MSDS 6371 FALL 2019 Midterm

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 II 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.**

Question 2

Suppose a hypothesis test is a performed and p-value of 0.523 is obtained. Which of the following is the correct interpretation of the p-value?

Select one:

**A. There is a 52.3% chance that the null hypothesis is true.**

**B. There is a 52.3% chance that the alternative hypothesis is true.**

**C. There is a 52.3% 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 52.3% 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**

Question 3

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

Select one:

**A. True**

**B. False**

Question 4

A 95% confidence interval of the difference of means  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 both** and  **are each between 22.3 and 25.6.**

**B. There is a 95% chance that both** or  **is between 22.3 and 25.6.**

**C. We are 95% confident that** is larger than

**D**. **We are 95% confident that** is larger than

**E. There is a 95% chance that** *x̄1 - x̄2*  **is between 22.3 and 25.6.**

**F. 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.**

Question 5

All else held constant, which increases the power of a one sample t-test?

Select one (the most appropriate answer):

**A. Increasing the Effect Size only**

**B. Increasing the Sample Size only**

**C. Increasing the significance level only**

**D. A and B will both increase the power of the test.**

**E. A and C will both increase the power of the test.**

**F. B and C will both increase the power of the test.**

**G. All will increase the power of the test.**

**H. None will increase the power of the test.**

Question 6

As long as subjects have been randomly selected from the population, a casual inference can be inferred. True or False.

Select one:

**A. True**

**B. False**

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.**

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 very 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 distributions?

Select one:

**A. Signed Rank Test**

**B. Rank Sum Test**

**C. Welch's T Test**

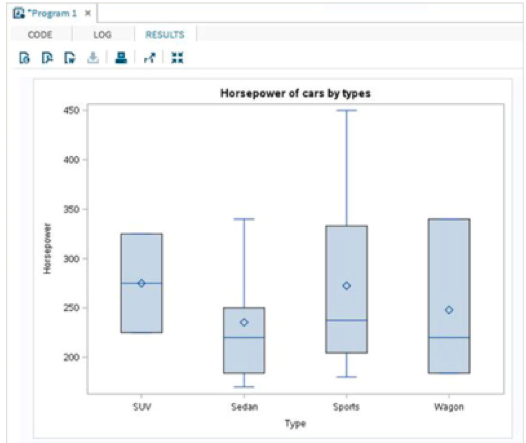
**D. Kruskal Wallis Test**

**E. Brown and Forsythe Test**

**F. Pooled T Test**

**G. 1-way ANOVA**

Question 9



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

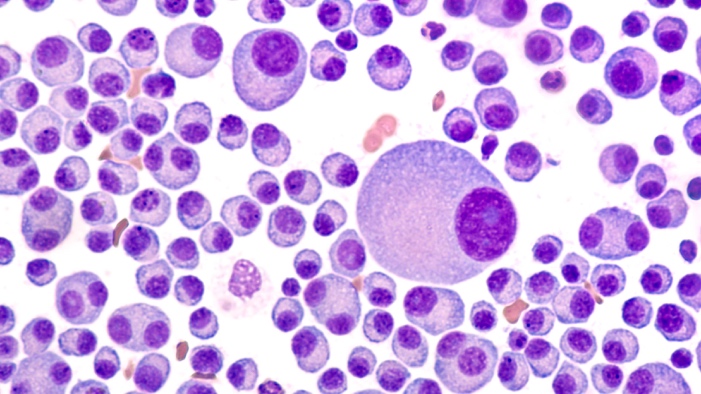
**A. Left Skewed Data**

**B. Symmetric Data**

**C. Uniformly Distributed Data**

**D. Right Skewed Data**

MSDS 6371 Analysis Question



Multiple Myeloma is a relatively rare form of blood cancer in which plasma cells reproduce uncontrollably in the bone marrow. This in turn crowds out read and white blood cells which leads to complications in the patients overall health. The cancer can be monitored by looking at a patient’s *Lambda and Kappa* protein counts. Higher counts indicate more / worse disease. There have been several drugs that have been developed to fight this type of cancer which is measured in the reduction of the *Lamda* and *Kappa* protein levels. Four of these drugs are: Revlimid, Velcade, Dex and Pomolyst.

Through prior studies, all of the drugs have been shown to decrease the Lamda and Kappa proteins in myeloma patients. This study is focused on comparing the four treatments to see which ones are more effective in reducing the Lamda and Kappa protein levels.

To test this, the researchers went to MD Anderson hospital in Houston and randomly selected 35 patient’s records that had been taking one of the drugs above. To be clear, there were 140 patient’s records in the study, 35 patients that took Revlimid, 35 that took Velcade, 35 that took Dex, and 35 that took Pomolyst. The researcher recorded the percentage drop in Lamda protein during the standard 15 month treatment of taking the drug. This data is recorded in the file myeloma.csv.

Histograms and Box Plots of the percent drops of **Lambda** protein levels are provided below.

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

|  |
| --- |
|  |

***Make sure and provide your SAS or R code for all questions below.***

1. The principal question of interest for the researchers was if there were significant differences in the mean or median percent drop of Lambda protein between the drugs and if there are, and estimate of the magnitude of the difference(s). Provide analysis that will best answer the principle question of interest above. Please address all assumptions needed to conduct your analysis and provide a scope of inference with your findings. Assume the researchers are interested in maintaining a family wise error rate of 5% (alpha\_family = .05).
2. A second question of interest centered on comparing Type A drugs to Type B drugs. Turns out that Revlimid and Velcade are Type A drugs while Dex and Pomolyst are both Type B drugs. Test the claim that the Type A drugs have a greater mean percent drop of Lambda protein than the Type B drugs by comparing the mean percent drop of Revlimid and Velcade to the mean percent drop of Dex and Pomolyst using a contrast. For this question you may assume all the assumptions are met to run a contrast but do show all 6 steps of the hypothesis test (t test) and provide a 95% confidence interval for the difference as well. Test at the alpha = .01 level of significance.
3. BONUS (5 pts) **KAPPA** PROTEIN ANALYSIS: The researchers thought that Revlimid would be more effective than the Velcade in reducing the **Kappa** protein levels. One problem they encountered was that the Kappa percent reduction was missing for many of the Revlimid and Velcade patients (as can be seen in the data set). Conduct a complete analysis (state the problem, address the assumptions, conduct the 6 step test and provide a scope of inference) that will test the claim that the Revlimid has a greater mean or median percent drop in **Kappa** protein levels than Velcade. Use an alpha = .01 level of significance and provide confidence intervals with your analysis.

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

