DS 6371 Midterm Analysis Question

|  |  |
| --- | --- |
|  |  |

Answer the following questions with the Global Warming Data. You will need to show the 6 steps and BONUS interpret the results in non-statistical terms as well. (You must show BOTH the statistical inferences and the non-statistical to get the Bonus.) Please have your answers in Bold.

*While based on real events, the data and scenarios below are made up … but could have happened.*

On June 14th 2019, Greenland lost more ice than at any other time in recorded history. In order to assess the level of concern around the world, a random sample of 188 total citizens from 6 different countries was taken and each citizen was asked to rank the seriousness of the potential issue of global warming. They were asked to use a 0 – 100 ranking system and were given the guidelines below:

***0*** *= not serious as all … global warming is the least of our concerns … even if it is real, it does not pose and economic or environmental threat in the least.*

***50*** *= it should constantly monitored and studied but there is no need to make any drastic changes to energy sources (aka fossil fuels) , or environmental choices (deforestation, beef production, airline travel) that are thought to be influencing the climate.*

***100*** *= needs urgent attention immediately … left unaddressed global warming will soon become an imminent threat to the survival of most life on Earth.*

The countries that were randomly sampled were the United States, China, Germany, South Africa, Sri Lanka, and Russia. In addition to the rating and country of citizenship, subjects in the study were also asked to record their gender and to rate their wealth. Their choices were, Low, Moderate, and Wealthy. The data can be found in the GWarming.csv file.

1. The first question the researchers were interested in was if there was a significant difference between men and women on rating of the seriousness of global warming. If so, by how much? Quantify any uncertainty with a confidence interval.

2. Researchers wanted to investigate if there was a significant difference between the rating of any of the countries. Test the claim that there is no difference in the ratings of any of the countries.

**YOU WILL ONLY NEED TO COMPLETE 3a OR 3b.**

3a. If there was significant evidence of a difference above, provide additional analysis as to which pairs of countries have different ratings of the seriousness of global warming. Although you should test all the pairs, only report the pair of countries that have the biggest difference in rating and provide a confidence interval to quantify your uncertainty.

3b. If there was not significant evidence of a difference found in the test in question 2, then perform the analysis again except this time only consider those that consider themselves wealthy. That is, of those who consider themselves wealthy, is there evidence of a significant difference in the mean or median rating of the seriousness of global warming between any of the 6 countries?

4. The researchers were also interested in investigating the difference in rating of the seriousness of global warming between communist and non-communist countries. Use a contrast to test for this difference. You may consider China and Russia to be communist and U.S., Germany, Sri Lanka and South Africa to not be communist. Again, quantify your uncertainty with confidence intervals where appropriate.

Multiple Choice Questions

**Q 1**

Question 1

Suppose the following hypotheses are tested:

H0: *μ*=1.4

Ha: *μ*≠1.4

Alpha (*α*) = 0.05

What is the definition of a Type I Error in this context?

Select one:

**A. A Type I Error is failing to reject the null hypothesis if the population mean is 1.5.**

**B. A Type I Error is rejecting the null hypothesis if the population mean is 1.4.**

**C. A Type II Error is failing to reject the null if the population mean is not 1.4.**

**D. A Type II Error is rejecting the null hypothesis if the population mean is not 1.4.**

**Q 2**

Question 2

Suppose a hypothesis test is a performed and p-value of 0.023 is obtained. Which of the following is the correct interpretation of the p-value?

Select one:

**A. There is a 2.3% chance that the null hypothesis is true.**

**B. There is a 2.3% chance that the alternative hypothesis is true.**

**C. There is a 2.3% chance that one would get a test statistic as extreme or more extreme than the observed**

**value by chance alone if the null is true.**

**D. There is a 2.3% chance that one would get a test statistic as extreme or more extreme than the observed**

**value by chance alone if the alternative is true.**

**E. A and C are True**

**F. B and D are True**

**G. None are True**

**Q 3**

Question 3

The signed rank test is an alternative to the paired samples t-test. True or False?

Select one:

**A. True**

**B. False**

**Q 4**

Question 4

A 95% confidence interval for a mean *μ* is found to be [22.3, 25.6]. Which is a correct interpretation of this confidence interval?

Select one:

**A. There is a 95% chance that the population mean***μ***is between 22.3 and 25.6.**

**B. The population mean***μ***is likely between 22.3 and 25.6. The procedure used gives a confidence interval containing the population mean***μ***for 95% of samples.**

**C. There is a 95% chance that the sample mean***x̄***is between 22.3 and 25.6.**

**D. The sample mean***x̄***is likely between 22.3 and 25.6. The procedure used gives a confidence interval containing the sample mean***x̄***for 95% of samples.**

**Q 5**

Question 5

All else held constant, increasing the effect size reduces, increasing the sample size **AND** increasing the significance level **ALL** increase the power of the test. True or False?

Select one:

**A. TRUE**

**B. False**

**Q 6**

Question 6

As long as subjects are randomly assigned to treatment groups, inference from hypothesis tests can be generalized beyond the sample to the population as a whole. True or False?

Select one:

**A. True**

**B. False**

**Q 7**

Question 7

Suppose a data set of continuous numbers consists of most of the numbers clustered together with a few outliers much higher than the others (and no other outliers). Choose the best answer that describes the skewness of the data.

Select one:

**A. The data set is skewed to the right.**

**B. The data set is skewed to the left.**

**C. The data set is not skewed. (It is symmetric.)**

**D. There is not enough information to determine skewness.**

**Q 8**

Question 8

A researcher for Car and Driver magazine was interested in if there was a difference between the MPGs (miles per gallon) of hybrid cars and their manufacturer. In order to test this, the magazine gained access to 3 Toyota Corolla Hybrid, 4 Ford Fusions and 5 Chevy Malibu Hybrids and recorded the MPGs from each of these cars. From a previous study, there is reason to believe that the distributions of mpgs from these cars are right skewed and that the standard deviations are similar. What is the best test to test for a difference in the centers between any pair of these distribution?

Select one:

**A. Signed Rank Test**

**B. Rank Sum Test**

**C. Welch's T Test**

**D. Kruskal Wallis Test**

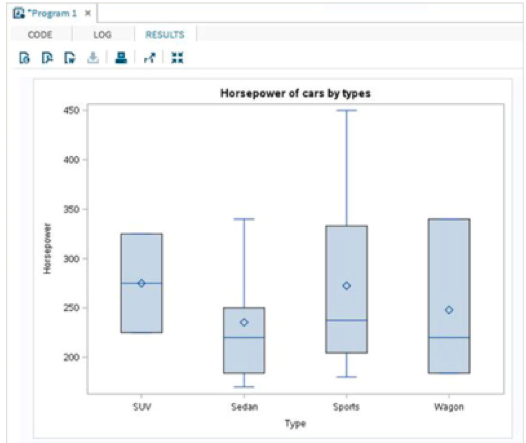
**E. Brown and Forsythe Test**

**F. Pooled T Test**

**G. 1-way ANOVA**

**Q 9**

Question 9



The box plots for the Sedan, Sports and Wagon types above are consistent with:

**A. Left Skewed Data**

**B. Symmetric Data**

**C. Uniformly Distributed Data**

**D. Right Skewed Data**