08/11/2020

Repair Cost Analysis Report

STATISTICS 500

FREDRICK RYANS

Section A: Explanation of Statistical Methods

We were asked by the company to examine if one of the provided repair shop companies were charging more than another company for repair estimates. Based on the data provided for this analysis, we conclude that the data is quantitative (or numerical) for the two repair shops (i.e., Shop A, Shop B). Thus, the statistical method that would be optimal for this analysis based on what we know about the data, would be the *2-sample t-test (pooled)*. The *2-sample t-test* would be effective in this case since the data was produced from two independent groups (i.e., Shop A, Shop B).

The other methods under consideration, were the *paired t-test* and *linear regression.* However, it was believed that the paired t-test would not be optimal for this problem since shop A is not dependent on shop B for services or estimations based on what is known, nor were two measurements taken from each group to satisfy paired testing requirements. Additionally, linear regression was also considered, but not optimal because we do not need to make a prediction on data, and data is not paired in such a way that this would be the most effective method to use in this case.

Section B: Detailed Analysis

To conduct our analysis, we first formulate our hypotheses. In this case we have a null and alternative hypothesis of the following (i.e., H_0, H_a):

H_0: there is no difference in price estimation between Shop A and Shop B.

$\boldsymbol{H_0}$: *Shop A – Shop B = 0*

H_a: there is a difference in price estimation between Shop A and Shop B.

$\boldsymbol{H_a}$: *Shop A – Shop B ≠ 0*

After setting up the hypotheses, we verified that the following conditions were satisfied for the *2-sample t-test*:

1. ***Parent populations from which samples are drawn are normally distributed.***
2. ***The two samples are random and independent of each other.***
3. ***Population variances are equal and unknown.***

Based on these conditions, we assume the data to be normally distributed, based on Figure 1c in Appendix A because the plot shows a p-value of 0.853. Since the p-value is greater than 0.05 (are assumed alpha, or 95% confidence), we can safely assume this data is normally distributed, and satisfies condition 1. We also know that shop A is independent from shop B, satisfying the second condition. Lastly, population of variance is assumed to be equal since the standard deviation of shop A divided by shop B is less than or equal to 2 ($3.22/ 2.94 \leq 2$), satisfying the last condition. It was also decided to conduct a pooled sample test, since it was determined that there was equal variance.

Moreover, the results from the 2-sample t-test produced a p-value of 0.602. Statistically speaking, we must fail to reject the null hypothesis. Since there is a significance level of %5, there is not enough evidence to conclude that there a difference in pricing for Shop A and B. Thus, we conclude that there is no significant price difference between Shop A and B that would make one shop significant enough to choose one over the other. However, it is still enough to possibly choose one shop over the other depending on the budget. The rest of the statistical data for this analysis can be found in *Appendix A, Figures 1-3*.

Section C: Conclusion & Recommendations

The question under concern by my supervisors is if there was a higher difference in pricing between both repair companies (*Shop A and Shop B*). From our analysis, we conclude that Shop A on average charges approximately $1,682.70 (Mu = 16.827 x 100), while the average pricing for repairs from Shop B charges around the same at $1,623.30 (Mu = 16.233 x 100). This is a difference of (Shop A – Shop B = 0.594 x 100 or $59.40), which may be significant enough to choose Shop B over A and reject our null hypotheses.

Moreover, since the company is looking for repairs, $59 may not be a significance difference for our company to do business with one shop over another. If the repairs were for a single individual, then it would make sense for them to choose shop B over A. However, since we are a company, I conclude that we would be fine to choose shop B over A for its lower cost, if it is in range with the yearly budget of the company. Otherwise, It is safe to conclude, that there is no significance difference between shop A and B overall.

Appendix A

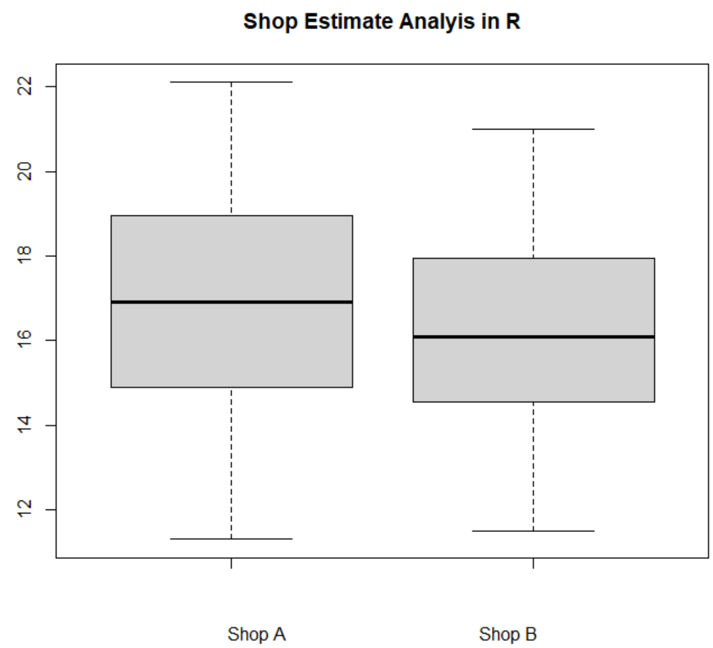

Figure 1a. This shows a box plot of the two shops with the mean.

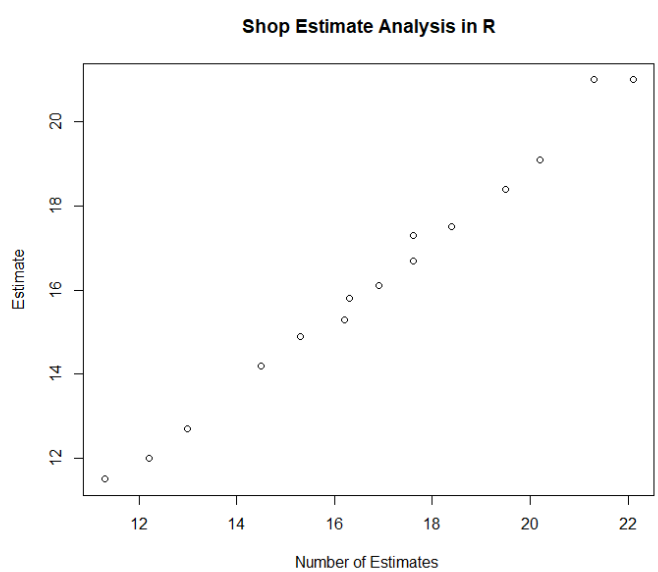

Figure 1b. This plot shows a scatterplot of shops A and B.

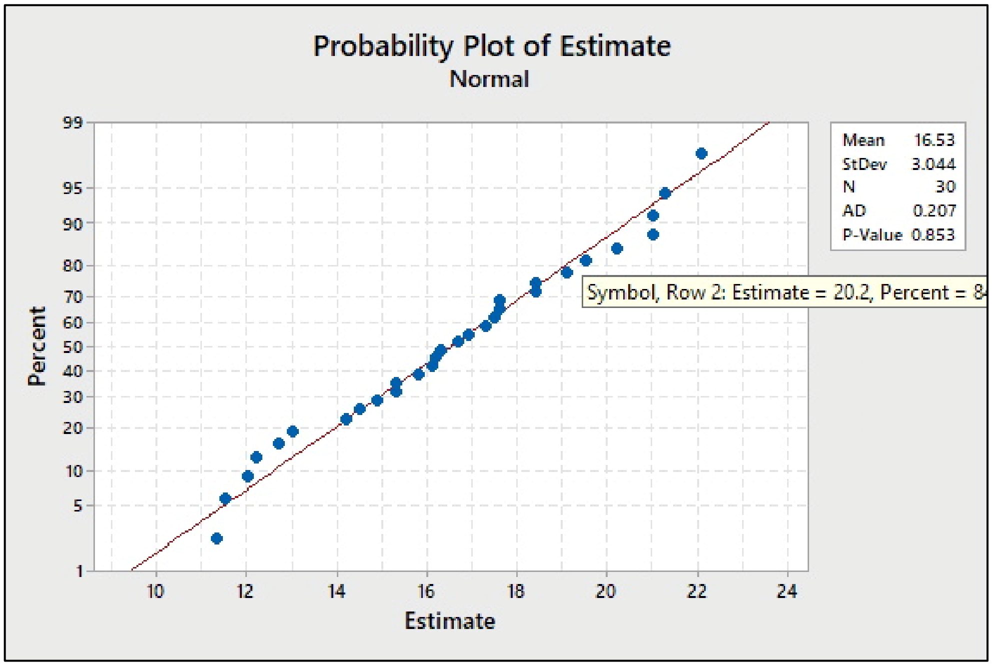

Figure 1c. This plot shows the test for Normality for the dataset using the Anderson-Darling Test.

Descriptive Statistics: Shop A, Shop B

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Shop A	15	15	16.827	0.831	3.219	11.300	14.500	16.900	19.500	22.100
Shop B	15	15	16.233	0.759	2.941	11.500	14.200	16.100	18.400	21.000

Figure 2. Descriptive Statistics Output for the dataset.

Two-Sample T-Test and CI: Estimate, Shop

Method

μ_1: mean of Estimate when Shop = A
μ_2: mean of Estimate when Shop = B
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics: Estimate

Shop	N	Mean	StDev	SE Mean
A	15	16.83	3.22	0.83
B	15	16.23	2.94	0.76

Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
0.59	3.08	(-1.71, 2.90)

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
0.53	28	0.602

Figure 3. Output for *2-sample t test*

Appendix B

To complete this project, I collaborated with classmates via Zoom. The following people were collaborators for helping complete this project: *Jim Mahoney, Rosana Burkey, Audrey Simard, Brittani Bedford, Kyle Hawkey*