**Part 1. (5 Points)**

### A paint manufacturing company claims the following in one of its advertisements.

### “*We have the fastest drying paint! In a designed, randomized experiment, our paint dried faster than our competitors paint.”*

### Below is the data from the experiment:

### 

|  |  |  |  |
| --- | --- | --- | --- |
| Dry time in minutes for Manufacturer’s Paint | 33.4 | 29.2 | 35.7 |
| Dry time in minutes for Competition’s Paint | 33.5 | 35.8 | 29.4 |

### (1 point) What are the average dry times for each company’s paint?

The average dry time for the manufacturer’s paint is 32.767 minutes. The average dry time for the competitor’s paint is 32.9 minutes.

### (2 points) Is the manufacturers claim truthful? Either way, is the advertisement misleading? Why or why not?

The manufacture’s claim is not truthful and is misleading because the average time for the paints of both the manufacturer and competitor is almost the same except the average time varies only by a couple seconds between both paint companies.

### (2 points) Suppose the company advertising the faster drying paint performed the experiment themselves. Why could this be a potential problem?

If the company advertising the faster drying paint performed the experiment themselves, this could be a potential problem because in one way, it could lead to developing a biased solution because of the company’s interest to develop faster drying paints compared to their competitors. Also, the company picked only one competitor out of many other potential competitors, and they chose to develop faster drying paints compared to their other competitor. Hence, this could be potentially misleading if the company perform the experiment themselves to produce a faster drying paint.

### Part 2. (18 Points)

### On Canvas, you’ll find the R script, One\_Variable\_Display\_and\_Summary\_Stats.R and the ST314 student survey dataset, st314\_student\_survey.csv. You’ll use both of these to explore one categorical and one quantitative variable from the survey. Download the R script and the dataset, open the R script and follow the command instructions. Then answer the following questions:

### Categorical Variable

The variable “Major” describes the individual’s major in school. The variable “Phone” identifies the type of phone the individual has (iOS, Android, other). Both of these variables are categorical. Select one of the two categorical variables just mentioned and answer the following three questions.

### (1 point) Choose a categorical variable to explore. Which variable did you choose?

I chose the categorical variable named “Major”.

### (2 point) Paste the table of counts and bar chart for the categorical variable of your choosing. Include color and an appropriate title on your plot.

Table of Counts:

Text

Description automatically generated

Bar Chart:

Chart

Description automatically generated

### (2 point) Briefly, describe the distribution in context. Recall, categorical variables are summarized by counts and/or percents.

The distribution of the above bar chart is summarized by a table of counts and not proportions. Of the categories above, “Mechanical Engineering” category has the highest major count followed by “CS” and “Civil/CEM”. “Chemical Engineering” seems to have to lowest major count. The difference between the above bar chart and a histogram is that in the bar chart above, we can reorganize the categories above (ex: rank from highest to smallest); however, this is not possible in a histogram.

### Quantitative Variable

The variable “Credit Hours” indicates the number of credit hours the individual was enrolled in during the term the survey was completed. The variable “Gaming Hours” describes approximately how many hours a week the survey participant games. Both variables are quantitative. Select one of the two quantitative variables just mentioned and answer the following three questions.

### (1 point) Choose a quantitative variable to explore. Which variable did you choose? Is the variable discrete or continuous?

I chose the quantitative variable called “Credit Hours”, which is a discrete variable.

### (2 point) Create a histogram of the variable. Include color and an appropriate title on your plot. Paste plot.

Chart, histogram

Description automatically generated

### (2 point) Create a boxplot of the variable. Include color and an appropriate title on your plot. Paste plot.

Chart, box and whisker chart

Description automatically generated

### (1 point) Which plot do you prefer (histogram or boxplot) to visualize the variable? Why?

I would prefer the histogram to visualize the variables the because the boxplot above has many outliers and the median is shifted to the far right to Q3 indicating 75% of the data lies below the median. It does not also show the distribution compared to the histogram. Histogram helps to show the frequency of the credit hours.

### (2 points) Give a table that includes the mean, standard deviation, minimum, 1st quartile, median, 3rd quartile, maximum and IQR.

Text

Description automatically generated with medium confidence

### (3 points) Use the plots and summary statistics to describe the data in the context of the problem. Include the shape, center, and spread in your description. State whether there are any outliers.

The shape of the data is almost symmetrically distributed because the mean is close to the median. The mean of the data is 14.56; however, the median gives a more accurate description of credit hours which is 15 credit hours. The standard deviation is 1.79. Therefore, the average deviation from the mean is approximately 1.79. There are seven outliers from the above box plot.

### (2 points) Given the shape of the data which measure, the mean, median or either, would be a more appropriate to represent the center of the data? Explain your reasoning.

Mean of 14.56 does not mean a whole lot in terms of the number credits students in the survey took. Median of 15 indicates 50% of the sample population took at least 15 credits. Since the graph is symmetrically distributed, either the mean or median in this case would be appropriate to represent the center of the data.

### Gradescope Page Matching (2 points)

When you upload your PDF file to Gradescope, you will need to match each question on this assignment to the correct pages. Video instructions for doing this are available in the Start Here module on Canvas on the page “Submitting Assignments in Gradescope”. Failure to follow these instructions will result in a 2-point deduction on your assignment grade. Match this page to outline item “Gradescope Page Matching”.