# Search for Cause II

## 1 Question

The objective of this investigation is to determine whether the dominant causes act in the assembly or component family. From the previous investigation, we found that the process in between y100 and y200 and, y200 and y300 have a high SD ratio indicating that part of the process does not have a dominant cause. Therefore, the dominant cause must be in the components or assembly.

## 2 Plan

We will investigate the assembly and then the component (if necessary). To investigation those parts of the process, we need to do another baseline investigation with y100 as the output. The new baseline data has 18 observations over 15 shifts (3 Shifts x 5 days) using random systematic sampling which will cost $270. Based on the new baseline data, we will select 2 parts with opposite and extreme output values relative to the full extent of variation and implement a disassembly/reassembly investigation that disassembles and reassembles the chosen parts 3 times. This will cost $300. If the assembly variation is large then the variation is within the assembly, however, if the assembly variation is small we have to conduct a component swap investigation. The component swap investigation will have components A and B in one group and C, D, and E components in another group. We will swap one component from one group to the other group until we see which component(s) causes large variation. The cost of component swap investigation will vary on the number of swaps required before identifying the component(s) with the dominant cause.

## 3 Data

Below is a view our whole dataset.

| **part** | **y100** |
|---|---|
| 27208 | 10.2 |
| 27208 | 8.2 |
| 27208 | 10.2 |
| 27208 | **9.6** (Original) |
| 27897 | -3.4 |
| 27897 | -4.0 |
| 27897 | -5.4 |
| 27897 | **-12.4** (Original) |

# 4 Analysis

## 4.1 Assembly / Reassembly Investigation

### 4.1.1 Stripchart

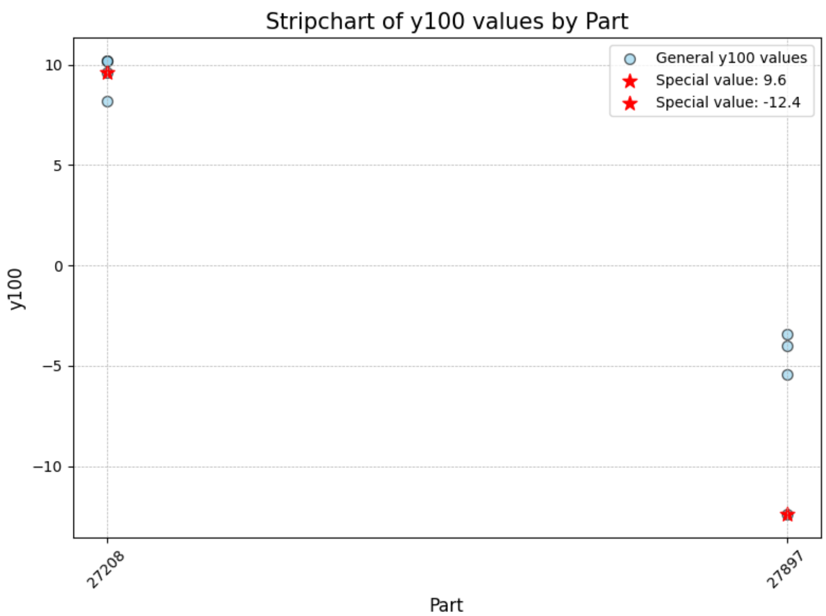
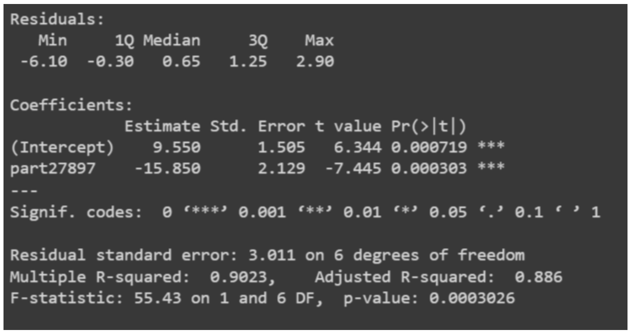



Figure 1: Stripchart By Partnumber

The variability exhibited by the two part numbers, 27208 and 27897, appears to be moderately low. For part number 27208, the data points are closely clustered around a value of 10, which indicates a narrow range of variation. In the case of part number 27897, the majority of the data points are centered around -4, but this group includes an outlier at -12.4. Excluding this outlier, part number 27897 would display a similarly low level of variability as observed in part number 27208, suggesting a generally uniform performance across both sets of measurements.

Given the minimal variation observed within each part number, it's implied that this variability is unlikely attributed to the Reassembly/Disassembly Investigation Stage. Rather, the source of variation might be identified at an earlier stage. This hypothesis will be further corroborated by the subsequent analysis outlined below.

### 4.1.2 Model Summaries



```
Residuals:
   Min     1Q Median     3Q    Max
 -6.10  -0.30   0.65   1.25   2.90

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.550      1.505   6.344 0.000719 ***
part27897    -15.850      2.129  -7.445 0.000303 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.011 on 6 degrees of freedom
Multiple R-squared:  0.9023,    Adjusted R-squared:  0.886
F-statistic: 55.43 on 1 and 6 DF,  p-value: 0.0003026
```

Figure 2: Linear Regression Model Summary

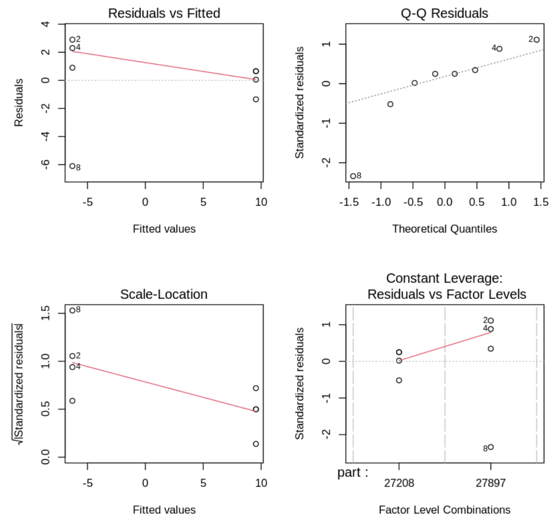



Figure 3: Linear Model Between $y_{200}$ and $y_{300}$

The regression analysis indicates significant variability in part 27897 compared to the baseline, evidenced by a strong coefficient and a very low p-value. The model explains a high proportion of the variation in the data, as shown by the multiple R-squared value of 0.9023. However, the diagnostic plots suggest some potential issues with the model assumptions: there may be non-linearity, as indicated by the pattern in the Residuals vs Fitted plot, and possible heteroscedasticity, as the Scale-Location plot shows a change in spread of residuals. Despite these concerns, there are no apparent influential outliers distorting the model's results.

### 4.1.3 ANOVAs

| Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|
| part | 1 | 502.445 | 502.445 | 55.42692 | 0.0003025625 |
| Residuals | 6 | 54.390 | 9.065 | NA | NA |

Table 1: ANOVA Table

The ANOVA table indicates significant variation between the parts being analyzed. The part factor, with a single degree of freedom, accounts for a considerable sum of squares (502.445), suggesting that the mean values across the parts differ substantially. The mean square, which is a measure of the average variation due to the part number, is also high (502.445). In contrast, the residual variation, which captures the variability within part groups, has a lower mean square of 9.065. The high F value of 55.42692 for the part suggests this variation between parts is statistically significant. The corresponding p-value of 0.0003025625, far below the typical alpha level of 0.05, further confirms that the observed variation is unlikely to be due to random chance. Therefore, the data strongly suggest that the variation in measurements is influenced by the part number rather than being within-part variations or due to the Reassembly/Disassembly Investigation Stage

### 4.1.4 Numerical Values

The standard deviations for each part number are as follows:

$$\sigma_{\text{part\_27208}} = 0.9434$$
$$\sigma_{\text{part\_27897}} = 4.1521$$

If Part 27897 does not include the value $-12.4$:

$$\sigma_{\text{part\_27897}} = 1.0263$$

The overall standard deviation is:

$$\sigma_{\text{overall}} = 8.9189$$

The provided standard deviations reflect the extent of variability within the measurements of each part number as well as overall. For part 27208, the standard deviation is 0.9434, which indicates a low variability among its measurements. This suggests that the values for part 27208 are tightly clustered around the mean.

In contrast, part 27897 shows a higher standard deviation of 4.1521 when considering all data points, including the value of -12.4, which points to a wider spread of the data and hence more variability. However, if we exclude the -12.4 outlier, the standard deviation drops significantly to 1.0263 for part 27897, bringing its level of variability closer to that of part 27208. This highlights the impact of the outlier on the perceived variability of part 27897.

The overall standard deviation for all parts combined is 8.9189, which is substantially higher than the individual standard deviations of each part. This overall higher variability suggests that when considering all parts together, there is a much wider range of values in the dataset.

## 4.2 Component Swap Investigation

### 4.2.1 2x2 Tables

- We create a set of 2 by 2 Tables to visualize the variance of each part number between each component. We split our components into two initial groups: (A,B) and (C,D,E).

| **Part** | **A** | **B** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | 12 |
| 27897 | -1.6 | -3.4, -4, -5.4, -12.4 |

Table 2: Component A vs B

- We see that the variation from A vs B to A vs C remains fairly constant. This implies that component A is not home to a dominant cause

| **Part** | **A** | **C** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | 10.2 |
| 27897 | -0.2 | -3.4, -4, -5.4, -12.4 |

Table 3: Component A vs C

- We see that the variation from A vs C to B vs C remains fairly constant. This implies that component C is not home to a dominant cause

| **Part** | **B** | **C** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | 8.9 |
| 27897 | -1.8 | -3.4, -4, -5.4, -12.4 |

Table 4: Component B vs C

- We see that the variation from B vs C to B vs E remains fairly constant. This implies that component B is not home to a dominant cause

| **Part** | **B** | **E** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | 12.2 |
| 27897 | -1.4 | -3.4, -4, -5.4, -12.4 |

Table 5: Component B vs E

- We see that the variation from B vs E to A vs E remains fairly constant. This implies that component E is not home to a dominant cause

| **Part** | **A** | **E** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | 10.2 |
| 27897 | -3.6 | -3.4, -4, -5.4, -12.4 |

Table 6: Component A vs E

- We see that the variation from A vs E to A vs D is significantly different. This suggests that D is home to a dominant cause, however we have to further investigate.

| **Part** | **A** | **D** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | -1.4 |
| 27897 | 7.2 | -3.4, -4, -5.4, -12.4 |

Table 7: Component A vs D

- We see that the variation from A vs B to B vs D is significantly different, and that A vs D and B vs D have fairly similar variations. This suggests that D is home to a dominant cause.

| **Part** | **B** | **D** |
|---|---|---|
| 27208 | 10.2, 8.2, 10.2, 9.6 | -0.8 |
| 27897 | 7.6 | -3.4, -4, -5.4, -12.4 |

Table 8: Component B vs D

We conclude from this component swap investigation that D is home to a dominant cause, while A, B, C, and E are not home to any dominant causes.

# 5 Conclusion

In our Assembly and Reassembly Investigation, we established that the assembly variation for each of the two examined part numbers is minimal. This indicates that the primary source of variation stems from the components family rather than the assembly process itself.

Further, our Component Investigation revealed a significant insight. Through a component swap analysis, we identified that Component D is the locus of a predominant variation. Given that Component D encompasses inputs x10, x11, and x12, it is evident that one or more of these inputs are the principal contributors to this variation