Brandi Jones

DAT-475-T1310

Applied Data Analysis

September 30, 2022

Project Two

When looking at our dataset for Project Two, we can see that there are three different models, Model1, Model2, and Model3. Each of these three models represents a production line. We can use this data to determine which production line shows the highest percentage of defects. In our dataset, we also have columns for Defects and Percentage. Our goal is to run a One-Way ANOVA to determine if there is a significant difference between the means of the three production lines. A One-Way ANOVA “compares the means of three or more independent groups to determine if there is a statistically significant difference between the corresponding population means” (Zach, 2020).

Table

Description automatically generated

The parameters we will be using in this test include Model and Percentage. The Models are grouped by 1, 2, or 3, and the mean of each model’s percentage will be used in our ANOVA. Determining if there is a significant difference in the means of the percentage values of the three models will lead us to creating our Null and Alternative Hypotheses.

* Independent Variables – Model1, Model2, Model3
* Dependent Variable - Percentage

Null - H0: There is no significant difference between the three means (μ1=μ2=μ3)

Alternative - Ha: There is at least one difference between the three means (μ1≠ μ2≠ μ3)

Significance Level - α = 0.05

Table

Description automatically generated

Looking at the descriptives of our dataset shows us that there are a total of 15 observations, five for Model1, five for Model2, and five for Model3. The descriptives table also gives us the mean, standard deviation, standard error, lower bound, upper bound, minimum, and maximum of each model.

Chart, line chart

Description automatically generated

I also plotted the means of the three individual models to make it easier to view. It can be seen in this line chart that Model1 has the highest mean, followed by Model3 and then Model2.

Table

Description automatically generated

After running the ANOVA, we can see that the Total Sum of Squares is equal to 789.096. The Sum of Squares Between Groups is 369.554 and the Sum of Squares Within Groups is 419.542. The F-Statistic is 5.285 and the significance (p-value) is .023. Since the p-value is less than the alpha of 0.05, we can determine that this test shows a significant difference, and we would reject the null hypothesis. This leads us to accept the alternative hypothesis, that there is at least one difference between the three means (μ1≠ μ2≠ μ3).

Table

Description automatically generated

“The aim of a post-hoc analysis is to find patterns after the study has been completed, and to find results that weren’t the primary objective. Tukey’s test determines if your sample consists of groups that differ from each other. Every mean is compared with the mean of all other groups using the “Honest Significant Difference,” which represents how far apart the groups are” (Zaveri, 2022). Running this Post Hoc test allows us to compare the mean of each individual model against another individual model to check for significance.

* Model 1 vs Model 2 – Significant – Reject the null hypothesis
* Model 1 vs Model 3 – Significant – Reject the null hypothesis
* Model 2 vs Model 3 – Not significant – Fail to reject the null hypothesis

After running the Post Hoc test, I can determine that there is a significant difference between Model 1 vs Model 2 and Model 1 vs Model 3. Since Model 1 had the largest mean out of the three, this makes sense that it shows the most significance between the models. There is not a significant difference between Model 2 vs Model 3, which also makes sense as the means for these two models were much closer together than with Model 1.

In conclusion, our ANOVA test was successful in showing us that the null hypothesis (There is no significant difference between the three means (μ1=μ2=μ3)) should be rejected and that we should accept the alternative (There is at least one difference between the three means (μ1≠ μ2≠ μ3)). Our Post Hoc test has shown that the Model with the most significance is Model 1, meaning that production line 1 is the line with the highest percentage of defects, and more focus should be placed on this line to determine the root causes of these defects.

**References**

Zach. (2020, April 6). *How to find the F critical value in Excel*. Statology. Retrieved from https://www.statology.org/f-critical-value-excel/

Zaveri, A. (2022, August 4). *Post Hoc Analysis: Process and types of tests*. Mind the Graph Blog. Retrieved from https://mindthegraph.com/blog/post-hoc-analysis/