STAT 435: Final Report

Due on April 21, 2023

University of Waterloo
Faculty of Mathematics
Department of Statistics and Actuarial Science

Prepared by:

Gunchica Bhalla

Introduction

Watfactory is a computer-based model of an automotive camshaft manufacturing process that includes 60 inputs that can vary and 30 inputs that are fixed. The output can be measured in three locations, and the objective of the project is to minimize the variability of the final output (y300) to be within the range of -10 to 10, given a budget of $10,000. At the end of the midterm investigation remaining balance is $3,840.

In summary, the midterm investigation yeilded the variation belongs to the day-to-day time family, the measurement system is capable, the variation transmission take place in the heat treatment phaze and that x_{51}seems to be the dominant cause. In this report I will be continuing the investigation to verify the dominant cause, search for the solution and test it to see if I can achieve the goal set in the baseline. For the investigation covered in this report, I started with $3,840.I spent $606 on verification of the dominant cause, $728 to investigate for an adjuster and another $168 to verify the adjuster, $728 in search of a desensitizer, $150 on the implementation of feedforward control and lastly $215 on the the implementation of feedback control. I spent around an additional $312 for data collection for the feedforward control implementation. At the end remaining budget is $130.

Throughout the investigation I used the processes outlined in Statistical Engineering (Steiner & MacKay, 2005).

I will be following the QPDAC approach for each of the 5 parts of the investigation. Each of the steps will be defined in great detail in the report.

Part 1: Verification of Cause

Question

In this part, it will be determined whether the , x_{51}, is the dominant cause in our Watfactory process. I wiall also be looking into x_{48} as suggested by the feedback on my midterm.The question I want to answer by the end of this investigation is:

1. Which of the x_{51} or x_{48} is the true dominant cause?

Plan

The plan for this investigation is to collect data for at least 3 replicates.Since our dominant cause has a linear trend, the 2 levels we chose were essentially the min and max of the input value range. Since the guide did not specify the range of values the input can take, we used the identified min and max from our previous investigation, Search For the Dominant Cause III which we identified to be 2 and 4. The plan is to randomize the runs to reduce the risk of confounding in the experiment between and other varying inputs, it also helps to balance other varying inputs over the runs at the high and low level of the suspect. The cost for the experiment for 5 replicates is $303. I repeated the process with x_{48} as to incorporate the feedback from the midterm, as I had already conducted the verification investigation with one suspect I repeated it for x_{48} to make sure that it does not present new information. The total amount spent on the investigation was $606.

Data

The investigation was conducted according to the plan. I collected a total of 6 parts for this investigation, 6 from one level (2) and 6 from the other (4). y_{300} was collected to see how the level / input value impacts y_{300}. The data/parts used in the experiment were randomised as detailed above. I also x_{48} collecting 6 parts for each level -10.6 and 16 each. Essentially, the data consists of y_{300}, x_{51}, x_{48} values. For x_{48} I choose the minimum and maximum for x_{48} observed in my data rather than the limits mentioned in the introduction file because I did not suspect the impact changing with other linits and also the limited funds prohibit me from collecting more data.

Analysis

With the aim to verify that which of x_{51} and x_{48} is the cause for variation, I plotted the output y_{300} at the two levels to observe the difference in the output when the operator is changed from 2 to 4 as shown in the Figure below.

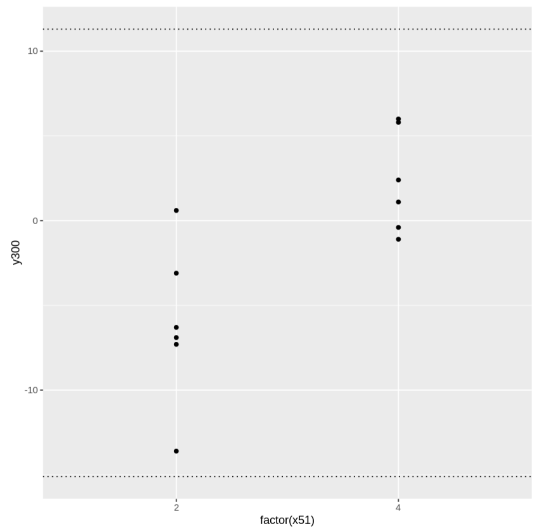

Figure 1. y_{300} when operator 2 and 4 are on shift

I also plotted x_{48}, the induction level the figure below shows the change in the y_{300} value when the induction value changes from -10.6 to 16.

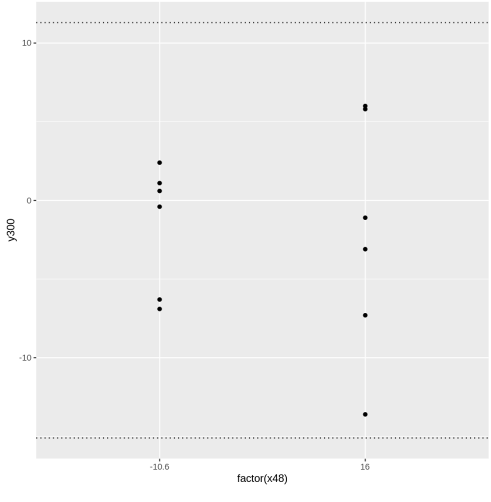

Figure 2. Y_{300} when the induction level is -10.6 and 16

In the above figures it can be noted that y_{300} is has more variation when the operator is 2 and variation increases in y_{300} increases at induction level is 16. To continue the investigation I plotted an interaction as shown in the Figure below.

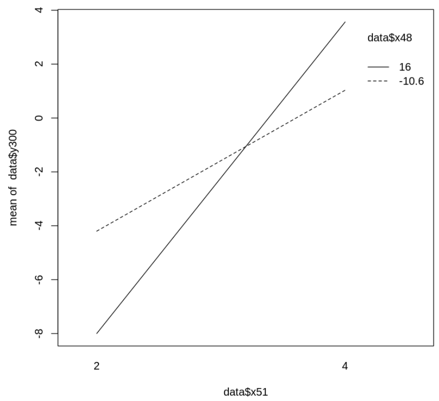

Figure 3. Interaction between y_{300}, x_{51} and x_{48}

The above figure shows that there is an interaction between x_{51} and x_{48}. Continuing in the investigation I plotted a pareto chart to investigate the impact of x_{51}, x_{48} and the interaction as shown in the Figure below.

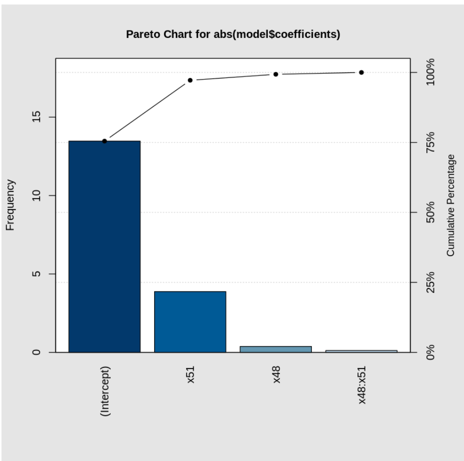

Figure 4. Pareto Analysis

The pareto analysis helps to identify the most important factors that contribute to a problem or opportunity. In this situation I want to identify the contribution of x_{51}, x_{48} and the x_{51} x_{48} interaction to the variation seen in y_{300}. The pareto analysis shows that frequency and the cumulative percentage for each factor. After the intercepts, x_{51} has the highest frequency and percentage. According to the chart x_{51} with a frequency of 3.88 and a cumulative percentage of 97.2%. This suggests that x_{51} has the greatest impact on the outcome variable y_{300}.

Looking at Figure 1 it can be noted that variation in y_{300} reduces when the x_{51} value is changed from 2 to 4. By calculating the mean of y_{300} at the two different levels as shown in the image below, it can observed that changing the operator from 2 to 4 causes the mean to change by 8.4 as shown in Figure 5. It should be noted that the change in mean covers the FEoV calculated in the baseline investigation.

```
library(dplyr)
df_mean <- data %>%
  group_by(factor(x51)) %>%
  summarize(mean_value = mean(y300))

df_mean
```

A tibble: 2 × 2

factor(x51)	mean_value
<fct>	<dbl>
2	-6.1
4	2.3

```
abs(-6.1 -2.3)
```

8.4

Figure 5. Impact of change on y_{300}

From the previous investigations we know that the mean for y_{300} is -1.41. Given that the mean y_{300} value for each level is very different, we can say that the value of x_{51} impacts the y_{300} output. Between the two means, there is a range of 8.4, which is significantly large than the mean of y_{300} proving that the change in operator can lead to significant variation in the y_{300}.

Conclusion

In conclusion, we can see that as the value of x_{51} changes, the output value y_{300} significantly changes as well. This confirms the suspicions that the input x_{51} in the heat treatment phase is the main cause of variation in the process.

Part 2: Search for Solution I

Question

In this part, I begin the remedial journey. The questions I want to answer by the end of this investigation is:

1. What are the main effects and interaction effects from the fixed inputs?
2. Is there an adjuster we can incorporate amongst the fixed inputs?

Plan

The plan for this investigation is to identify how we can reduce variation at a reasonable cost. We want to collect data for the output y_{300}.The fixed inputs I planned to collect include all those between z_{19} and z_{22}, inclusive. I used two levels for each of these inputs - the minimum and maximum feasible value. The levels for each input were identified from the Fixed Input Information table in the Watfactory guide. The plan is to design a fractional factorial experiment. However, given that I have 4 fixed inputs I am considering that, a full factorial experiment would be far too large and costly to conduct and measure. The experiment could either be of size 16 or 32 - due to having 4 inputs to consider. 32 was too expensive to conduct so we went with the following experiment design with 16 runs and 4 factors using the FrF2 package in R. As show in the image below.

	z19	z20	z21	z22
1	25	20	20	13
2	53	20	20	13
3	25	22	20	13
4	53	22	20	13
5	25	20	48	13
6	53	20	48	13
7	25	22	48	13
8	53	22	48	13
9	25	20	20	21
10	53	20	20	21
11	25	22	20	21
12	53	22	20	21
13	25	20	48	21
14	53	20	48	21
15	25	22	48	21
16	53	22	48	21

Figure 6. Full Factorial Analysis Data

Data

The investigation was conducted according to the plan. y_{300} was collected to see how the level / input value impacts y_{300}. The data consists of 16 rows where the relevant columns are y_{300} (response variable) and z_{19} - z_{22} (the inputs). The inputs take on values based on their specified levels.

Analysis

For this analysis, we will start with a visual investigation to see how the inputs may impact the value of y_{300}.

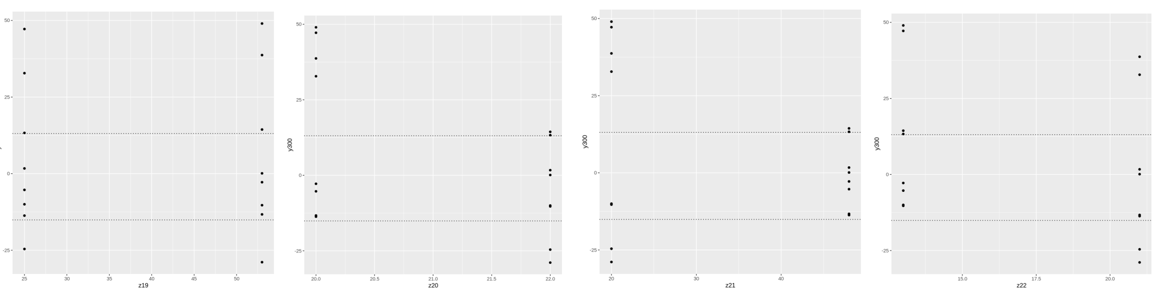

Figure 7. Plots of z inputs vs y_{300} output

As it can be seen form the above figure it can be inferred that z_{20} and z_{21} have an impact on the y_{300}. The impact of change in the value of z_{20} on y_{300} is visible. When the value of z_{20} changes the variation in the y_{300} reduces noticeably. Similarly for z_{21} the change in variation of y_{300} at the two values is very noticeable. To better understand the impact I conducted a linear regression as shown in Appendix A. I incorporated 2-way interaction in the linear regression model to check for interactions. The low residual numbers suggest that there is no three way interaction. The linear regression model suggests that z_{20}, z_{21} and the interaction between the are the most significant contributors for the variation seen in y_{300}. This is confirmed by the anova analysis shown in Appendix B. The high sum of square value suggests that the interaction between z_{20} and z_{21} is the largest contributor to the variation. For visual analysis I plotted the interaction plots for all the fixed inputs as shown in the figure below.

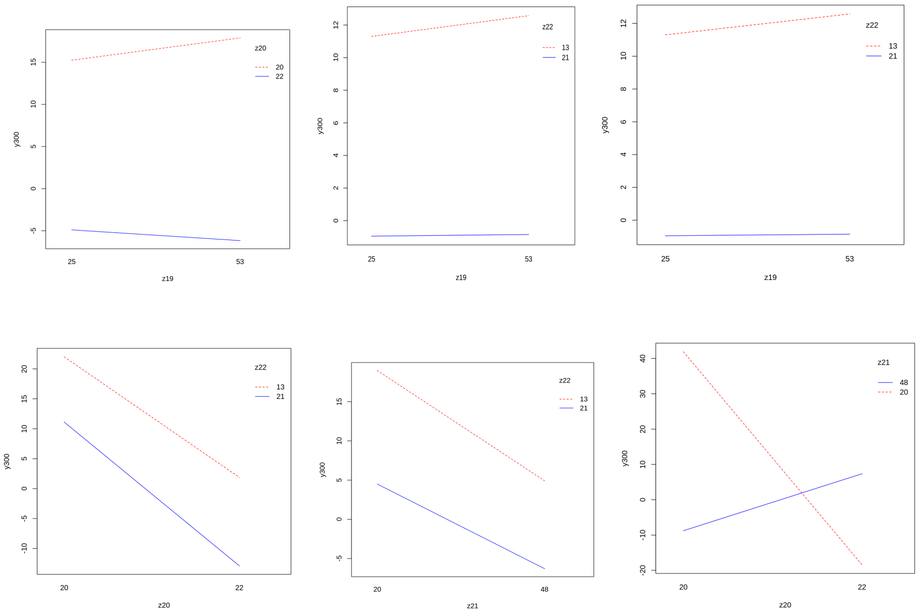

Figure 8. Interaction Plots for fixed inputs

The above images show slight interaction between all the fixed inputs but the interaction plot between z_{21} and z_{20} has an intersection signifying higher interaction. These graphs add evidence to the linear regression model which reveals that this interaction is responsible for the variation seen in y_{300}. At this point in the investigation it can be seen that z_{20}, z_{21}, and the interaction between z_{20} and z_{21} are significant contributors but choosing which one as an adjuster is not clear, so I plotted a pareto analysis chart as shown in Figure below.

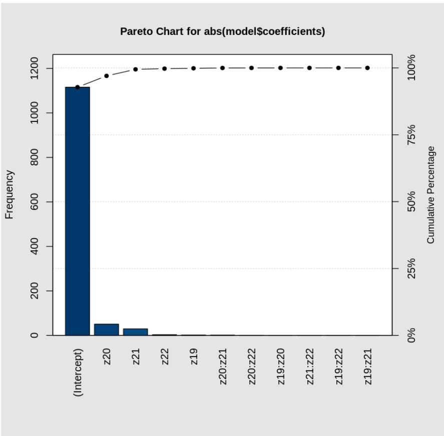

Figure 9. Pareto Chart

In the pareto chart it can be seen that the interaction between z_{20} and z_{21} has a very low frequency and cumulative percentage. z_{20} has the second highest cumulative percentage i.e it is the most impactful variable after the intercept. To verify the this I collected data for z_{20} at 3 different levels as shown in the table below and did 4 random replications. This cost me $168.

Level	1	2	3
z_{20}	20	21	22

Using the data from the second graph I plotted the y_{300} at different valued of z_{20} as shown in the image below.

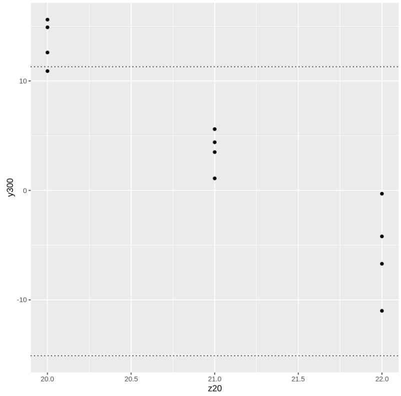

Figure 10. $z_{20} \sim y_{300}$

As it can be seen from the above image the variation seen in y_{300} changes with the changes in value of z_{20}. There seems to be a negative relationship between y_{300} and z_{20}. To verify the visual analysis I conducted a linear regression as shown in Appendix C. The high Adjusted R-squared of 0.884 indicates a strong relation between y_{300} and z_{20}. This is confirmed by the anova analysis showed in Appendix D.

Conclusion

In conclusion, there is sufficient evidence that this experiment has succeeded. I conclude that the z_{20} is the most significant cause of variation for y_{300}. This finding should be used to proceed to the next stage of the feedforward control in effort to reduce variation to our goal as defined in our baseline investigation. It should be noted that cost of changing the value of z_{20} is $14 so it is a economically feasible option.

Part 3: Search for Solution I I

Question

In this part, we continue the remedial journey. The questions we want to answer by the end of this investigation is:

1. Is there any interaction between the fixed inputs and the identified dominant cause that are responsible for the variation seen in y_{300}?

Plan

The plan for the investigation is to identify a potential desensitizer (ie: if there is a fixed input that has an interaction with the dominant cause x_{51}). As it was found that the variation lies in the heat treatment phase meaning majority variation between y_{200} and y_{300}.To do this I will run an experiment where the output y_{300} against extremes in fixed inputs and extremes in the dominant cause values. I measured the fixed inputs that affect the dominant cause, however, I do not measure the categorical variables. Thus I measured z_{19}, z_{20}, z_{21}, and z_{22}.

The design for a fractional factorial experiment since it would be far too costly to run a full factorial investigation. I designed the experiment is designed as follows:

	z19	z20	z21	z22	x51
1	25	20	20	13	2
2	53	20	20	13	2
3	25	22	20	13	2
4	53	22	20	13	2
5	25	20	48	13	2
6	53	20	48	13	2
7	25	22	48	13	2
8	53	22	48	13	2
9	25	20	20	21	4
10	53	20	20	21	4
11	25	22	20	21	4
12	53	22	20	21	4
13	25	20	48	21	4
14	53	20	48	21	4
15	25	22	48	21	4
16	53	22	48	21	4

Figure 11: Experiment design

The above experiment will allow us to see how the fixed inputs and the dominant cause interact to affect the output value. I randomised the order of runs to reduce the risk of confounding in the experiment between and inputs. The cost of this investigation is $670.

Data

The data consists of 5 columns corresponding to the fixed input levels, a column for the dominant cause level and the output y_{300}. The data is made up of 16 rows of data, 8 rows consisting of different treatment combinations and then are run twice to total 16 rows, one at each level of the dominant cause.

Analysis

First, we plotted the data stratified by the fixed input. There appear to be subtle differences between the two extreme levels in each plot, however, no significant differences in the plots were noted.

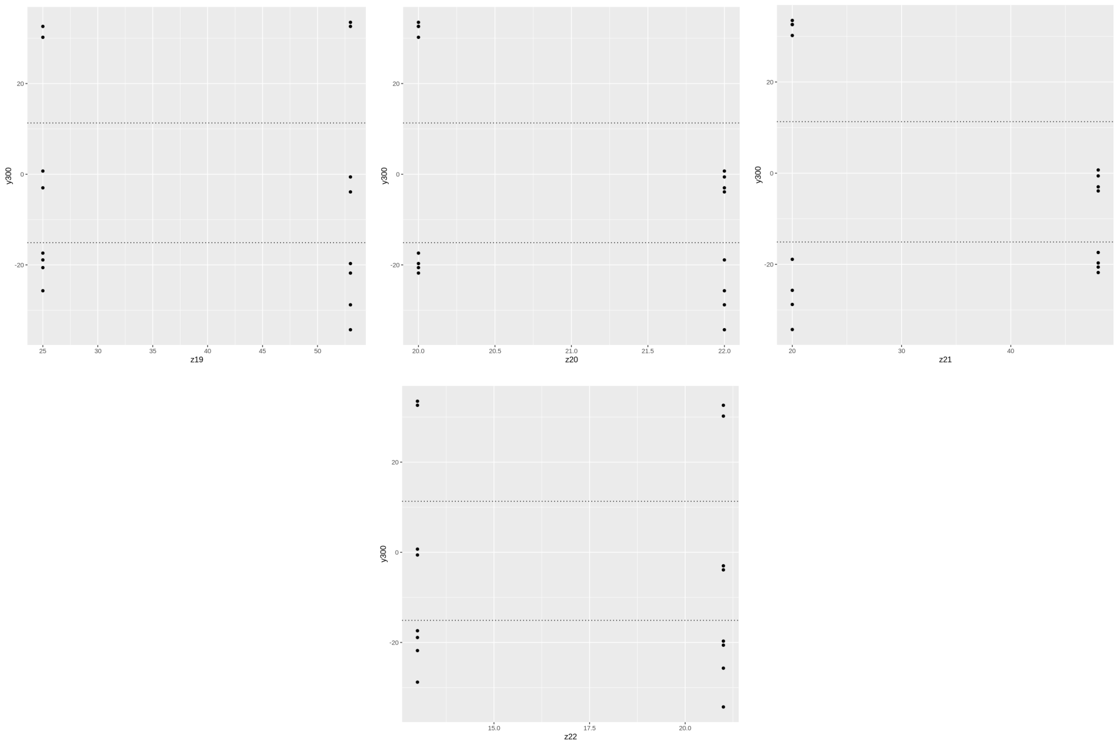

Figure 12: Fixed Input vs y_{300}

Next, I ran a linear model which accounts for the fixed inputs and interactions between the fixed inputs and the dominant cause as shown in Appendix E. The interaction plots showed that there is no relationship between z_{22} and x_{51}. Intuitively that makes sense as the coil length has no impact on the operator. The interaction plots can be seen below.

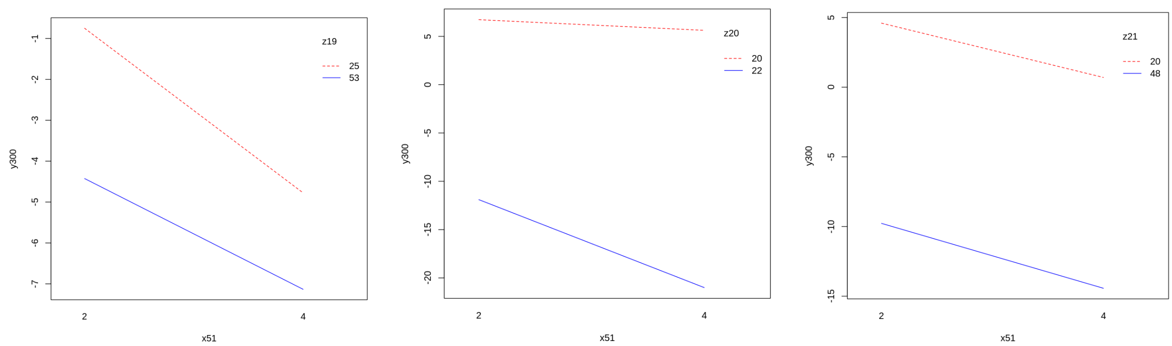

Figure 13: Interaction between fixed inputs and x_{51}

The interaction plots show very little interaction between the fixed inputs and the dominant cause. Similarly the linear regression analysis shown in Appendix E shows the Multiple R-squared value is 0.3836, which means that the model explains 38.36% of the variance in y_{300}. However, the Adjusted R-squared value is negative (-0.2328), which indicates that the model may be overfitting the data. The F-statistic tests whether at least one of the independent variables is significantly related to y_{300}, and the associated p-value of 0.7267 suggests that there is no significant relationship. To confirm the lack of a relationship I conducted an ANOVA analysis as shown in Appendix F. The analysis reveals none of the variables or interaction terms have p-values less than 0.05, which suggests that none of them have a significant effect on y_{300}. The interaction plots, linear regression and the ANOVA analysis together confirm that there is not interaction between the fixed variables and the identified dominant cause.

Conclusion

In conclusion, from the above analysis, it can be seen that there is no interaction between the dominant cause and fixed inputs. The investigation did not reveal a desentizer meaning moving forward using a desentizer for variation reduction is not a possibility for me.

Part 4: Testing Solution

Question

In this part, we continue the remedial journey. The questions we want to answer by the end of this investigation is:

1. Which solution is the most promising based on the feasibility and cost investigation?
2. Can the chosen solution achieve the variation reduction goal defined in the baseline investigation?

Plan

The plan for the investigation is to select a solution that would help to achieve the desired P_{pk} and variation outlined in the baseline.

I consider a cost benefit analysis on various approaches as follows.

Fix the Obvious
We could permanently fix x_{51}, the dominant cause of variation. Fixing x_{51} has an infinite cost. Thus, this is an infeasible approach for this investigation.

100% Inspection
With the 100% inspection approach, all parts outside the inspection limits are scrapped. Inspection on categorical inputs is not possible. As the variation lies in the welding stage, it could possibly do inspection on z_{19}, z_{20}, z_{21} and z_{22}. The cost for the inspection is essentially the scrapping cost. As the cost is $100/part and from our baseline investigation we know that around 4.8% of 600 samples fall out of the required range, keeping this number in mind the cost of conducting is very expensive for each of z_{19}, z_{20} and z_{21}. Thus, this approach is not beneficial given its high costs and continued for other methods for solution validation.

Make the Process Robust
Changing the process centre involves running the process with changed values of the fixed input. This approach is generally used in cases with unknown dominant cause. The previous investigation revealed x_{51} to be the domiant cause for the variation seen in y_{300} so this approach is not feasible for this experiment.

Change the Process Center
As the baseline investigation shows that the process is centered around 0, this would not help with the variation reduction thus I decided to skip this approach.

Feedback Control

This approach specifies the dominant cause (varying input), adjuster (fixed input), adjustment rule and a sampling plan. In the previous part, it was found z_{20} to be a possible adjuster for Feedback control. This is a feasible approach for this process, I implemented this approach as this approach specifies the output to measure, adjuster (fixed input), adjustment rule and a sam-
pling plan and there is strong time structure in output variation, that is, I have shift to shift variation. It was implemented after conducting some further investigation outlined later in this part. This approach costs \$5.76/part to adjust the variation. I also tried a different threshold to see if they yield better results. For threshold 1 the approach cost \$ 0.5/part and for threshold 3 it cost \$ 14.5/part.

Feedforward Control

In this approach,an adjuster (fixed input) in our case, z_{20}, as found in the previous parts is applicable here since we have shift to shift variation. This approach requires different values of the adjuster for different levels of the dominant cause. I attempted to implement this approach but did not see good results. I attempted it twice and did not see any significant improvement. The implementation would be expensive as changing the fixed value will cost a multiple of \$50/part (b = 50 from the given chart)as calculated from the formula in the figure below.

$$\text{Cost/part} = b\left(\frac{\text{new level} - \text{current level}}{\text{upper-lower}}\right)^2$$

Figure 14. Cost Calculation for changing the level of z_{20}

Desensitise the Process

Desensitising a process involves changing the value of a desensitizer (fixed input) to a level that reduces the variation. In a previous part, it was found that there is no interaction between the fixed inputs and the dominant cause. As I was unable to find a desentisiter this approach is redundant so I decided to skip it.

Data

The data was collected according to the plan. The data is the same as in the baseline investigation where we have the output value y_{300} along with columns such as the day, shift and part numbers. There are rows, 1 per part measurement.

Analysis

Feedforward Control

As previously identified that x_{51} is the dominant cause I wanted to use the following predicting equation (linear regression) found by conducting a linear regression analysis as shown in Appendix F. The linear regression reveals the following equation as shown in the figure below.

$$y_{300} = 3.150 - 12.062(x_{51}{=}2) - 4.444(x_{51}{=}3) - 1.837(x_{51}{=}4) + \varepsilon$$

Figure 15. Linear Regression Equation

The above image has the prediction equation required by the feedforward controlt approach but unfortunately the interface does not accept this format of equation. For categorical variable it requires the expected level of the adjuster for each category of the categorical variable. This requires knowing which value of the adjusted z_{20} for which operator during their shift will reduce the variable. I used the image below to calculate the optimal level of z_{20}.

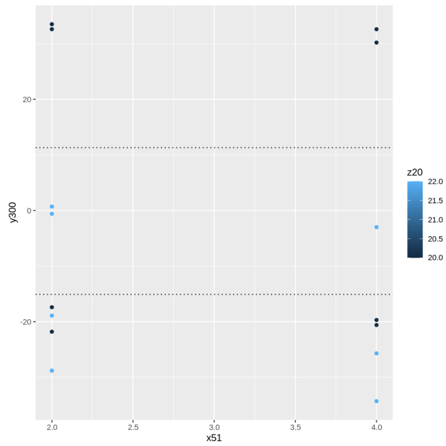

Figure 16. Scatterplot of impact of z_{20} on x_{51}

As it can be seen from the scatterplot z_{20} at 22 for the second operator (x_{51}= 2) has lesser variation impact on y_{300} than z_{20} at 22. Similarly for the fourth operator (x_{51}= 4) z_{20} at 22 shows lesser variation in y_{300}. Using this visual analysis I conducted the feedforward adjustment with z_{20} value adjusted to 22 from 21 for each of the operators. This adjustment cost me $12.5/part and did not return a good result as shown in Appendix G. The histogram shows that the adjusment did not reduce the variation enough for the process to become "capable" as there are still a few outliers in the histogram. The p_{pk} calculations confirm it as the variation is 5.56 and the p_{pk} value has not increased satisfactorily.

Due to the failure of this implementation, regardless of the price I tried a second implementation. This time I wanted to strategize the values of z_{20} for each operator. I spent $312 and collected data with z_{20}, x_{51} and y_{300} and plotted a scatterplot as shown in the figure below.

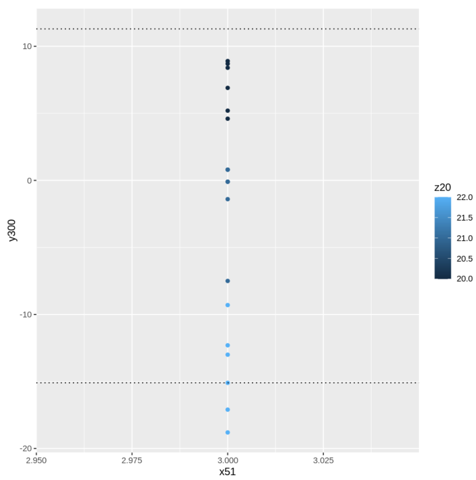

Figure 17. Scatterplot of impact of z_{20} on x_{51}= 3

As it can be seen that for the third operator (x51= 3) , z_{20} at 21.5 has the least variation, thus this time I implemented the solution with values as shown in the figure below.

x_{51}	z_{20}
1	22
2	22
3	21.5
4	22

Figure 18. Implementation 2 values

Unfortunatley even the second implementation did not work well as shown in Appendix H. Surprisingly the outliers increased along with the standard deviation and the P_{pk}. In conclusion the feedforward control process did not result in significant variation reduction as expected because feedforward control is not recommended , thus I moved on to implement feedback control as that is the other feasible option.

Feedback Control

Similar to Feedforward adjustment, feedback control requires an adjustment plan. The plan aims to know when the adjuster value should be adjusted and by how much. I tried two implementations of feedback adjustment. The equation was as shown in the figure below.

$$z_{new} = z + w$$
$$w = \begin{cases} 0 & if\ |\bar{y} - y_{target}| \leq g \\ m(\bar{y} - y_{target}) & otherwise \end{cases}$$

Figure 19. Feedback Control Adjustment Plan

From the above figure we cab see that g is the tolerance threshold, i.e the difference from the average that y_{300} variation model can handle and m is the adjustment amount i.e the intercept of the linear model signifying the change in z with one unit increase in y_{300}. The plan is to implement the adjustment at two different g values to see which one gets a better solution. Before the implementation it was crucial to calculate the m value. To define m, it is the inverse of the observed amount increase in y_{300} with an one unit increase in the adjuster. As shown in Appendix I we know that an unit increase in z_{20} decreases y_{300} by 10.1784. The inverse of that amount is approximately -0.098 thus we set m to be -0.098. Before running the feedback adjustment I plotted the adjusted z_{20} values and predicted y_{300} using the linear regression model in Appendix I to see the impact. The figure below shows the comparison between the adjusted and non adjusted values.

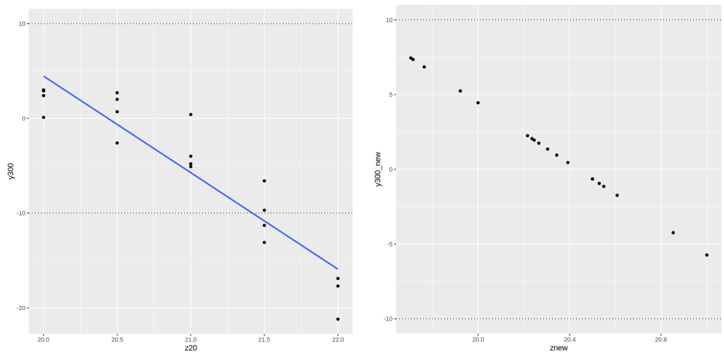

Figure 20. Predicted adjustment Impact

As it can be seen from the figure above the adjustment looks promising thus I implemented , this implementation cost me $5.8/per part. Data was collected for solution validation and I started the analysis with a visual analysis bu plotting a histogram as shown in figure below.

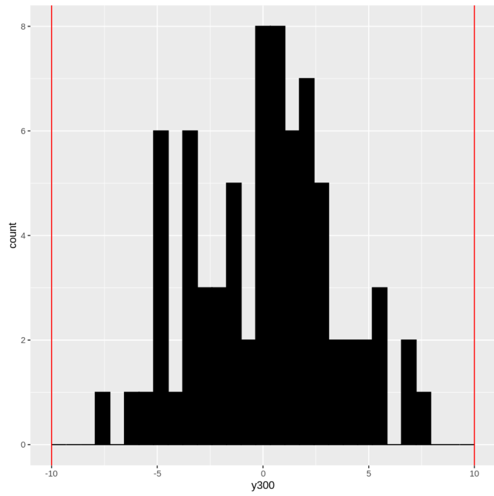

Figure 21. Histogram of the Solution

It is evident in the histogram that there are no outliers for y_{300}. This solution seems to have worked, to verify that I calculated the P_{pk} value as shown in the figure below.

```r
# Ppk Calculations
mean <- mean(solution2$y300)
mean
sd <- sd(solution2$y300)
sd
CpkL <- (mean-(-10)) / (3*sd)
CpkU <- (10-mean) / (3*sd)

# calc Ppk
Ppk <- min(CpkL, CpkU)
Ppk
```

```
0.124
3.36202320039586
0.979172303038357
```

Figure 22. P_{pk} calculations for the Solution

As it can be seen that the P_{pk} value is approximately 0.98, that is a large improvement from the baseline but the value is not one. So I decided to implement it again with a different g value. I implemented it with g = 1 and g = 3, the resulting histogram were as shown in the figure below.

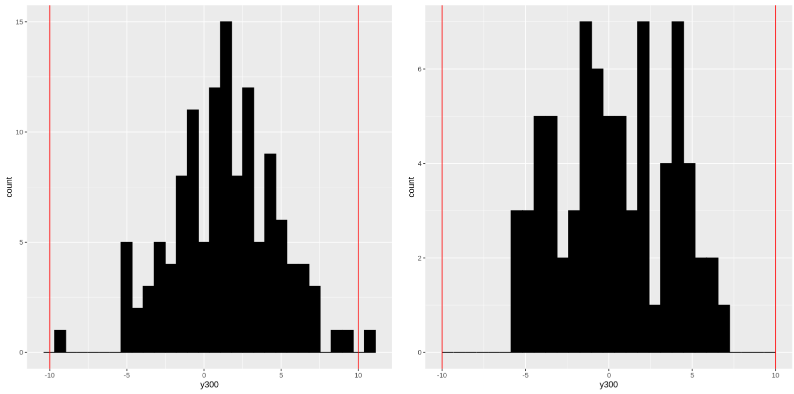

Figure 23.Histograms for g = 1 and g = 3

The histogram revealed that at g = 1 not all y_{300} values fall in the desired range for y_{300} and for g = 3 the histogram is not very centered around 0. The P_{pk} values for these experiments were lower that previous one. The calculations are shown in the figure below.

```
[146] # Ppk Calculations
mean <- mean(solution3$y300)
mean
sd <- sd(solution3$y300)
sd
CpkL <- (mean-(-10)) / (3*sd)
CpkU <- (10-mean) / (3*sd)

# calc Ppk
Ppk <- min(CpkL, CpkU)
Ppk
```

```
0.333333333333333
3.34984200245944
0.96190274641505
```

```
# Ppk Calculations
mean <- mean(solution5$y300)
mean
sd <- sd(solution5$y300)
sd
CpkL <- (mean-(-10)) / (3*sd)
CpkU <- (10-mean) / (3*sd)

# calc Ppk
Ppk <- min(CpkL, CpkU)
Ppk
```

```
1.348
3.34754604391471
0.861526611483847
```

Figure 24.P_{pk} calculation for g = 1 and g = 3

The solutions that I implemented, none of the implementations have a P_{pk} value above 1 but can be compared on other standards. For this comparison I did not consider the feedforward control due to their high cost and poor results.

In the image below there is a comparison table for the solutions.

	Feedback Control		
	g=1	g=2	g=3
FEoV	12.6	15.5	19.8
Range	[-5.7, 6.9]	[-7.7,7.8]	[-9,10.8]
Mean	0.33	0.124	1.348
SD	3.35	3.36	3.35
P_{pk}	0.96	0.98	0.86

Figure 25.Solution Comparison Table

It can be seen that all three versions for this approach yield good results. The aim of this investigation was to reduce the variation in y_{300} and increase the P_{pk} value. In the baseline investigation it was found that the variation in y_{300} was 5.55 which means that feedback control with g = 2 reduced 40% of the standard deviation and increases the P_{pk} by approximately 78%. The goal set in the baseline was to reduce the variation by 44.73% which was not met but a 40% reduction is very close to the set goal. The reduction in variation in shift to shift can be seen in the figure below.

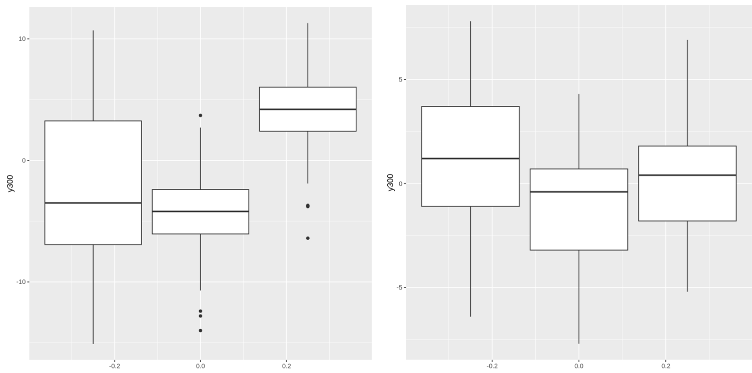

Figure 26.Shift-Shift Variation before and after adjustment

The other time family that had significant impact on the variation in y_{300} was day-to-day. The figure below shows the reduction in variation in the day-to-day family.

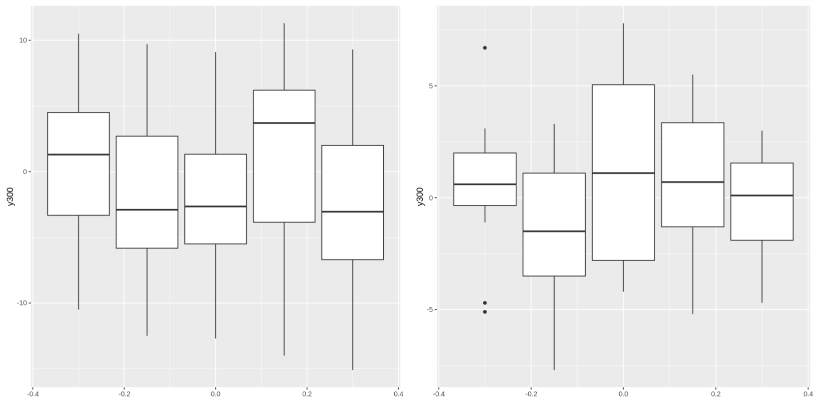

Figure 27.Day-Day Variation before and after adjustment

It is evident that the variation still exists but can be seen that there is less variation than previously seen.

Conclusion

In conclusion, I learnt that the feedback control method yielded better results than feedforward in this use case. Using z_{20} as an adjuster I was able to reduce the variation to 3.35 from 5.55 in the baseline. This process also increased the P_{pk} from 0.55 to 0.98. Eventhough the process is still not capable the varaition in y_{300} has reduced significantly (~40%). No other process was disrupted or any new variation added. I can successfully say that the goal to reduce variation in y_{300} was successfuly achieved.

Appendix

Appendix A: Linear regression for search of an adjuster

```
lm(y300 ~ (z19 + z20 + z21 + z22)^2, data) -> fitted_model
summary(fitted_model)
```

```
Call:
lm(formula = y300 ~ (z19 + z20 + z21 + z22)^2, data = data)

Residuals:
       1        2        3        4        5        6        7        8
 0.66875 -0.85625 -0.05625  0.24375  0.41875 -0.23125 -1.03125  0.84375
       9       10       11       12       13       14       15       16
-1.78125  1.96875  1.16875 -1.35625  0.69375 -0.88125 -0.08125  0.26875

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.116e+03  5.231e+01  21.326 4.20e-06 ***
z19          1.593e+00  6.640e-01   2.399   0.0617 .
z20         -5.067e+01  2.464e+00 -20.567 5.03e-06 ***
z21         -2.938e+01  6.654e-01 -44.158 1.12e-07 ***
z22          3.189e+00  2.298e+00   1.387   0.2240
z19:z20     -7.009e-02  3.076e-02  -2.278   0.0717 .
z19:z21     -2.232e-04  2.197e-03  -0.102   0.9230
z19:z22     -5.246e-03  7.691e-03  -0.682   0.5255
z20:z21      1.367e+00  3.076e-02  44.422 1.09e-07 ***
z20:z22     -2.422e-01  1.077e-01  -2.249   0.0743 .
z21:z22      1.462e-02  7.691e-03   1.901   0.1157
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.723 on 5 degrees of freedom
Multiple R-squared:  0.9984,	Adjusted R-squared:  0.9951
F-statistic: 307.8 on 10 and 5 DF,  p-value: 2.474e-06
```

Appendix B: ANOVA Analysis for search of adjuster

```
aov <- aov(y300 ~ (z19 + z20 + z21 + z22)^2, data)
summary(aov)
```

```
            Df Sum Sq Mean Sq  F value   Pr(>F)    
z19          1      2       2    0.637   0.4610    
z20          1   1956    1956  659.063 1.67e-06 ***
z21          1    619     619  208.505 2.87e-05 ***
z22          1    659     659  222.132 2.46e-05 ***
z19:z20      1     15      15    5.191   0.0717 .  
z19:z21      1      0       0    0.010   0.9230    
z19:z22      1      1       1    0.465   0.5255    
z20:z21      1   5856    5856 1973.321 1.09e-07 ***
z20:z22      1     15      15    5.060   0.0743 .  
z21:z22      1     11      11    3.614   0.1157    
Residuals    5     15       3                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix C: $z_{20} \sim y_{300}$ Linear Regression

```
lm(y300 ~ z20, data)-> fitted_model
summary(fitted_model)
```

```
Call:
lm(formula = y300 ~ z20, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.3417 -1.4042  0.0833  1.5646  5.3583 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  203.892     21.729   9.383 2.84e-06 ***
z20           -9.525      1.034  -9.212 3.35e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.924 on 10 degrees of freedom
Multiple R-squared:  0.8946,	Adjusted R-squared:  0.884 
F-statistic: 84.87 on 1 and 10 DF,  p-value: 3.354e-06
```

Appendix D: ANOVA Analysis for z_{20}~y_{300}

```
aov <- aov(y300 ~ z20, data)
summary(aov)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
z20          1  725.8   725.8   84.87 3.35e-06 ***
Residuals   10   85.5     8.6                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix E: Linear Regression for Fixed Inputs and Dominiant Cause

```
model <- lm(y300 ~ x51*z19+x51*z20+x51*z21, data)
summary(model)
```

```
Call:
lm(formula = y300 ~ x51 * z19 + x51 * z20 + x51 * z21, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-22.250 -17.400  -8.375  19.137  29.125 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  59.5098   478.4845   0.124    0.905
x51          78.0196   158.5380   0.492    0.638
z19           0.1473     1.5862   0.093    0.929
z20          -2.4875    22.2075  -0.112    0.914
z21          -0.2259     1.5862  -0.142    0.891
x51:z19      -0.1393     0.5192  -0.268    0.796
x51:z20      -3.4125     7.2691  -0.469    0.653
x51:z21      -0.1437     0.5192  -0.277    0.790

Residual standard error: 27.41 on 7 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.3836,	Adjusted R-squared:  -0.2328 
F-statistic: 0.6223 on 7 and 7 DF,  p-value: 0.7267
```

Appendix E: ANOVA Analysis for Fixed Inputs and Dominiant Cause

```
aov <- aov(y300 ~ x51*z19+x51*z20+x51*z21, data)
summary(aov)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
x51          1     38    38.2   0.051  0.828
z19          1     35    34.9   0.047  0.835
z20          1   1901  1900.5   2.529  0.156
z21          1   1067  1067.0   1.420  0.272
x51:z19      1     29    28.8   0.038  0.850
x51:z20      1    146   146.4   0.195  0.672
x51:z21      1     58    57.6   0.077  0.790
Residuals    7   5261   751.5
1 observation deleted due to missingness
```

Appendix F: Linear Regression Analysis for Feedforward Adjustment

```
model <- lm(y300 ~ as.factor(x51), search3)
summary(model)
```

```
Call:
lm(formula = y300 ~ as.factor(x51), data = search3)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.7875 -1.4063  0.1125  1.8938  6.1875 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)        3.150      1.090   2.889  0.00650 ** 
as.factor(x51)2  -12.062      1.542  -7.823 2.82e-09 ***
as.factor(x51)3   -4.444      1.335  -3.328  0.00203 ** 
as.factor(x51)4   -1.837      1.542  -1.192  0.24117    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.084 on 36 degrees of freedom
Multiple R-squared:  0.6639,	Adjusted R-squared:  0.6359 
F-statistic:  23.7 on 3 and 36 DF,  p-value: 1.21e-08
```

Appendix G: Feedforward Control Adjusment Implementation 1

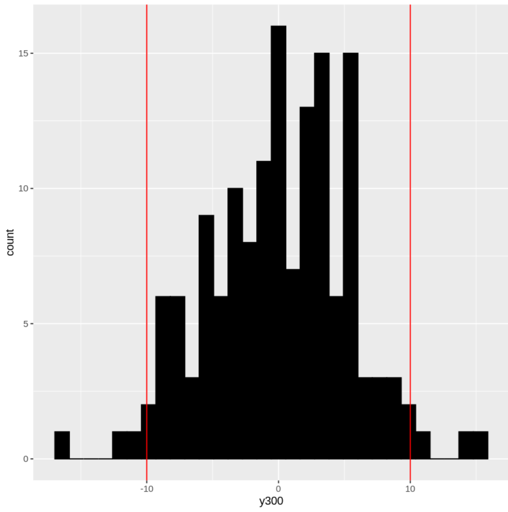

Figure G1. Histogram of the Solution

```
# Ppk Calculations
mean <- mean(solution4$y300)
mean
sd <- sd(solution4$y300)
sd
CpkL <- (mean-(-10)) / (3*sd)
CpkU <- (10-mean) / (3*sd)

# calc Ppk
Ppk <- min(CpkL, CpkU)
Ppk
```

```
0.0633333333333333
5.36481098157325
0.617397748699601
```

Figure G2. P_{pk} *calculations for the Solution*

Appendix H: Feedforward Adjusment Solution Implementation 2

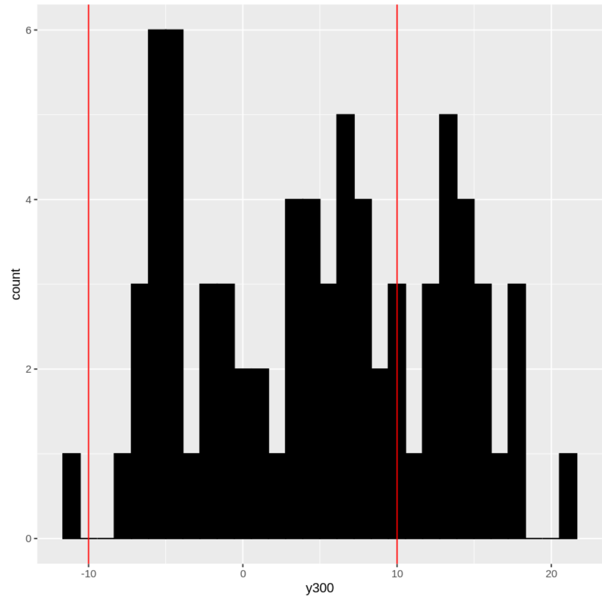

Figure H1. Histogram of the Solution

```r
# Ppk Calculations
mean <- mean(solution5$y300)
mean
sd <- sd(solution5$y300)
sd
CpkL <- (mean-(-10)) / (3*sd)
CpkU <- (10-mean) / (3*sd)

# calc Ppk
Ppk <- min(CpkL, CpkU)
Ppk
```

```
4.84133333333333
7.97586042691004
0.215594990824299
```

Figure H2. P_{pk} *calculations for the Solution*

Appendix I : Adjustor Impact Analysis

```
model <- lm(y300 ~ z20,adj)
summary(model)
```

```
Call:
lm(formula = y300 ~ z20, data = adj)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.2765 -1.8670 -0.4701  1.5393  6.1362 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 208.1982    20.8049   10.01 1.53e-08 ***
z20         -10.1874     0.9927  -10.26 1.06e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.969 on 17 degrees of freedom
Multiple R-squared:  0.861,	Adjusted R-squared:  0.8528 
F-statistic: 105.3 on 1 and 17 DF,  p-value: 1.057e-08
```

Appendix J : Code

https://colab.research.google.com/drive/1z_XRsBWMOuLkJxjl3ielZDqAbkNJiR6r?usp=sharing

Appendix K : Within team feedback

Team Member	Score
Gunchica Bhalla	5
Caitlan Krasinski	5
Olivia You	5