Pretzel Process Quality Control Analysis

University of Toronto: Faculty of Applied Science and Engineering

Jonah Ernest

1.0 Introduction and Manufacturing Process

With over $550M of pretzels sold each year in the United States, one would think that pretzel production has been optimized to reduce errors and minimize out of control samples [1]. Furthermore, the pretzel manufacturing process is almost completely automated, with approximately 90% of pretzels never being touched by humans [2]. To ensure that pretzels are meeting high quality standards, it is important to implement quality control methods and monitor various quality characteristics throughout and after the production of goods. This report will first provide an overview of the manufacturing process of Snyder's Pretzel Rods (see Table 1) [3]. Then, data was collected including both continuous variable and attribute characteristics, which were further analyzed using control charts, statistics, and quality control methods. These findings are then discussed to suggest the most appropriate methods, allowing pretzel manufacturers to more efficiently detect quality shifts and out of control samples.

Table 1: The various steps involved in the manufacturing process of Snyder's Pretzel Rods pretzels

Manufacturing Process	**Description**
Mixing dough	The first step in the manufacturing process is to create the dough by combining and mixing raw ingredients into factory compounders in steel tanks such as the flour, yeast, shortening, water, salt, sugar and more. Since pretzel dough is mixed much less than other bread products, it is not sticky and is then fermented.
Forming the pretzels	An extrusion device is used to shape it into the rod shape using a wire cutter. Once in their proper form, they are then transferred to a conveyor belt where they are rolled to create a flat shape. This step in the process is what determines the length of the pretzel as it ensures relatively uniform size across pretzels. The target length of the pretzel rods are 10cm with a specified acceptable range of 8.5cm to 11.5cm.
Dipping and salting	The pretzels are now given a few minutes to continue fermenting. They are then placed in a heated alkaline bath for up to 20 seconds or until they start to float, allowing them to gelatinize to help the salt stick to their surface.
Cooking	Pretzels are most commonly cooked in gas convection tunnel ovens anywhere from 350-550ºF for about 4-8 minutes, and then again at 250ºF for 20-40 minutes. As the pretzels move along the conveyor belt and through the oven, the temperature gradually decreases allowing it to initially caramelize. This change in temperature also helps prevent them from developing a weak structure which often leads to the pretzels breaking, chipping and splitting while being shipped.
Packaging	After cooling, they are moved into the packaging machines, where they are weighed and put into their air tight packaging.

2.0 Data Collection

The team purchased a box of packaged pretzel stick snacks from the grocery store which contained 20 individually wrapped bags that each contained 6 large pretzel rods. The team measured the length of each pretzel in the bags, focusing on one bag at a time. Since there were three team members, two members measured the length of each pretzel in centimeters, to the nearest millimeter while one was tasked with recording the found length in a spreadsheet. For each individual pretzel, the two members measured each pretzel separately to find discrepancies and reduce the risk of human error. If there was discrepancy among the two measurements for a particular pretzel, both teammates remeasured until consensus to the nearest millimeter was found and recorded. In addition to the length of the pretzel, it was also noted which subgroup (bag) the unit came from, the number of chips in the rod (integer amount), and whether the pretzel was broken or not in any manner, represented by a binary variable (1 if broken). This broken attribute and number of chips will later be used to construct the control charts for attributes. The format of the data can be seen in the excerpt of the spreadsheet in Figure 2.

Observation	Sample Number	Length of Pretzel (cm)	Broken Attribute	Number of Chips
1	1	9.6	0	0
2	1	10.3	0	1
3	1	10.3	0	0
4	1	9.9	1	2
5	1	10.2	0	2
6	1	9.9	0	3
7	2	10.5	0	1

Figure 2. Excerpt of collected data.

2.1 Data Visualisation and Basic Diagrams

Diagrams were created to summarize and better visualize the collected data. This includes a stem and leaf plot which is helpful to organize data as its collected, a histogram to summarize the length of the pretzels by grouping them into bins, a cumulative frequency plot to show the number of observations above (or below) any value in the dataset (such as the quartiles), and a time series plot to visualize any trends in the data (see Appendix A). A cause and effect diagram was created in order to identify the possible causes leading to defective observations, primarily measurement errors, variations in materials and ingredients, transporting the product, and faulty machinery can all possibly lead to broken or chipped pretzel rods (see Appendix B). Finally, an out-of-control-action plan (OCAP) was created for the pretzel manufacturing process, which includes many of the potential activating events related to Figure 1, and is used later to assign possible causes to out-of-control samples (see Appendix C).

2.2 Statistical Analysis

In addition to visualizing the data graphically, basic statistics were calculated to gather an understanding of the distribution of the data and identify key values such as the minimum and maximum observations, mean, standard deviation, and other values shown in Figure 3 below.

Statistics

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median
Length of Pretzel (cm)	120	9.9430	0.0317	0.3477	0.1209	8.9000	9.7000	9.9000

Variable	Q3	Maximum	Range	IQR	Mode	N for Mode
Length of Pretzel (cm)	10.2000	10.7000	1.8000	0.5000	9.7, 9.8, 10.1	14

Figure 3. Basic statistics of the length of pretzels.

While the data appeared to be normally distributed in the histogram, the team wanted to confirm this observation by using Minitab to identify the best-fitting distribution. A goodness of fit test was conducted on the length of pretzel data and it did in fact identify the normal distribution to be the best fit with a p-value of 0.064 which is greater than 0.05 meaning it is an "acceptable" fit. The results of the test can be found in Appendix D .

To explore the data further, the team calculated the 95% confidence interval of the mean of the data using both the normal and student-t distributions. Given that the number of observations is 120, the normal test statistic should be a sufficient measure, however, the t-test is always appropriate so it was calculated for both. With 95% confidence, the mean interval rounds to [9.88,10.01] for both distributions as shown in Appendix E.

Furthermore, a hypothesis test was conducted to test the manufactured target mean of 10cm against the sample mean to draw conclusions about the entirety of the pretzel population. The t-test and z-test failed to reject the null hypothesis as the test statistics did not fall within the rejection region and the p-value of the tests were both greater than 0.05. This is reassuring for the manufacturing process because it means that with 95% confidence, the mean will adhere to the 10cm target. The test and the distribution of the sample mean against the target mean (using the standard deviation of the sample) can be referred to in Appendix F.

3.0 Control Charts for Variables

The team first constructed four control charts, using minitab, to analyze the difference among ranges, means, standard deviations, and variances of each of the 20 sample subgroups. In order to simulate a Phase 1 and a Phase 2 manufacturing process, the team used a 50-50 split on the data to create a training and testing set. These simulations of a Phase 1 process will suggest whether the process is in control or not. The team then used the calculated control limits, as well as the statistical values from the Phase 1 charts to create four second control charts using the testing sample sets. In addition to these values, the software required the standard deviation to be calculated from the center line statistic of the Phase 1 chart. The control limit formulae as well as the constants pulled from the table of control chart constants were used to calculate the new standard deviations for the testing set. These Phase 2 processes will help confirm if the process was in control or not.

3.1 X-Bar & R Charts

The team first constructed an $\overline{X}$ & R chart. The subgroup average means and ranges with their respective upper and lower control limits and formulas are shown in Appendix G. For the $\overline{X}$ chart the LCL = 9.534, CL = 9.959, and UCL = 10.385, and for the R chart LCL = 0, CL = 0.88, and UCL = 1.763. As can be seen, there are no samples in the training sets that are out of control. This simulation of a Phase 1 process suggests that the process is in control.

The team then created a Phase 2 $\overline{X}$ & R chart using the testing sample sets. The formula for the calculation of the standard deviation provided to the software are also seen in Appendix G. This Phase 2 process confirms that it was in control as the original control limits were used and no out of control points were detected. These charts can be extremely useful when talking about the process of the pretzels' manufacturing as ranges between these pretzel samples and their means allows for analysis of continuous production processes.

3.2 X-Bar & S Charts

The next chart the team constructed is an $\overline{X}$ & S. The subgroup average means and standard deviations with their respective upper and lower control limits can be seen in Appendix H. For the $\overline{X}$ chart the LCL = 9.533, CL = 9.959, and UCL = 10.386, and for the S chart LCL = 0.0101, CL = 0.3315, and

UCL = 0.6529. There are no samples in the training set that are out of control which suggests that the Phase 1 process is in control.

The team then used the calculated control limits to create a second $\overline{X}$ & S chart, but replaced the training sets with the testing sample sets. This Phase 2 process confirms that the process is in control since the original control limits were used and no out of control points were found. These charts are used when the subgroup sizes are large. In the team's case, each subgroup has a size of 6, making this chart's application less useful than the $\overline{X}$ & R charts.

3.3 X-Bar & S^2 Charts

When constructing the $\overline{X}$ & S^2 charts, the team performed the necessary calculation on Excel as there was no functionality to create these charts on minitab. The team manually calculated the variances and means of all samples and subgroups accordingly as well as the $\overline{X}$ & $\overline{S}^2$ values. These were used to find the LCL, CL, and UCL values for both the $\overline{X}$ chart and the S^2; the team assumed $\alpha = 0.05$ for a two standard deviation control limit. The subgroup average means and variances with their respective upper and lower control limits can be seen in Appendix I. For the $\overline{X}$ chart the LCL = 9.793, CL = 9.959, and UCL = 10.125, were calculated using $A_3 = 1.287$ and the following formula $\overline{\overline{X}} \pm A_3 \cdot \overline{S}^2$. For the $\overline{S}^2$ chart, the LCL = 0.0659, CL = 0.1238, and UCL = 0.1964. Several samples were found to be out of control in the Phase 1 implantation of the $\overline{X}$ & S^2 charts. Referring to the OCAP diagram, this is likely a result of measurement error in the team's data collection strategy, or possibly inconsistent rolling of the pretzel dough in the manufacturing process.

When the Phase 2 process was performed, less of the subgroup points were out of control. This could be due to the better and more consistent measurements of the pretzels in the second half of the samples.

3.4 I - MR Charts

The team constructed an I-MR chart using Minitab. This chart was constructed to analyze the difference between each of the 120 individual values and their moving ranges. The values of the 120 samples and their moving ranges can be seen in Appendix J, as well as the upper and lower control limits and their formulas. For the individual values, the control limits are LCL = 8.959, CL = 9.959, and UCL = 10.960, and the moving range (MR) values have control limits of LCL = 0, CL = 0.376, and UCL = 1.229. It can be seen that the individual values chart has one sample out of control, number 36, and the MR chart has sample 37 out of control. This may be due to the large difference between the measurements of sample 36 and 37, and the fact that sample 36 is the shortest pretzel of the 120 samples.

The team then calculated the new standard deviation for the Phase 2 chart using the formulas in Appendix J. It can be seen that there are many more out of control samples in the individual and the moving range chart. This may be a result of graphing the values with the old control limits. Though the Phase 2 process conducted on the testing set was out of control, the I-MR chart is used on samples that had very small variances and the process is slow. The pretzel manufacturing process can have large variances in pretzels, up to more than 1 cm, and the production process is quite fast. These aspects of the manufacturing process make the I-MR chart of little importance.

3.5 Average Run Length for the X-Bar S Chart

The team calculated the probability of not detecting a shift in the sample mean. In other words, given that the process is out of control, what is the probability that the shift in the sample mean is not detected. This probability is represented by:

$$\beta = P\left(LCL \leq \bar{x} \leq UCL \mid \mu_{new} = \mu_0 + k\sigma\right) = \Phi\left(\frac{UCL-\mu_0+k\sigma}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{LCL-\mu_0+k\sigma}{\sigma/\sqrt{n}}\right) = \Phi\left(L - k\sqrt{n}\right) - \Phi\left(-L - k\sqrt{n}\right)$$

The number of samples per observation in n = 6. As the team was attempting to find a 3σ limit, L=3 was assumed. In addition to these values, the constant k was set equal to 2, a common value used in industry when assessing shift in sample means. Substituting these values results in the following β = 0.02938. This means that, if the process is out of control, there is a 0.02938 probability of not detecting the shift in the process. Therefore, it can be concluded that in the overwhelming majority of the cases when the process is out of control, the shift will be detected. The team then proceeded to calculate the average run length of the mean and standard deviation chart using the following formula.

$$ARL = \frac{1}{1-\beta} = \frac{1}{1-0.02938} = 1.03$$

This means that the expected number of samples to detect a shift of 1σ with n = 6 is around 1. This once again supports the conclusion that the system of defect detection implemented is extremely effective.

4.0 Control Charts for Attributes

In order to construct control charts for attributes, the team collected two additional pieces of data, the broken attribute and the number of chips, both of which can be seen in Figure 2 or in the attached Excel file under the "Attributes" tab. The broken attribute takes on a value of one if the sample in observation *i* was a broken pretzel rod, and zero otherwise. The chips attribute is the number of chips or cracks in the pretzel rod, and if the pretzel is broken, then it counts as an additional crack/chip. Furthermore, similarly to control charts for variables, this section will utilize a training and testing data set. The training set is the first ten samples, and is used to create each of the control charts, whereas the testing set is comprised of the last ten samples and are monitored using the control charts that were created to determine if the process is in control or not.

4.1 Fraction Nonconforming (p) Chart

Fraction nonconforming is the ratio between the number of nonconforming items in the population, with the total number of items in that population. In order to construct the P-chart, the sample fraction nonconforming and average of the sample fractions nonconforming were calculated using the broken attribute data and the following formulae:

$$\hat{p}_i = \frac{D_i}{n}, \text{ and } \bar{p} = \frac{\sum_{i=1}^{m} D_i}{mn} = \frac{\sum_{i=1}^{m} \hat{p}_i}{m} = 0.15$$

for any sample i, where there are n=6 observations in each of the 20 samples, and D_i is the total number of nonconforming (broken) observations) and has a binomial distribution with parameters n and p. After finding the D_i's, the P-chart was constructed in Appendix K along with the calculations for the control limits, which are as follows: UCL = 0.5873, CL = 0.15, and LCL = 0.

From the first chart, we can see that in Phase I, there were no out of control samples, with the majority being centered at zero (the LCL), so it indicates that the process was in control. Even after simulating Phase II by using the data from samples 11-20 but with the control limits that were established from the previous phase's data, we can see that there were again no outliers, indicating that the process is stable and controlled.

4.1.1 P-chart with Varying n

A modified p-chart was created using the modified dataset with varying n to account for the pretzel samples having different numbers of rods in each one. The smallest sample only had three rods, and the maximum remained at six rods per sample. The variable-width control limits method was used to create the p-chart with varying samples sizes, and the average fraction nonconforming was calculated as follows: $\bar{p} = \frac{\sum_{i=1}^{20} D_i}{\sum_{i=1}^{20} n_i} = 0.1277$ (the center line). The control limits were then calculated, and the control chart was assembled (see Appendix L). Once again, all the samples in both the training and testing sets were outside of the control limits, so it can be said that the system is in control.

4.2 Number Nonconforming (np) Control Chart

This control chart is now monitoring the number nonconforming rather than the fraction nonconforming. The control limits are UCL = 3.524, CL = 0.9, and LCL = 0. The full control charts for the training and testing sets, as well as the calculations for the control limits can both be found in Appendix M. From these charts we can see that all the samples in both the testing and training sets are within the limits, so it can be said that the process is in control. Additionally, it is observed that the np-charts and p-charts have the same trends and shapes, the only difference being the y-axis is in terms of conformities rather than proportion.

4.3 Control Chart for Nonconformities (c) Chart

The c-chart monitors the number of chips in each pretzel unit by assuming the number of these nonconformities in each sample is modeled by a Poisson distribution. The parameter for the poisson distribution is $\bar{c}$, which is calculated by taking the sum of all the defects across all units in each sample, by the number of samples. This is then used to find the control limits for the c-chart which are as follows: UCL = 10.69, CL = 4.4, and LCL = 0. See Appendix N for the charts and calculations of the control limits). Similarly to the previous control charts for attributes, the nonconformities of each sample were all within the control limits, and so even after using the chart to monitor the testing set, the process was in control.

4.4 Control Chart for Average Number of Nonconformities per Unit (u) chart

The u-chart is based on the average number of nonconformities per inspection unit. In this case, the u-chart will be used to monitor the average number of chips/cracks per pretzel rod. This was done using the formula $u = \frac{x}{n}$, where x is the total number of conformities in a sample of n inspection units, and similarly to the c-chart, x is a Poisson random variable. The control chart was then created with the following control limits UCL = 1.782, CL = 0.733, LCL = 0. See Appendix O for the control charts and calculations). After creating the chart with the training set of samples, none of them were outside of the control limits, and even after using this chart to detect out of control samples in the testing set, none were beyond the control limits, so it can be said that the process is in control. Similarly to how the trends observed in the np-chart trends were the same as that of the p-chart, the trends in the u-chart were the same as that of the c-chart, the only difference being the y-axis is now the average number of nonconformities rather than the number of nonconformities per inspection unit.

4.5 Operating Characteristic (OC) Curves

An operating-characteristic function of the fraction nonconforming control chart acts as a visual representation for a type II error against the process fraction nonconforming. This acts as a measure for the sensitivity of the p-chart to detect a shift in the process from $\bar{p}$ to some other value p. This can be seen through the calculation of the type II error:

$$\beta = P\{\hat{p} < UCL|p\} - P\{\hat{p} \leq LCL|p\} = P\{D < nUCL|p\} - P\{D \leq nLCL|p\}$$

Using the parameters from the p-chart, in addition with a significance level of 0.0027, sample size of n=6, and comparing proportions of 0.01, 0.015 and 0.03, the power curve for one proportion was obtained and can be seen in Appendix P.

5.0 Process Capability Analysis

Process capability describes the uniformity of the process and the variability of critical-to-quality characteristics. These characteristics are a measure of how uniform the output of the process is. Analyzing process capability can provide many interesting and significant insights on which phase the process is in, and how the process can be shifted to achieve an in control process.

5.1 Anderson-Darling Normality Fit

Before being able to calculate the process capability ratios, the team had to ensure that the observations in Phase 2 were normally distributed. First a Phase 2 I-MR chart was constructed to identify and remove out of control points (see Appendix Q). The remaining observations were ten used to conduct the Anderson-Darling fit test. This test helps identify which distribution best fits the observations from Phase 2. The graphs of various distributions can be seen in Appendix R where it can be observed that the data is best fit by a normal distribution. In addition, when comparing the P-values of the various distributions, the normal distribution had the largest P-values making it the best fit for the observations. This can also be confirmed by the Anderson-Darling statistics and the P-value of the normal distribution. This fit test results in a P-value of 0.063, which is larger than 0.05, once again reaffirming that the data fits a normal distribution. From the Anderson-Darling fit test, the team was able to extrapolate the mean and variance for the normal distribution; $\hat{\mu} = 9.94$, $\hat{\sigma} = 0.347$. These can be seen, along with other statistics, in Appendix R.

5.2 Cp, Cpk, Cpm, and Process Fallout

In order to express process capability in a quantitative manner, the team used the process capability ratio, C_p. For this ratio, the team needed to use the standard deviation found using at the beginning of the report, $\hat{\sigma} = 0.347$, as well as the previously stated target pretzel length of 10 cm and lower and upper specification limits; LSL = 8.5 cm, USL = 11.5 cm. Using the following formula the team was able to determine the extent to which units are produced within the given specifications.

$$C_p = \frac{USL - LSL}{6 \cdot \hat{\sigma}} = \frac{11.5 - 8.5}{6 \cdot 0.347} = 1.44$$

This value can also be confirmed by looking at the process capability summary graphics in Appendix S. Having a $C_p = 1.44$ suggests that the process is currently under statistical control as it exceeds the industry standard threshold of 1.33 [4]. Furthermore, the team determined the percent of availability capacity that is being used, P, through the following formula:

$$P = \frac{1}{C_p} \cdot 100 = \frac{1}{1.44} \cdot 100 \simeq 69\%$$

This suggests that 69% of the available capacity is being used in the pretzel manufacturing process. This may be due to a shift in the process causing it to be off-center. For cases where the process may be experiencing such a trend, the C_{pk} process capability parameter can be used instead to better reflect the true process capability ratio.

$$C_{pk} = min\{C_{pu}, C_{pl}\} = min\{\frac{USL - \mu}{3\cdot\sigma}, \frac{\mu - LSL}{3\cdot\sigma}\} = min\{\frac{11.5 - 9.94}{3\cdot0.347}, \frac{9.94 - 8.5}{3\cdot0.347}\} = min\{1.49, 1.38\} = 1.38$$

Generally, if $C_p = C_{pk}$, the process is said to be centered at the midpoint of the specification limits. If $C_{pk} < C_p$ then the process is said to be off center. Their relative magnitudes are a description of how off-center the process really is. As C_p measures the potential capability and C_{pk} measures the actual capability, the ratio between these two characteristics also supports the conclusion that 69% of the available capacity is being used in the process.

Yet another issue arises, as sometimes the process mean becomes variable, making it difficult to calculate the process capability characteristics. To tackle this issue, the team calculated the C_{pm}, which depends on the variability of the sample mean over time, where the target T is defined as :

$$T = \frac{USL + LSL}{2} = \frac{11.5 + 8.5}{2} = 10 \quad , \quad C_{pm} = \frac{USL - LSL}{6\cdot\sqrt{\sigma^2 + (\mu - T^2)}} = \frac{11.5 - 8.5}{6\cdot\sqrt{0.347^2 + (9.94 - 10^2)}} = undefined$$

The C_{pm} takes the bias of the target directly into consideration, and helps identify the bias in the process which is the difference between the process average and the target. Here we can see that the bias is 10 - 9.94 = 0.06. This means that the current in control process is relatively unbiased, and is aligned with the center line 94% of the time.

Using all this information we can then calculate the process fallout, which is the probability that a unit will be outside of the specification bounds, using the following formula:

$$PF = 1 - P(LSL \leq x \leq USL) = 1 - P\left(\frac{LSL - \hat{\mu}}{\hat{\sigma}} \leq z \leq \frac{USL - \hat{\mu}}{\hat{\sigma}}\right) = 1 - P\left(\frac{8.5 - 9.94}{0.347} \leq z \leq \frac{11.5 - 9.94}{0.347}\right)$$

$= 1 - P(-4.15 \leq z \leq 4.49) = 0.00004\, ppm$ (parts per million)

This process fallout rate indicates that for each one million batches of sample, 4 of the units produced will be outside the specification limits. This is a very good result for Snyder's pretzels, as it has relatively little out of control units. This, along with the other process capability characteristics that were calculated, can be seen in the process capability report in Appendix S.

5.3 Six Pack Chart

The six pack process capability chart is a summary of the $\overline{X}$ & R chart, the distributions of value among sample groups, the normal distribution of the observations, and the overall capability plot where the different process capability characteristics and ratios are presented. The potential within capacities on this graph, seen in Appendix T, such as C_p, C_{pk}, C_{pm}, C_{pu}, and C_{pl} are all shown along with the distribution of the Phase 2 observation along their possible pretzel length values. As it can be seen the process is almost if not exactly centered between the specification limits allowing for a very efficient and non-wasteful manufacturing process.

6.0 CUSUM & EWMA Charts

The cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts are more advanced techniques used for statistical process monitoring and control. They are useful in phase II after assignable causes of large shifts in the parameters being measured have been addressed. For example, had a significant change been made to the pretzel manufacturing process, the Shewhart charts explored in earlier sections would have been beneficial to process the data collected in phase I. However, the Shewhart control charts only consider previous sample observations rather than the entirety of samples collected in sequence over time. Therefore, they lack the ability to detect very small shifts in the process. In general, the Shewhart charts are unreliable in detecting shifts of 1.5σ or less, highlighting the importance of alternatives that are more sensitive to shifts.

6.1 CUSUM Charts

The CUSUM chart is a plot of the cumulative sum of the deviation of the sample parameter from a value of interest. Considering the data being analyzed, the CUSUM chart can be used to detect very small process shifts, acknowledging that the pretzels being measured were manufactured and gathered at phase II of the statistical process. As discovered earlier, the Shewhart charts had difficulty detecting out-of-control samples, likely due to the fact that clear assignable causes of variation would have been eliminated before the product reached the shelves of grocery stores (in phase 1).

The team built a CUSUM chart to monitor the mean and variance of samples of size n=1 as well as for the entire data set where the size of each bag (sample) is n=6. To build a chart for n=1, the first sample measured from each bag was extracted for a total of 20 observations (1 from each of the 20 samples). The target value to measure deviation from the length against for n=1 was the mean of the 20 observations and the mean of all observations for n=6. Using this target mean (μ_0) for the process, the quantity plotted is $C_i = \sum_{j=1}^{i} (\bar{x}_j - \mu_0)$ for sample sizes of n>1 and $C_i = \sum_{j=1}^{i} (x_j - \mu_0)$ for sample size n=1.The cumulative sum of the deviation between the pretzel length of each observation and this target mean (μ_0) was plotted after transforming the data into an interpretable format with control limits. This was achieved using the tabular (algorithmic) CUSUM by separating the accumulations into two statistics called one-sided upper and lower CUSUMs. The C_i^+ statistic accumulates deviations above the target mean and C_i^- accumulates those below the target mean. Each statistic considers a slack variable (K=kσ) so that only deviations beyond this slack are accumulated. The tabular CUSUM equations are given by $C_i^+ = max\left[0,\ x_i - (\mu_0 + K)\ + C_{i-1}^+\right]$ and $C_i^- = max\left[0, (\mu_0 - K) - x_i + C_{i-1}^-\right]$ where $C_i^+ = C_i^- = 0$. K was selected to be 0.5σ representing a shift that would be considered out-of-control. Subsequently, a decision interval (H=hσ) is added to the chart as a control limit to see if C_i^- or C_i^+ exceeds it, at which point the process is considered out-of-control. An H value of 4.77σ was selected as it yields an ARL_0 of 370 ($\alpha = 0.0027$) when K is equal to 0.5σ. The step-by-step process of building the CUSUM chart to monitor a shift in mean for n=1 is displayed in Appendix U. It also includes an adaptation of the chart with Fast Initial Response (FIR) where $C_i^- = C_i^+ = H/2$ to detect a shift in the process earlier. A standardized CUSUM was also created, which in practice can be helpful in maintaining consistent K and H values across charts, but is not applicable in this analysis. The standardized CUSUM also guides the creation of a CUSUM for monitoring process variability when n=1 which is illustrated in Appendix U as well. A quantity that is more sensitive to variance changes is $v_i = \frac{\sqrt{|y_i|} - 0.822}{0.349}$ where y_i are the standardized x_i values. Therefore, the one-sided standardized standard deviation CUSUM equations to monitor process variability with a decision interval of h are $S_i^+ = max\left[0,\ v_i - k\ + S_{i-1}^+\right]$ and

$S_i^- = max\left[0, -k - v_i + S_{i-1}^-\right]$ where $S_i^+ = S_i^- = 0$. The CUSUM chart monitoring mean for n=6 was created in the same manner as that for n=1 (using $\overline{x}_i$ instead of x_i) and can be seen in Appendix U.

As seen in the CUSUM charts, the process is consistently in control and there is no shift in the mean or variance detected. This is expected given the likelihood of strict manufacturing standards for the pretzels. Additionally, it is important to note that the measurements of the data are not in sequential or timely fashion which prevents the charts from detecting a pattern as time progresses.

6.2 EWMA Charts

The EWMA chart is another good alternative for detecting small shifts in deviations of a parameter throughout a process. Its sensitivity and performance is comparable to that of the CUSUM chart while also requiring less steps to build. The exponentially weighted moving average takes into consideration present and past information, and is given by the equation $z_i = \lambda x_i + (1 - \lambda)z_{i-1}$ where $z_0 = \mu_0$ and the weight is a constant $0 < \lambda \leq 1$ which decreases geometrically for the preceding values z_i accumulates. While the CUSUM charts assumed normality of the observations, the EWMA is insensitive to normality when monitoring mean and assumes a symmetric distribution with mean μ_0 and variance σ_0^2 where $\sigma_{z_i}^2 = \sigma^2\left(\frac{\lambda}{2-\lambda}\right)(1 - (1 - \lambda)^{2i})$. Therefore, the upper control limits, center line and lower control limits can be defined as

$UCL = \mu_0 + L\sigma_0\sqrt{\left(\frac{\lambda}{2-\lambda}\right)(1 - (1 - \lambda)^{2i})}$	$CL = \mu_0$	$LCL = \mu_0 - L\sigma_0\sqrt{\left(\frac{\lambda}{2-\lambda}\right)(1 - (1 - \lambda)^{2i})}$

where L is the width of the control limits of the symmetric distribution. Smaller values of λ are helpful in detecting smaller shifts and therefore, $\lambda = 0.1$ was selected for use in accordance with common practice. As per the table in Appendix V, the corresponding L = 2.814 was used to build the EWMA charts, those of which for n=1 can be seen in Appendix V. To monitor the process standard deviation, an exponentially weighted mean square error (EWMS) chart was created defined by $S_i^2 = \lambda(x_i - \mu_0)^2 + (1 - \lambda)S_{i-1}^2$. Now assuming normality of the mean and standard deviation, plotting $\sqrt{S_i^2}$ on an exponentially weighted root mean square (EWRMS) chart where the target value is σ_0 yields

$UCL = \sigma_0\sqrt{\frac{\chi^2_{v,\alpha/2}}{v}}$, where $v = \left(\frac{\lambda}{2-\lambda}\right)$	$CL = \sigma_0$	$LCL = \sigma_0\sqrt{\frac{\chi^2_{v,1-\alpha/2}}{v}}$, where $v = \left(\frac{\lambda}{2-\lambda}\right)$

The charts for n=6 in Appendix V were created similarly, however, x_i becomes $\overline{x}_i$ and the standard deviation is divided by the root of the number of observations in the sample (n=6). The same conclusion can be drawn from the EWMA charts as from the CUSUM charts. The process is consistently in control and there is no shift in the mean or variance detected.

7.0 References

[1] H. Cohick, “Pretzel facts,” *Positively Pennsylvania*, 01-Oct-2019. [Online]. Available: https://positivelypa.com/pretzel-facts/. [Accessed: 14-Apr-2023].

[2] “Pretzel,” *How Products Are Made*. [Online]. Available: http://www.madehow.com/Volume-4/Pretzel.html. [Accessed: 14-Apr-2023].

[3] *Snydersofhanover.com*. [Online]. Available: https://www.snydersofhanover.com/rods/. [Accessed: 14-Apr-2023].

[4] 1factory, “Process capability analysis CP, CPK, PP, PPK - A Guide,” *1factory*. [Online]. Available: https://www.1factory.com/quality-academy/guide-to-process-capability-analysis-cp-cpk-pp-ppk.html. [Accessed: 14-Apr-2023].

8.0 Appendix

Appendix A: Basic Data Visualization Diagrams

Stem-and-leaf of LengthOfPretzel   N = 120

```
1      8    9
3      9    01
5      9    23
11     9    444455
36     9    6666666666777777777777777
(26)   9    88888888888888999999999999
58     10   000000000011111111111111
34     10   22222222233333333333
14     10   44444455555
3      10   677

  Leaf Unit = 0.1
```

Figure 4: Stem and leaf plot for length of pretzels

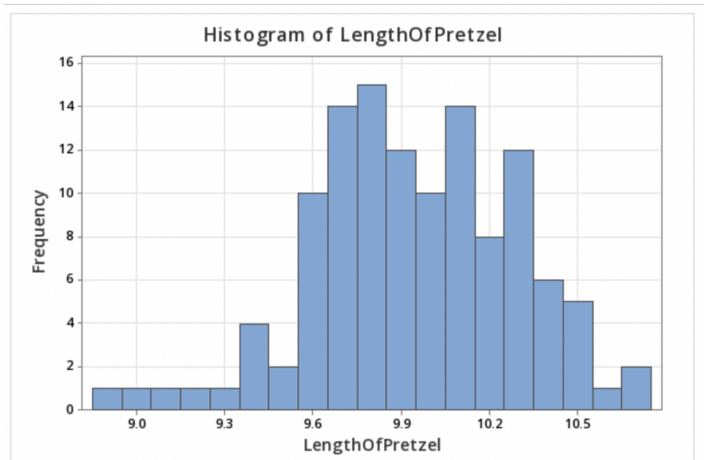


Figure 5: Histogram of length of pretzels

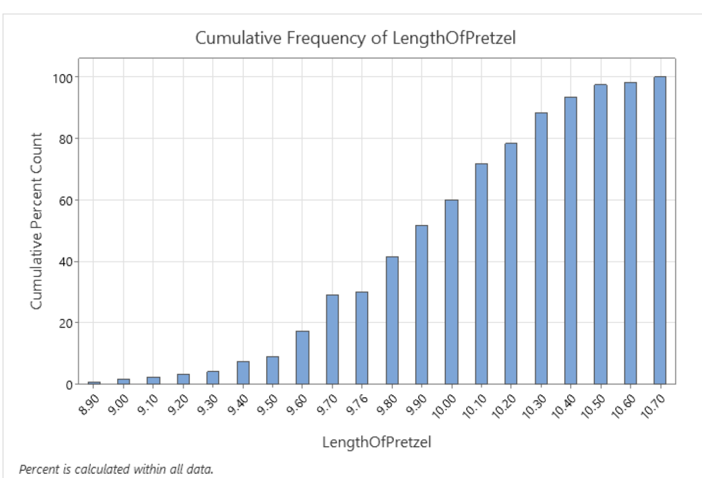


Figure 6: Cumulative Frequency Plot for length of pretzels

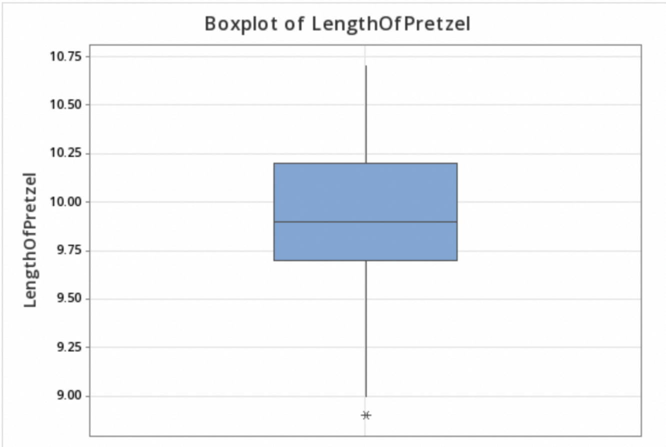


Figure 7: Boxplot

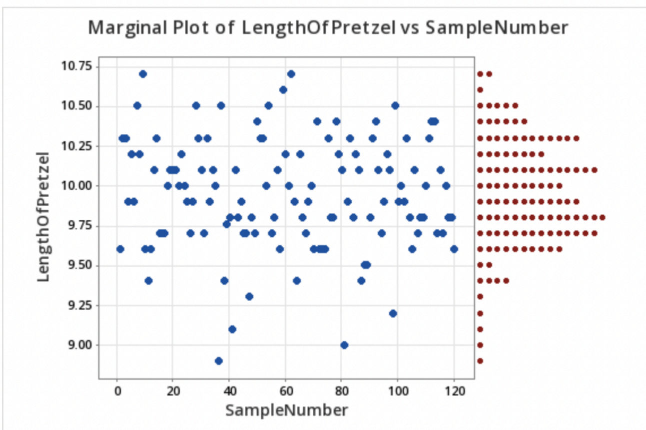
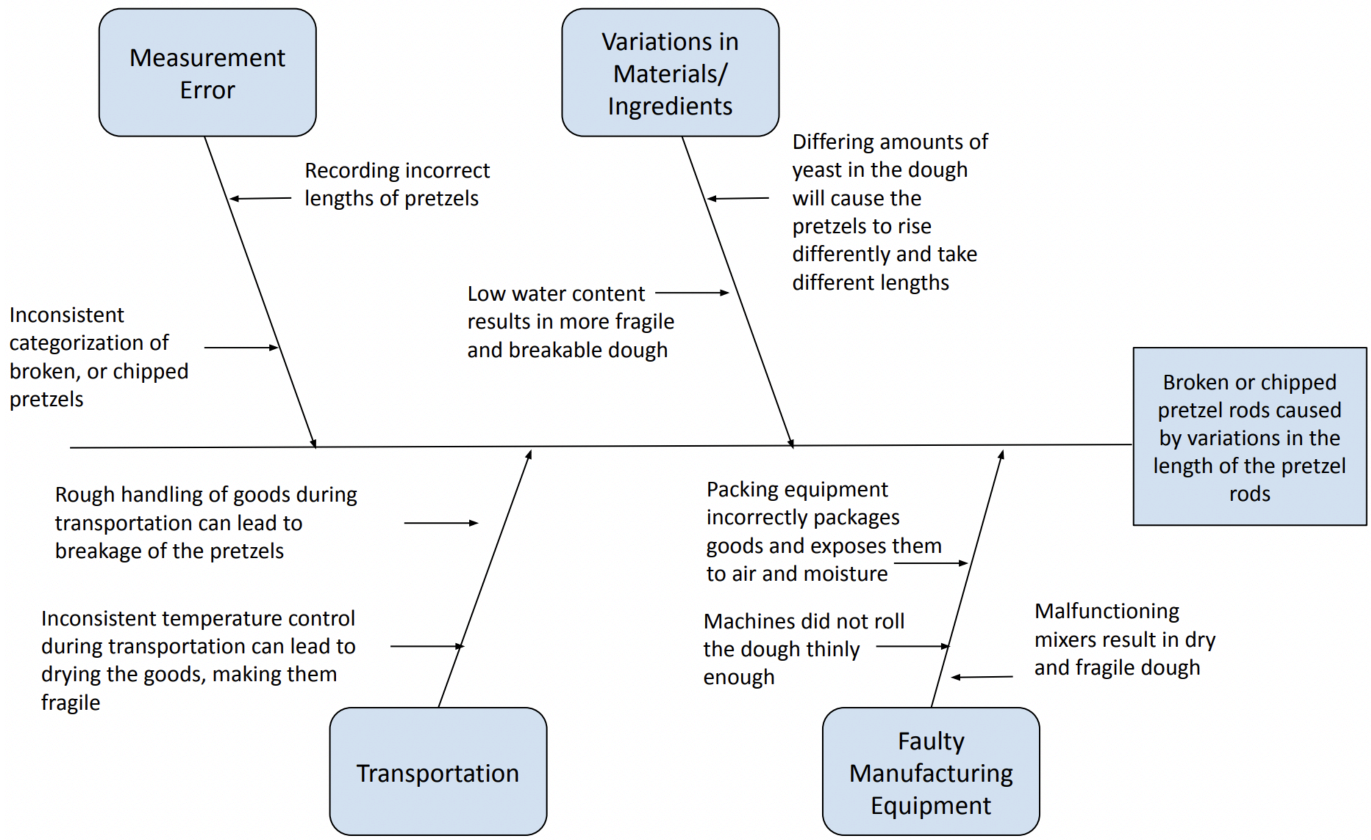
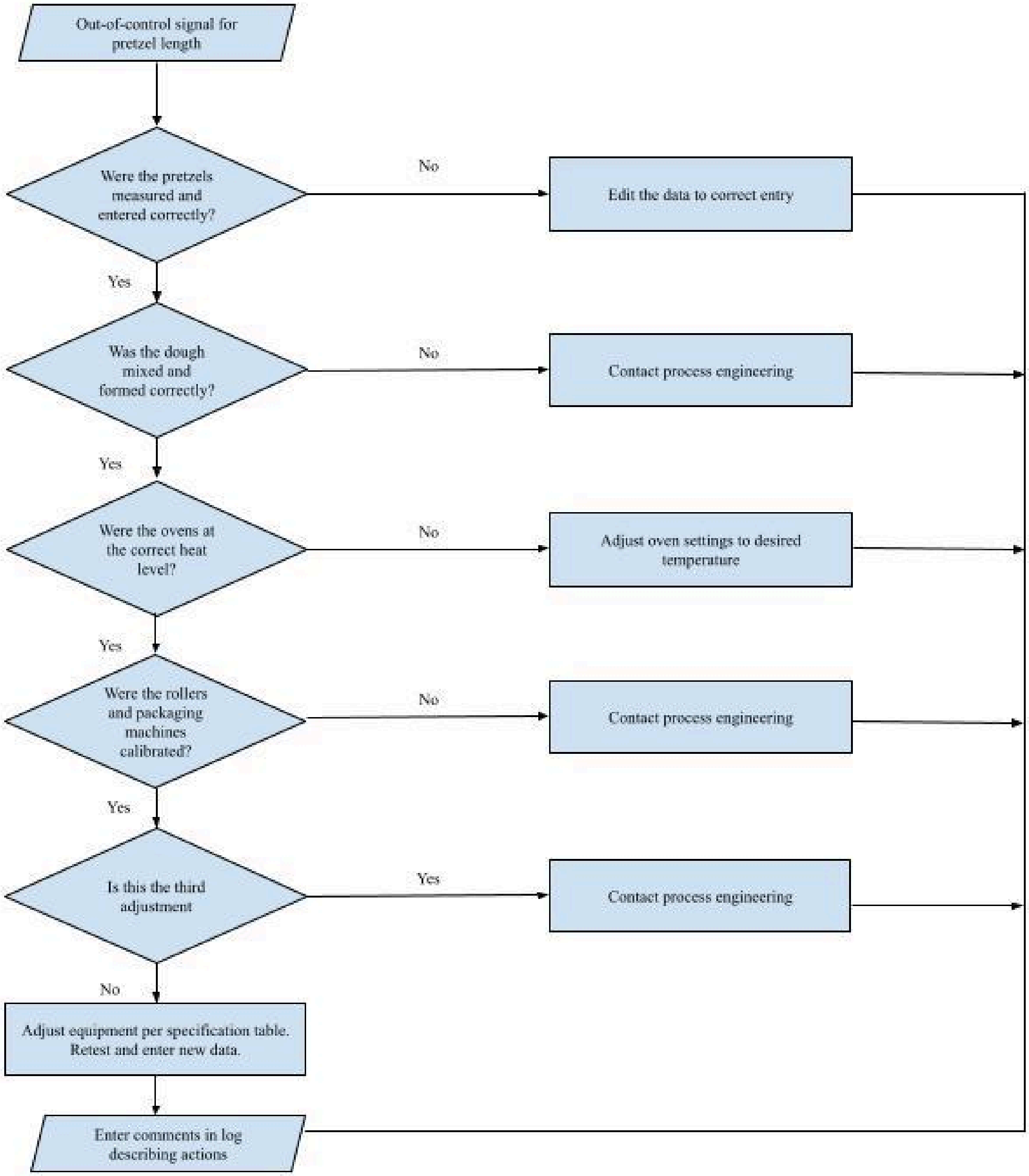


Figure 8: Time Series Plot

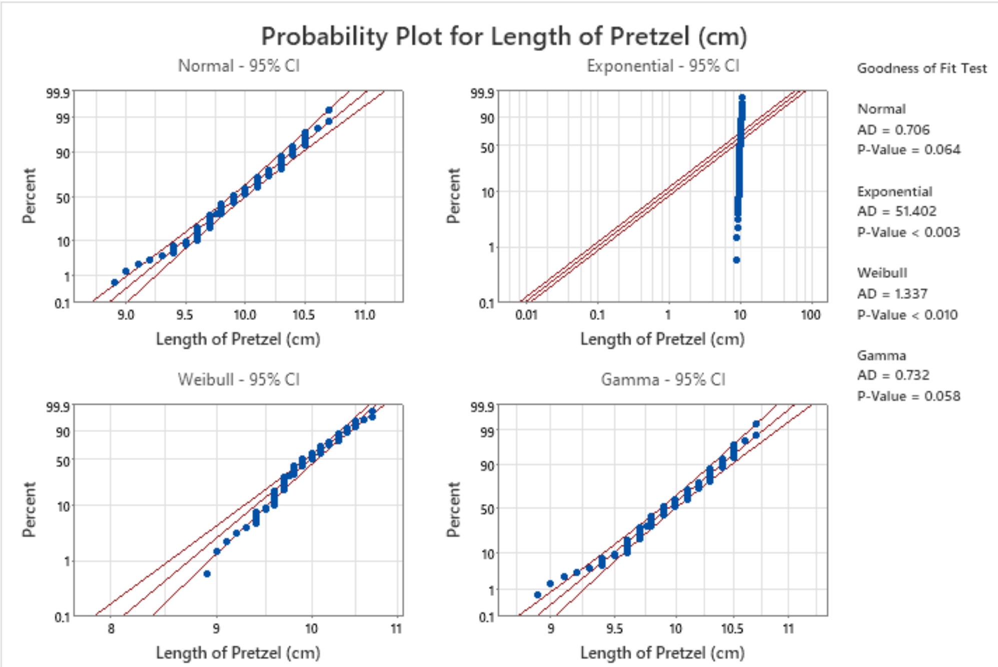
Appendix B: Cause and effect diagram



Appendix C: Out-of-control-action plan for the pretzel rod manufacturing process

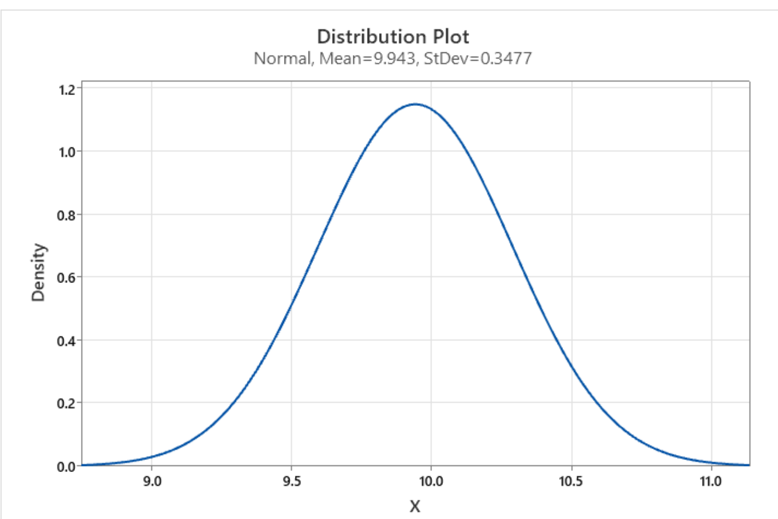


Appendix D: Analyzing the probability distribution of pretzel length (finding the best-fitting pdf)



Goodness of Fit Test

Distribution	AD	P
Normal	0.706	0.064
Exponential	51.402	<0.003
Weibull	1.337	<0.010
Gamma	0.732	0.058



Appendix E: 95% Confidence intervals on mean of pretzel length using normal and student-t distributions

$$P\left(\mu - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \bar{x} < \mu + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 0.95 \qquad \text{where } \alpha = 0.05,\ z_{0.025} = 1.96$$

$$P\left(\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 0.95 \qquad \text{where } \alpha = 0.05,\ z_{0.025} = 1.96$$

$$P\left(9.943 - 1.96\left(\frac{0.3477}{\sqrt{120}}\right) < \mu < 9.943 + 1.96\left(\frac{0.3477}{\sqrt{120}}\right)\right) = 0.95$$

$$P(9.943 - 0.0622 < \mu < 9.943 + 0.0622) = 0.95$$

$\therefore$ With 95% confidence, the mean falls in the interval [9.8808,10.0052]

$$P\left(\bar{x} - t_{n-1,\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + t_{n-1,\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)\right) = 0.95 \qquad \text{where } \alpha = 0.05,\ t_{119,0.025} = 1.98$$

$$P\left(9.943 - 1.98\left(\frac{0.3477}{\sqrt{120}}\right) < \mu < 9.943 + 1.98\left(\frac{0.3477}{\sqrt{120}}\right)\right) = 0.95$$

$$P(9.943 - 0.0628 < \mu < 9.943 + 0.0628) = 0.95$$

$\therefore$ With 95% confidence, the mean falls in the interval [9.8801,10.0059]

Appendix F: Hypothesis test on the mean

Null Hypothesis H_0: $\mu = 10$

Alternate Hypothesis H_1 $\mu \neq 10$

Since n is "large":

CLT -> $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

Standardized $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Test Statistic $z_0 = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{9.943-10}{\frac{0.3477}{\sqrt{120}}} = -1.7958$

$z_{\frac{0.05}{2}} = 1.96$

Using a t-test sampling distribution:

Test Statistic $t_0 = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \frac{9.943-10}{\frac{0.3477}{\sqrt{120}}} = -1.7958$

$t_{119,\frac{0.05}{2}} = 1.98$

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
120	9.9430	0.3477	0.0317	(9.8808, 10.0052)

μ: population mean of Length of Pretzel (cm)
Known standard deviation = 0.3477

Test

Null hypothesis H_0: $\mu = 10$
Alternative hypothesis H_1: $\mu \neq 10$

Z-Value	P-Value
-1.7958	0.073

Test

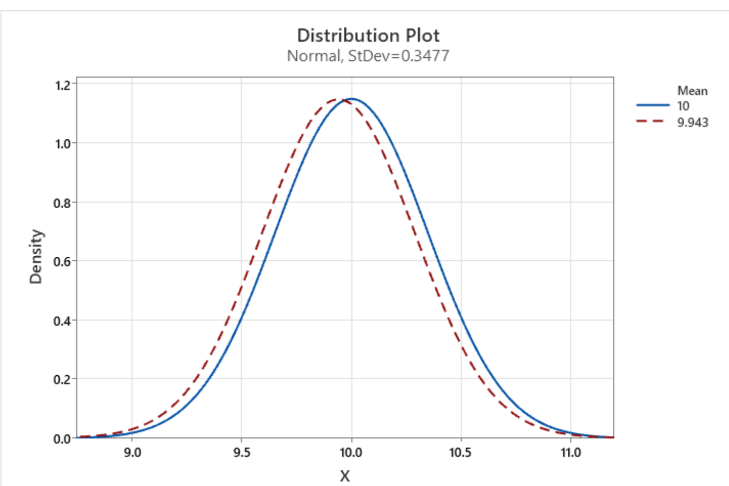
Null hypothesis H_0: $\mu = 10$
Alternative hypothesis H_1: $\mu \neq 10$

T-Value	P-Value
-1.7957	0.075

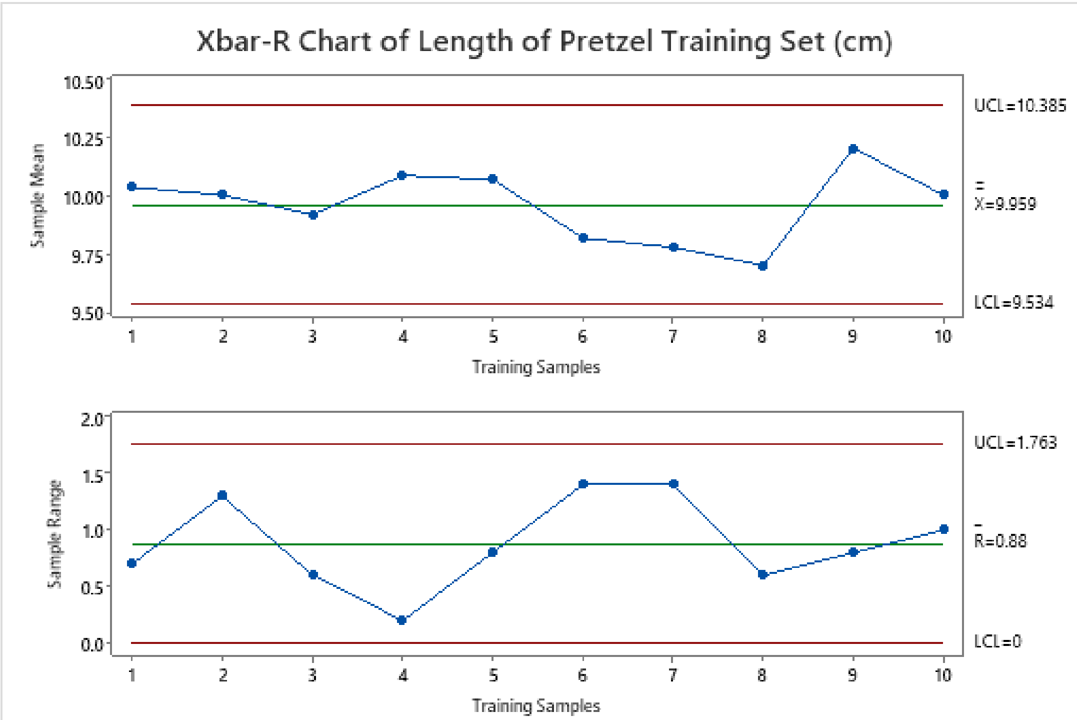
Descriptive Statistics

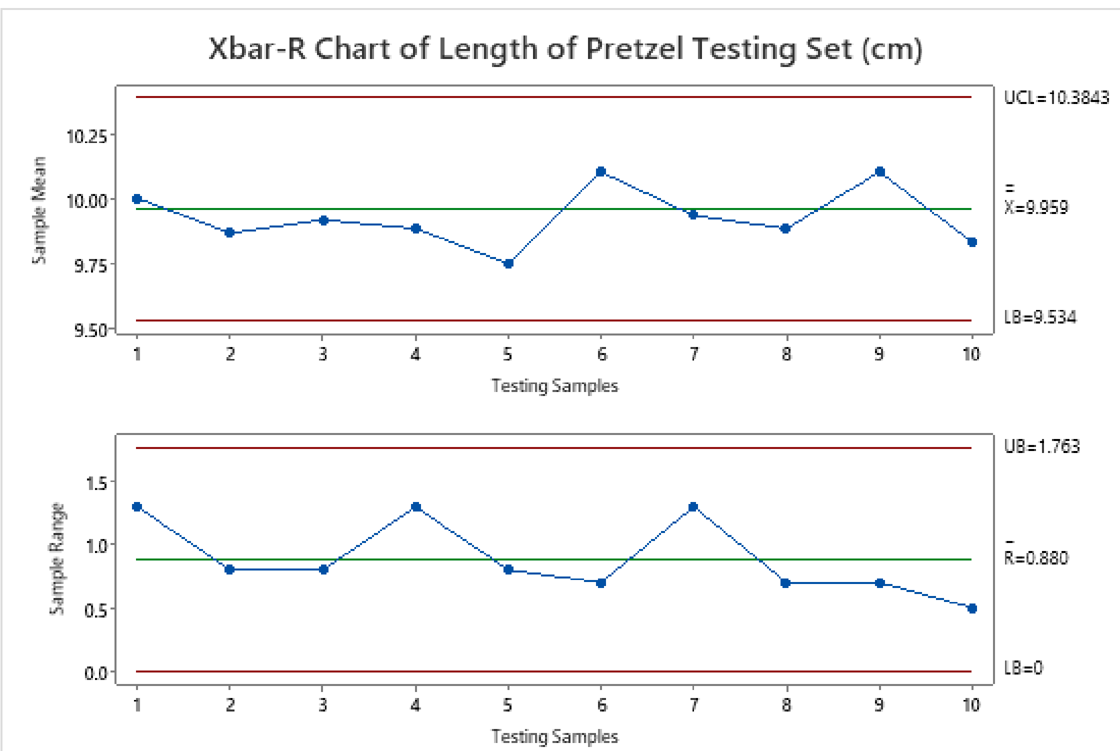
N	Mean	StDev	SE Mean	95% CI for μ
120	9.9430	0.3477	0.0317	(9.8801, 10.0059)

μ: population mean of Length of Pretzel (cm)



Appendix G: $\overline{X}$ & R Charts





For $\overline{X}$:

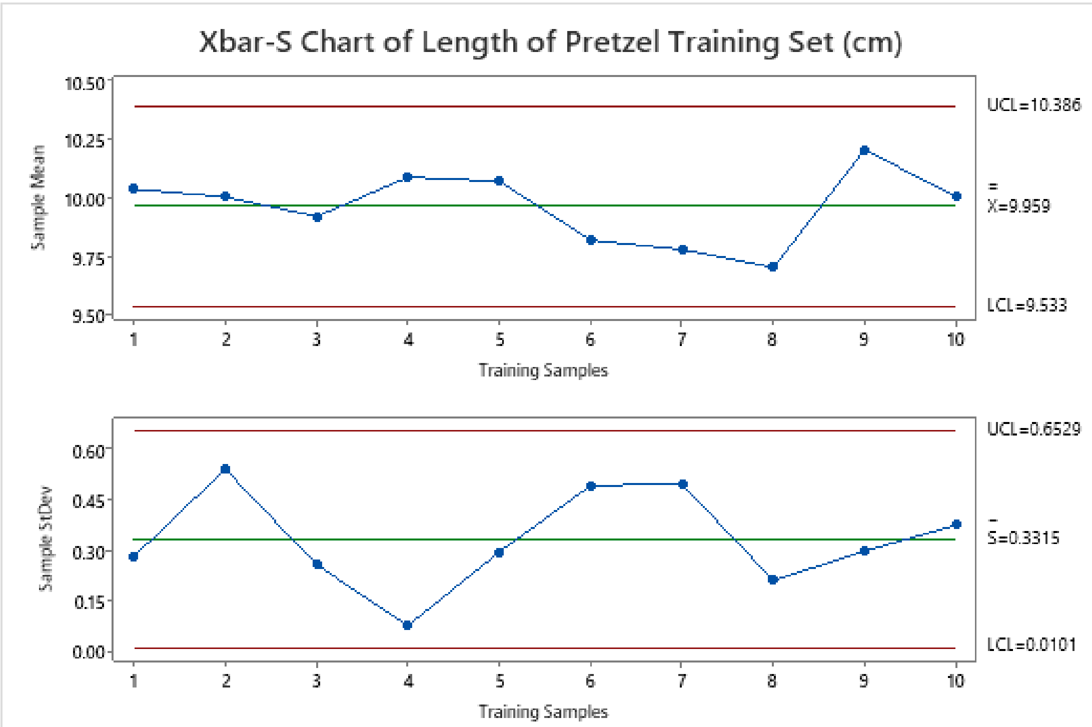
$LCL = \overline{\overline{X}} - A_2 \cdot \overline{R} = 9.534$	$CL = \overline{X} = 9.959$	$UCL = \overline{\overline{X}} + A_2 \cdot \overline{R} = 10.385$

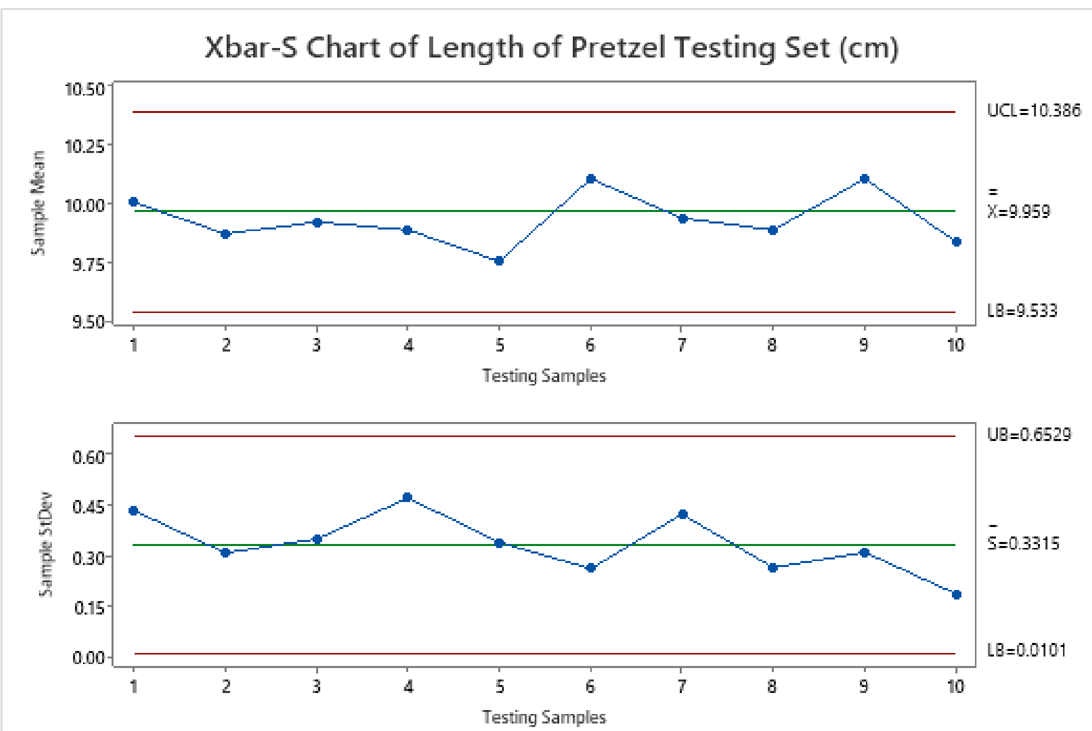
For R:

$LCL = D_3 \cdot \overline{R} = 0$	$CL = \overline{R} = 0.88$	$UCL = D_4 \cdot \overline{R} = 1.763$

For σ: $\sigma = \frac{\overline{R}}{d_2}$

Appendix H: $\overline{X}$ & S Charts





For $\overline{X}$:

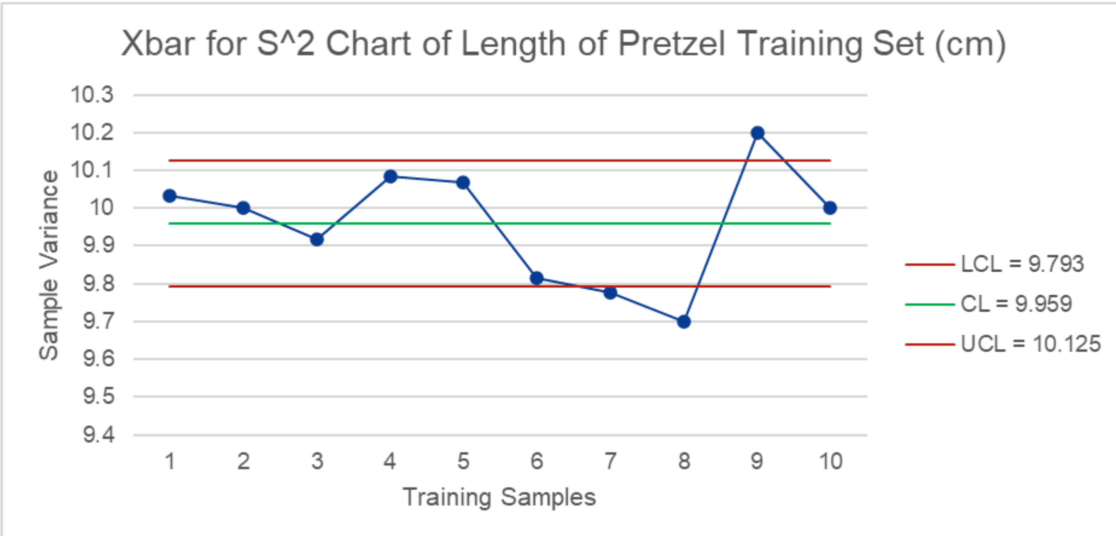
$LCL = \overline{\overline{X}} - A_3 \cdot \overline{S} = 9.533$	$CL = \overline{X} = 9.959$	$UCL = \overline{\overline{X}} + A_3 \cdot \overline{S} = 10.386$

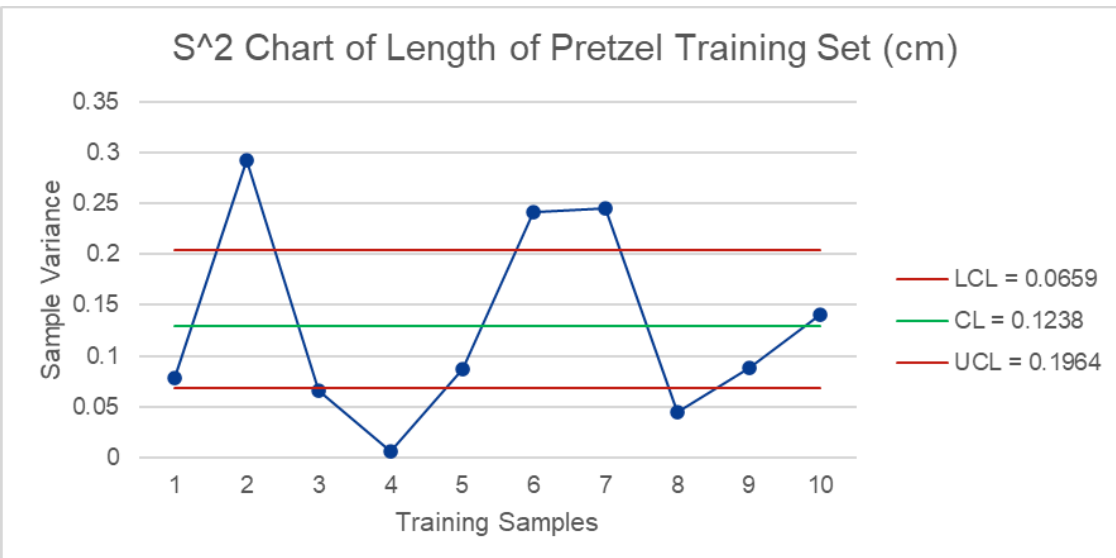
For S:

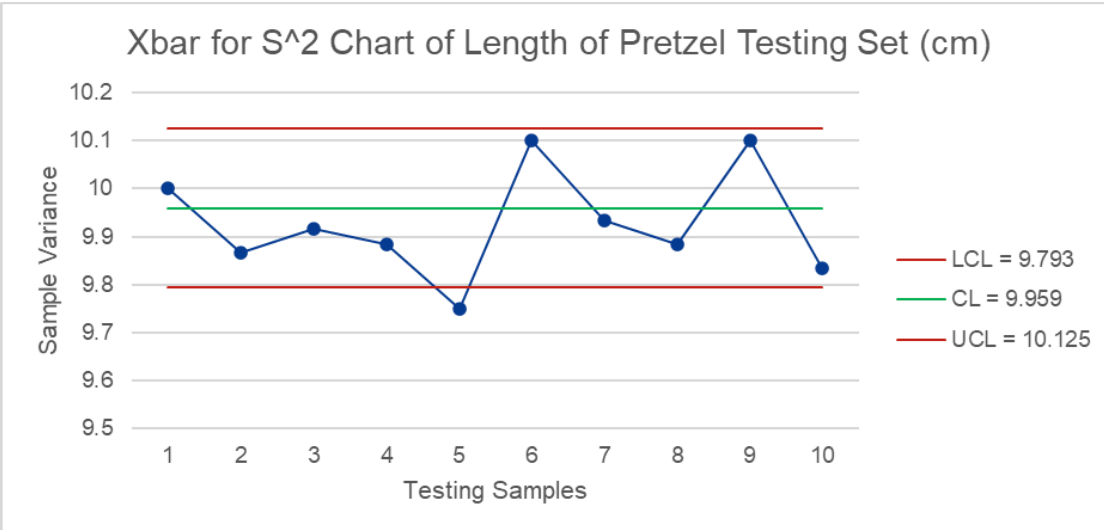
$LCL = B_3 \cdot \overline{S} = 0.0101$	$CL = \overline{S} = 0.3315$	$UCL = B_4 \cdot \overline{S} = 0.6529$

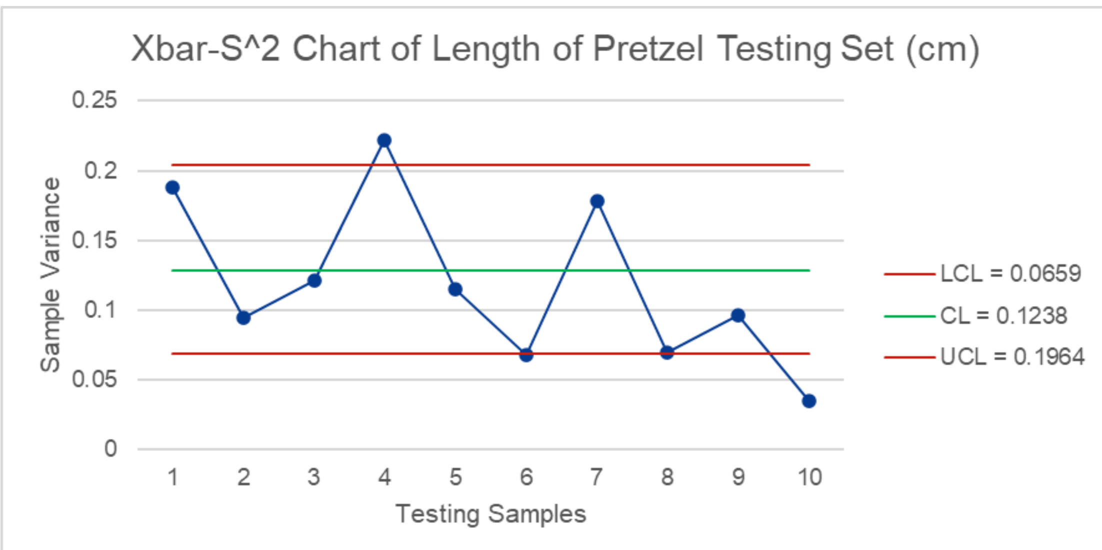
For σ: $\sigma = \frac{\overline{R}}{d_2}$

Appendix I: $\overline{X}$ & S^2 Charts

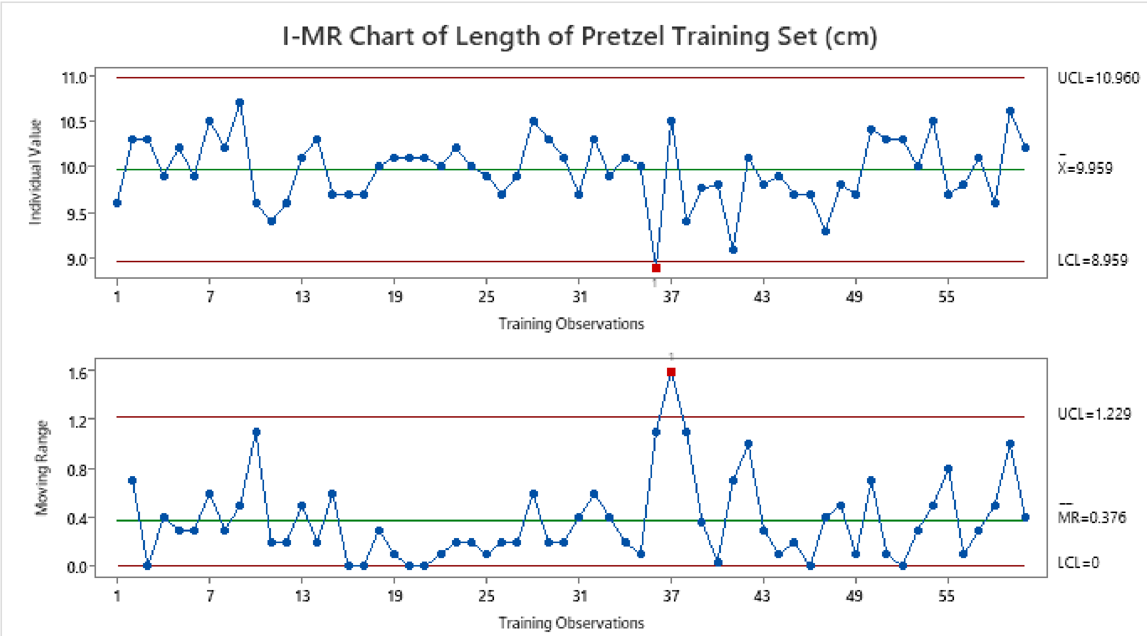


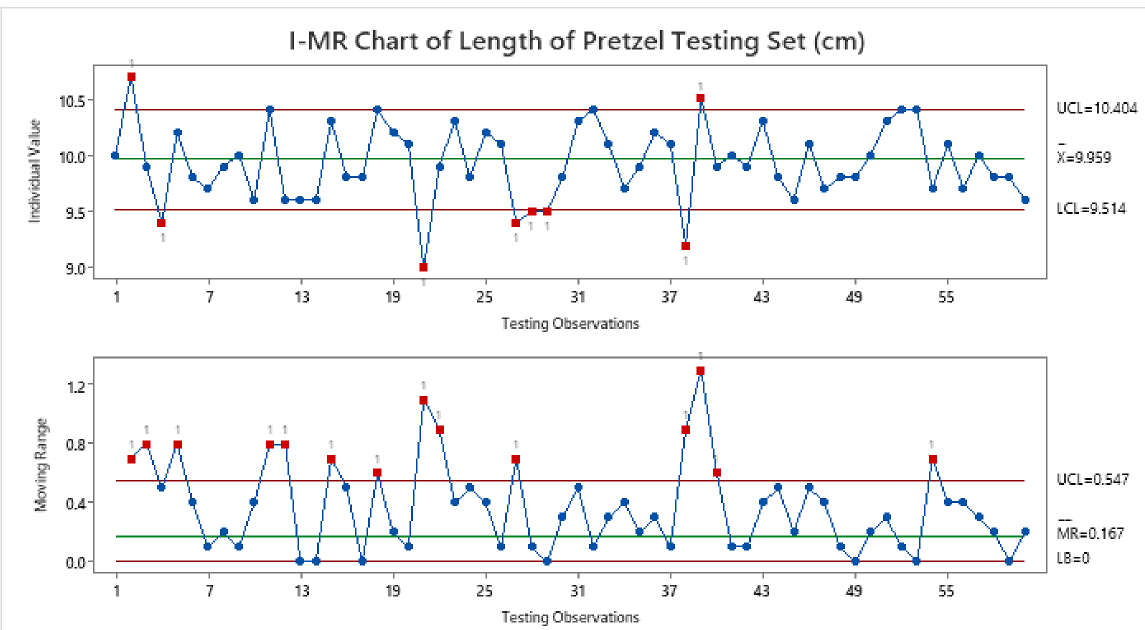



Xbar for S^2 Chart of Length of Pretzel Testing Set (cm)
Sample Variance
10.2
10.1
10
9.9
9.8
9.7
9.6
9.5
1
2
3
4
5
6
7
8
9
10
Testing Samples
LCL = 9.793
CL = 9.959
UCL = 10.125


Xbar-S^2 Chart of Length of Pretzel Testing Set (cm)
Sample Variance
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
Testing Samples
LCL = 0.0659
CL = 0.1238
UCL = 0.1964

Appendix J: $I - MR$ Charts





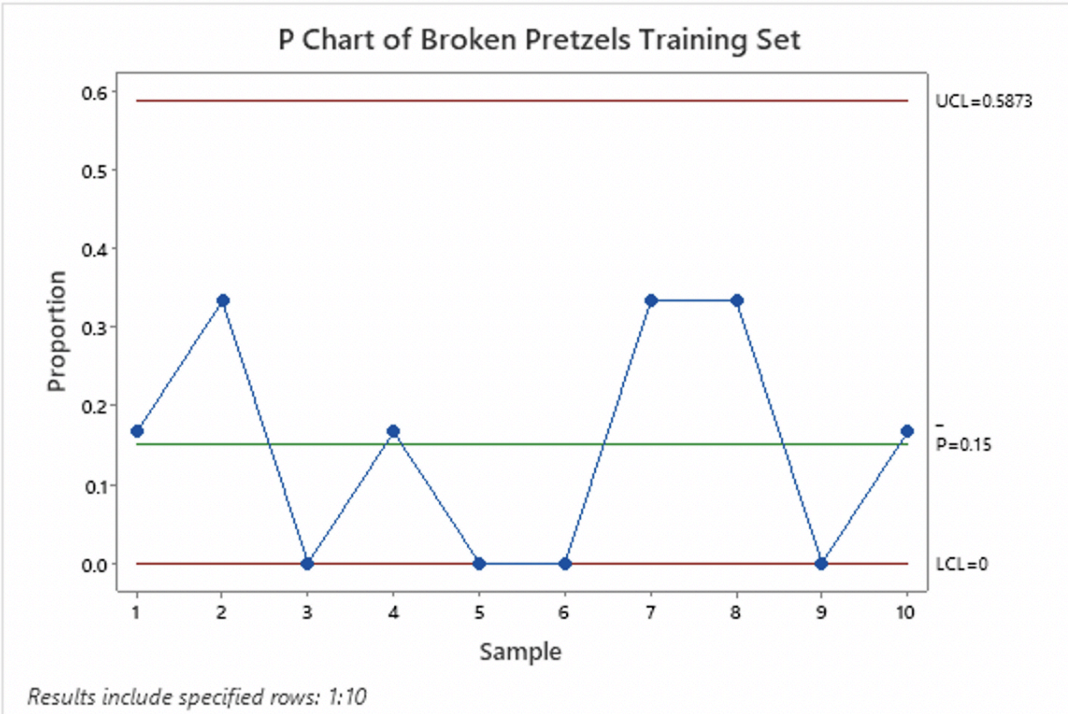
For I:

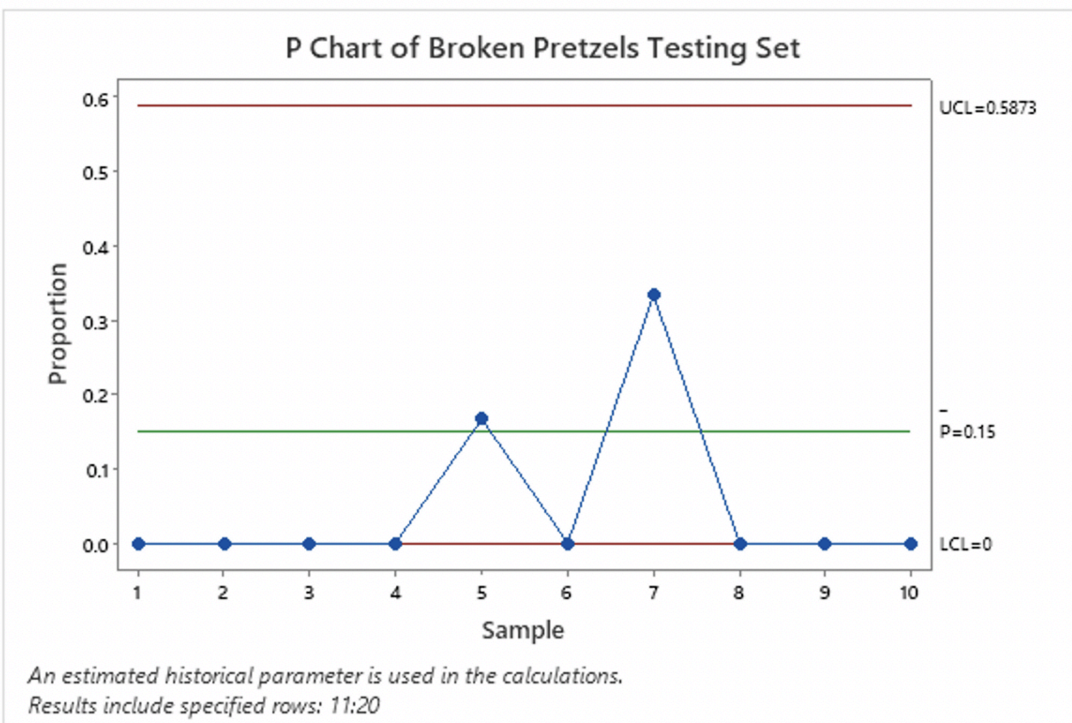
$LCL = \overline{\overline{X}} - \frac{3}{d_2} \cdot \overline{MR} = 8.959$	$CL = \overline{X} = 9.959$	$UCL = \overline{\overline{X}} + \frac{3}{d_2} \cdot \overline{MR} = 10.960$

For MR:

$LCL = D_3 \cdot \overline{MR} = 0$	$CL = \overline{MR} = 0.376$	$UCL = D_4 \cdot \overline{MR} = 1.229$

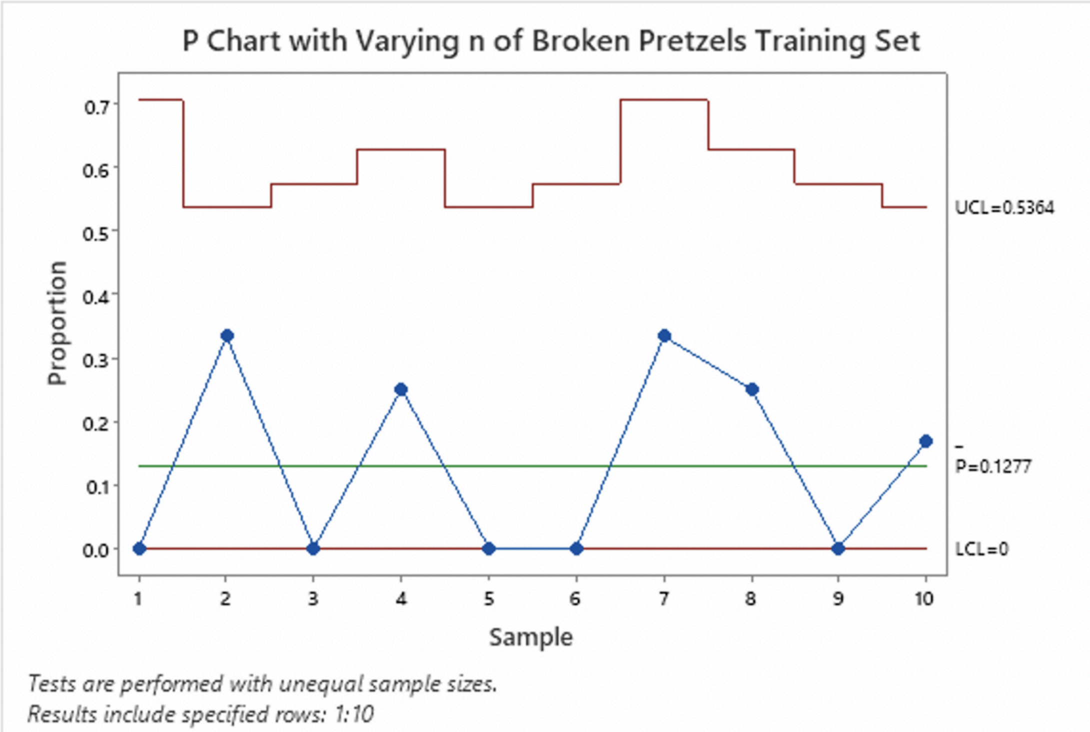
Appendix K: P-Chart for Non-Conformities

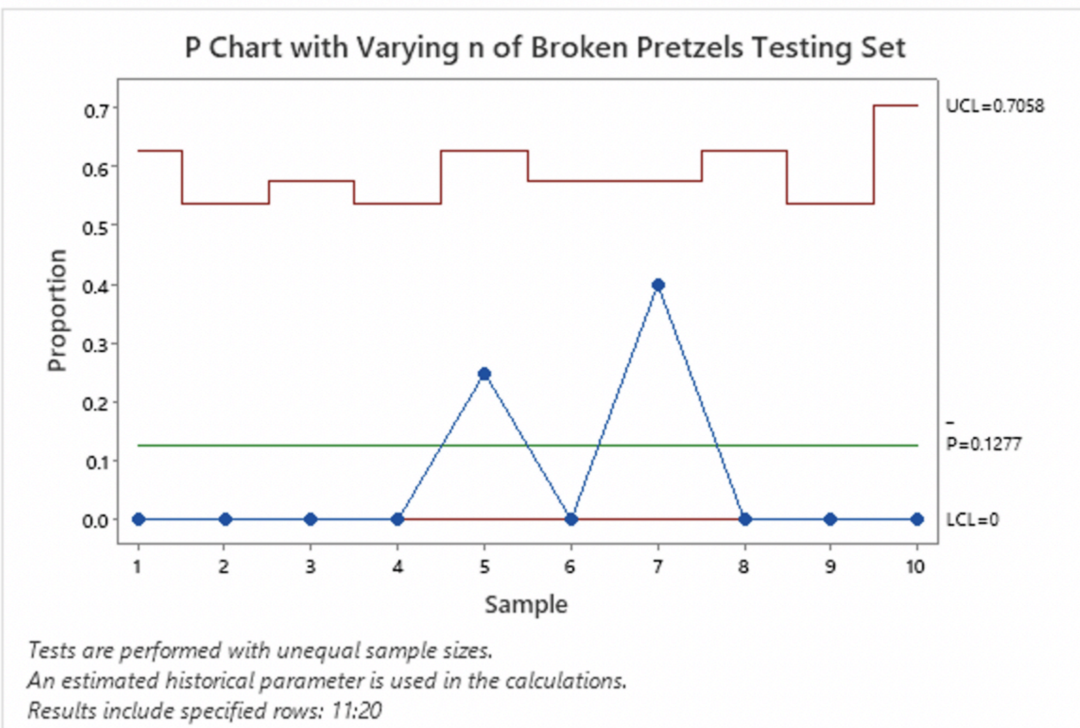




$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} =$ -1.162 → 0	$CL = \bar{p} = 0.15$	$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ = 0.5873

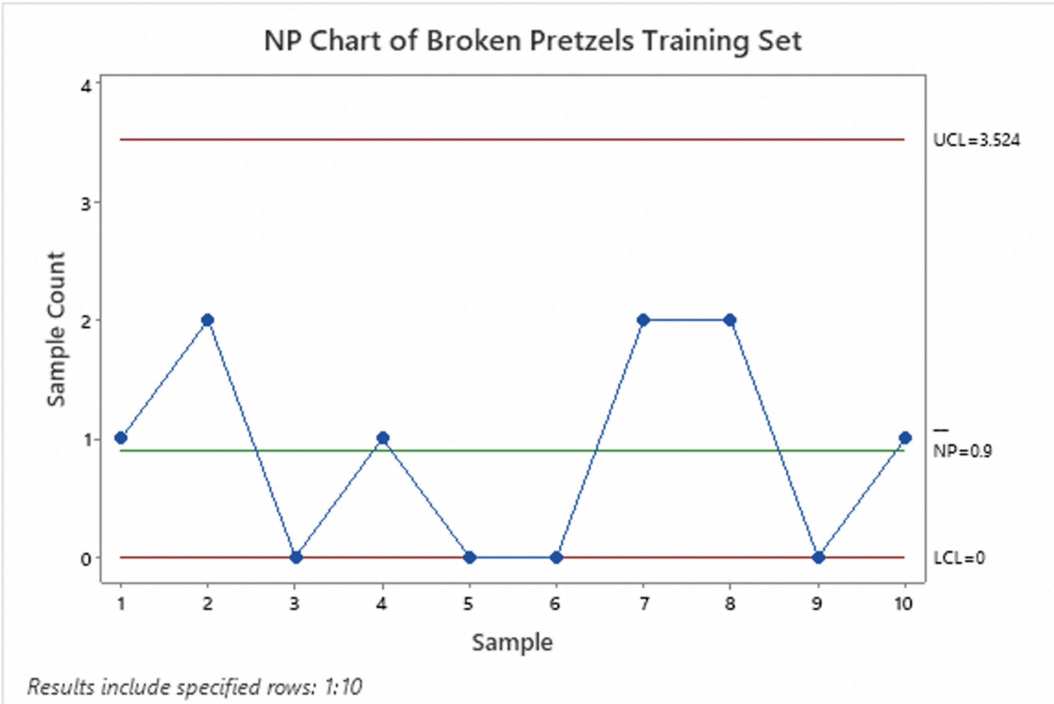
Appendix L: P-chart with varying n using variable limits approach

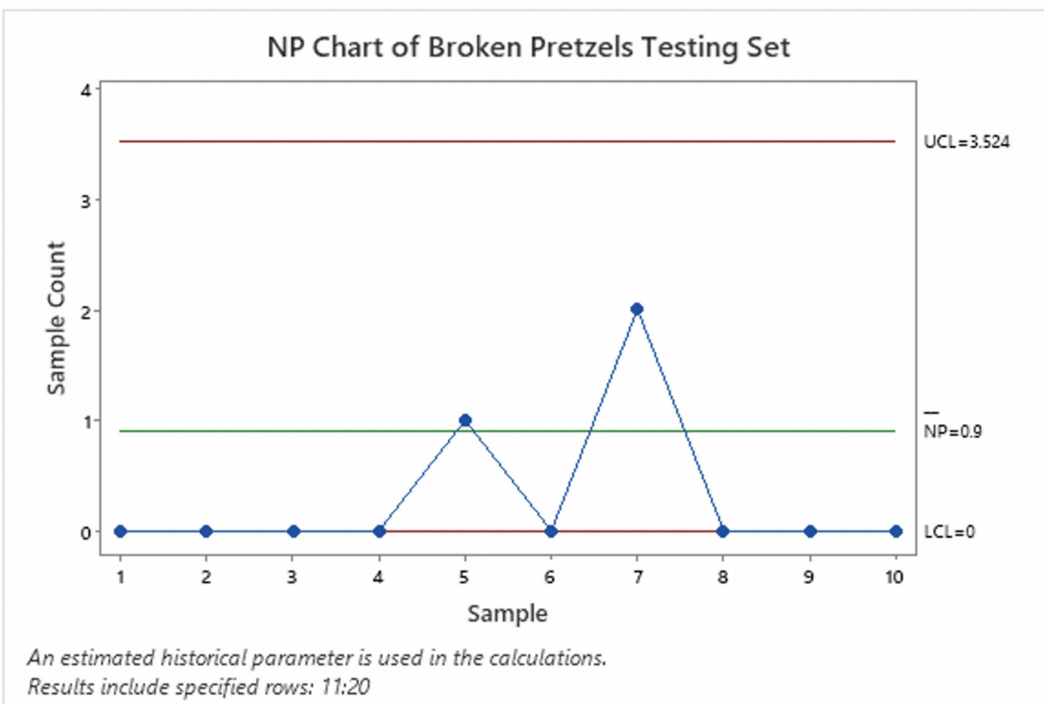




$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} = 0$	$CL = \bar{p} = 0.1277$	$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$ $= 0.5364$

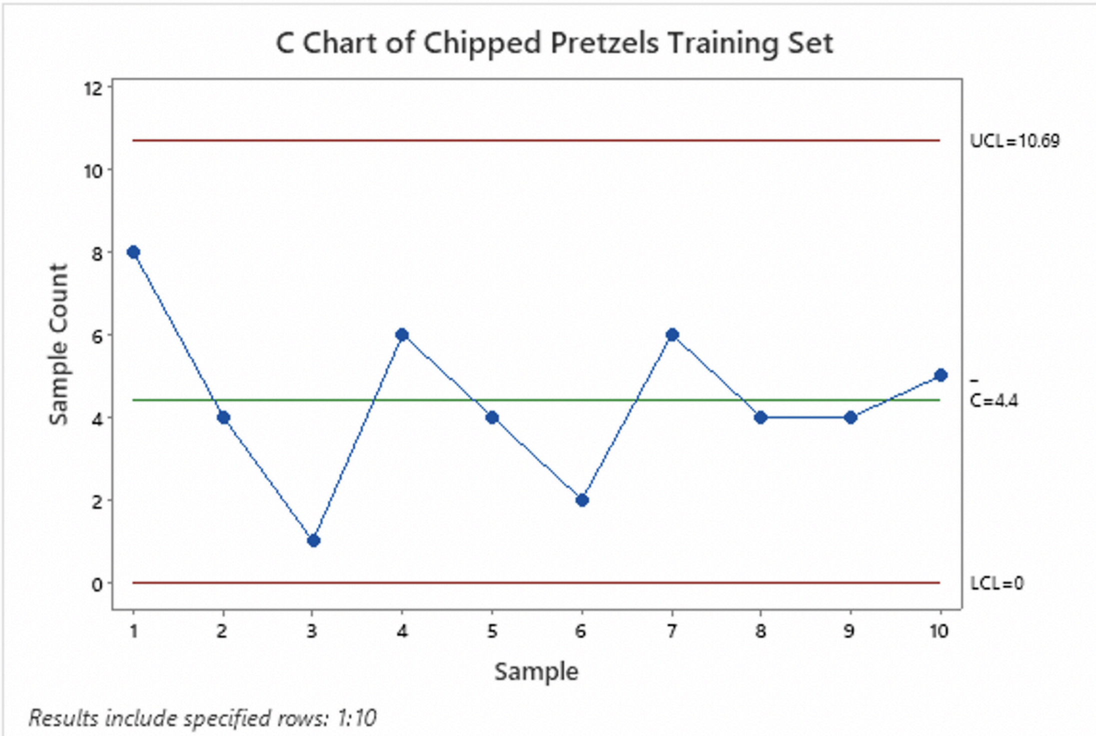
Appendix M: np-Chart for Non-Conformities

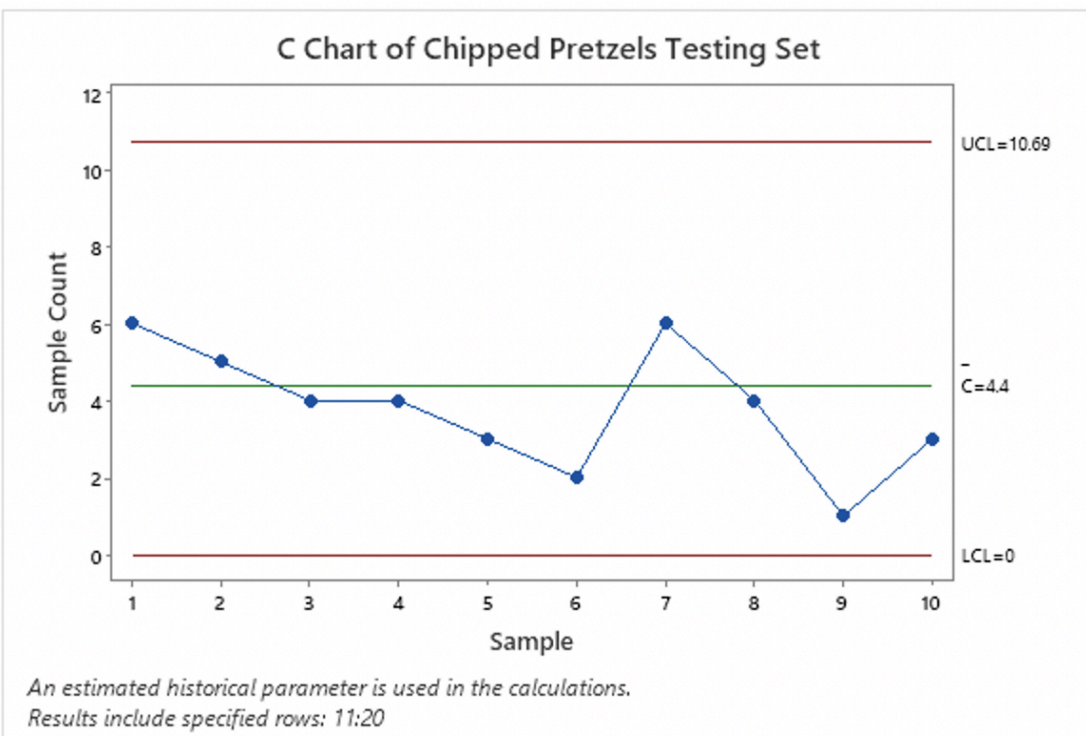




$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})} = 0$	$CL = np = 0.9$	$UCL = np + 3\sqrt{np(1-p)} = 3.524$

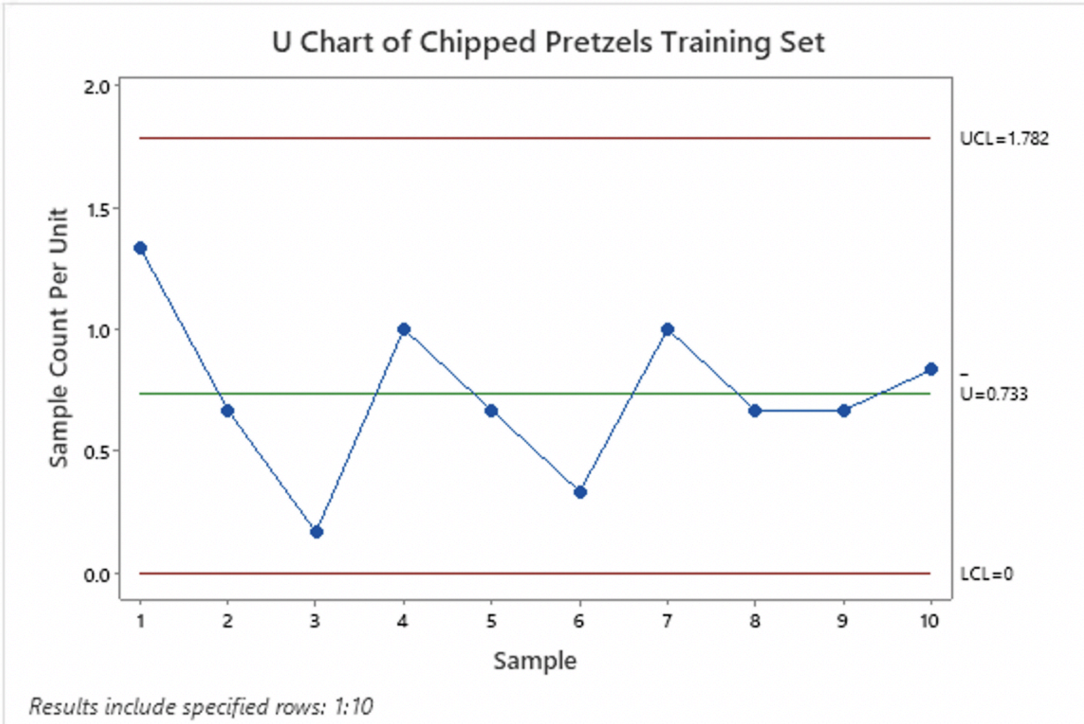
Appendix N: C-chart for Nonconformities

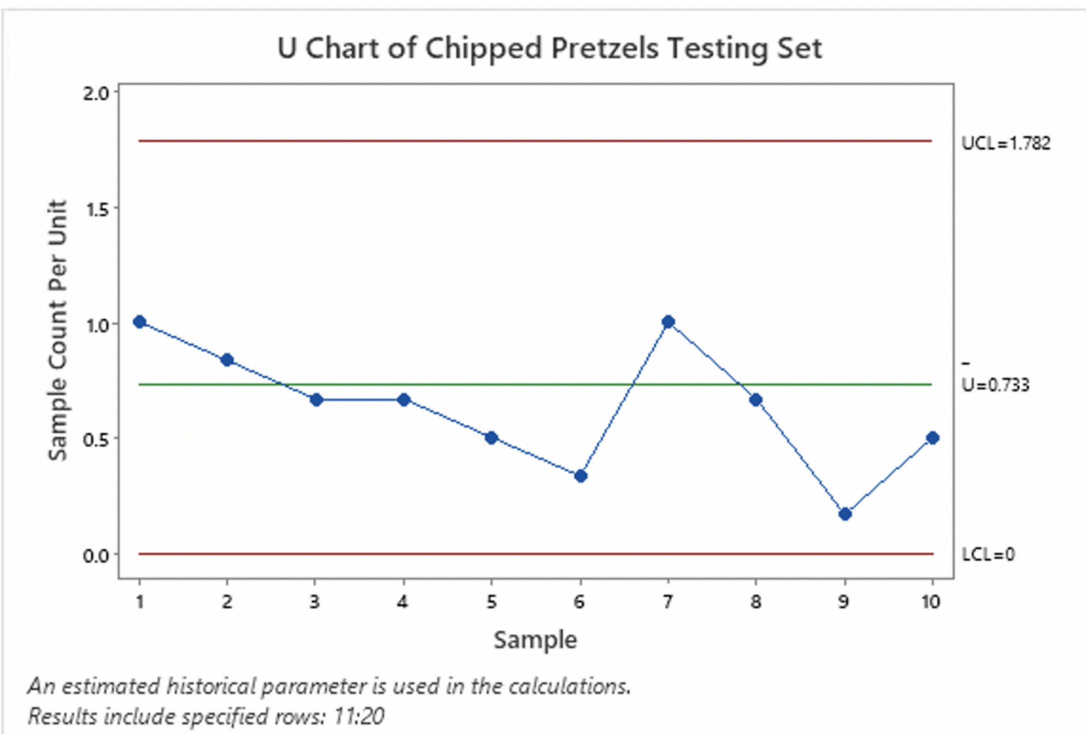




$LCL = \bar{c} - 3\sqrt{\bar{c}}$ =-1.89 →0	$CL = \bar{c} = 4.4$	$UCL = \bar{c} + 3\sqrt{\bar{c}} = 10.69$

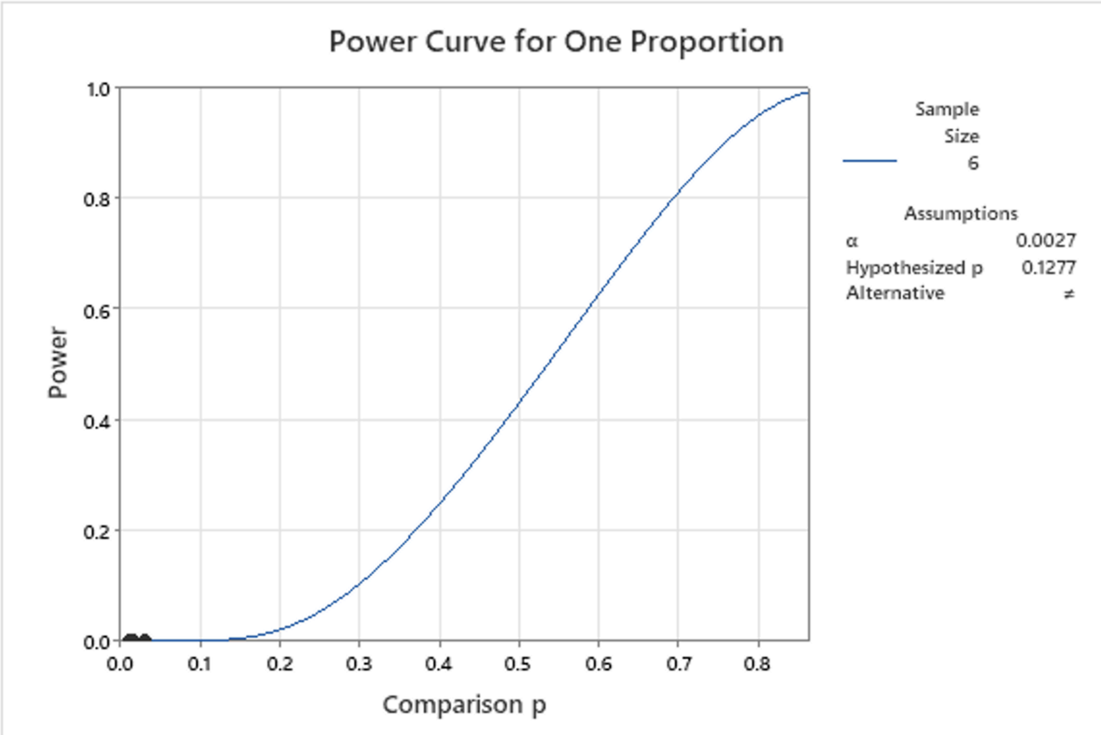
Appendix O: u-Chart





$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$ = -0.316 → 0	$CL = \bar{c} = 0.733$	$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$ = 1.782

Appendix P: Power Curve for one proportion (based on p-chart)



Test for One Proportion

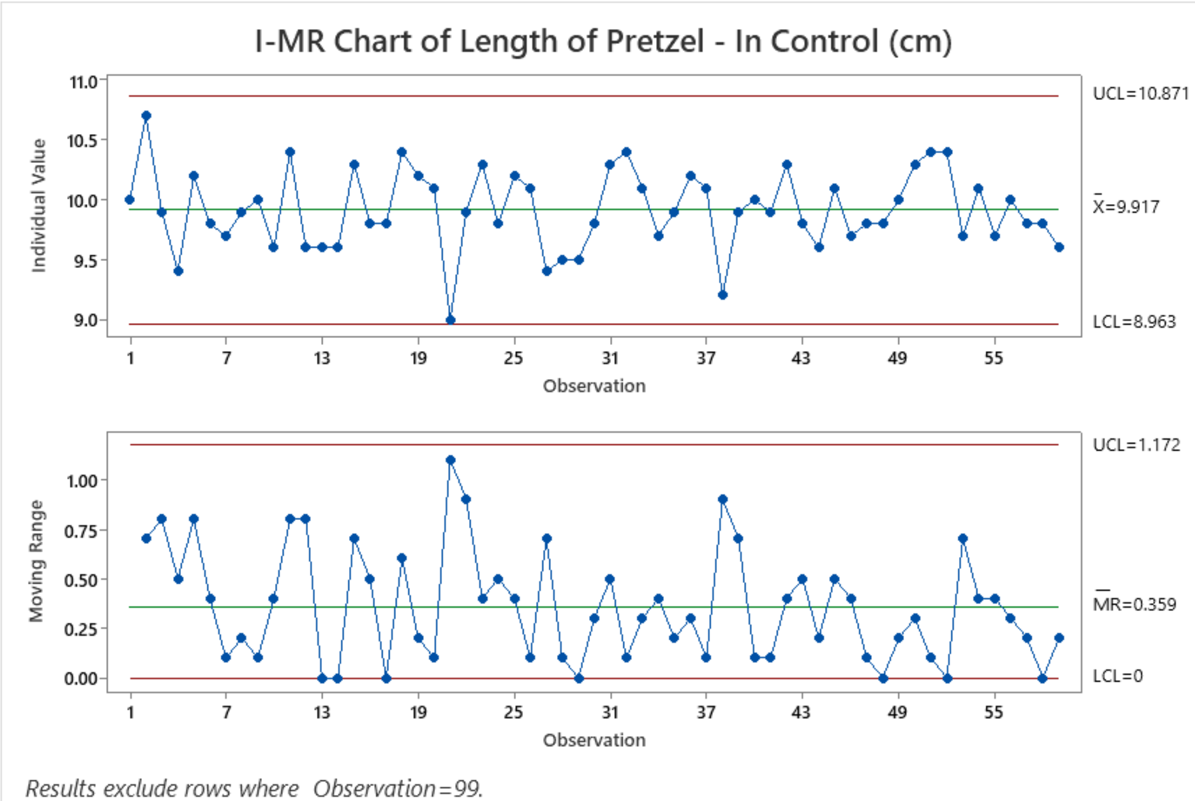
Testing p = 0.1277 (versus ≠ 0.1277)

α = 0.0027

Results

Comparison p	Sample Size	Power
0.010	6	0.0000000
0.015	6	0.0000000
0.030	6	0.0000040

Appendix Q: I-MR Phase 2 Chart



Appendix R: Anderson-Darling Normalization Charts

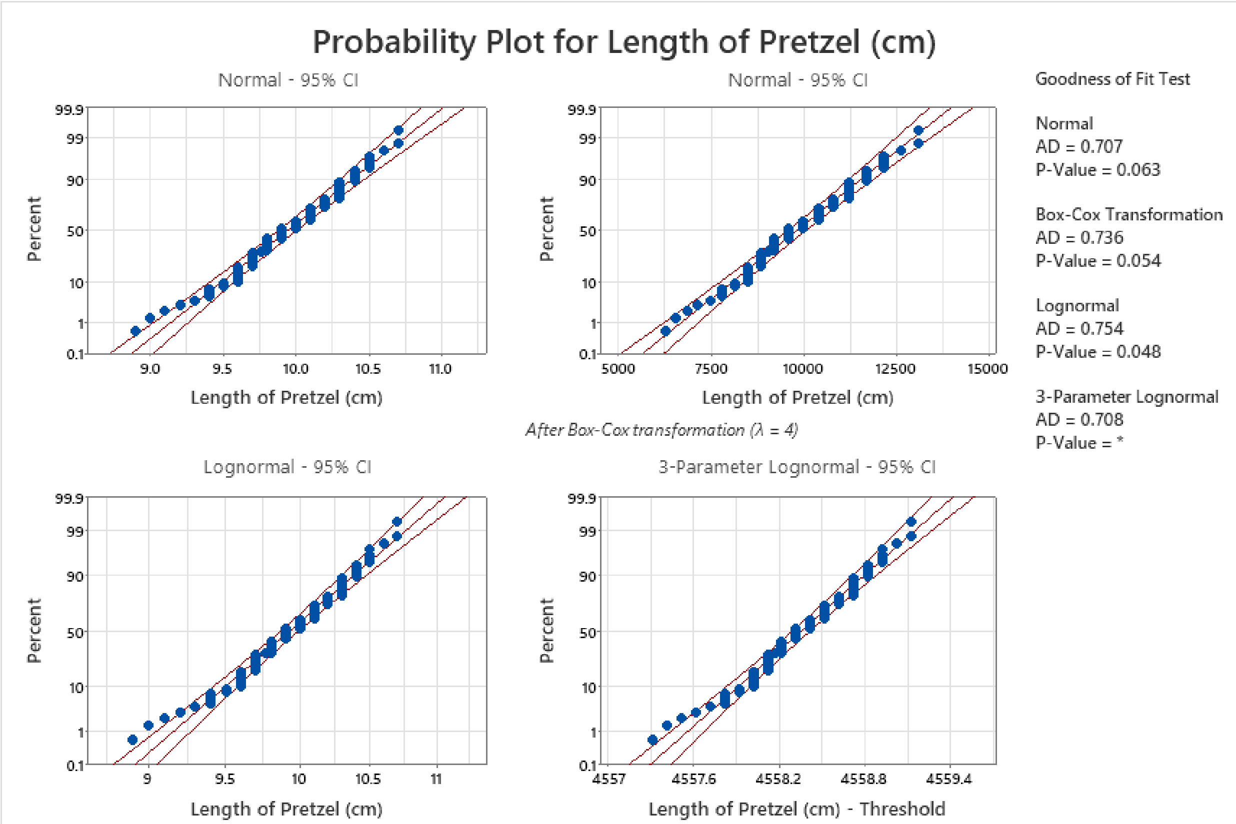


Figure 9: Samples graphed along side various distributions of the Anderson-Darling normalization

Goodness of Fit Test

Distribution	AD	P	LRT P
Normal	0.707	0.063	
Box-Cox Transformation	0.736	0.054	
Lognormal	0.754	0.048	
3-Parameter Lognormal	0.708	*	0.237
Exponential	51.000	<0.003	
2-Parameter Exponential	24.515	<0.010	0.000
Weibull	1.327	<0.010	
3-Parameter Weibull	0.762	0.023	0.011
Smallest Extreme Value	1.502	<0.010	
Largest Extreme Value	2.834	<0.010	
Gamma	0.735	0.057	
3-Parameter Gamma	0.779	*	1.000
Logistic	0.807	0.021	
Loglogistic	0.809	0.020	
3-Parameter Loglogistic	0.809	*	0.445

Figure 10: Anderson-Darling and P-value for various distributions

ML Estimates of Distribution Parameters

Distribution	Location	Shape	Scale	Threshold
Normal*	9.93832		0.34537	
Box-Cox Transformation*	9825.23211		1345.01245	
Lognormal*	2.29579		0.03498	
3-Parameter Lognormal	8.42472		0.00008	-4548.42309
Exponential			9.93832	
2-Parameter Exponential			1.04712	8.89120
Weibull		31.71204	10.10115	
3-Parameter Weibull		5.29377	1.73022	8.34355
Smallest Extreme Value	10.10643		0.31835	
Largest Extreme Value	9.76208		0.36899	
Gamma		828.23994	0.01200	
3-Parameter Gamma		289.82923	0.02048	4.00073
Logistic	9.94302		0.19611	
Loglogistic	2.29659		0.01977	
3-Parameter Loglogistic	8.51913		0.00004	-4999.73420

** Scale: Adjusted ML estimate*

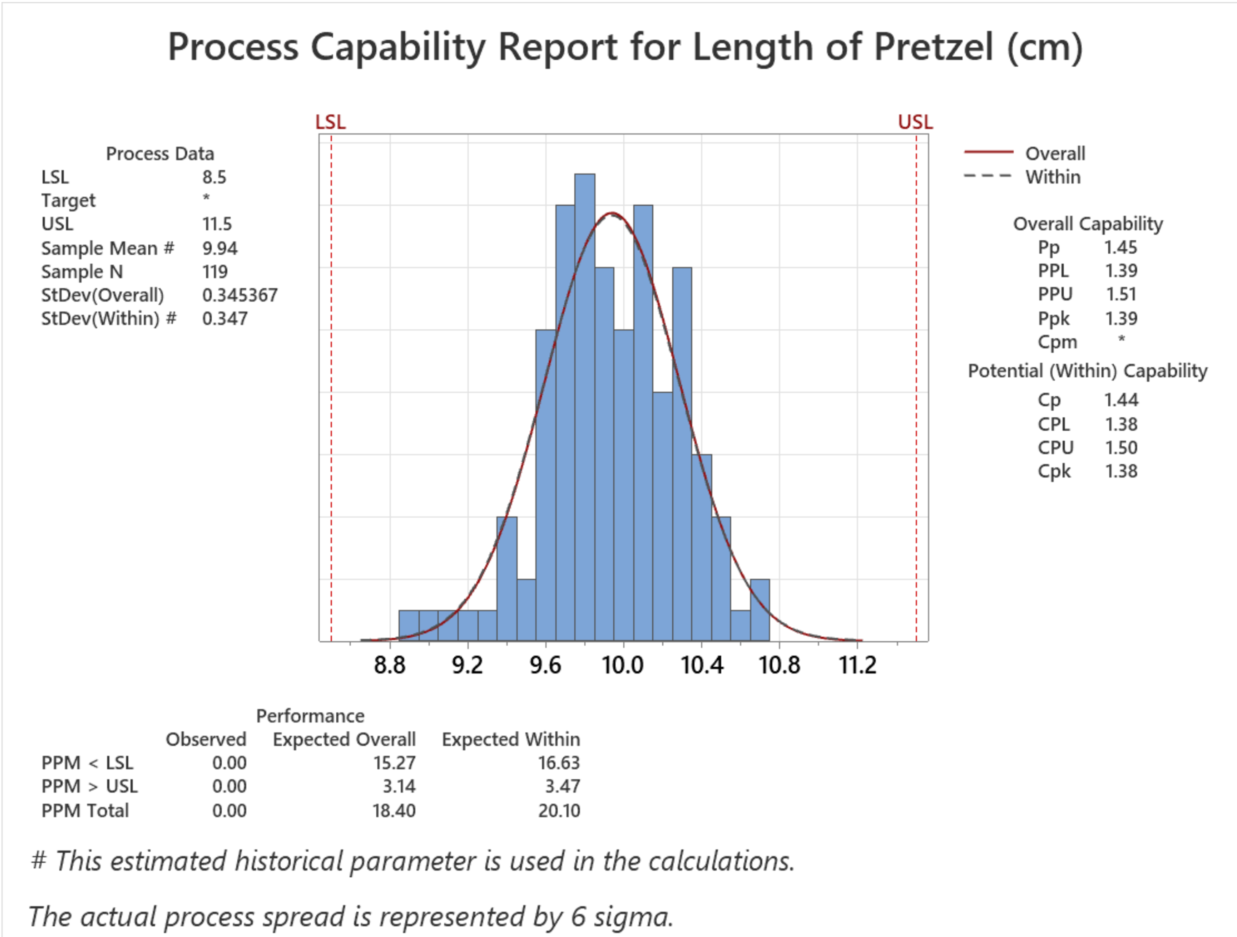
Figure 11: Mean and standard deviation parameter for various distribution of the Anderson-Darling normalization

Descriptive Statistics

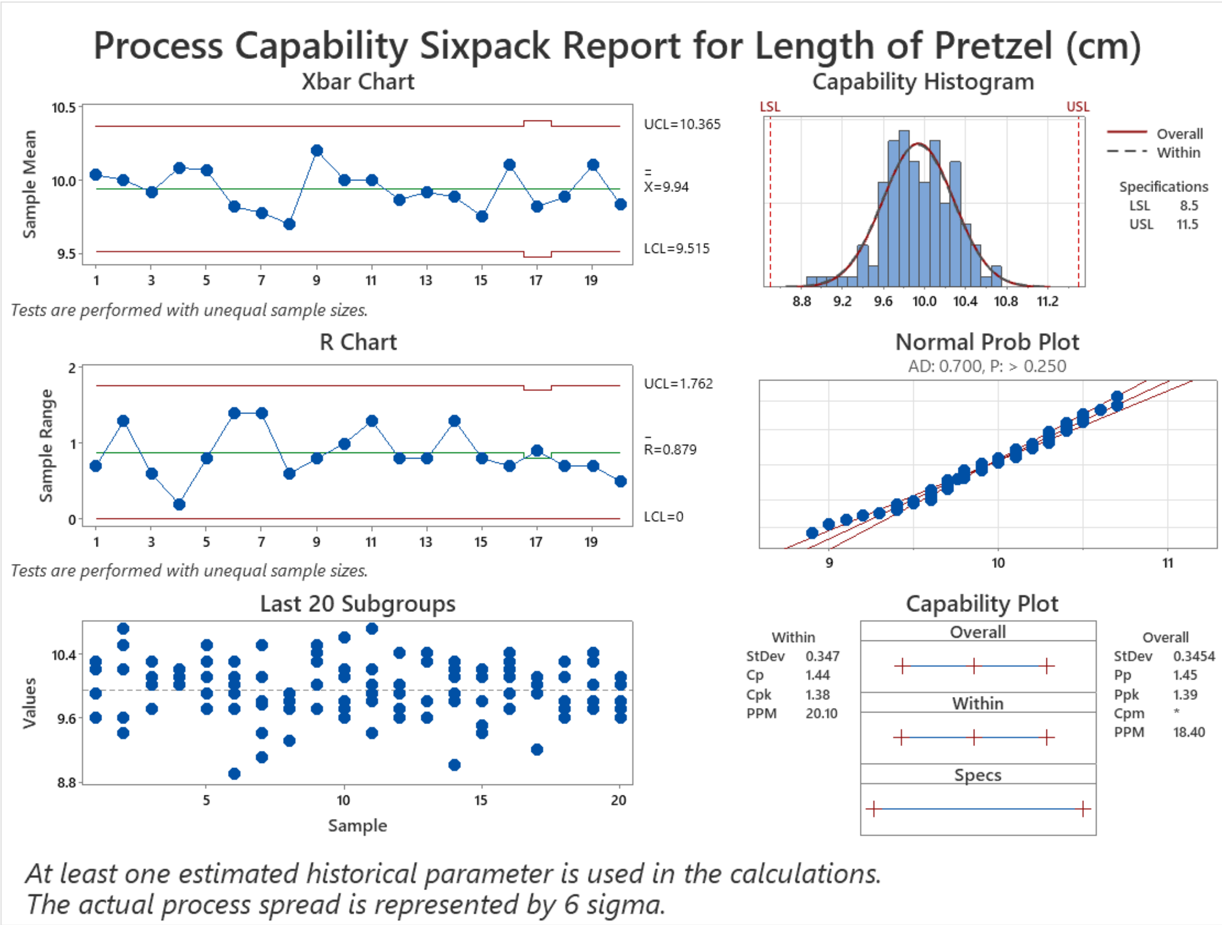
N	N*	Mean	StDev	Median	Minimum	Maximum	Skewness	Kurtosis
119	0	9.93832	0.345367	9.9	8.9	10.7	-0.285493	0.228027

Figure 12:Parameters for Phase 2 Anderson-Darling normalization

Appendix S: Process capability summary report

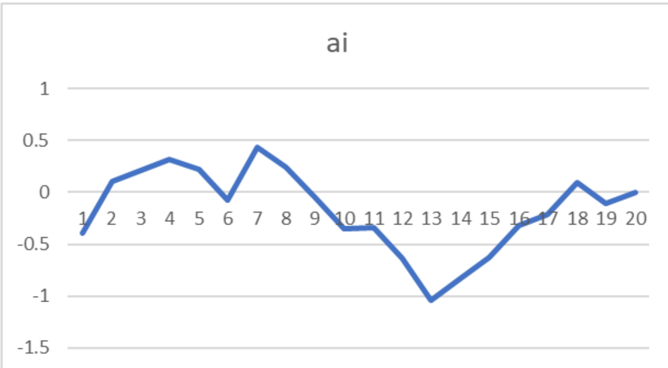


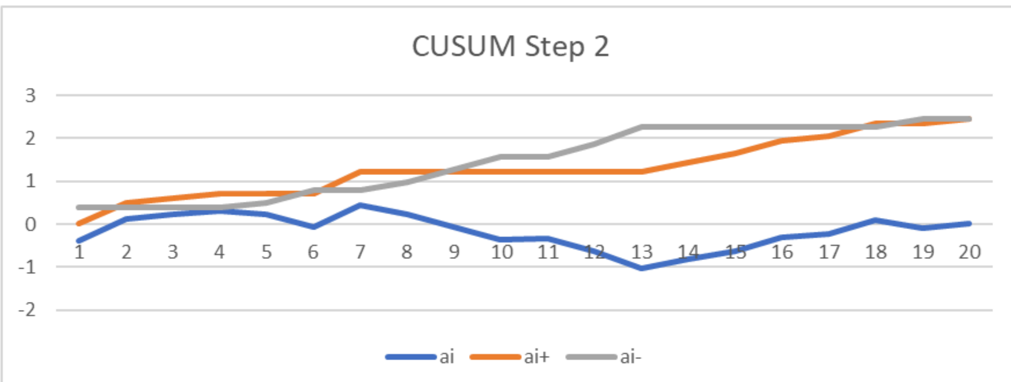
Appendix T: Process capability sixpack chart

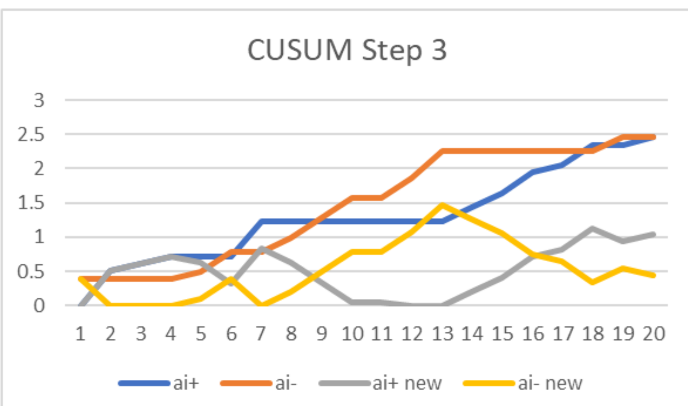


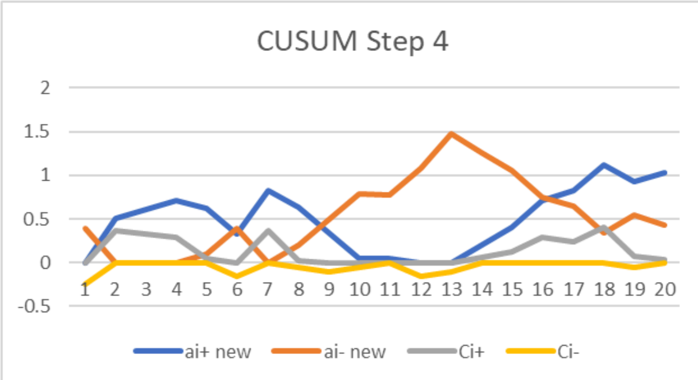
Appendix U: CUSUM Charts

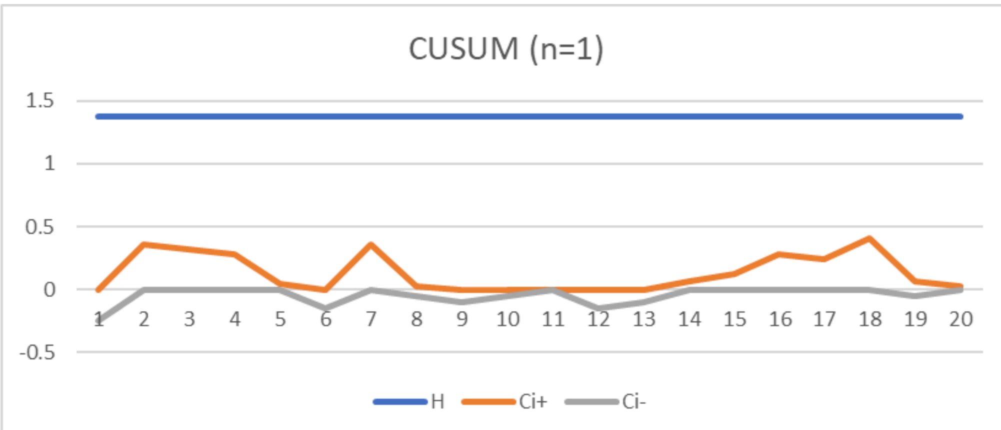
Please refer to the "CUSUM" Excel file for detailed steps of building the charts.

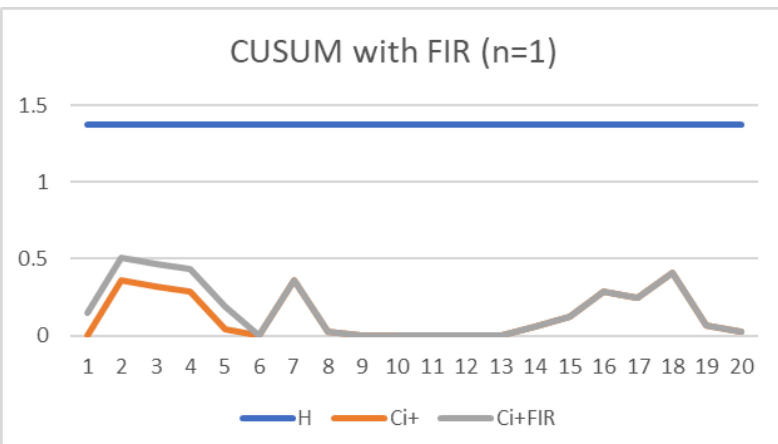


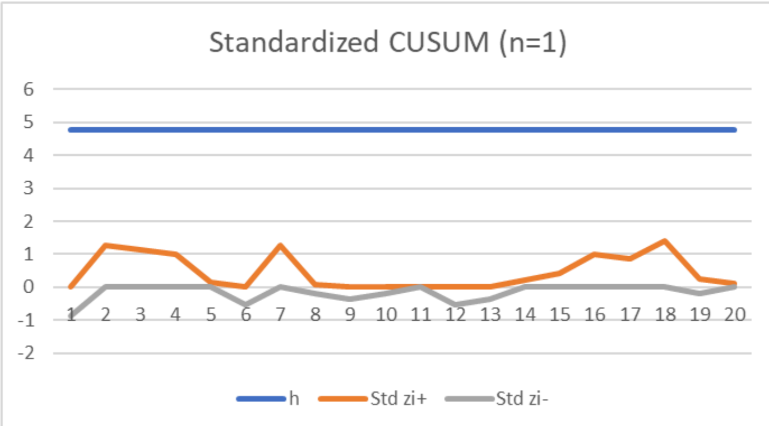


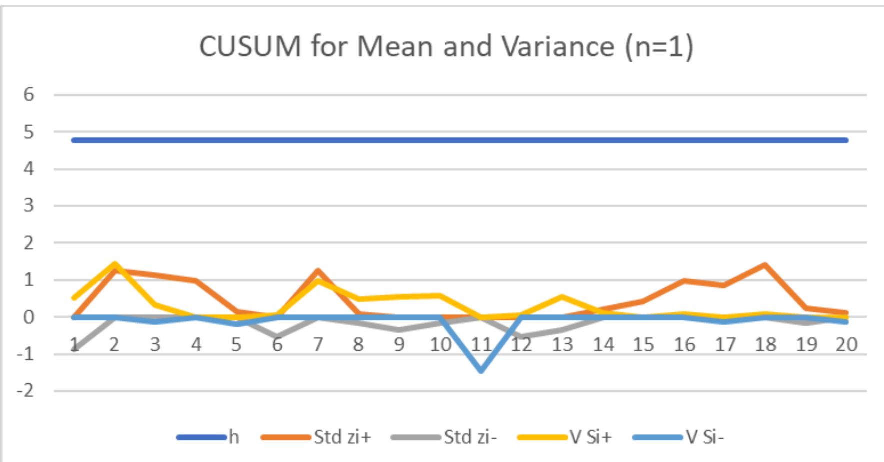


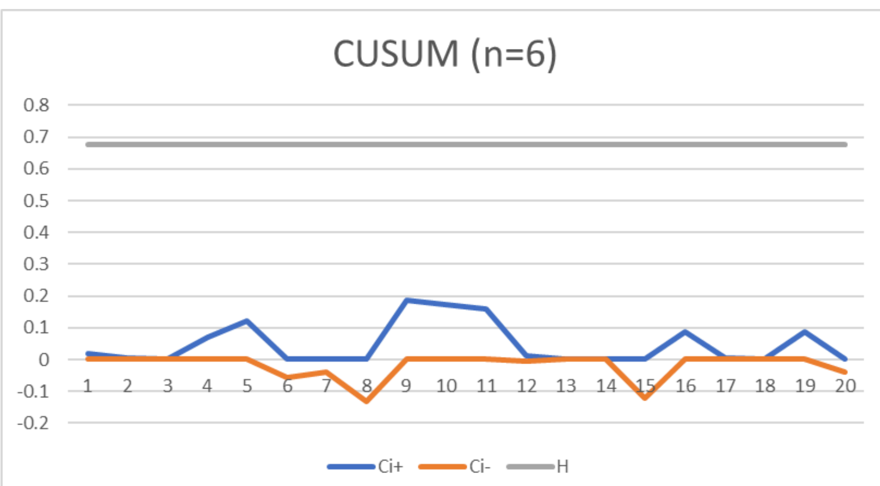

CUSUM Step 4
2
1.5
1
0.5
0
-0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
ai+ new
ai- new
Ci+
Ci-


CUSUM (n=1)
1.5
1
0.5
0
-0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
H
Ci+
Ci-


CUSUM with FIR (n=1)
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
H
Ci+
Ci+FIR

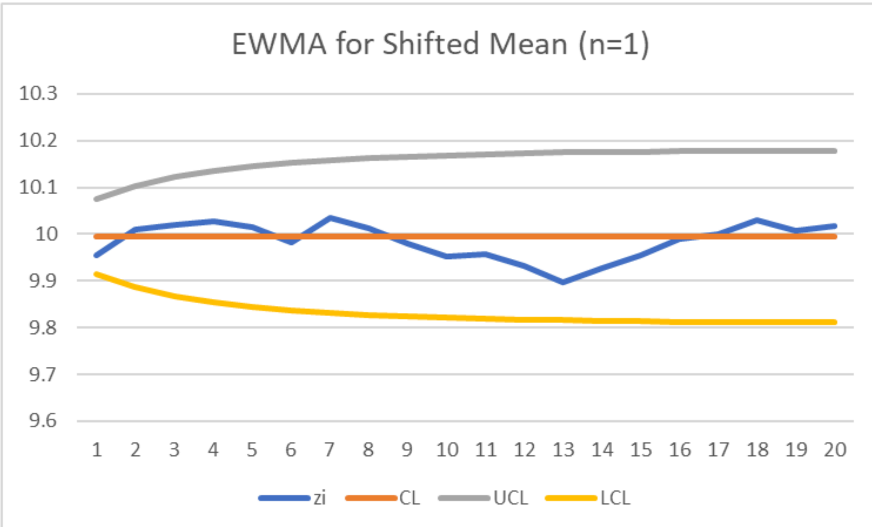

Standardized CUSUM (n=1)
6
5
4
3
2
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
h
Std zi+
Std zi-

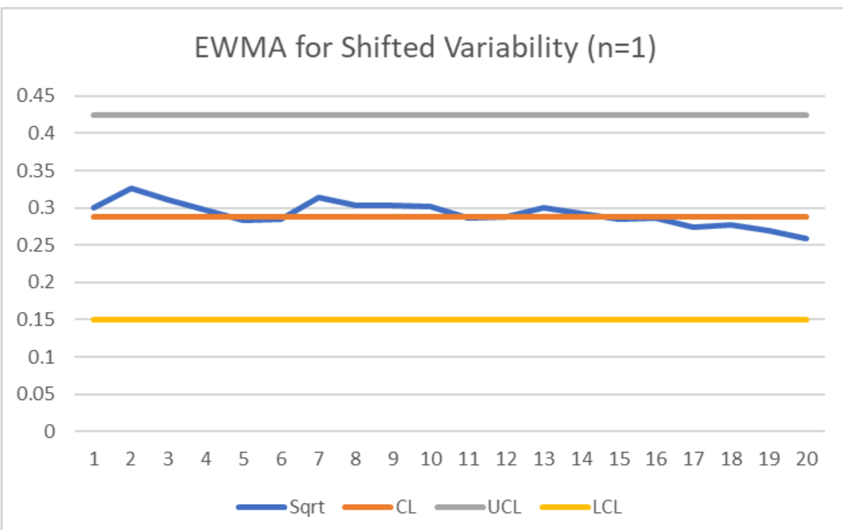

CUSUM for Mean and Variance (n=1)
6
5
4
3
2
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
h
Std zi+
Std zi-
V Si+
V Si-

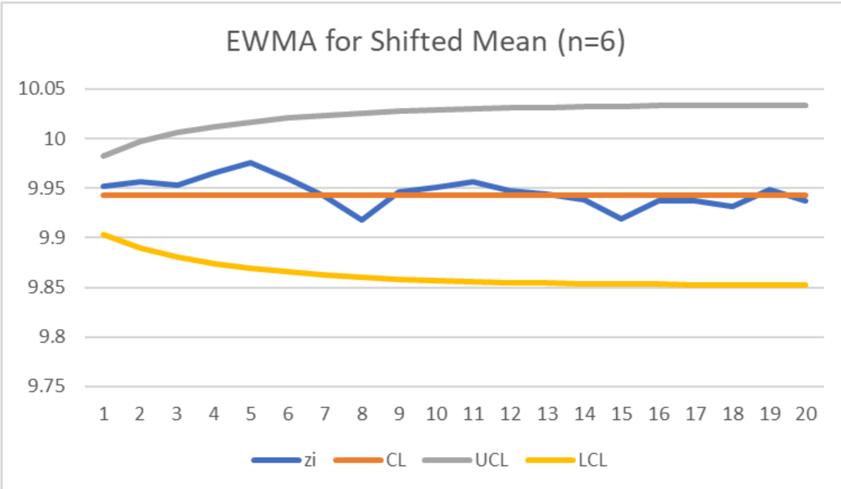

CUSUM (n=6)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Ci+
Ci-
H

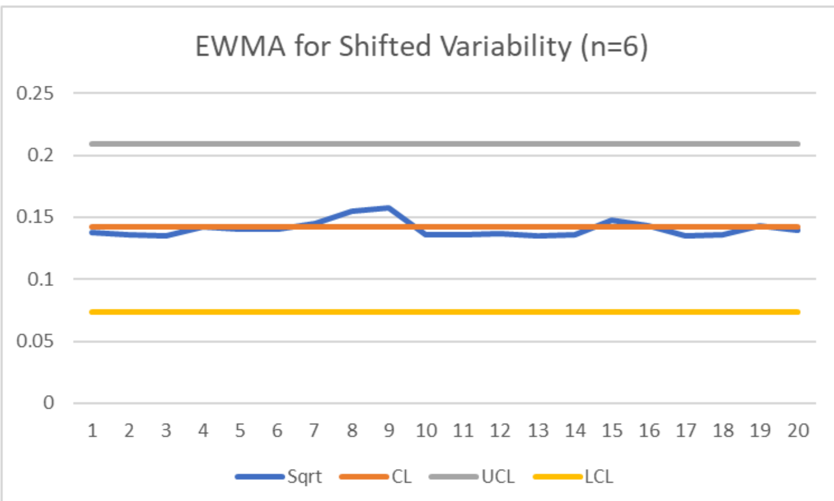
Appendix V: EWMA Charts

Please refer to the "EWMA" Excel file for detailed steps of building the charts.






EWMA for Shifted Mean (n=6)
10.05
10
9.95
9.9
9.85
9.8
9.75
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
zi
CL
UCL
LCL


EWMA for Shifted Variability (n=6)
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Sqrt
CL
UCL
LCL