**Project Two: Hypothesis Testing for Production Line Data using One-Way ANOVA**

Hom Dahal

Southern New Hampshire University

DAT-475 Applied Data Analysis

07/31/2022

In our first project, we analyzed the manufacturing defects of a company in Tijuana, Mexico. We used Pareto chart analysis and found faults in production lines. We also performed root cause analysis using a fishbone diagram and determined the root cause of the flaws in production lines. Our next step is to perform a hypothesis test using data from the production line with the highest number of defects. The hypothesis test will help us determine if there is a statistically significant variation in the number of faults between the three-production line. Suppose we become successful in finding the statistically significant differences. In that case, we, as data analysts, can convince stakeholders that the production lines with the most significant defects need to have them fixed. (Project two Guidelines and Rubric, Project Case Study, n.d.)

Three variables make up our dataset: model, defects, and percentage. We will use the model as a factor and percentage as the dependent variable to do a hypothesis test. The data for the three models are displayed in the following snapshot and are ready for hypothesis testing.Table

Description automatically generated

From the findings from our first project, we believe that the manufacturing process has a defect which is our research hypothesis. So, the null hypothesis is that the manufacturing process has no defects. From the three models provided to us, if we find enough evidence to show a significant difference between the mean of the three models, we will reject the null hypothesis in favor of the research hypothesis. This way, we can convince the stakeholders that there is a problem in the manufacturing process.

1. **Hypothesis Statement**

If M1, M2, and M3 represent the mean of model1, model2, and model3, respectively then we define our null and alternative hypothesis mathematically as follows:

Null Hypothesis, H0 : M1 = M2 = M3.

Alternative Hypothesis H1 : Not ( M1 = M2 = M3).

The main parameters for this hypothesis are the means of each model and the variances. We use these parameters because means and variances are used for hypothesis tests using one-way ANOVA. (Singh, 2020)

The below screenshot shows the result of our one-way ANOVA descriptive statistics for each model.

Table

Description automatically generated

1. **Degree of Freedom and F-critical**

The screenshot below shows the degree of freedom between and within groups which is 2 and 12 respectively. Using these two values we find the F-critical value of 3.89 from the chart.

Table

Description automatically generated

1. **F-statistic**

Also, from the above output screenshot from the SPSS report we see that the F-statistic value is 5.28. This F-statistic value is not in the 95% region of acceptance: [-∞, F-Critical Value]

Chart, line chart

Description automatically generated

F-critical = 3.89 ( from chart) , 5.285 > 3.89

For our test, an F-statistic value greater than the F-critical value provides evidence to reject the null hypothesis in favor of the research hypothesis at the level of significance we chose (0.05). We can conclude that the means of at least one of the three models deviate significantly. We must do a post hoc test to pinpoint which models differ. The multiple comparison test is another name for this assessment. The test screenshot is shown below.

Table

Description automatically generated

From the above table, we can see the following information:

1. Model1 Vs Model2 🡪 Null hypothesis is rejected with a significance of 0.036
2. Model1 Vs Model3 🡪Null hypothesis is rejected with a significance of 0.041
3. Model2 Vs Model3 🡪 Failed to reject the null hypothesis with a significance of 0.997

The means of the following pairs are significantly different: Model1 – Model2 and Model1- Model3. From the table above, we see that there is no significant difference between the mean of Model2 and Model3, so the mean of Model1 is significantly different.

1. **Conclusions:**

We have seen from section 3 above that the F-statistic value is more than F-critical. Thus, we reject the null hypothesis in favor of the alternative hypothesis and conclude there is a significant difference between the mean of at least one of the three models. We also performed the multiple comparison test to determine that model with the highest number of defects is the model1.

**Chart, bar chart

Description automatically generated**

Resources

Project two Guidelines and Rubric, Project Case Study(n.d.). Retrieved September 19, 2021,

From <https://learn.snhu.edu/content/enforced/1119486-DAT-475-T6752-OL-TRAD-UG.22EW6/course_documents/DAT%20475%20Project%20Case%20Study.pdf?_&d2lSessionVal=CUeEJqmMxTlAUDLiq08W6Aphn&ou=1119486>

Singh, G. (2020, April 1). *Analysis of variance (ANOVA): Introduction, types & techniques*. Analytics Vidhya. Retrieved July 31, 2022, from <https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/>