which also means less variability outside the 50% of the data.
* The whiskers for Spirit are longer, meaning more variability outside the 50%.
* The circles are the outliers, which Delta has less than Spirit; meaning there were some exceptionally long delays.

|  |
| --- |
| A diagram of a box plot  Description automatically generated |

**Robustness –** Since the Welch’s t-test was an appropriate t-test to conduct in this situation, because it is specifically designed to be robust when this assumption is violated, which was in this case.

1. If you do not believe the assumptions are met and the test is not robust to any violated assumptions, then conduct an alternative test that tests for difference in center (mean or median.) ***If you believed the assumptions were not violated or that the test was robust to any assumptions that were violated, then there is nothing to do here***

***.***

I believe from the standard two-sample t-test, there was a violation of normality and with the equality of variances. I will then conduct Wilcoxon Rank-Sum Test to handle the normality of the data. I will also reiterate Welch’s t-test for the handling of the equality of the variances.

Addressing assumptions in the Wilcoxon Rank-Sum test:

* **Independence of observations:** It is a given assumption that these observations are independent of each other under the Wilcoxon Rank-Sum test.
* **Normality:** This non-parametric test does not assume the data follows a normal distribution. This makes this test robust to deviations from normality. No need to evaluate or assume normality.
* **Equality of Variances:** The spread of the variances is also not required by the Wilcoxon Rank-Sum test. The test compares the ranks of the data and does not make assumptions of the variances of the two groups. Therefore, also, no need to evaluate or assume equal variances.

Wilcoxon rank sum test with continuity correction

data: delta and spirit

W = 5181, p-value = 5.12e-08

alternative hypothesis: true location shift is not equal to 0

Addressing the assumptions in the Welch’s test:

* + **Independence of observations:** It was a given assumption as both airlines’ data are set to be independent.
  + **Normality:** The Welch’s t-test is robust enough to meet this assumption with a p-value of 1.065 x 10-7, and with confidence in my inference about the difference on the means of the two airlines.
  + **Equality:** The Welch’s t-test is specifically designed to handle unequal variances, which makes appropriate to conduct the t-test with this data set.

Welch Two Sample t-test

data: airline\_data$Delta and airline\_data$Spirit

t = -5.4964, df = 220.86, p-value = 1.065e-07

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-3.885658 -1.834622

sample estimates:

mean of x mean of y

2.818182 5.678322

1. You either conducted a test in a) or b) above, create a “complete analysis” around that test. Remember a “complete analysis” is: i) State the Problem, ii) address the assumptions, iii) perform the test paying careful attention to step 6: the conclusion, provide a scope of inference. (You may copy and paste the 6 step test for a) or b) above, you don’t have to redo the test.)
   * + - 1. **State the problem:**

**Ho:** There are no significant differences between the two airlines.

**Ha:** There are significant differences between the two airlines.

* + - * 1. **Address** the assumptions in the Wilcoxon Rank-Sum test:
* Independence of observations: It is a given assumption that these observations are independent of each other under the Wilcoxon Rank-Sum test.
* Normality: This non-parametric test does not assume the data follows a normal distribution. This makes this test robust to deviations from normality. No need to evaluate or assume normality.
* Equality of Variances: The spread of the variances is also not required by the Wilcoxon Rank-Sum test. The test compares the ranks of the data and does not make assumptions of the variances of the two groups. Therefore, also, there is no need to evaluate or assume equal variances.

Wilcoxon rank sum test with continuity correction

data: delta and spirit

W = 5181, p-value = 5.12e-08

alternative hypothesis: true location shift is not equal to 0

**Step 1 – Define the problem:**

**Ho:** There are no significant differences between the two airlines.

**Ha:** There are significant differences between the two airlines.

**Step 2 – Identification of Critical Values:**

With α = 0.05 for a two-tailed test, the critical values are ±1.96 (0.1/2 = 0.025 for a 95% confidence level). A t-value with a magnitude greater than 1.96 will lead to rejecting the null hypothesis at the 0.05 significance level.

**Step 3 – Value of Test Statistic:**

t-value: t = 1.7232

**Step 4 – Give p-value:**

p-value = 0.09551

**Step 5 – Decision:**

With a new significance level α=0.05 and the p-value of 0.09551, which is greater than the α, I fail to reject the null hypothesis (Ho).

**Step 6 – Conclusion:**

There was no evidence that to suggest significant differences between the two airlines. The data does not support the claim the two airlines significantly differ in their delays. I base my conclusion on the p-value (0.09551) being greater than the level of significance of 0.05

**Scope of Inference:**

The analysis did not produce enough evidence to conclude the presence of significant disparities in the performance metric between the two airlines. Within the scope of this study and the available dataset, there was no statistically significance distinction in the performance metric under scrutiny.

1. (20 points) The dataset *airline\_delay2* has a random sample of flights from 5 airlines: Southwest, American Alaska, Lufthansa and Fiji; these names are stored in a column called *airline*. It also has the departure delay from each of these flights in a column called *delay*.
2. Is there significant evidence to suggest that any of the airlines have different mean departure delays from one another? ***You may assume any assumptions you need here are met … you do not need to address the assumptions here. For this question, you just need the 6 step test.***

**Performing a Wilcoxon Rank-Sum Test:**

**Step 1 – Define the problem”**

**Ho:** There are no significant differences between the means departure delays among the five airlines.

**Ha:** There is at least one difference between the means departure delays between the five airlines.

**Step 2 – Level of Significance:**

α = 0.05

**Step 3 – Value of test statistic:**

**t = 94.5**

**Step 4 – Perform the one-way ANOVA test**

summary(result\_anova)

Df Sum Sq Mean Sq F value Pr(>F)

American 1 3 2.76 0.052 0.8202

Alaska 1 249 249.18 4.681 0.0313 \*

Luftanza 1 0 0.40 0.008 0.9306

Fiji 1 6 5.99 0.113 0.7375

Residuals 296 15756 53.23

**Step 5 – Decision**

The ANOVA results indicate there is a significant difference in departure delays between Alaska and the rest of the airlines, but no significant differences among American, Luftanza, and Fiji.

**Step 6 – Conclusion**

Based on the one-way ANOVA, I conclude there is no significance difference between American, Luftanza, and Fiji when considering a significance level of 0.05. Alaska stands out alone on having a significant difference among the other airlines.

1. If there was not significant evidence of a difference then simply say that there was no evidence of a significant difference and go on to the next problem.

If there was significant evidence of a difference in the mean departure delay for at least one pair of airlines,

1. test which pair or pairs of airlines have significant evidence of being different?

**Alpha is 0.05.**

**Southwest vs. American**

t-statistic= 34634

p-value: 5.467e-07

The p-value less than alpha, suggesting strong evidence there is a significant difference in departure delay times between Southwest and American airlines.

**Southwest vs. Alaska**

t-statistic= 34977

p-value: 1.239e-06

The p-value is less than alpha, indicating strong evidence of a significant difference in departure delay times between Southwest and Alaska airlines.

**Southwest vs. Luftanza**

t-statistic= 26651

p-value: < 2.2e-16

The p-value is less than alpha, indicating a highly significant difference in departure delay times between Southwest and Luftanza airlines.

**Southwest vs. Fiji**

t-statistic= 25952

p-value: < 2.2e-16

The p-value is less than alpha, indicating a highly significant difference in departure delay times between Southwest and Fiji airlines.

**American vs. Luftanza**

t-statistic= 35406

p-value: 3.368e-06

The p-value is less than alpha, suggesting strong evidence of a significant difference in departure delay times between American and Luftanza airlines.

**American vs. Fiji**

t-statistic= 34723

p-value: 6.824e-07

The p-value is less than alpha, indicating strong evidence of a significant difference in departure delay times between American and Fiji airlines.

**Alaska vs. Luftanza**

t-statistic= 34724

p-value: 6.76e-07

The p-value is less than alpha, suggesting strong evidence of a significant difference in departure delay times between Alaska and Luftanza airlines.

**Alaska vs. Fiji**

t-statistic= 33901

p-value: 8.652e-08

The p-value is less than alpha, indicating strong evidence of a significant difference in departure delay times between Alaska and Fiji airlines.

1. if the FAA (Federal Airline Administration) allows an airline to market themselves as having a smaller average departure delay if there is strong evidence that their mean departure delay is 2 minutes or more less than another airline, are there any airlines that can make this claim and if so, against which competitor(s)?

Mean Difference = (t \* (n1 + n2)) / 2

t = t-statistic from Wilcoxon rank-sum test

* **Southwest vs. American:** Southwest has a significantly smaller average departure delay than American by approximately 16.29 minutes.
* **Southwest vs. Alaska:** Southwest has a significantly smaller average departure delay than Alaska by approximately 13.97 minutes.
* **Southwest vs. Luftanza:** Southwest has a significantly smaller average departure delay than Luftanza by approximately 29.27 minutes.
* **Southwest vs. Fiji:** Southwest has a significantly smaller average departure delay than Fiji by approximately 30.63 minutes.
* **American vs. Luftanza:** American has a significantly smaller average departure delay than Luftanza by approximately 12.98 minutes.
* **American vs. Fiji:** American has a significantly smaller average departure delay than Fiji by approximately 13.77 minutes.

1. (18 points) Using the airline\_delay2 dataset again, note that American and Alaska airlines are affiliates of one another. We would like to use a contrast to test the claim that the average departure delay of American and Alaska airlines is different than that of the non-US airlines Lufthansa and Fiji airlines. As always, we would like to conduct the most powerful test possible. You only need to show the Ho and Ha, the test statistic, pvalue, and the conclusion including a confidence interval. Please also include your code.

**Ho:** The average departure delay for US airlines is the same as of those non-US airlines.

**Ha:** The average departure delay for US airlines is not the same as of those non-US airlines..

**t-statistic:**

f-value for the contrast "US vs Non-US" is 16.88

p-value: <0001

**p-value:**

p-value for the contrast "US vs Non-US" is < .0001

**Confidence Interval for least squares means (99% confidence level):**

US Airlines:

Alaska: (8.696146,10.998206)

American: (8.453621,10.755681).

Southwest: (6.247641,8.549701)

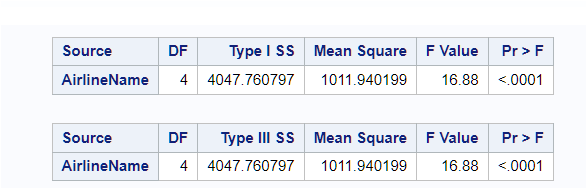
Non-US Airlines:

Fiji: (10.732691,13.034751)

Lufthansa: (10.579867,12.881927)

**SAS code:**

proc glm data=airlines\_with\_id;  
class AirlineName;  
model DepartureDelay = AirlineName;  
lsmeans AirlineName / cl;   
estimate 'US vs Non-US' AirlineName 1 1 1 -1 -1 / cl;   
run;

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generatedA screen shot of a graph

Description automatically generated

