Elijah Cruz

April, 2025

DAT-475

Project Two

This report summarizes the results of a One-Way ANOVA conducted to evaluate whether there are significant differences in defect percentages among three production lines—Model1, Model2, and Model3. The analysis indicated that Model1 shows a statistically significant difference in defect rates compared to both Model2 and Model3, highlighting the need for focused improvement initiatives on Model1.

I will be using the data set provided to perform a one-way Anova which will allow for hypothesis testing.

To examine the differences in defect percentages across the production lines, a One-Way ANOVA test was conducted. This statistical method is suitable for comparing the means of more than two groups, making it ideal for analyzing the three models in question. The test was configured with the following parameters:

- Dependent Variable: Defect Percentage
- Independent Variable: Model
- Significance Level ($\alpha$): 0.05

These parameters were selected because the primary objective is to determine whether the three distinct models yield different defect rates—a scenario involving a categorical independent variable. The significance level of 0.05 is a standard benchmark for identifying statistically meaningful differences.

The hypotheses for the analysis are as follows:

- Null Hypothesis ($H_0$): There is no significant difference in defect percentages among the production lines (Model1 = Model2 = Model3).
- Alternative Hypothesis ($H_1$): At least one production line has a significantly different defect percentage compared to the others.

These hypotheses aim to identify whether certain production lines are more prone to defects, which could point to underlying process issues requiring attention.

The hypothesis testing was performed in R using the supplied dataset. The analysis involved the following steps:

1. The data was loaded into R using the `read.csv` function.
2. A One-Way ANOVA was executed with the `aov` function to evaluate differences in defect percentages among the three production models.
3. Upon finding a significant result from the ANOVA, a Post Hoc analysis was conducted using the `TukeyHSD` function to determine which model comparisons were statistically different.

ANOVA Results:

```
> df<-read.csv("C://Users//Public//Desktop//DAT-475//DATAAnova2.csv")
> anovaResult <- aov(Percentage ~ Model, data = df)
> summary(anovaResult)
            Df Sum Sq Mean Sq F value Pr(>F)
Model        2  369.6  184.78   5.285 0.0226 *
Residuals   12  419.5   34.96
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The ANOVA test produced an F-statistic of 5.285 with a p-value of 0.0226, indicating a statistically significant difference in defect percentages across the production models.

Post Hoc Analysis:

```
> TukeyHSD(anovaResult)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Percentage ~ Model, data = df)

$Model
                 diff        lwr       upr     p adj
Model2-Model1 -10.654 -20.630789 -0.677211 0.0362961
Model3-Model1 -10.400 -20.376789 -0.423211 0.0409467
Model3-Model2   0.254  -9.722789 10.230789 0.9974603
```

Tukey's HSD post hoc analysis revealed significant differences between Model1 and both Model2 and Model3 (adjusted p-value < 0.05), while no significant difference was found between Model2 and Model3 (adjusted p-value = 0.9975).

As a result, the null hypothesis is rejected in favor of the alternative hypothesis—indicating a statistically significant difference in defect percentages, specifically between Model1 and the other two models. The test results confirm significant differences for Model1 vs. Model2 and Model1 vs. Model3, suggesting that corrective measures should be prioritized for Model1. In contrast, the similarity in performance between Model2 and Model3 suggests no immediate need for intervention in those lines.

Based on these findings, it is recommended to conduct a focused analysis of Model1's manufacturing process to identify the root causes of its higher defect rate. Targeted corrective actions should be implemented, followed by continuous monitoring and quality control procedures to ensure that improvements are effective and sustained.