Mini-Project duo group 7

Venkata Bapanapalli

Karthik Meyyappan

Contributions: Both members equally contributed to analytically solve and implement the code of the given two questions.

1. Question:

Chart, scatter chart

Description automatically generated

We can infer from the scatterplot that the slope is positive. The linear relationship is weak since there is positive relationship between GPA values and ACT values.

The t\* value is the point estimate of the population correlation we get from the sample of the bootstrap, which is 0.274. The bias we get is 0.00998. The standard error is 0.106.

The Confidence Interval is [0.0693, 0.4858] which was computed using percentile bootstrap.

The point estimate was very close to the correlation of the sample values by only being different by 0.01. Since the Point Estimate is 0.27 there is a positive correlation between GPA values and ACT values.

1. Chart, box and whisker chart

   Description automatically generatedChart, scatter chart

   Description automatically generatedQuestion:
   1. From the boxplots it can be observed that the voltage readings at remote locations is higher compared to the voltage readings at local locations. Both remote and local locations have medians that are greater than the mean. This indicates that both graphs are left skewed. From the QQ plots it can be observed that outliers do exists in both of the datasets. Moreover, it can be concluded that the datasets are normalized as majority of the data points coincide with the line.
   2. According to the problem the manufacturing process can be established locally if there is no difference in the population means of voltage readings at the two locations. With this we can set up a hypothesis in the following form:

Difference = sample mean of remote - sample mean of local

null hypothesis: Difference = 0

alternate hypothesis: Difference != 0

Result of t test:

Welch Two Sample t-test

data: remote\_data and local\_data

t = 2.8911, df = 57.16, p-value = 0.005419

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

95 percent confidence interval: 0.1172284 0.6454382

sample estimates:

mean of x mean of y

[9.803667, 9.422333]

The resulting 95% confidence interval is (0.1172284, 0.6454382). From this is can be concluded that the null hypothesis is rejected as 0 does not lie withing the interval. This means that the difference between the mean is not zero and that the manufacturing process cannot be established at local locations.

* 1. Base on the two parts above it can be concluded that it would be ideal to set up manufacturing processes at remote locations as it has higher voltage reading which is preferred to power heavy equipment.

1. Chart, scatter chart

   Description automatically generatedQuestion:

Chart, scatter chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Both plots are identical since they have similar boxplots. A further look into the IQR and the 5-plot summary also prove that they are similar.

We can determine the null hypothesis as:

We can determine the Alternative Hypothesis as:

The confidence interval is (-0.006887, 0.008262). Since the value 0 is in the CI the null hypothesis is accepted.