DAT-475-12550-M01 Project 2

Cadden Buist

Southern New Hampshire University

Professor Keane

12/01/2024

Lately at a manufacturing company in Tijuana, Mexico there has been an issue of components having defects since the company’s demands for their products has increased. This issue is the reason that I was brought in to find out what is going on and use data to try and find the root cause for this issue to hopefully solve this problem. Within this analysis I will be looking at production lines to try and narrow down if there is a significant difference from 1 production line compared to the other 2 production lines. To set this hypothesis up we must first figure out what we want to test and as was discussed we are testing whether or not there is a significant difference from 1 production line compared to the other 2. Next, we need to discuss which parameters we plan on using for this test and after studying the data the data that will be used will be model or production lines 1, 2, and 3 along with the percentages of the defects each production line experiences. These parameters were chosen because they show us how often a defect happens for each production line making it easier for us to get a solution to this issue being experienced.

Our Null hypothesis for this analysis will be H0 = There is no significant difference in the percentage of defects from any of the production lines to the other 2. Our Alternative hypothesis will be Ha = There is a significant difference from one of the production lines than from the other 2. These two hypothesis formulas will look like this H0 = A1=A2=A3 and Ha = A1=A2\=A3. As we can see only one of the production lines can be significantly different for us to be able to reject the null hypothesis.

A screenshot of a computer

Description automatically generated

From the image above we can see that that every p-value is significant since we are using a 5% significant value or 0.05. We do see that the model does have a low fit with R-squared and Adjusted R-square both being under 50% or under 0.5. Our F-statistic does give us a value of 5.285 and looking at the F-critical table we get a value of 3.37 and this tells us that we can reject the null hypothesis because our F-statistic is greater than our F-critical or 5.285>3.37. Then again we did discuss how this model is a bad fit so I want to also look at another ANOVA test and the Tukey test.

A screenshot of a computer

Description automatically generated

In the image above we see that the only p-value that is being shown is for model which is 0.0266 and that value is less than the significant value 0.05. This tells us that our Model variable is significantly different or in other words we can reject the null hypothesis.

A screenshot of a computer

Description automatically generated

Finally, we see that our Model2 and Model3 tested against our Model 1 is significantly different with both equally less than the significant value at 0.05. We also see that the P-value when Model2 and Model3 are compared to one another it is 0.997 which is way higher than 0.05 significant value making those values insignificant to one another.

A screenshot of a computer

Description automatically generated

Last but not least here is a visual we can use to see the difference from model1 compared to model2 and model3. As we can see model2 and model3 are very similar to one another and look like it could be the exact same variables in both models. When we look at model1 and its low number appears to be higher than the other 2 model’s high values. This boxplot visual puts everything talked about in this analysis into perspective.

**Conclusion**

We can conclude that we can reject the null hypothesis as we have been able to see all the info collected points to model1 or production line 1 is significantly having defect issues compared to the other production lines. We can now focus on production line 1 and figure out why there are so many defects coming out of production line one. This will likely bring the defects problems way down getting us close to our goal bringing the defect problem down 20%