SANDIA REPORT
SAND2026-16541
Printed January 2026

Sandia National Laboratories

An Overview of Manufacturing Controls for Production of High-Consequence, Single-Use Systems

Talia G. Duffy

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Extended Testing (Crowder, et al. (2025)) is a reliability demonstration technique that can be used to dramatically reduce sample size requirements. Manufacturing controls are needed to supplement extended testing by identifying production issues that a reduced sample sizes might overlook, especially built-in, or latent, manufacturing defects. This report focuses on some of the most-commonly used, yet most impactful, manufacturing control tools that are used to limit production-related defects and efficiently screen any remaining defects at final inspection. These tools include statistical process control (SPC), acceptance sampling, environmental stress screening (HASS and ESS), and mistake proofing. The goal is to minimize the probability that built-in defects ever reach the customer. In terms of nuclear weapons (NW), the goal is to prevent defective units from ever entering the nation's NW stockpile. Examples of each of the techniques are illustrated with case studies.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND TERMS

Acronym/Term	Definition
SPC	Statistical Process Control
EWMA	Exponentially Weight Moving Average
HASS	Highly Accelerated Stress Screen
ESS	Environmental Stress Screen
NW	Nuclear Weapons
NWE	Nuclear Weapons Enterprise
WR	War Reserve
Extended Testing	Methodology Used to Demonstrate Reliability of WR Components

1. INTRODUCTION

A manufacturing process should be efficient, controllable, and high-quality. Achieving all three can be a significant challenge, especially when working with high-consequence, single-use systems. In these cases, the cost of low quality can be extreme. At Sandia National Laboratories (SNL), there is extreme interest in minimizing the probability that a unit with manufacturing defects makes its way into the stockpile.

Methods of manufacturing control can supplement extended testing by catching issues that smaller sample sizes might overlook. This is especially true when it comes to built-in, or latent, manufacturing defects.

The idea of quality arises in all business ventures, whether it be manufacturing high-consequence single-use systems or carrying out a service. Many famous industrial engineers, business professionals and statisticians, notable for their work on developing process control methodologies and philosophies, have shared the following definitions of quality:

- Joseph Juran: "Fitness for use"
- Phillip B. Crosby: "Conformance to requirements"
- Myron Tribus: "Giving people what they have a right to expect"
- Armand Feignenbaum: "Meeting the expectations of the customer"
- W. Edwards Deming: "Good quality and the right uniformity have no meaning except with reference to the consumer's demands."

Some of these thoughts focus on the product's ability to serve its intended purpose. Others prioritize engineering guidance. Feigenbaum and Deming support an essential idea: many great figures in the field of quality control can try to define quality, but the real power is in the hands of the customers.

These thoughts are relatively nontechnical and may give the impression that quality is a vague, high-level goal. Over the years, industrial engineers and statisticians have tried to formally solve quality-related manufacturing problems. The fields of statistics, operations management, industrial engineering, and more contribute to a diverse collection of manufacturing controls methods and philosophies.

This report covers some of the most important manufacturing control methods, including statistical process control, acceptance sampling, environmental stress screening, and mistake proofing.

2. INTRODUCTION TO STATISTICAL QUALITY CONTROL

2.1. Introduction

The general idea behind quality control is best illustrated with an example.

The image below features one of the original Star Wars action figures released by Kenner Products in the late 1970s. The figure is assembled from multiple smaller, independently produced components, including the head piece.

Figure 1. Luke Skywalker in action figure form. Flickr Creative Commons photo.

If the head piece of this figure has an inner radius that is too large, it won't fit securely onto the torso. The entire figure must be scrapped; even if the arms, legs, and torso were produced correctly, the final product cannot be shipped out without a head. This wastes material and production time.

Since this figure has a defective head piece, it stands to reason that there is an error in the production process affecting other head pieces. Indeed, four out of ten Luke Skywalker heads are currently produced with inner radii too large to snap onto the torso. Figuring out what to do with these defective parts and reactively attempting to get this process back on track expends labor and time. The problem has progressed to the point where production must be completely halted.

Worst of all, before the issue was detected, thousands of figures with loosely attached heads were packaged and shipped out to stores. The heads easily toppled off in transit and during play, scaring children around the world. Kenner lost its good reputation with families, and sales dropped.

Clearly, letting the radii of these Luke Skywalker heads fall out of control had consequences at many steps in the production process. Star Wars action figures are not high-consequence, single-use systems, but it is not hard to imagine just how important quality control can be in producing airbag deployment systems or nuclear weapons systems.

2.1.1. Measuring Quality and the Statistical Basis for Quality Control

Statistical quality control (SQC) is a set of statistical techniques and visualizations used to monitor the output of a manufacturing process. For any product, requirements or target values are set by process engineers, customers, or other entities. Requirements focus on specific quality characteristics of the output, like height, voltage, or even simple classification into groups – defective versus non-defective, for example.

Production processes are subject to inherent and inevitable variation, and each quality characteristic's level of conformance relative to the requirements can be measured. The goal of SQC is to minimize the variance of key quality characteristics as much as possible.

For example, a quality characteristic of interest to Kenner is the inner radius of Luke Skywalker's head. Suppose Kenner toy engineers set a requirement that the inner radius must be 0.25 cm. Due to factors beyond anyone's control, no radius will be exactly equal to 0.25 cm, and each radius will be slightly different than the next. An acceptable Luke Skywalker head will have a radius close enough to 0.25 cm that it still snaps snugly onto the torso. This is called natural variation or "chance causes of variation."

However, an issue might develop in the Luke Skywalker Head Machine, causing four in ten heads to have an inner radius of about 0.35 cm. This is much too large; parts with such a measurement are defective. This excess variance in the radii suggests a specific problem in the machine should be eliminated. This is called unnatural variation or "assignable causes of variation" (Montgomery 2020).

The key to SQC is determining how much variation is too much variation. A process is considered "in control" if it is subject only to natural variation. A process is considered "out of control" once the output variation starts showing signs of an assignable cause. To be more specific, a process is in control if the quality characteristic falls within allowable limits, and out of control if it exceeds those limits (Shewhart 1931). This is often visualized with a Shewhart Control Chart, first developed by statistician Walter A. Shewhart in the 1920s (Best and Neuhauser 2006).

For example, suppose $m = 10$ sample groups, each of size $n = 4$ Luke Skywalker heads are taken from Kenner Production Plant #66. The radius of each head is measured, and the average radius of each sample is recorded. We will assume, for sake of the example, that the process is currently in control. The requirement for each radius is still 0.25. The upper specification limit (USL) set by process engineers is 0.251, and the lower specification limit (LSL) is 0.249. If manufactured within this range, the heads will still snap onto the torso.

In the current operating state of the process, the radii of Luke Skywalker heads are normally distributed:

$$H \sim N(\mu = 0.25, \sigma = 0.0005).$$

So, the sample average radius will have distribution:

$$\bar{H} \sim N\left(\mu = 0.25, \sigma = \frac{0.0005}{\sqrt{4}} = 0.00025\right).$$

The following samples are collected:

Table 1. The average radii of 10 head samples of size $n = 4$

Sample	Average Radius $\bar{H}$
1	0.250177
2	0.250009
3	0.250106
4	0.249798
5	0.250192
6	0.249557
7	0.249998
8	0.249647
9	0.250281
10	0.249896

First, we will plot the data serially, in the order it was collected. This allows us to view how the average radius of each head sample changes over time.

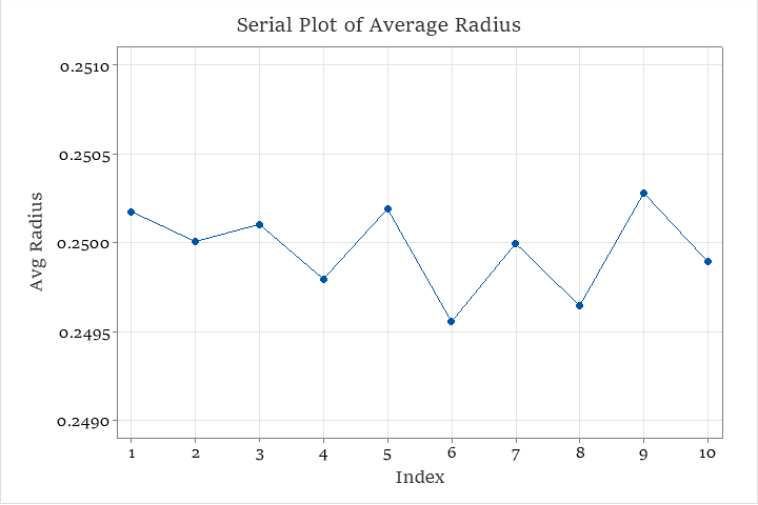

Figure 2. Serial plot of the data.

To finish this basic control chart, we calculate control limits (different from the specification limits!), plot the ten observations in order, and draw horizontal lines for the process average and control limits.

Here, the distribution of head radius is known, so we can use the given parameters to calculate the control limits. The center of the chart is, naturally, the mean $\mu = 0.25$. Assuming three-sigma control limits, our normally distributed process gives:

$$LCL = \mu + 3\frac{\sigma}{\sqrt{n}} = 0.25 + 3 * 0.00025 = 0.25075$$

$$UCL = \mu - 3\frac{\sigma}{\sqrt{n}} = 0.25 - 3 * 0.00025 = 0.24925.$$

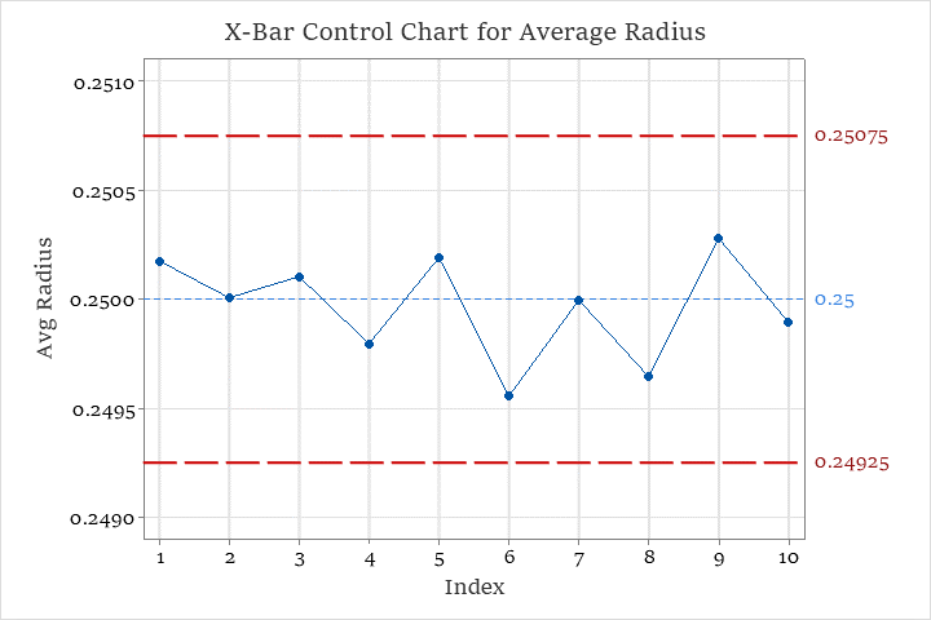

Figure 3. Completed control chart for average radius

This is called an $\bar{x}$ chart (pronounced "x-bar chart"), since it tracks the sample mean over time. The various types of control charts and how to calculate control limits will be discussed in the following sections. As in this chart, limits are often some chosen as some multiple k of the standard deviation away from the mean.

An overarching quality control plan consists of some statistical elements, like these charts, and some not so statistical elements. Organizations are advised to approach quality control and improvement not only as a set of tools, but as a general philosophy adopted at all levels (Montgomery 2020).

The leading dogma is known as Six-Sigma, after the standard Greek symbol σ, which usually represents the standard deviation of the Normal distribution. Six-Sigma was first established in 1986 at Motorola as a program to remedy manufacturing issues. Today, it is a comprehensive plan for

quality improvement practiced by organizations around the world. Companies and individuals can receive official Six Sigma training and certification. The central example of Six Sigma theory is known as DMAIC: Define, Measure, Analyze, Improve, and Control (Six Sigma US 2025).

Figure 4. The DMAIC process, central to Six-Sigma. Wikicommons photo.

Statistically, under Six Sigma the process should be controlled such that the control limits are six standard deviations above and below the process mean. Using the standard normal distribution, the probability of a point lying within six sigma control limits is 99.9999998%.

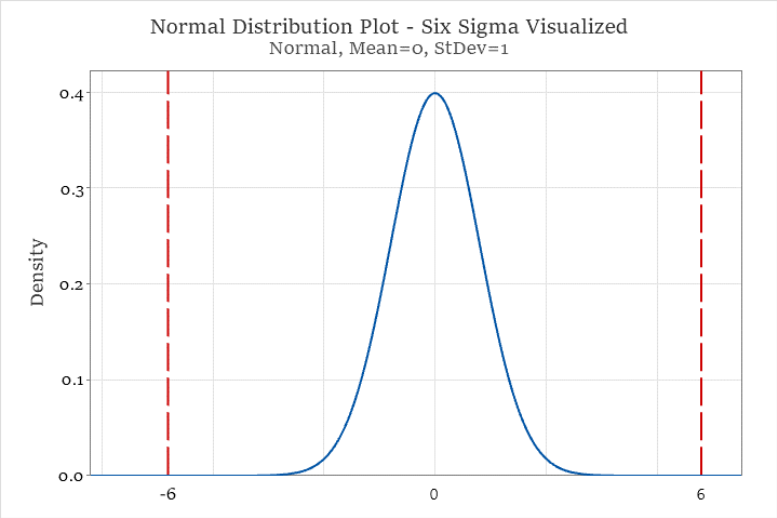

Figure 5. Normal distribution plot showing the near-zero probability in the tails at 6-sigma.

This results in a very low defect rate–in practice, it is essentially zero (Vardeman and Jobe 2016). With statistics, we can pin down the definition of quality into something measurable, controllable, and actionable.

2.2. Control Charts

The implementation of control charts is typically broken down into two phases.

Phase I, also known as the "off-line" phase, entails the initial setup of the control chart. Preliminary samples are taken, quality characteristics are measured, and "trial" statistics and control limits are estimated. After creating a control chart with these preliminary values, out-of-control points are identified, and their assignable causes are eliminated. Once the problem is taken care of, the control

limits are re-estimated without the offending points. Naturally, this process is iterative; Vining (2009) recommends repeating the process four to five times with 20 samples each before moving on with phase II.

Once we have reliable knowledge of what the process looks like when it's in control, we begin Phase II, also known as the "on-line" phase. This phase entails monitoring new output of the process and judging the quality characteristic against the control limits from Phase I. As new samples are taken, new points are plotted on the control chart. Operators watch for out-of-control points or suspicious patterns, and if a shift is detected, the search for an assignable cause begins.

Zwetsloot et al (2023) proposes two additional phases: Phase 0, data collection and defining measurement, and Phase 3, regular maintenance of the model. This four-phase plan codifies steps that are typically taken to support the standard two-phase plan and may be useful if the reader prefers more detailed guidance.

2.2.1. Control Charts as Hypothesis Tests

Statistically, control charts are a visual representation of the results of a hypothesis test for control over time.

Each point on a control chart is analogous to a test statistic. Our null hypothesis is that our process is in control, and our alternative hypothesis is that our process is out of control. If a point on the control chart falls between the control limits, we accept that we have a controlled process. But if a point falls outside the control limits, say there's enough evidence of an uncontrolled process and reject the null hypothesis.

Hypothesis tests are not the end-all-be-all of decision making, and the same is true for control charts. Statistician Acheson Duncan said it best in a warning to readers of Quality Control and Industrial Statistics (1965):

> "If no points fall outside control limits and if there is no evidence of nonrandom variation within the limits, it does not mean that assignable causes are not present. It simply means that the hypothesis that chance causes are alone at work is a tenable hypothesis and that it is likely to be unprofitable to look for special assignable causes."

The β-risk of a control chart is the probability that a shift in the process is not detected by a control chart on the next sample. This is analogous to type II error in a hypothesis test – the null hypothesis of control should be rejected, but it is not. The value of β is calculated with the new distribution of the control chart statistic following the shift.

The α-risk of a control chart is the probability that a point plots out of control when it was generated by an in-control process. This is analogous to type I error or the significance level in a

hypothesis test – the null hypothesis was rejected when it should not have been. In the long run, at least one point will plot out of control just by chance, when no assignable causes are present. The value of α is calculated with the estimated (or given) in-control distribution of the chart statistic.

We can drive the β-risk closer and closer to 0 by narrowing our control limits, but this would raise our α-risk to closer and closer to 1, and vice versa. Both situations are useless, so instead, the problem should be viewed as a careful trade-off between β-risk and α- risk.

In the United Kingdom, it is popular to choose control limits not by multiples of the standard deviation, but with a quantile that gives a desired type I error probability. A popular design is choosing limits such that the probability of a point falling outside the control limits is 0.002. However, Montgomery (2020) warns that the true distribution of the characteristic is not always known, and estimated sigma-based limits generally lead to good results that are close enough to what the probability limits would be anyway.

2.2.2. Control Chart Diagnostics and Design Concepts

One way of evaluating the performance of a control chart is the operating characteristic (OC) curve. OC curves plot the β-risk (probability of not detecting a shift) against the magnitude of the shift. Multiple curves can be drawn on the same graph to show how chart performance varies with sample size.

In the middle of the night, a mysterious small green figure breaks into Kenner Production Plant #66 and whacks the Luke Skywalker Head Machine with a brown cane. The machine falls out of its calibrated state and starts producing Luke Skywalker heads with a shifted mean. The small figure quickly escapes through a window, never to be seen again.

Figure 6. A mysterious small green figure. Flickr Creative Commons Photo.

Recall that the in-control process has distribution

$$H \sim N(\mu = 0.25, \sigma = 0.0005)$$

$$\bar{H} \sim N\left(\mu = 0.25, \sigma = \frac{0.0005}{\sqrt{4}} = 0.00025\right)$$

and control limits

$$LCL = \ 0.25075$$

and

$$UCL = 0.24925.$$

Let's say the shift was one standard deviation above the process mean, so the mean of H and $\bar{H}$ is now 0.2505. The standard deviation does not change.

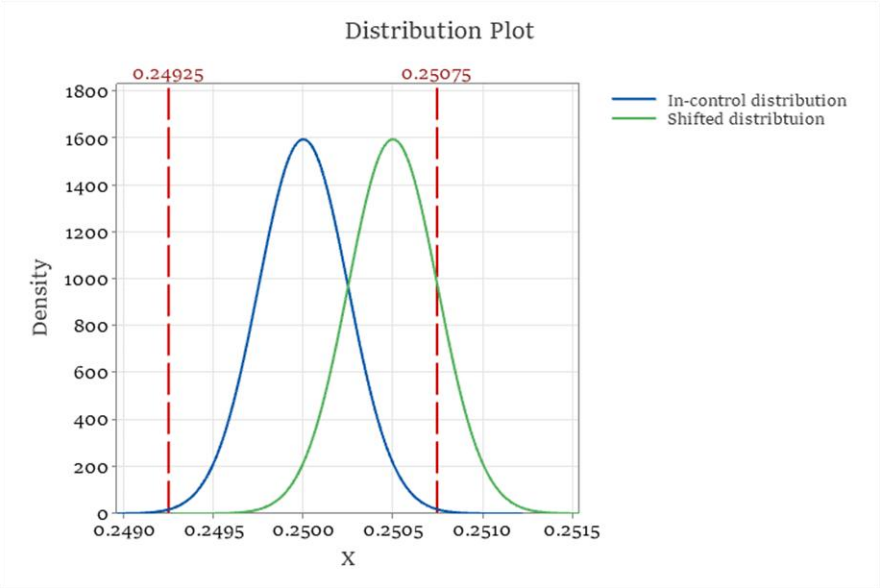

Figure 7. The distribution of average radius before and after the shift

The probability of $\bar{H}$ falling between the control limits in this state can be calculated with the normal cumulative distribution function:

$$\beta = P(0.24925 \leq \bar{H} \leq 0.25075 | \mu = 0.2505)$$

$$= \Phi\left(\frac{0.25075 - 0.2505}{0.00025}\right) - \Phi\left(\frac{0.24925 - 0.2505}{0.00025}\right)$$

$$= 0.84.$$

With this control chart, there's an 84% chance the shift will not be detected on the next sample.

The following table displays the result of the above process for different shifts:

Table 2. β risk for different shifts

Shift	Mean	β
0.5σ	0.25025	0.9772182
1σ	0.2505	0.8413445
1.5σ	0.25075	0.5
2σ	0.2510	0.1586553
2.5σ	0.25125	0.02275013
3σ	0.2515	0.001349898

And below is the corresponding OC curve. Note that the curve will be different for differ rent sample sizes.

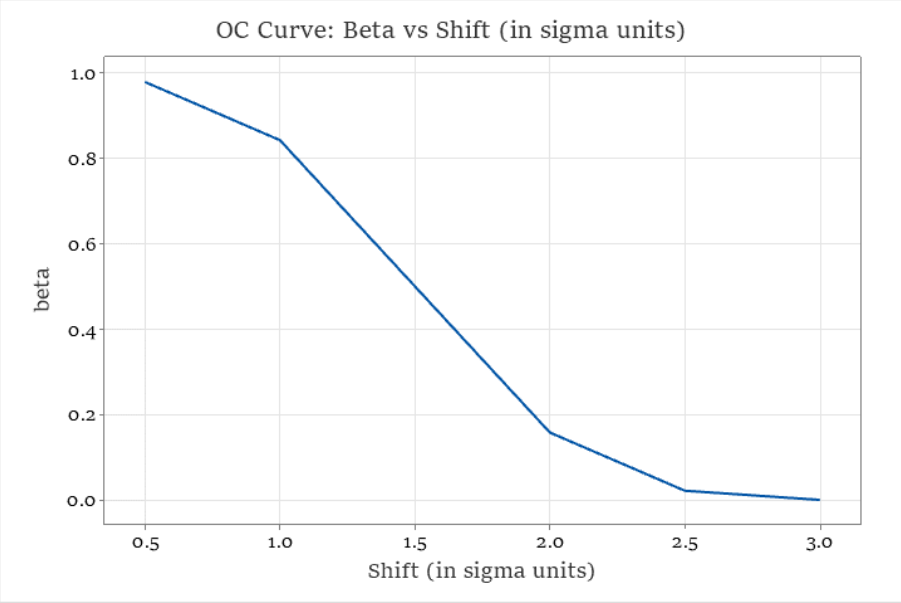

Figure 8. The OC curve for varying sigma shifts.

Another control chart diagnostic is known as the average run length (ARL), which is the number of samples measured, on average, until the first point plots out of control. For an in-control process, this can be interpreted as the number of samples measured until the first false alarm, and it is calculated as:

$$ARL = \frac{1}{\alpha}$$

Here, α is the α-risk probability as previously discussed.

If N is the number of samples until the first false alarm, then N follows a geometric distribution, see Vardeman and Jobe (2016). Specifically,

$$N \sim Geom(\alpha = \frac{1}{ARL}),$$

and we have $E(N) = ARL$, which gives the above formula. For three-sigma limits, assuming a normal distribution, the ARL is roughly 370 samples. This value is sometimes used as a benchmark to help set up appropriate control limits for other types of charts and distributions (Montgomery 2020).

For an out-of-control process, we are no longer dealing with "false alarms" since there really has been a shift. The ARL in this case is the average number of samples until we get a signal of the shift, and is calculated as:

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

For an out-of-control process, the probability of a point falling outside of the control limits is no longer α. Usually, it is higher, and given as $1-\beta$, where β is the β-risk previously discussed. This leads to a much shorter ARL – which is what we'd want (Vardeman and Jobe (2016)). The out-of-control ARL can be thought of as a measure of the power of a hypothesis test.

For example, if the Luke Skywalker Head Machine is operating in control, the α risk is

$$\alpha = P(\overline{H} \leq 0.24925 | \mu = 0.25) + P(\overline{H} \geq 0.25075 | \mu = 0.25)$$

$$= \Phi\left(\frac{0.24925 - 0.25}{0.00025}\right) + \left(1 - \Phi\left(\frac{0.25075 - 0.25}{0.00025}\right)\right)$$

$$= 0.0027$$

Which gives us an in-control ARL of:

$$ARL = 1/\alpha = 1/0.0027 = 370.37 \approx 370$$

Note that we have employed three-sigma limits in this example. The ARL we calculated matches the 370 samples previously discussed for three-sigma limits and a normal distribution.

Let's say there was a shift in the process mean of 1σ. The process is no longer in control. From the values in Table 2, the β-risk in this situation is 0.8413445. This gives us:

$$ARL = \frac{1}{1-\beta} = \frac{1}{1-0.8413445} = 6.30296 \approx 6$$

Which is a much shorter ARL, as expected. It will only take around 6 samples on average to detect a shift of 1σ.

Another key element of control chart design is the sampling strategy. The theory behind sampling for control charts is known as rational subgrouping. The idea is that groups or samples should be *rationally* chosen to maximize between group variation rather than within group variation (Wetherill and Brown, 1991). Rational subgrouping is essential if we hope to get accurate estimates of our process parameters – for example, a single sample taken over the course of a shift in the process mean would result in a very inflated $\hat{\sigma}$.

Since we are attempting to distinguish natural variation from unnatural variation, it's best to collect each sample over a short period of time. Vardeman and Jobe (2016) elaborate:

> "If what one calls 'samples' often contain data from genuinely different process conditions, the apparent level of background noise will be so large that it will be hard to see important process changes."

The frequency and size of samples are strongly influenced by knowledge of the process, both physically and economically. Frequency and size should be chosen to balance economic restrictions and effective construction of the chart. Sizes of $n = 3$, 4, or 5 are popular (Montgomery 2020).

2.2.3. Shewhart $\bar{x}$ *and* R Control Charts

Perhaps the most essential rendition of control charts, and one of the first Shewhart developed in the 1920s and 30s, are the $\bar{x}$ and R charts. They are used together to monitor the central tendency and variation of our quality characteristic over time. The construction of the $\bar{x}$ and R charts – and other control charts covered in this report – is, unless otherwise stated, summarized from Montgomery (2020), and his work should be referenced for further details.

Recall the $\bar{x}$ chart produced earlier in this chapter. It was constructed with known mean and standard deviation, which is rarely the case in real life. Typically, careful work must occur in Phase I of the control chart process to estimate the parameters of production.

The center line is estimated with the average of our sample averages – more succinctly called the grand average. If m is the number of samples taken, we have:

$$\bar{\bar{x}} = \frac{1}{m}\sum_{j=1}^{m} \bar{x_j}$$

Where each $\bar{x_j}$ is the sample average of sample j. Since we are pairing the $\bar{x}$ and R charts, it is natural to estimate the process standard deviation using the range. This is done by scaling the average sample range with some predefined constant. The range of each sample is calculated as:

$$R_j = x_{max_j} - x_{min_j}$$

And we calculate the average sample range as usual:

$$\bar{R} = \frac{1}{m}\sum_{j=1}^{m} R_j$$

Finally, we choose the appropriate constant A_2 based on the size of each sample, and construct final formulas for the control limits of the $\bar{x}$ chart:

$$UCL = \bar{\bar{x}} + A_2\bar{R}$$
$$LCL = \bar{\bar{x}} - A_2\bar{R}$$

The values of A_2 are calculated assuming three-sigma control limits. For adjustments to this formula, refer to Chapter 6 of Montgomery (2020). The table of values for this constant as it appears in Montgomery is included as an appendix to this chapter.

Note that the control limits of the $\bar{x}$ depend on the average range. Because of this, if both charts are plotting out of control, it is sensible to investigate the R chart first.

We revisit the previous example, with the data now expanded to include the range of each sample. Suppose we no longer know the exact distribution of the inner radius of each Luke Skywalker head.

Table 3. Sample information updated with sample range

Sample	Average Radius $\bar{H}$	Range R
1	0.250177	0.0015701
2	0.250009	0.0014646
3	0.250106	0.0023924
4	0.249798	0.0015639
5	0.250192	0.0006340
6	0.249557	0.0013881
7	0.249998	0.0008633
8	0.249647	0.0013025
9	0.250281	0.0012952
10	0.249896	0.0009532

For sample size $n = 4$, the appropriate value of A_2 is 0.729. The grand average is 0.249966. The average sample range is 0.00134273. The estimated control limits are then

$$UCL = 0.249966 + 0.729 * 0.00134273 = 0.250945$$

$$LCL = 0.249966 - 0.729 * 0.00134273 = 0.248987$$

Which are very similar to the control limits calculated when the distribution was known.

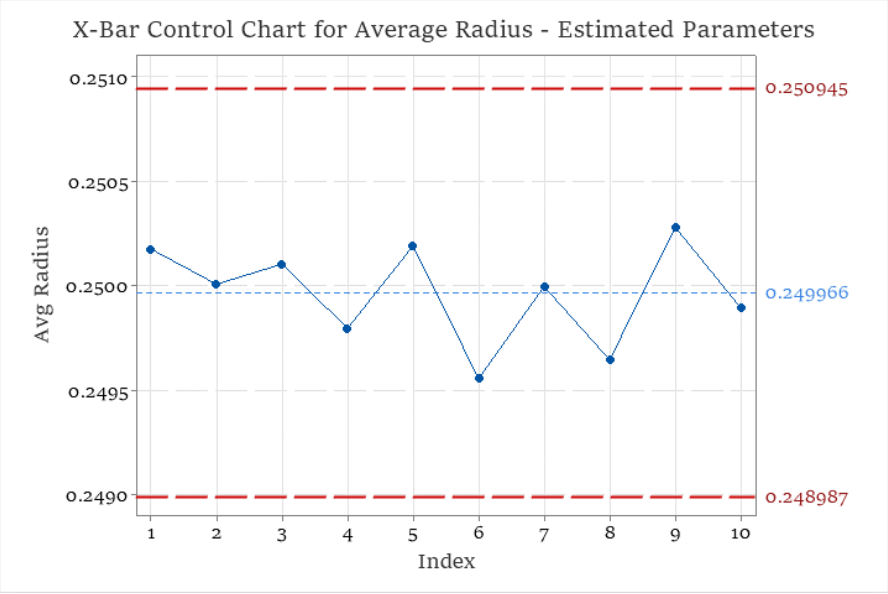

Figure 9. Completed control chart for average radius with estimated parameters.

The next step is to construct the corresponding R chart. The center line of the R chart is, naturally, the average sample range we already calculated. We multiply the sample range by predefined constants to get the upper and lower control limits.

$$UCL = D_4\bar{R}$$
$$LCL = D_3\bar{R}$$

The values of D_4 and D_3 are defined by the sample size. For $n = 4$, we have that $D_4 = 2.282$ and $D_3 = 0$. These values are also tabulated in Montgomery (2020). The control limits for our R chart are then:

$$UCL = 2.282 * 0.00134273 = 0.003064$$
$$LCL = 0 * 0.00134273 = 0$$

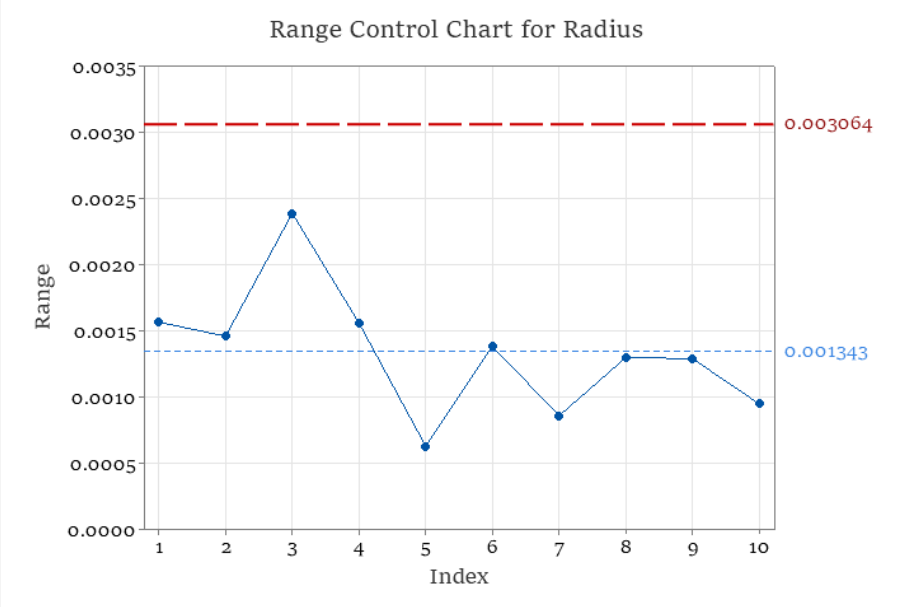

Figure 10. Range control chart for radius

As we can see, both the $\bar{x}$ and R charts indicate that our process is in control! Children will be spared from the horrors of toppling Luke Skywalker heads.

On $\bar{x}$ and R charts, points plotted outside of the control limits are the most obvious indicators of an out-of-control process. However, it is important to be wary of other patterns that may appear in Phase II. Examples include a long series of points moving in the same direction ("trends"), periodic movement ("cycles"), or too many consecutive points on one side of the center line ("runs") (Noskievičová 2013).

Anything that departs from what appears to be a truly random pattern of points should be considered a warning of an assignable cause.

Montgomery (2020) stresses the importance of distinguishing between specification limits and control limits when it comes to the $\bar{x}$ chart. Specification limits are always externally determined and should not be featured on control charts. Control limits are the k-sigma or probability limits that arise naturally from the variation of an in-control process.

One major weakness in $\bar{x}$ and R charts is that they are insensitive to small changes in the process; if a process is slightly out of control, the charts may not produce a signal. Some scholars promote the addition of warning limits (Wetherill and Brown, 1991) at smaller tolerances than the control limits. Another option is to add more sensitive chart types to the cadre, such as the Exponentially Weighted Moving Average (EWMA) chart, which is discussed in a later section.

2.2.4. Shewhart $\bar{X}$ *and* s Control Charts

Sometimes, the $\bar{x}$ chart is paired with a chart monitoring the sample standard deviation instead of the sample range of the process. This is called an s chart. This section will cover the construction of the s chart and new matching control limits for the $\bar{x}$ chart.

It is well known that

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

is an unbiased estimator of variance. However, taking the square root of the above to obtain s does not result in an unbiased estimate of standard deviation. To use these estimates on our control charts, we must scale the results with constants that depend on the sample size, n.
First, we calculate the standard deviation of each sample with:

$$s_j = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Naturally, the center line of the s chart is the average sample standard deviation:

$$\bar{s} = \frac{1}{m}\sum_{j=1}^{m} s_j$$

And the control limits are found, like the R chart, by scaling our center line with a constant that depends on sample size:

$$UCL = B_4\bar{s}$$
$$LCL = B_3\bar{s}.$$

Since we're using an s chart, the control limits for the $\bar{x}$ should be adjusted to

$$UCL = \bar{\bar{x}} + A_3\bar{s}$$
$$LCL = \bar{\bar{x}} - A_3\bar{s}$$

These constants are listed as they appear in Montgomery (2020), but depending on the reference and the estimator used for s, the values may be different.

2.2.5. Individual-Moving Range (I-MR) Control Charts

For some applications, it may be difficult or impractical to collect samples of multiple units. In this case, each sample will be of size 1, which poses the question: how do we calculate and monitor within-group variation when no such variation exists? Enter the Shewhart control chart for individuals, or the individuals-moving range (I-MR) control chart.

Since no range can be calculated for the sample, we instead calculate a moving range, usually of span $n = 2$:

$$MR_j = |x_j - x_{j-1}|$$

Once the moving range is calculated for each unit, we find the average moving range:

$$\overline{MR} = \sum_{j=1}^{m} MR_j$$

We can recalculate the control limits for the $\bar{x}$ chart as

$$UCL = \bar{x} + 3 * \frac{\overline{MR}}{d_2}$$
$$LCL = \bar{x} - 3 * \frac{\overline{MR}}{d_2}$$

It's easy to see that including 3 creates our 3-sigma limits, and $\overline{MR}/d_2$ is the new estimate of process standard deviation, with the constant d_2 corresponding to the span of the moving range.
We can also create a control chart for the moving range. The center line is $\overline{MR}$, and the control limits are found using the same constants as the R chart, except now they are chosen based on the span length instead of sample size:

$$UCL = D_4\overline{MR}$$
$$LCL = D_3\overline{MR}.$$

The span of the moving range can be increased. This will decrease the variance but increase the bias in our estimate of standard deviation.

Each instance of the moving range is dependent on the previous observation, so points on the moving range chart will be autocorrelated. Keep this in mind during Phase II interpretation. Montgomery (2020) warns that the Shewhart control chart for individuals is even worse at detecting small changes than the basic R chart. It is also highly sensitive to non-normality in the distribution of the quality characteristic. The following section presents a better alternative with EWMA control charts.

But first, an example. Kenner Production Plant #2187 has a Luke Skywalker Head Machine that produces heads at a much slower rate than Plant #66. The engineers there have decided that collecting samples of size greater than 1 would be impractical. As a result, they track the inner radius of Luke Skywalker Heads with Shewhart control charts for individuals.

The following data on $m = 10$ heads was collected, and the moving range of span $n = 2$ was calculated for each point:

Table 4. Data for the Shewhart individuals chart

Head	Inner Radius	Moving Range
1	0.25053	
2	0.24983	0.00070
3	0.25007	0.00024
4	0.25020	0.00013
5	0.25083	0.00063
6	0.24983	0.00100
7	0.25001	0.00018
8	0.24999	0.00002
9	0.24988	0.00011
10	0.25084	0.00096

The distribution is unknown here, so we will use the data to estimate the center line and control limits of our charts. For the $\bar{x}$ chart, our center line is the average inner radius of the 10 heads, $\bar{h} = 0.2502$. To find our control limits, we need average moving range. Here, $\overline{MR} = 0.000441$. For a span of $n = 2$, the constant $d_2 = 1.128$. This gives us control limits

$$UCL = 0.2502 + 3 * \frac{0.000441}{1.128} = 0.251373$$

$$LCL = 0.2502 - 3 * \frac{0.000441}{1.128} = 0.249027$$

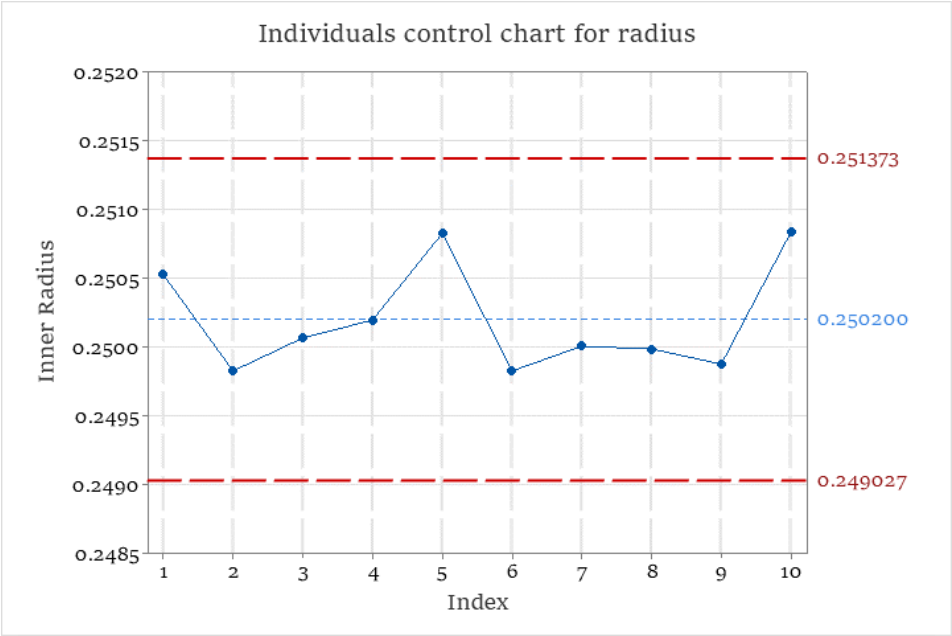

Figure 11. Shewhart individuals control chart for radius

We have already calculated $\overline{MR}$, which is the center line of the moving range chart. For a span of $n = 2$, the constants $D_3 = 0$ and $D_4 = 3.267$. This gives us control limits:

$$UCL = 3.267 * 0.000441 = 0.001441$$

$$\text{LCL} = 0 * 0.000441 = 0$$

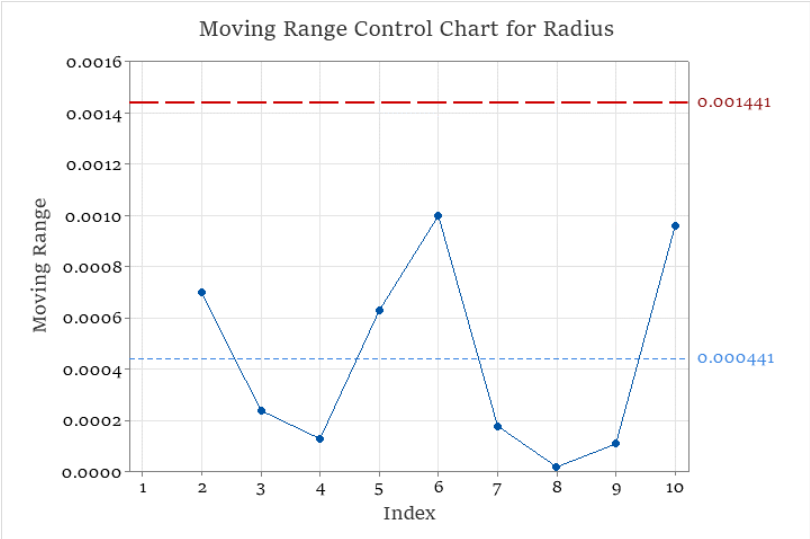

Figure 12. Individuals moving range control chart for radius

2.2.6. Exponentially Weighted Moving Average Control Charts

The Exponentially Weighted Moving Average (EWMA) chart is a popular option in the current SQC landscape because it improves upon several weaknesses found in classical Shewhart control charts. The biggest advantage: EWMA charts are better at detecting small shifts without sacrificing in-control ARL. These charts can be used for both individual observations and sample means.

The design steps in this chapter will largely follow Crowder (1989) and Montgomery (2020). Over the sample number or "age," EWMA charts plot the exponentially weighted moving average of the quality characteristic:

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1}$$

The parameter λ, known as the "smoothing parameter," is chosen between 0 and 1 (not including 0) and is the weight of the current observation.

The initial value z_0 is either some prespecified target value or the average of a preliminary sample (like setting up the control limits for a Shewhart control chart in Phase I). We also use this target value or sample mean as the center line of the EWMA chart and call it μ_0.

At $\lambda = 1$, the weight of past observations is sent to zero; the EWMA depends only on the current observation and the chart is equivalent to the $\bar{x}$ chart. At λ close to zero, the current observation has little weight. It can be demonstrated that large values of λ are better for detecting large shifts and smaller values of λ are better for detecting small shifts.

To visualize this weighting scheme, we leverage the recursiveness of the EWMA formula.
Let x_t be observation number t. The EWMA for observation t is

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}$$

The EWMA for observation $t - 1$ is

$$z_{t-1} = \lambda x_{t-1} + (1 - \lambda) z_{t-2}$$

Which can be plugged into z_t:

$$z_t = \lambda x_t + (1 - \lambda)[\lambda x_{t-1} + (1 - \lambda) z_{t-2}]$$

$$= \lambda x_t + \lambda(1 - \lambda) x_{t-1} + (1 - \lambda)^2 z_{t-2}$$

The EWMA for observation $t - 2$ is

$$z_{t-2} = \lambda x_{t-2} + (1-\lambda) z_{t-3}$$

Which can again be plugged into z_t:

$$z_t = \lambda x_t + \lambda(1-\lambda) x_{t-1} + (1-\lambda)^2 [\lambda x_{t-2} + (1-\lambda) z_{t-3}]$$

$$= \lambda x_t + \lambda(1-\lambda) x_{t-1} + \lambda(1-\lambda)^2 x_{t-2} + (1-\lambda)^3 z_{t-3}$$

And again:

$$z_t = \lambda x_t + \lambda(1-\lambda) x_{t-1} + \lambda(1-\lambda)^2 x_{t-2} + \lambda(1-\lambda)^3 x_{t-3} + (1-\lambda)^4 z_{t-4}$$

This can be repeated until we reach the "first" observation x_0. As we can see, for the current EWMA at observation x_t, the weight of observations x_{t-j} decreases geometrically according to the formula

$$\text{weight of } x_{t-j} = w(j) = \lambda(1-\lambda)^j$$

Where j can be thought of as the "age" of the sample. For example, if $\lambda = 0.1$, and we are currently at time t, the weights of previous samples can be plotted as follows:

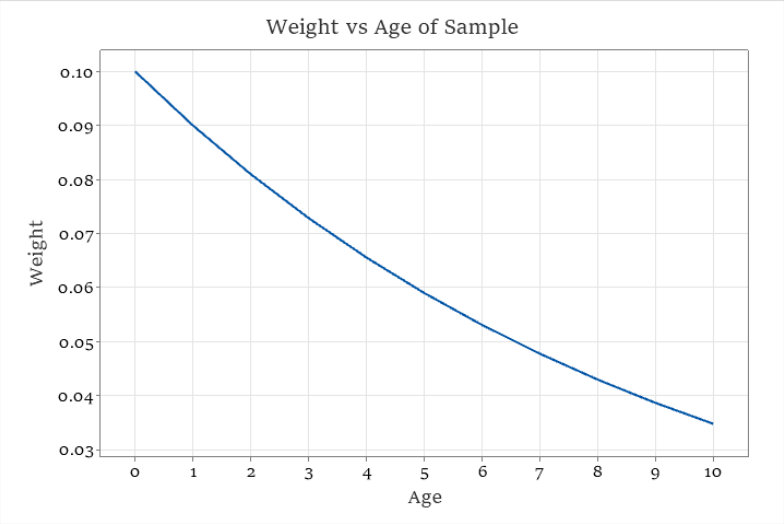

Figure 13. Geometrically decreasing weight of previous samples.

For independent observations x_i with variance σ^2, the variance of the EWMA is

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda}\right)\left[1 - (1-\lambda)^{2i}\right]$$

which gives control limits:

$$UCL = \mu_0 + K\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]}$$

$$LCL = \mu_0 - K\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]}$$

Again, we have control limits defined by K multiples of the standard deviation. The variance of z_i approaches a constant value over time, and this asymptotic variance gives stabilized control limits:

$$UCL = \mu_0 + K\sigma\sqrt{\frac{\lambda}{2-\lambda}}$$

$$LCL = \mu_0 - K\sigma\sqrt{\frac{\lambda}{2-\lambda}}$$

Crowder (1989) outlines a clear process for choosing K and λ:

1. Fix the false alarm rate (in-control ARL) to a desired value. A common choice is 370, to match a Shewhart chart with three-sigma limits.
2. Fix the shift level that is of greatest interest. For example, engineers may feel a shift of 0.5 standard deviations should be detected quickly. Crowder's paper features several "lambda versus shift" curves for different in-control ARLs. Use the appropriate curve to locate the optimal λ for the chosen shift.
3. Crowder's paper also features several "K versus lambda" curves for different in-control ARLs. Now, use the appropriate curve to locate the optimal K for the chosen λ.
4. Perform a sensitivity analysis: choose different combinations of λ and K that create the same in-control ARL and see how they perform with various shift magnitudes. Choose the combination that works best for your needs.

The operators at Kenner Production Plant #66 believe the consequences of an out-of-control inner radius are so great that small changes should be detected early – specifically, a shift of 0.5σ. They decide to start using an EWMA chart and change their sampling scheme accordingly. Now, individual inner radii are measured and recorded at frequent intervals.

Suppose we know the in-control distribution of the inner radius, defined as before:

$$H \sim N(\mu = 0.25, \sigma = 0.0005)$$

Right under the operators' noses, a mysterious small furry creature crawls up to the Luke Skywalker Head Machine, opens it up, and steals a shiny gear. The creature pockets the gear and sneaks out of the production plant. Without this gear, the machine starts producing Luke Skywalker heads with a shifted mean.

Figure 14. A mysterious small furry figure. Flickr Creative Commons Photo.

Suppose the shift was 1σ, so the mean of the inner radii after sample 5 is 0.2505. The following 15 measurements are taken:

Table 5. New individuals samples of the inner radii at Kenner Production Plant #66

Sample	Average Radius $\overline{H}$
1	0.24946
2	0.25027
3	0.25030
4	0.25013
5	0.24976
6	0.250563
7	0.250434
8	0.250159
9	0.250473
10	0.250497

11	0.249142
12	0.250788
13	0.249790
14	0.251201
15	0.250563

The operators have chosen $\lambda = 0.05$ and $K = 2.5$, estimating optimal values for an in-control ARL of 370 from Crowder's charts. Using the recursive formula given above and the target 0.25 for z_0, we can calculate:

$$z_1 = 0.05 * 0.24946 + (1 - 0.05) * 0.25 = 0.249973$$

$$z_2 = 0.05 * 0.25027 + (1 - 0.05) * 0.249973 = 0.249988$$

And so on.

The control limits are:

$$UCL_1 = 0.25 + 2.5 * 0.0005\sqrt{\frac{0.05}{2 - 0.05}[1 - (1 - 0.05)^{2*1}]} = 0.2500625$$

$$LCL_1 = 0.25 - 2.5 * 0.0005\sqrt{\frac{0.05}{2 - 0.05}[1 - (1 - 0.05)^{2*1}]} = 0.2499375$$

$$UCL_2 = 0.25 + 2.5 * 0.0005\sqrt{\frac{0.05}{2 - 0.05}[1 - (1 - 0.05)^{2*2}]} = 0.2500862$$

$$LCL_2 = 0.25 - 2.5 * 0.0005\sqrt{\frac{0.05}{2 - 0.05}[1 - (1 - 0.05)^{2*2}]} = 0.249913793$$

And so on. If the observed data x_t is a single measurement, its standard deviation is σ. If x_t is instead the average of n measurements, its standard deviation is $\frac{\sigma}{\sqrt{n}}$.

The resulting EWMA control chart is shown in Figure 15.

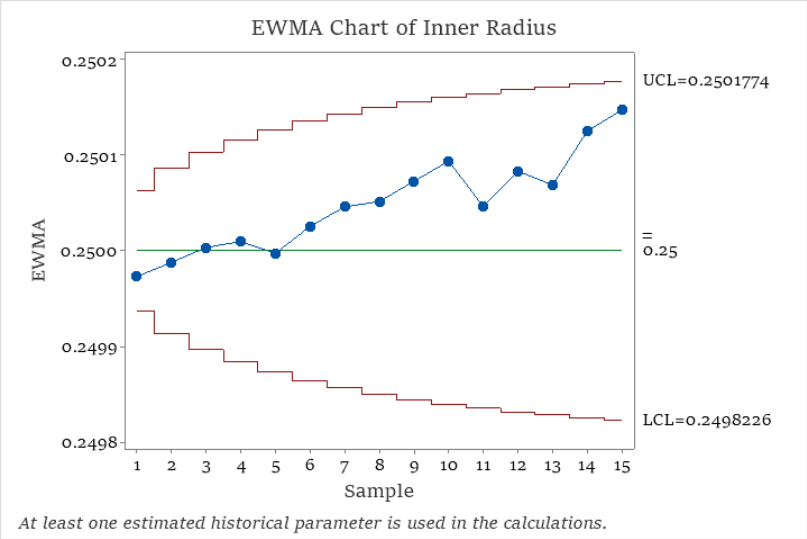

Figure 15. Minitab-generated EWMA chart for the inner radius

Another advantage of the EWMA chart over Shewhart charts is that it works well for non-normally distributed quality characteristics, in terms of both shift detection and in-control ARL. EWMA charts can be used in conjunction with the standard $\bar{x}$ and R charts so that both small and large shifts can be detected.

Although popular for monitoring process means, EWMA charts can be created to monitor process variability as well, as outlined by Crowder and Hamilton (1992). They propose a one-sided EWMA chart (with an upper control limit only) tracking the quantity:

$$y_i = ln(s_i^2)$$

For each sample, i. They apply this transformation because the natural log of the variances more closely resembles a normal distribution. With some probability theory not shown here, they also note that changes in variance σ^2 will cause a shift in the location parameter of the distribution of $ln(s_i^2)$, which is preferred for an EWMA chart.

Since this is a one-sided chart, they tweak the EWMA:

$$z_i = max[(1-\lambda) * z_{i-1} + \lambda y_i\,, ln(\sigma_0^2)]$$

Where $ln(\sigma_0^2)$ is some target or sample-calculated value. This value is also used as the starting z_0.

Crowder and Hamilton note their reason for establishing a minimum value of the EWMA at $ln(\sigma_0^2)$:

> "Since we are primarily interested in detecting increases in process variability, the EWMA is reset to $ln(\sigma_0^2)$ if its value ever falls below $ln(\sigma_0^2)$. Decreases in process variability, which are also important, will be suggested by the frequent occurrence of values equal to $ln(\sigma_0^2)$ on the EWMA chart."

To begin creating the control chart, they show that

$$Var\left(ln(s_i^2)\right) \cong \frac{2}{n-1} + \frac{2}{(n-1)^2} + \frac{4}{3(n-1)^2} - \frac{16}{15(n-1)^5}$$

And, as a result,

$$\sigma_{z_i}^2 = \left(\frac{\lambda}{2-\lambda}\right) * Var\left(ln(s_i^2)\right)$$

This gives the control limit:

$$UCL = ln(\sigma_0^2) + K\sqrt{\frac{\lambda}{2-\lambda} * Var\left(ln(s_i^2)\right)}$$

Where λ and K are chosen with a scheme similar to Crowder's scheme for EWMA mean charts, described above.

2.2.7. Control Charts for Attribute Data

Sometimes, the quality of a product is not judged by measurement of a specific variable; rather, they are classified as conforming versus nonconforming or pass versus fail. For single-use high consequence systems, this situation is common. Quality data in this form are known as attribute data. With a little probability theory, we can create attribute control charts to monitor and improve these processes, too.

p Chart

First, we will cover the p chart, or the fraction nonconforming chart. If D_i is the number of nonconforming units in each sample and n is the size of each sample, then the quantity we track on a p chart is:

$$\widehat{p_i} = \frac{D_i}{n}$$

Which is the ratio of nonconforming units to total units in each sample. This p_i is judged over time against a given target value or an estimated mean.

Previous charts were based on the normal distribution, but the p chart is based on the binomial distribution. Let D be the number of units that are nonconforming in a sample of size n, and let p be the probability that any unit is nonconforming. Then D has distribution:

$$D \sim binom(n, p)$$

With probability mass function

$$P(D = d) = \binom{n}{d} p^d (1-p)^{n-d}, \qquad d = 0,1,2,3 \ldots n$$

If we take a sample of size n, D is a random variable, and, therefore, the sample fraction nonconforming is also a random variable defined as

$$\hat{p} = \frac{D}{n}$$

with expected value and variance

$$E(\hat{p}) = p$$

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

Suppose the distribution of D is known and defined as above. Then, naturally, to set up a control chart for the sample nonconforming $\widehat{p_i}$ of each sample i, the center line is the known p. The control limits are still defined in terms of standard deviation, with 3σ as a popular choice:

$$UCL = p + 3\sqrt{\frac{p(1-p)}{n}}$$

$$LCL = p - 3\sqrt{\frac{p(1-p)}{n}}$$

If a negative control limit results from these equations, set it to zero, since counts, and therefore proportions, cannot be negative.

If the true value of p is not known, it can be estimated with the average sample fraction nonconforming:

$$\bar{p} = \frac{1}{m}\sum_{i=1}^{m} \hat{p}_i$$

Where m is the number of samples taken. Replacing all instances of p in the previous equations with this estimate $\bar{p}$ gives us trial control limits that should be refined in Phase I.

To avoid false alarms, it's important to pay attention to the sample size for p chart design. If the probability of defects is small, it's possible that a single nonconforming unit would cause a sample to plot out of control. The value of n should be large enough that there's a few nonconforming units in each sample.

For example, Rice (1947) suggests solving for n such that $P(X < 1) = 0.1$, where X is the number of nonconforming units. This can be done with the binomial distribution directly or the Poisson approximation to the binomial. Duncan (1965) describes solving for n such that the probability of detecting some specified shift on the next sample is at least 0.5.
Let us return to Kenner Production Plant #66.

Once all the individual parts are created, a Luke Skywalker action figure is considered nonconforming if it cannot be effectively assembled. Suppose $m = 10$ samples of $n = 25$ figures are taken, and the fraction of nonconforming figures is recorded.

Table 6. Fraction nonconforming for $m = 10$ samples of $n = 25$ Luke Skywalker figures

Sample	Fraction Nonconforming $\hat{p}_i$
1	0.20
2	0.16
3	0.24
4	0.16
5	0.20
6	0.36
7	0.28
8	0.20
9	0.36
10	0.16

The distribution is not known here, so we will have to estimate our parameters from the data. The sample average nonconforming $\bar{p} = 0.232$.

The resulting control limits:

$$UCL = 0.232 + 3\sqrt{\frac{0.232(1-0.232)}{25}} = 0.48527$$

$$UCL = 0.232 - 3\sqrt{\frac{0.232(1-0.232)}{25}} = -0.02127 \rightarrow 0$$

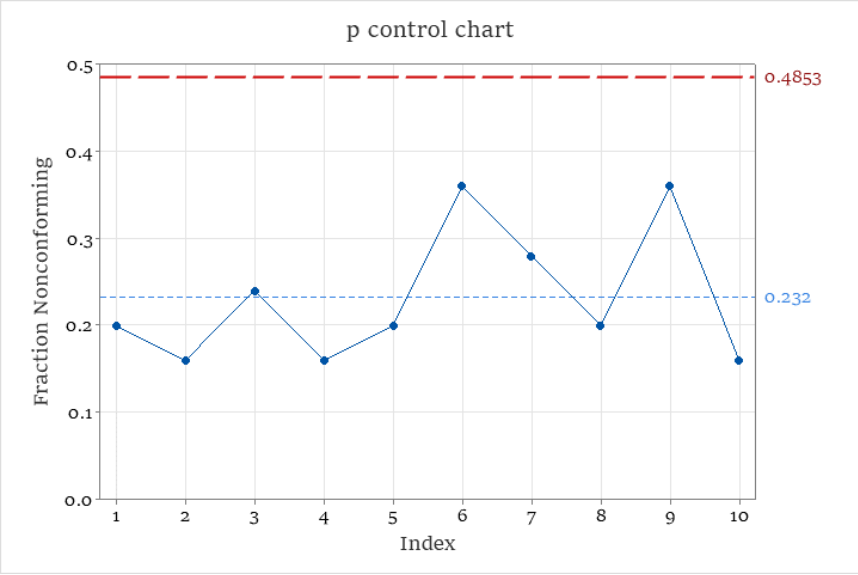

Figure 16. Fraction nonconforming control chart for action figures

c Chart

Another type of attribute chart is called a c chart. Sometimes, we may want to count the total number of nonconformities on a single unit rather than the number or fraction of nonconforming units in a sample. In this case, the number of nonconformities is not bounded by a constant n – it can theoretically extend towards infinity. Imagine a countably infinite number of cracks in an eggshell, or a countably infinite number of scratches on a phone screen. This idea can be modeled with the Poisson distribution. Z

Suppose C is the number of nonconformities on a single unit, then C has distribution:

$$C \sim Poisson(\omega),$$

with probability mass function:

$$P(C = c) = \frac{e^{-\omega}\omega^{c}}{c!}.$$

The mean and variance of the Poisson distribution are quite simple:

$$E(C) = \omega$$

$$Var(C) = \omega$$

If the distribution is known, the center line of the c chart is ω. The three-sigma control limits are simply

$$UCL = \omega + 3\sqrt{\omega}$$

$$LCL = \omega - 3\sqrt{\omega}$$

If the LCL falls to a negative value, it should be set to 0. If the distribution is not known, it should be estimated in a Phase I process by taking the average number of nonconformities of a sample of units over time:

$$\bar{\omega} = \sum_{j=1}^{m} \omega_j$$

Where ω_j is the number of nonconformities on sample unit j. Control limits are then found by replacing all instances of ω in the previous equations with this estimate $\bar{\omega}$.

With all these mysterious creatures running around, sometimes the Luke Skywalker action figures at Kenner Production Plant #66 get scratched up. Theoretically, it's possible to have one scratch, 10 scratches, 100 scratches or more on a single figure (although the latter is highly unlikely). Marketing analysts discovered that if an action figure has more than 5 scratches, a customer will be disappointed enough to buy their next action figure from one of Kenner's competitors.

In response, operators have controlled the process to follow a Poisson distribution with an average $\omega = 1.5$ scratches per unit. Thus, our center line is at 1.5 and our control limits are

$$UCL = 1.5 + 3 * \sqrt{1.5} = 5.1742$$

$$LCL = 1.5 - 3 * \sqrt{1.5} = -2.1742 \rightarrow 0$$

The following record of scratches on 10 assembled Luke Skywalker figures was taken just before some particularly suspicious, small furry creatures were spotted near the door of the plant.

Table 7. Number of scratches for $m = 10$ assembled Luke Skywalker figures

Unit	Number of scratches C
1	3
2	2
3	4
4	2
5	3
6	7
7	5
8	3
9	7
10	2

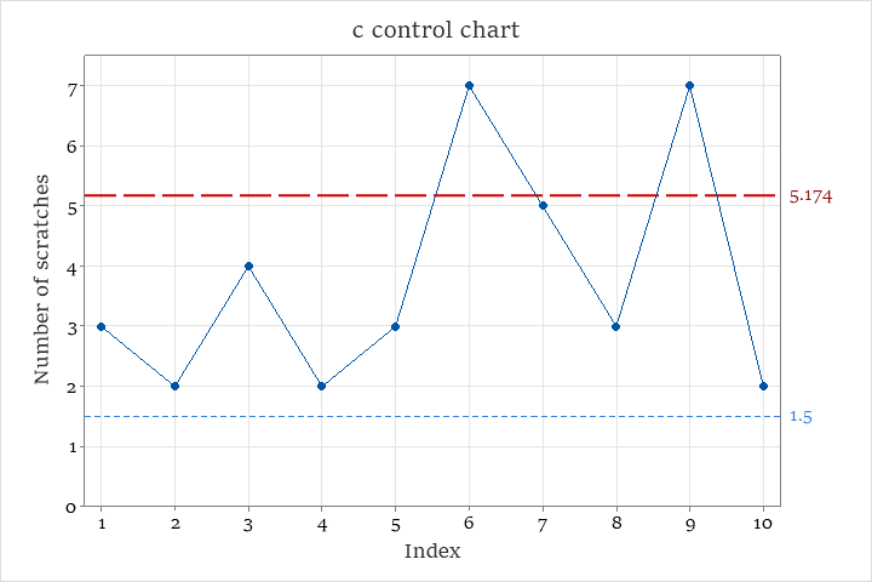

Figure 17. c control chart for attributes

Clearly, the process is out of control, as indicated by units 6 and 9, and the fact that all observations are above the center line. Something needs to be done about those little guys.

2.2.8. Use of Control Charts

Control charts can be used to identify process upsets and to maintain process stability within both non-manufacturing and manufacturing settings. In each case it is essential to have a well-defined action plan to respond quickly to any out-of-control indication. The action plan will be designed to fit the specific nature of the process that is being monitored.

3. INTRODUCTION TO ACCEPTANCE SAMPLING

3.1. Introduction

At various points in the production process, there may (and should) be an interest in determining the conformance of production materials and outputs. Often, product is organized into lots, which arrive to or leave the facility in a stream. Incoming or outgoing inspection of each lot in the stream can prevent subpar material from moving on to the next stage. However, it can be prohibitively costly, slow, or destructive to inspect every single element of product and decide whether it should be kept or discarded. Plans can be crafted to safeguard against low-quality product without 100% screening. Acceptance sampling is the subfield of SQC that defines these plans and their operational use.

Acceptance sampling does not directly contribute to quality estimation or inform specific changes to the manufacturing process (Duncan 1965). However, it still has indirect impacts on quality, so it is often placed within the SQC umbrella. This idea will again be illustrated with an example.

To gain an edge over competing toy companies, Kenner decides to start equipping its Luke Skywalker action figures with lightsabers that contain real Kyber crystals. They pilot the idea at Kenner Production Plant #66. Without their own harvesting methods, Kenner outsources their supply of Kyber to a company called Iluminate, which gets its name from the planet Ilum, where Kyber is most commonly found.

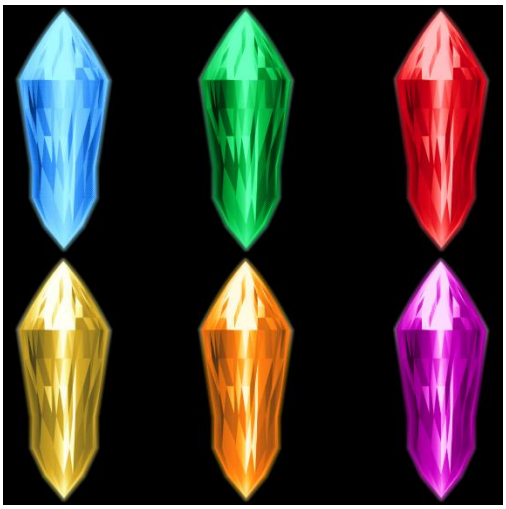

Figure 18. Various Kyber crystals, which give lightsabers their color and connection to the Force. Creative Commons photo.

After being harvested from the Crystal Cave on the planet Ilum, the Kyber crystals are cut and polished at Iluminate Crystal Refinery #1977. Iluminate, and their refinery, are the "producers" of the processed crystals. Iluminate ships finished batches, or "lots," of crystals to Kenner Production

Plant #66, who are the "consumers" of the crystals. Since Kenner doesn't yet know the quality of processed crystals from Iluminate, they have a strong interest in inspecting the incoming batches. Low quality would be dangerous – if a crystal is cracked, its connection to the Force becomes unstable. But the crystals are intricate – there's no way the personnel at Kenner Production Plant #66 could perform 100% inspection of every lot.

Kenner personnel decide, as a middle-ground solution, to inspect a sample of each incoming lot. If the number of cracked crystals in the sample exceeds some number, the entire lot will be rejected and sent back to Iluminate. If the number of cracked crystals stays below a specified limit, the lot is accepted and used in production. This decision – accept or reject – is called "lot sentencing" or "lot disposition," and it is the goal of acceptance sampling. Kenner can create acceptance sampling plans informed by statistics to meet their needs.

The National Institute of Standards and Technology emphasizes that the point of acceptance sampling is simply lot sentencing, and these plans should not be seen as a monitoring or control tool (Engineering Statistics Handbook). However, the indirect effects of acceptance sampling include psychological pressure on the producer to increase or maintain quality and motivation to "make it right the first time" (Duncan 1965).

Most acceptance sampling plans focus on attribute data, such as conforming versus nonconforming. We will discuss single-sampling plans, double-sampling plans, and sequential-sampling plans for attributes, along with continuous sampling plans, in this section. Acceptance sampling can also be performed for variables data; such methods are briefly described at the end of this section.

3.2. Single Sampling Plan

The design of single-sampling plans can be inferred from the name: From a lot of size N, we take one sample of size n and inspect that sample to make a decision about the entire lot. The methods outlined here are taken from Montgomery (2020), Schilling and Neubauer (2009), and Duncan (1965) unless otherwise indicated.

There are two parameters in a single sampling plan: the sample size n and the acceptance number c. The number of defects d in each lot is variable, and the lot sentencing is as follows:

$$d \leq c\text{: accept lot}$$

$$d > c\text{: reject lot}$$

For example, suppose Iluminate ships Kyber crystals to Kenner in lots of 500. Kenner decides to sample a fifth of each lot, so $n = 100$, and reject the lot if they find more than 2 cracked crystals in the sample, so $c = 2$.

The probability theory behind single-sampling plans is similar to the theory used to create p control charts for attributes. Let the lot size N be large, approaching infinity. Then the number of defects D follows a binomial distribution:

$$D \sim binom(n, p)$$

$$P(D = d) = f(d) = \frac{n!}{d!\,(n-d)!} p^d (1-p)^{n-d}$$

Where p is the fraction of defective units in the entire lot.

Under the previous plan, the acceptance probability, P_a, is the probability of finding only 0, 1, or 2 defects in the sample. If Iluminate Crystal Refinery #1977 consistently produces lots with fraction defective $p = 0.03$, then this comes out to:

$$P_a = P(D = 0) + P(D = 1) + P(D = 2)$$

$$= \sum_{d=0}^{2} \frac{100!}{d!\,(100-d)!} 0.03^d (0.97)^{100-d}$$

$$= 0.419775$$

To judge the performance of an existing sampling plan or use statistics to form a new one, we return to our old friend, the OC curve. In acceptance sampling, the OC curve plots the probability of accepting a lot against the fraction of defective units in the lot. Two types of OC curves are used in acceptance sampling; the most common is called Type B, which assumes we are working with a continuous stream of large lots. Duncan (1965) provides an excellent interpretation of the Type B OC curve:

> "The fractions defective of the lots will follow a binomial distribution… The probability of accepting a lot will be the proportion of lots from the given process that will in the long run be accepted under the plan."

We already have one point on our OC curve: (0.03, 0.419775). To find another point, suppose now that $p = 0.02$ instead. The calculations are the same as before:

$$P_a = \sum_{d=0}^{2} \frac{100!}{d!\,(100-d)!} 0.02^d (0.98)^{100-d}$$

$$= 0.6766856$$

Using the same process, we can find:

Table 8. Acceptance probabilities as p varies under current plan.

Fraction Defective p	Acceptance Probability P_a
0.01	0.920626
0.02	0.676685
0.03	0.419775
0.04	0.232142
0.05	0.118263
0.06	0.056612

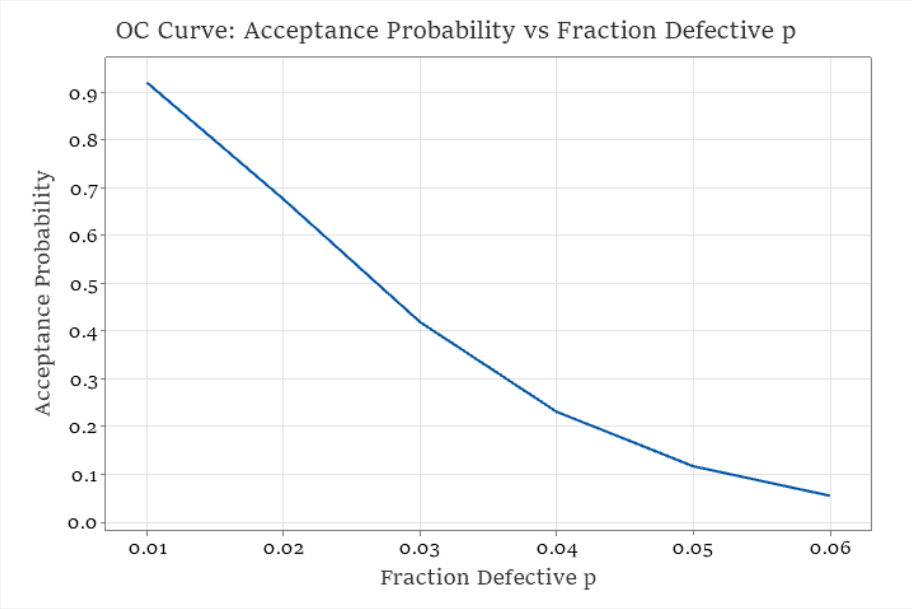

Figure 19. OC Curve for the sampling plan n = 100, c = 2

If an acceptance sampling plan is viewed as a test of the hypothesis that a lot is acceptable, then the OC curve is a visualization of the power of this test (or, more specifically, the complement of the power) at various levels of p.

A popular way to design acceptance sampling plans is to work backwards from a desired OC curve. The idea is to set two points on the OC curve (just one is not enough); in other words, we choose acceptance probabilities $1 - \alpha$ and β for fractions defective p_1 and p_2. For example, Kenner decides that it wants to accept lots with fraction defective $p_1 = 0.01$ with probability $1 - \alpha = 0.95$, and accept lots with higher fraction defective $p_2 = 0.04$ with lower probability $\beta = 0.1$.

This process is often connected to two terms in acceptance sampling. The first is acceptable quality level (AQL), which is considered the lowest quality level, on average, that is acceptable in the long run. Consumers hope that incoming lots have fraction defective at the AQL or better (lower). The second is rejectable quality level (RQL), which is considered the lowest quality level we'd ever want to accept in an individual lot. Consumers don't want to accept many lots with fraction defective at the RQL or worse (higher).

We generally want a high acceptance probability at the AQL and a low acceptance probability at the RQL. Therefore, AQL is a popular choice for p_1 and RQL is a popular choice for p_2. Technically, we are solving the following system of equations for n and c:

$$1 - \alpha = \sum_{d=0}^{c} \frac{n!}{d!\,(n-d)!} p_1^d (1-p_1)^{n-d}$$

$$\beta = \sum_{d=0}^{c} \frac{n!}{d!\,(n-d)!} p_2^d (1-p_2)^{n-d}$$

Without an exact solution, we can use a binomial nomograph, as in Montgomery (2020), or tabulated values, as in Duncan (1965) to estimate a solution.

There may be a few plans that fall near the intersection point. Analyze them separately and choose based on the needs of the company – in which direction is it okay to have some "give"? At SNL, it is often predetermined that our acceptance number $c = 0$. With one of our parameters fixed, we no longer need to specify two points on the OC curve; only one point is needed to find the optimal sample size. This point is typically the AQL.

Computer software like Minitab can plot OC curves for a few similar plans that go through the specified point. Usually, the plan with the smallest sample size is chosen due to financial and practical limits. But, if possible, we can increase power with a more discriminatory, larger sample size plan.

Under the "Stat" menu in Minitab, select "Quality Tools" and then "Acceptance Sampling by Attributes." If we know we need $c = 0$ defectives, we can find the OC curve for multiple sample sizes:

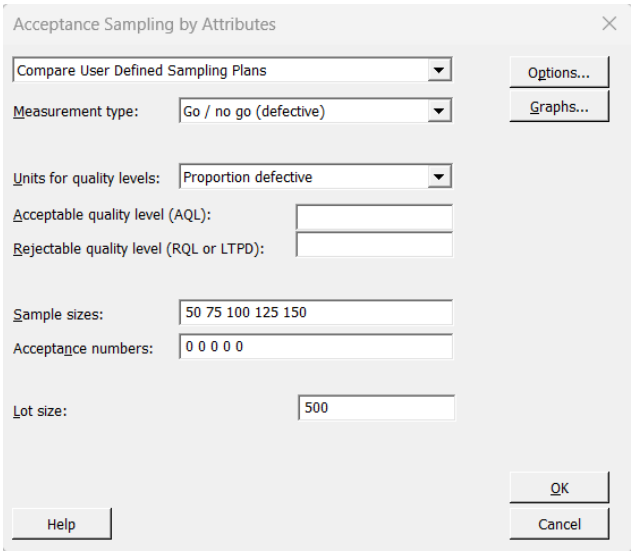

Figure 20. Minitab popup menu for comparing acceptance sampling plans.

Minitab produces multiple OC curves on the same plot for easy comparison. It also generates other helpful statistics to help narrow down a viable plan.

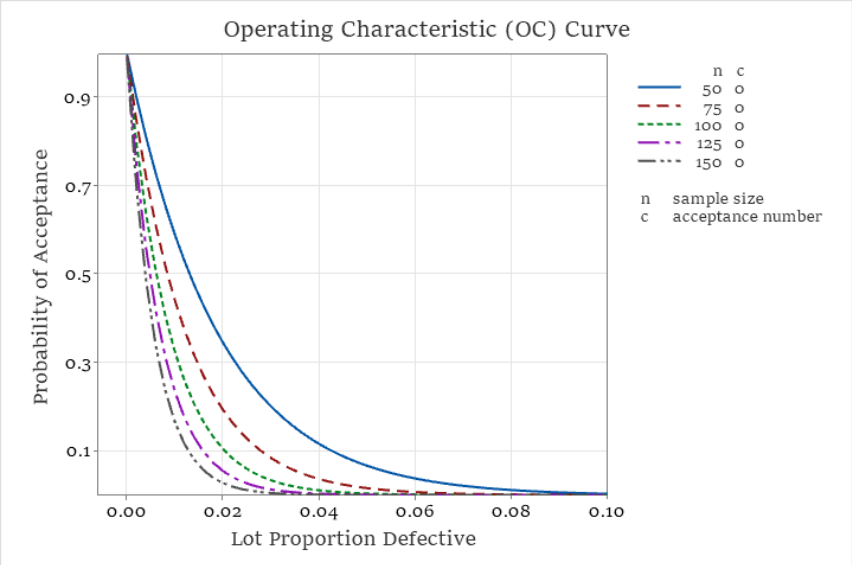

Figure 21. Minitab-generated OC curves for $c = 0$ acceptance sampling plans.

An alternative, but similar, method found in Schilling and Neubauer (2009) suggests finding the operating ratio:

$$R = \frac{p_2}{p_1}$$

And, based on the ratio, choosing tabulated values of other plan parameters based on the Poisson approximation to the binomial distribution.

Another way to evaluate acceptance sampling plans is the average outgoing quality (AOQ). Lots enter the plan with a quality level defined by fraction defective p. The AOQ is the quality level, on average, of lots that were inspected, accepted, and made it through to the next step in production. Notice that if each lot has the same fraction defective, the same proportion of defective units will enter the production process whether we perform acceptance sampling or not. Accepted lots have the same number of defective units as lots that were rejected; it's just that, by chance, our sample contained enough defectives to surpass the acceptance number c.

This seems to indicate that acceptance sampling is useless. However, acceptance sampling can increase AOQ when dealing with lots that have unstable proportions of defective units. And when paired with techniques like rectifying inspection, acceptance sampling will only increase (or maintain) AOQ.

Rectifying inspection is the process of removing or fixing all defects that are found. Rejected lots go through 100% screening and are returned to the consumer with 0 defects. Accepted lots do not go through screening, but any defects found in the sample of size n are fixed as well.

To calculate AOQ under these conditions, we find the "expected value" of defects. Lots are size N, the fraction defective is p, and samples of size n are inspected. With probability $1 - P_a$, a lot is rejected. Under rectifying inspection, a lot will end up with 0 defects with probability $1 - P_a$. With probability P_a, a lot is accepted. All defects found in the sample of size n are remedied, but since the lot was accepted without screening, $N - n$ units are accepted without inspection. A lot will end up with $p(N - n)$ defects with probability P_a.

$$\text{Expected number of defects} = (1 - P_a) \ * \ 0 + P_a * p(N - n) = P_a p(N - n)$$

To find the average outgoing quality level, we divide the expected number of defects in a lot by the number of items in a lot:

$$AOQ = \frac{P_a p(N - n)}{N}$$

Note that as N gets larger compared to n, $(N-n)/N$ approaches one. So, for large N, we have the approximation:

$$AOQ = P_a p$$

For a given plan, we can plot a curve of the AOQ over the lot fraction defective. This shows the relationship between incoming quality and outgoing quality.

To demonstrate, we will build on the example OC curve we created earlier in this section. Suppose Kenner's acceptance sampling plan for Kyber crystals now includes rectifying inspection. Recall that $N = 500,\ n = 100$ and $c = 2$. Suppose we say lot size N isn't quite large enough compared to n to warrant the approximation.

Table 9. AOQ as p varies under given plan.

Fraction Defective p	Acceptance Probability P_a	AOQ
0.01	0.9206268	0.0073650144
0.02	0.6766856	0.0108269696
0.03	0.4197751	0.0100746024
0.04	0.2321426	0.0074285632
0.05	0.118263	0.0047305200
0.06	0.05661278	0.0027174134

For $p = 0.01$, we have:

$$AOQ = \frac{0.9206 * 0.01 * (500-100)}{500} = 0.007365.$$

The rest of the values in the table are calculated with the same formula.

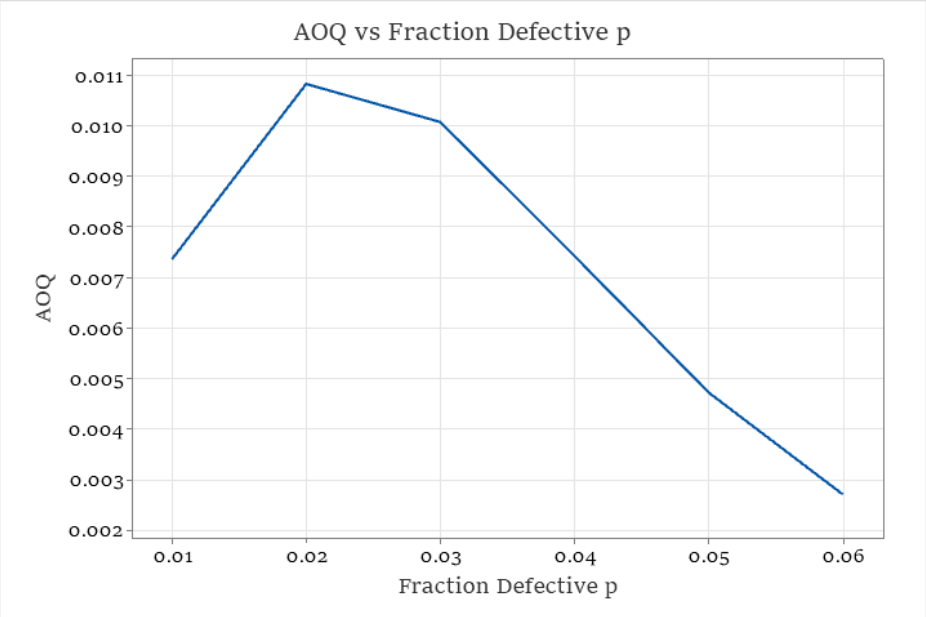

Figure 22. AOQ Curve for the sampling plan $n = 100, c = 2$

The maximum of the AOQ curve is known as the average outgoing quality limit (AOQL). Under a given plan, the fraction defective of lots that enter the production process, on average, will never be worse than the AOQL.

Another metric is the average total inspection (ATI), which is the expected number of units that will be inspected under a sampling plan with rectifying inspection. The formula for ATI is given by:

$$ATI = n + (1 - P_a)(N - n)$$

Since n items in each lot will be inspected regardless, and the remaining $N - n$ items will be inspected for rejected lots.

Under Kenner's plan, if $p = 0.01$, this leads to:

$$ATI = 100 + (1 - 0.9206) * (500 - 100) = 131.75$$

Suppose the lot size N is not sufficiently large to warrant use of the binomial distribution. In this case, the acceptance sampling plan is instead anchored in the hypergeometric distribution. The number of defects D in a lot of finite size N is a random variable, and

$$D \sim Hypergeom(N, n, p),$$

$$P(D = d) = f(d) = \frac{\binom{pN}{d}\binom{N-pN}{n-d}}{\binom{N}{n}}$$

Note that the hypergeometric distribution is usually parameterized with N, n, and K, the latter representing total number of "successes" in the population N. In our case, a "success" is a defective unit. However, for consistency, we include the fraction defective as a parameter, so the number of defectives $K = pN$. Acceptance probabilities for lots are found in the same manner as before, but now we use the hypergeometric probability mass function.

The second type of OC curves – called the Type A OC curve – are constructed according to this finite-size situation. Duncan (1965) again provides an excellent interpretation:

> "A Type A OC curve gives the probability of accepting an isolated lot. This may be interpreted as the proportion of lots that would be accepted in an infinite series of lots all identical to the one in question."

In other words, the fraction defective in each lot is no longer variable – it's the same for each lot. Type A curves lie "below" Type B curves on a plot, but as the lot size N gets larger, the Type A curve approaches the Type B curve. The Type B curve can be thought of as the Type A curve with $N = \infty$.

3.3. Single Sampling Plans, Schemes, and Systems

Before we go any further, we should cover some more terminology. We have just described a type of sampling plan, which is the most basic sampling structure. Sampling plans specify the sample size, acceptance number, and other parameters we need to sentence a single lot.

Next on the organizational pyramid are acceptance sampling schemes. Sampling schemes are groups of various sampling plans with "switching rules" to determine which plan to use under different circumstances. Usually, sampling schemes include plans of varying strictness. If quality seems to be increasing, we might switch to a less strict plan with lower inspection; if quality problems are detected, we might switch to a stricter plan with high or 100% inspection until the conditions improve.

Acceptance sampling systems take those same ideas one step further. Systems are sets of schemes to be used at different times or for different purposes, along with criteria for selecting them.

We have specifically noted that acceptance sampling is not a direct process control method. However, schemes and systems can have more impact on quality, if designed correctly. Schilling (2009) used the analogy:

> "An individual sampling plan has much effect of a long sniper, while the sampling scheme can provide a fusillade in the battle for quality improvement."

In the industry, some sampling systems have stood the test of time and are now considered a standard. For example, a military standard acceptance sampling system for attribute data called MIL-

STD 105A was designed during World War II, and its current version, MIL-STD 105E, continues to be widely used today.

Different systems might emphasize different criteria. The AQL is central to MIL-STD 105E, as it indexes plans based on a specified AQL for different lot sizes and reduced, normal, and tightened inspection.

More details about operating MIL STD 105E can be found in Montgomery (2020). Another standard for attributes sampling, considered the "civilian counterpart" of the military standard, is ANSI/ASQC Z1.4.

3.4. Double and Multiple Sampling Plans

Double sampling plans, as the name again implies, involve taking two separate samples from the lot N and using results from both samples to sentence the lot.

We now have four parameters: n_1, the size of the first sample; c_1, the acceptance number for the first sample; n_2, the size of the second sample, and c_2, the cumulative acceptance number for both samples.

The procedure is as follows.

1. Take the first sample of size n_1 from a lot of size N and count the number of defects d_1.
    a. If $d_1 \leq c_1$, immediately accept the lot.
    b. If $d_1 > c_2$, immediately reject the lot.
    c. If $c_1 < d_1 \leq c_2$, move to step 2.
2. Take the second sample of size n_2 from the same lot and count the number of defects d_2.
    a. If $d_1 + d_2 \leq c_2$, accept the lot.
    b. If $d_1 + d_2 > c_2$, reject the lot.

Basically, double-sampling plans establish a middle ground where we say more information is needed before we accept or reject the lot. Note that double sampling plans, when it comes to statistics, are not inherently better or worse than single sampling plans. For each single sampling plan, we can create a double sampling plan that has an identical OC curve.

The major benefit of double sampling plans is the psychological effect on the producer: we've given each lot a "second chance" at acceptance. There's also the possibility of decreasing ATI if we often sentence the lot before taking the second sample and n_1 is lower than the sample size may have been in a single sampling plan. We can also decrease ATI if we implement curtailment, which means we sentence the lot once $d_1 + d_2$ exceeds c_2 even if not all units in the second sample have been inspected.

The OC curve for double sampling is just like the OC curve for single sampling, which tracks overall acceptance probability against fraction defective. Two related curves are sometimes plotted alongside the main OC curve: the first-sample acceptance probability and the first-sample rejection probability.

Iluminate Crystal Refinery #1977 is starting to feel a bit agitated with Kenner Production Plant #66; they think lot inspectors are being too harsh on their product. Not wanting to lose their only source of Kyber – and realizing that curtailment could help reduce inspection costs – Kenner decides to switch to a double sampling plan.

In this plan, Kenner chooses $n_1 = 50$, $c_1 = 1$, $n_2 = 150$, and $c_2 = 4$. To create the OC curve, we find the acceptance probability for different fractions defective. Since two samples are possible, we need to sum the probability of acceptance on the first sample with the probability of acceptance on the second sample.

If $p = 0.03$, the probability of acceptance on the first sample is found, as before, with the binomial distribution. Recall that the number of defective units in a sample follows a binomial distribution. We accept at this point only if the number of defects is less than or equal to 1:

$$D_1 \sim binom(n = 50, p = 0.03)$$

$$P_{a_1} = P(D_1 = 0) + P(D_1 = 1)$$

$$= \sum_{d=0}^{1} \frac{50!}{d!\,(50-d)!}(0.03)^d(0.97)^{50-d}$$

$$= 0.55528$$

The number of defective units in the second sample also follows a binomial distribution:

$$D_2 \sim binom(n = 150, p = 0.03)$$

We take a second sample only if the number of defective units in the first was greater than 1 but less than or equal to 4. Thus, the probability of acceptance on the second sample should be broken up into three cases.

1. If $d_1 = 2$, we take the second sample and accept if $d_2 = 0$, 1, or 2.

$$P_{a_{2_1}} = P(D_1 = 2) * P(D_2 \leq 2)$$

$$= \left(\frac{50!}{2!\,(50-2)!}(0.03)^2(0.97)^{48}\right) * \left(\sum_{d=0}^{2}\frac{150!}{d!\,(150-d)!}(0.03)^d(0.97)^{150-d}\right)$$

$$= 0.04326$$

2. If $d_1 = 3$, we take the second sample and accept if $d_2 = 0$ or 1.

$$P_{a_{2_2}} = P(D_1 = 3) * P(D_2 \leq 1)$$

$$= \left(\frac{50!}{3!\,(50-3)!}(0.03)^3(0.97)^{47}\right) * \left(\sum_{d=0}^{1}\frac{150!}{d!\,(150-d)!}(0.03)^d(0.97)^{150-d}\right)$$

$$= 0.00739$$

3. If $d_1 = 4$, we take the second sample and accept if $d_2 = 0$.

$$P_{a_{2_3}} = P(D_1 = 4) * P(D_2 = 0)$$

$$= \left(\frac{50!}{4!\,(50-4)!}(0.03)^4(0.97)^{46}\right) * \left(\frac{150!}{0!\,(150-0)!}(0.03)^0(0.97)^{150}\right)$$

$$= 0.0004765$$

All of these probabilities are added together to find the total acceptance probability:

$$P_a = P_{a_1} + \left(P_{a_{2_1}} + P_{a_{2_2}} + P_{a_{2_3}}\right)$$

$$= 0.55528 + 0.04326 + 0.00739 + 0.0004765$$

$$= 0.606414$$

The same process can be followed for different fractions defective. We obtain the following points on the OC curve, which is plotted below.

Table 10. Acceptance probabilities for the double sampling plan.

Fraction Defective p	Acceptance Probability P_a
0.01	0.978904
0.02	0.826582
0.03	0.606414
0.04	0.419736
0.05	0.285125
0.06	0.191429

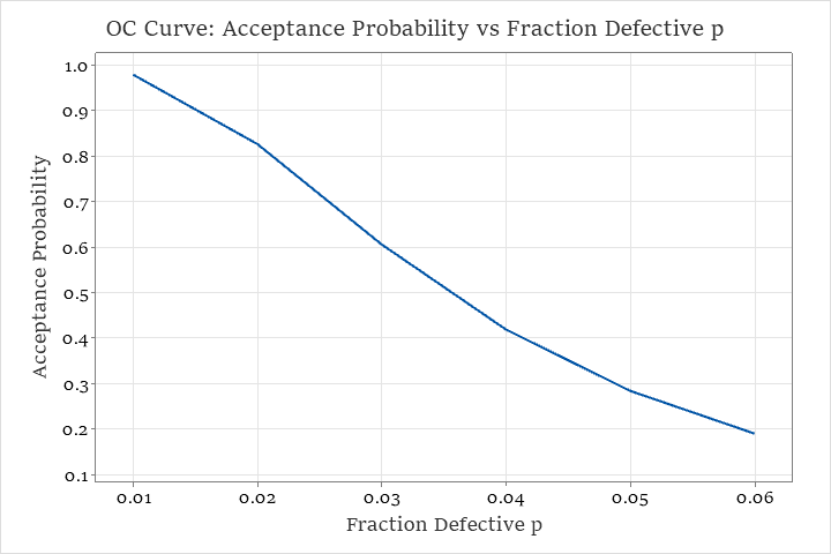

Figure 23. OC curve for the double sampling plan

To design a double sampling plan, we choose two specified points on the OC curve, just as before: $(p_1, 1-\alpha)$ and (p_2, β). However, because we now have more total parameters, we must also define a relationship between the sample sizes: $n_2 = m * n_1$, with chosen m.

For consumers, one of the key reasons to adopt a double sampling plan is the possibility of inspecting less units overall, especially if n_1 is smaller than the sample size would be in a single sampling plan. This is judged with the average sample number (ASN). Without curtailment, we have:

$$ASN = n_1 P_1 + (n_1 + n_1)(1 - P_1)$$

Where p varies, and results at different values of p can be used to draw out the average sample number curve.

With curtailment, the formula is much more complicated:

$$ASN = n_1 + \sum_{j=c_1+1}^{c_2} P(n_1, j)\left(n_2 P_L(n_2, c_2 - j) + \frac{c_2 - j + 1}{p} P_M(n_2 + 1, c_2 - j + 2)\right)$$

Where:

- $P(n_1, j)$ is the probability of finding j defective units in the first sample (with j lying between c_1 and c_2)
- $P_L(n_2,\ c_2 - j)$ is the probability of finding $c_2 - j$ or fewer defects in the second sample (which means the entire second sample is inspected)
- $P_M(n_2 + 1, c_2 - j + 2)$ is the probability of finding $c_2 - j + 2$ defects in a sample of size $n_2 + 1$.

Curtailment is a great strategy for the consumer and decreases ASN.

We can also define the evaluation metrics previously discussed for double sampling plans. Assuming all defective units are replaced with non-defective units, the average outgoing quality of a double sampling plan with rectifying inspection is:

$$\text{AOQ} = \frac{p\left(P_{a_1}(N - n_1) + P_{a_2}(N - n_1 - n_2)\right)}{N}$$

And the average total inspection is:

$$ATI = n_1 P_{a_1} + (n_1 + n_2) P_{a_2} + N(1 - P_a)$$

These same concepts can be extended further with multiple sampling plans, which operate similarly to double sampling plans, but with more stages and parameters.

3.5. Sequential Sampling Plans

In sequential sampling plans, units are drawn one at a time and evaluated. The cumulative number of defective units is recorded. This can be viewed as multiple sampling plans with no specified upper limit on the number of samples, and each sample size is 1.

These plans are based on Wald's sequential probability ratio test (SPRT), which is a type of hypothesis test that aims to accept or reject the null hypothesis "as soon as possible" (Nowak 2011). Indeed, sequential sampling plans have optimal efficiency and should be used when it's necessary or beneficial to keep the sample size small.

As each unit is inspected, there are three possible outcomes:

1. Accept the lot and stop inspection. This happens when the number of cumulative defects crosses the lower "acceptance bound."
2. Reject the lot and stop inspection. This happens when the number of cumulative defects crosses the upper "rejection bound."
3. Sample another unit. This happens when the cumulative number of defects lies between the bounds; there is not enough information to sentence the lot.

Duncan (1965) provides tabulated values for the upper and lower bounds as the sample grows. However, it is more intuitive to create a plot of cumulative defective units versus sample size. The acceptance and rejection bounds are drawn as sloped lines on the same plot.

The acceptance and rejection bounds take the form:

$$Y_A = -h_1 + sn$$
$$Y_R = h_2 + sn$$

Where n is the current sample size. For chosen values of p_1, p_2, $1 - \alpha$, and β, the parameters of the lines are found with formulas:

$$k = \log\left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)$$
$$h_1 = \frac{\log\left(\frac{1-\alpha}{\beta}\right)}{k}$$
$$h_2 = \frac{\log\left(\frac{1-\beta}{\alpha}\right)}{k}$$
$$s = \frac{\log\left(\frac{1-p_1}{1-p_2}\right)}{k}$$

So, these lines are parallel, but with different intercepts. The acceptance bound, naturally, lies below the rejection bound.

Let us refer to the previous example. Recall that Kenner aims to accept lots with fraction defective $p_1 = 0.01$ with probability $1 - \alpha = 0.95$, and accept lots with higher fraction defective $p_2 = 0.04$ with lower probability $\beta = 0.1$.

This would lead to a sequential sampling plan with parameters:

$$k = \log\left(\frac{0.04(1-0.01)}{0.01(1-0.04)}\right) = 0.61542$$

$$h_1 = \frac{\log\left(\frac{0.95}{0.1}\right)}{0.615} = 1.5887$$

$$h_2 = \frac{\log\left(\frac{0.9}{0.05}\right)}{0.615} = 2.03969$$

$$s = \frac{\log\left(\frac{1-0.01}{1-0.04}\right)}{0.615} = 0.021715$$

And acceptance / rejection bounds defined by the lines:

$$Y_A = -1.5887 + 0.021715n$$
$$Y_R = 2.03969 + 0.021715n$$

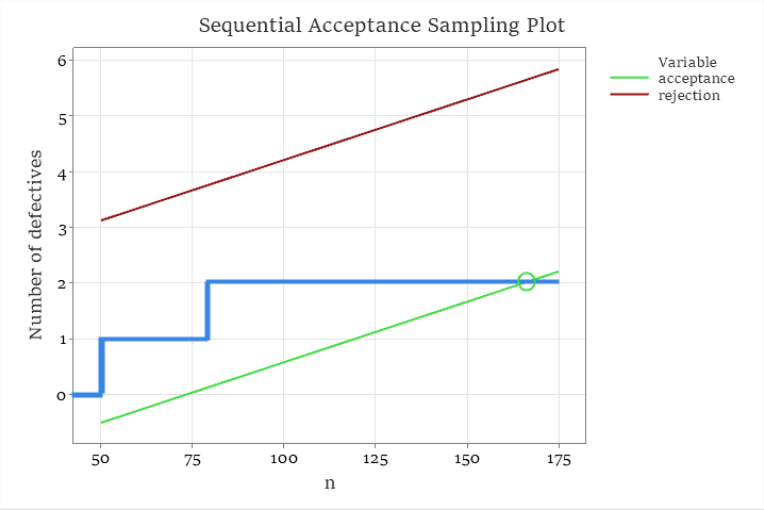

Figure 24. Example plot for a sequential acceptance sampling plan

On this plot, if we sample around 75 units with no defects, we can accept, because the trace of defective units crosses the lower acceptance boundary. But, for example, if we sample 75 units and find at least 4 defects, we reject the lot. The data displayed above shows acceptance with only 2 defects in around 170 units sampled, indicated by the green circle.

Evaluation metrics for sequential sampling plans are somewhat more complicated. For chosen values of $p_1, p_2, 1-\alpha$, and β, points on the OC curve can be found with the following formulas:

$$p = \frac{1 - \left(\frac{1-p_2}{1-p_1}\right)^h}{\left(\frac{p_2}{p_1}\right)^h - \left(\frac{1-p_2}{1-p_1}\right)^h}$$

$$P_a = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}$$

Where h is arbitrary and $-\infty \leq h \leq \infty,\ h \neq 0$. Several different values of h should be chosen to produce many ordered pairs of (p, P_a) and draw out the OC curve (Wald 1947).

From these points, we can find the ASN as the fraction defective varies:

$$ASN = P_a\left(\frac{\log\left(\frac{\beta}{1-\alpha}\right)}{p\log\left(\frac{p_2}{p_1}\right) + (1-p)\log\left(\frac{1-p_2}{1-p_1}\right)}\right) + (1-P_a)\left(\frac{log\left(\frac{1-\beta}{\alpha}\right)}{p\log\left(\frac{p_2}{p_1}\right) + (1-p)\log\left(\frac{1-p_2}{1-p_1}\right)}\right)$$

We can also calculate the ATI:

$$ATI = P_a\left(\frac{log\left(\frac{\beta}{1-\alpha}\right)}{p\log\left(\frac{p_2}{p_1}\right) + (1-p)\log\left(\frac{1-p_2}{1-p_1}\right)}\right) + (1-P_a)N$$

The AOQ is the simplest of all, and can be approximated by:

$$AOQ = P_a p.$$

3.6. Continuous Sampling Plans

If production units cannot easily be organized into lots, some of the plans previously described may not be easy to implement or provide the desired results. For continuous production processes – think conveyor belts – quality engineers and statisticians have developed continuous sampling plans. The most popular continuous sampling plan is known as Dodge's CSP-1, created by Harold Dodge. The steps are as follows:

0. Define a clearance number, i.
1. Begin by inspecting 100% of all units, called "screening." Determine whether each is defective or non-defective. Record the number of consecutive non-defective units, c.
    a. If we observe a defect and $c < i$, reset: $c = 0$.
    b. If $c = i$, move to step 2.
2. End 100% inspection and sample only a randomly selected fraction f of incoming units. Still determine whether each unit is defective or non-defective.
    a. If no defects are found, continue.
    b. If a single defect is found, return to step 1 and resume 100% inspection.

The values of f and i are chosen based on engineering and economic knowledge.

Paired with rectifying inspection, as it often is, CSP-1 often increases the quality of the product; the AOQL is tabulated in Montgomery (2020) for varying f and i.

Some important evaluation metrics for CSP-1 are based on inspection numbers. We may be interested in the average number of units that are inspected during screening before we reach $c = i$ and can switch to sampling:

$$n_1 = \frac{1 - (1-p)^i}{p(1-p)^i}$$

Where p is the in-control fraction defective in the process.

Similarly, we can find the average number of units that are inspected during sampling before we find a defective and must switch back to screening:

$$n_2 = \frac{1}{fp}$$

From there, we can find the average fraction of units inspected in the long run, during both the sampling and screening phases:

$$AFI = \frac{n_1 + fn_2}{n_1 + n_2}$$

And the average fraction of units accepted, which can be plotted as a function of p to obtain an OC curve for the CSP-1 plan:

$$P_a = \frac{n_1}{n_1 + n_2}.$$

3.7. Variables Acceptance Sampling

Acceptance sampling based on measurements of variable quality characteristics is called variables acceptance sampling. It is less common than attributes acceptance sampling but can sometimes allow smaller sample sizes and more insight into the production process. Two variables sampling techniques will be briefly described in this section.

These methods assume that the distribution of the quality characteristic is known; more specifically, it assumes the distribution is normal. If the quality characteristic has a known, non-normal distribution, the plans should be altered to reflect the correct probability theory. Our quality characteristic is some measurement, such as height. For example, each unit has a height, and the heights are normally distributed with process mean μ and process standard deviation σ.

A unit is considered defective if its measurement exceeds the upper specification limit (USL) or falls short of the lower specification limit (LSL). These limits are often set externally, and we may receive just one limit, or both. Just as before, we can specify points on the OC curve to choose a plan. Nomographs for variables sampling usually provide two options for the sample size: one for when σ is known, and another for when σ is unknown.

These test types are described as they appear in Montgomery (2020).

3.7.1. Test Type 1

First, we take a sample of size n from the lot and calculate the sample average and sample standard deviation. The quality characteristic of the units in this lot is assumed to be normally distributed with mean $\bar{x}$ and standard deviation s.

We then calculate the z-score of the specification limit, whichever is given:

$$Z_{LSL} = \frac{\bar{x} - LSL}{s}$$

$$Z_{USL} = \frac{USL - \bar{x}}{s}$$

These z-scores are the distance between the specification limit and the lot average, in standard deviation units. To define the rejection region, we choose some point k, called a critical distance, on the standard deviation scale. In other words, if Z_{LSL} or Z_{USL} is at least k standard deviations away from $\bar{x}$, then the fraction defective in the lot is small enough to warrant acceptance. If Z_{LSL} or Z_{USL}

is not k standard deviations away from $\bar{x}$, then the fraction defective in the lot is probably too high, and the lot should be rejected.

This procedure can be used only if we have one specification limit, not both.

3.7.2. Test Type 2

Again, we take a sample of size n from the lot and calculate the sample average and sample standard deviation. The z-scores of the USL are calculated with the same formulas provided above.

We can estimate the fraction defective in the lot based on the area under the standard normal curve that lies beyond Z_{LSL} and / or Z_{USL}. We call this $\hat{p}$, and if $\hat{p}$ is greater than a prespecified limit, p_L, we reject the lot; if it is less than p_L, we accept the lot.

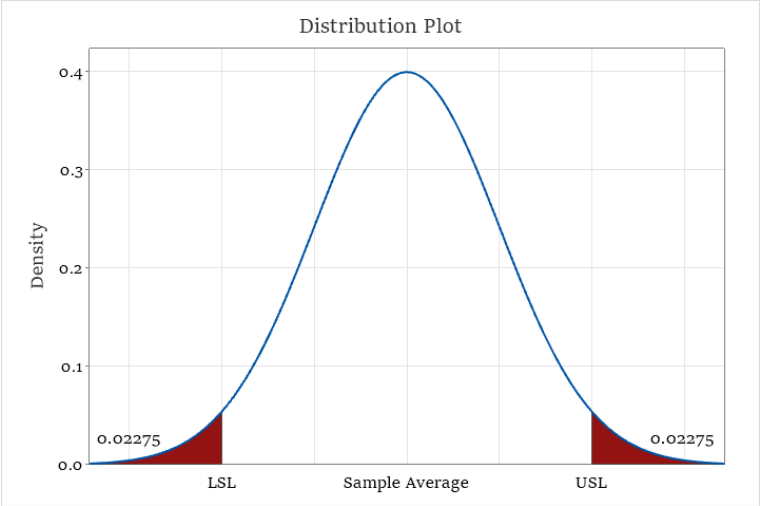

Figure 25. An example plot used to assist with a Type 2 variables acceptance sampling plan.

For example, in the sample distribution above, the probability of a unit lying outside the specification limits $\hat{p} = 0.0455$. If the limit $p_L = 0.05$, we can accept this lot. This plan can be used when just one of the specification limits are specified, or when both are specified.

3.7.3. Variables Military Standard

The military standard system for variables acceptance sampling is MIL STD 414. Just as the military standard for attributes sampling, MIL STD 414 is focused on the AQL of the acceptance sampling procedure. The civilian analog is ANSI/ASQC Z1.9.

3.8. Conclusion

Acceptance sampling is a compromise between blindly accepting all units, and painstakingly inspecting all of them. It's by no means a foolproof path to high quality. But it can be an appropriate safeguard or double-check on incoming materials or outgoing product, especially when used along with other manufacturing control methods.

4. INTRODUCTION TO HIGHLY ACCELERATED STRESS SCREENING (HASS) AND ENVIRONMENTAL STRESS SCREENING (ESS)

Product acceptance testing targets both patent and latent defects. Patent defects are clearly visible or otherwise detectable before the product is used. In complicated assemblies, especially those with electronic components, there may exist built-in defects that pass through simple inspection unnoticed. Instead, routine use or environmental conditions will cause these defects to manifest as failures in the field later on. These are known as latent defects (Kececioglu and Sun 1995). Theoretically, latent defects will reveal themselves early in the product's life (Department of Defense 1993). When a latent defect causes a failure or otherwise becomes apparent, we say the latent defect has been precipitated.

Environmental stress screening (ESS) is the component of manufacturing controls that attempts to precipitate as many latent defects as possible before units enter the stockpile. In ESS, a product is put through various environmental stress conditions, usually in the form of vibration and temperature cycles. The parameters of these environmental cycles should reflect the conditions it will experience in the field or during use.

The goal of ESS is to sort units into two groups: those with and those without latent defects. In a way, ESS can be viewed as 100% sequential acceptance sampling, but instead of merely inspecting each unit, we design specific screens to reveal functional defects. If a unit fails during a screen, it is either reworked, if possible, or discarded.

Highly accelerated stress screening (HASS) is a special case of ESS. In HASS, a product will go through cycles of environmental conditions more extreme than it would typically endure in the field, but for shorter amounts of time. HASS pushes the limits to precipitate certain defects that would not appear in a standard ESS procedure. As a result, HASS has a more direct goal of reliability improvement. It's more than just a screen; it's a search for systemic manufacturing issues.

While stressors should be more extreme in HASS than what the product would experience normally, we also must be careful not to cross the operational limits of the product. The operational limits are the stress levels that a defect-free unit can endure before it fails (Bahret 2016). If we go too far, we're not precipitating latent defects, we're just destroying perfectly good units. The concept of avoiding significant life reductions is often called "safety of screen" (Peterson 2004).

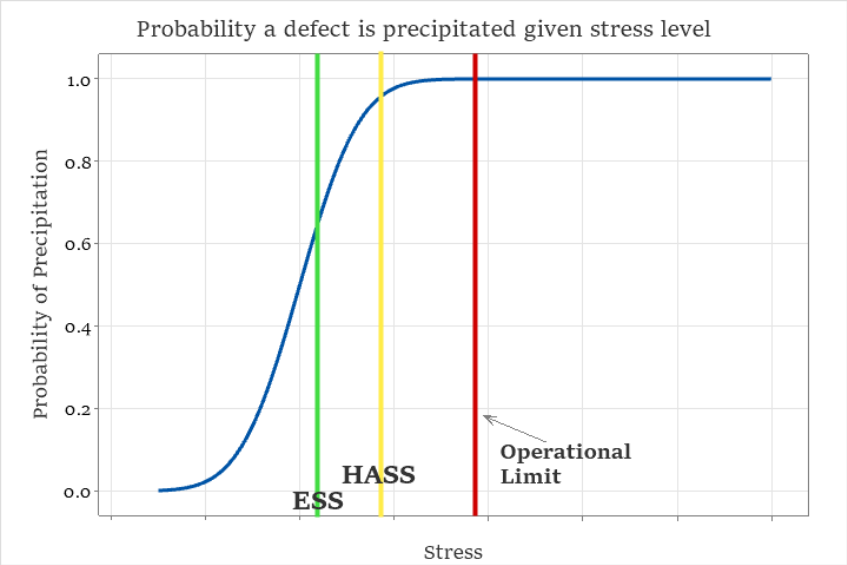

Figure 26. A visualization of the difference between ESS and HASS. The population is units that have latent defects. ESS, HASS, and Operational Limits are increasing levels of stress.

The objective of ESS/HASS is to quickly identify products that are predisposed to fail due to latent defects; done correctly, this can lead to a failure-free inventory (Kececioglu and Sun 1995). Producers will also see significant improvements in the products' mean time between failure (MTBF) and costs associated with the remediation of field failures.

Kececioglu and Sun (1995) emphasize that HASS/ESS is not to be confused with product testing. HASS/ESS is a process that aims to precipitate all latent defects and remove them from the product base. They also note that HASS/ESS is key to actualizing the reliability goals of electronic products and can promote quality improvement if precipitated latent defects can be traced to a root cause. Some sources do not mark a clear distinction between ESS and HASS. The definitions here are based on a rough consensus of the most recent sources and what is most applicable to SNL mission objectives.

The remainder of this section summarizes the HASS procedures and diagnostics as laid out in MIL-HDBK-344A (1993), a military standard for HASS of electronic equipment, and two documents produced by Crowder (2013, 2017) for SNL. For more complete descriptions, design details, and certain tabulated values, refer to the military standard text and the other listed references. Note that MIL-HDBK-344A uses the term ESS, but since it describes screens with increased stress levels, it is effectively describing HASS procedures.

4.1. General HASS Framework

The creation of a HASS program starts out with a heavy reliance on military standards and historical industry data. However, as more knowledge is accumulated, the HASS plan should be optimized for that specific product. Ideally, the plan will start big but narrow down over time to find the minimum cost screens that allow producers to meet reliability requirements.
The general process of creating HASS plans includes:

- identification of quantitative objectives, such as reliability requirements
- placement of screens at different points in the production line, also called levels of assembly
- evaluation of costs (monetary, etc.)
- determination of evaluation criteria for the screen (How do we know it's working?)
- development of guidelines for altering the plan if needed

In modern HASS procedures, screen placement early in production is preferred. The cost of fixing a precipitated defect only increases as the product continues through the phases of the production process; screens are cheapest at the component level (Safety and Reliability Society 2012).

There are several environmental parameters that must be defined for each screen, including maximum and minimum temperature of the cycles, the rate of change of the temperature, the spectrum shape of vibration, and the G root mean square (G-RMS), or energy level, of the vibration. Historical estimates show that around 20% of latent defects will precipitate due to vibration and around 80% will precipitate due to temperature cycling (Department of Defense 1993).

Besides the environmental parameters listed above, there are several other quantitative values that will be estimated at different stages to help design and evaluate the screens. The most important of these are D_{IN}, the average number of defects per unit that enter a screen, the screening efficiency or screening strength (SS), which represents the ability of the screen to precipitate and detect defects, and $D_{REMAINING}$, the average number of defects per unit that escape the screen undetected.

There are three phases to a HASS plan: the planning phase, the development phase, and the production phase.

4.2. Planning Phase

The planning phase begins with the proposition of multiple viable HASS procedures. Each plan consists of screens with different temperature and vibration parameters and different placements in the assembly process. The plans are then evaluated and compared based on cost, screening strength, $D_{REMAINING}$, and more, until the final plan is selected.

4.2.1. Reliability Goals

HASS design is driven by the demonstration of reliability goals. Usually, reliability requirements are specified in terms of the number or proportion of defects that escape into the field following the screen, $D_{REMAINING}$.

An expression for $D_{REMAINING}$ is:

$$D_{REMAINING} = D_{IN}\left(\frac{1 - SS}{1 - SS * D_{IN}}\right)$$

For a desired value of $D_{REMAINING}$, we can solve for the necessary screening strength:

$$SS = \frac{D_{IN} - D_{REMAINING}}{D_{IN} - D_{IN} * D_{REMAINING}}$$

Which can be approximated to:

$$SS \approx \frac{D_{IN} - D_{REMAINING}}{D_{IN}}$$

Crowder (2017) provides a table of values resulting from this relationship:

Table 11. Relationship between incoming defects, screening strength, and outgoing defects.

D_{IN}	SS	$D_{REMAINING}$
0.01	25%	0.008
0.01	50%	0.005
0.01	75%	0.003
0.01	80%	0.002
0.01	90%	0.001
0.02	25%	0.015
0.02	50%	0.010
0.02	75%	0.005
0.02	80%	0.004
0.02	90%	0.002
0.05	25%	0.038

D_{IN}	SS	$D_{REMAINING}$
0.05	50%	0.026
0.05	75%	0.013
0.05	80%	0.010
0.05	90%	0.005
0.10	25%	0.077
0.10	50%	0.053
0.10	75%	0.027
0.10	80%	0.022
0.10	90%	0.011
0.15	25%	0.117
0.15	50%	0.081
0.15	75%	0.042
0.15	80%	0.034
0.15	90%	0.017
0.15	95%	0.009

In short, for a given D_{IN}, the screening strength of the HASS must be tuned to achieve a goal value of $D_{REMAINING}$.

Obtaining an accurate initial estimate of D_{IN} for a specific product can prove challenging. MIL-STD-344A suggests using historical data, if it exists, or a military standard value.

Managers at Kenner Production Plant #66 realize that the new lightsabers designed for their Luke Skywalker action figures are rather complicated and include some electronic elements. They're concerned about latent defects, so they decide to implement a HASS plan.

After a short experiment, it can be estimated that around 5% of fully assembled lightsabers contain defects, either latent or patent. The managers want the outgoing defect rate reduced to no more than 1%. In other words,

$$D_{IN} \approx 0.05$$

$$D_{REMAINING} = 0.01$$

The screening strength that will allow Kenner personnel to achieve their reliability goals is:

$$SS = \frac{0.05 - 0.01}{0.05 - 0.05 * 0.01} = 0.8081$$

4.2.2. Safety of Screen

Once a HASS procedure is planned and implemented, its safety should be assessed. To do this, several units are put through the HASS screen multiple times in a row. Popular choices are 20 and 25 full screens. The survival or failure pattern of the product is used to conjecture how much the lifetime of a unit is reduced after enduring a single screen.

Let n be the total number of screens the unit endures for a safety of screen demonstration. If the unit fails after $x \leq n$ screens, then it is estimated that a single screen reduces the life of a unit by:

$$100\left(\frac{1}{x}\right)\%$$

If the unit does not fail throughout all n screens, we can say that a single screen reduces the life of a unit by no more than:

$$100\left(\frac{1}{n}\right)\%$$

For example, suppose personnel decide to start using the previously mentioned screen with around 80% strength. To demonstrate the screen's safety, they run a few lightsabers through 25 screens. No failures are observed. Therefore, under the current HASS plan, the lifetime of a given lightsaber is reduced by no more than:

$$100\left(\frac{1}{25}\right)\% = 4\%$$

Appropriate sample sizes for this demonstration can be determined based on reliability requirements and desired confidence levels.

4.3. Development Phase

In the development phase, the preliminary HASS procedure based on initial parameter estimates is, in a sense, on trial. Data of the product's response to the screens are collected, and the data are used to adapt the HASS plan to the specific product's needs.
Common data fields include:

- Test equipment used
- Duration of stress

- Test type
- Number of units tested
- Number of units that failed
- Assembly, sub-assembly, or part that failed
- Cause of failure
- Details of the screen
- Environment history

This information should be used to obtain better parametric estimates of D_{IN} and $D_{REMAINING}$. If $D_{REMAINING}$ is off target, the screening strength should be adjusted.

We might also change the placement of the screens at different levels in the assembly process. Increasing the screening strength may entail:

- Setting more extreme minimum and maximum temperatures
- Faster rate of change of temperature
- Higher G-RMS during vibration
- Longer overall screen duration

These changes can be made based on engineering knowledge and the reaction of the product to previous screens.

4.4. Production Phase

Finally, in the production phase, we have more complete data on the product's response to the screens, allowing us to perform extensive analysis.

The concept of process control comes back into play during the production phase. Fallout data from the screens should be monitored to ensure that reliability requirements are consistently being met. Process and quality improvements can be made in this phase if the root causes of fallout are properly diagnosed – or, in SQC terminology, the assignable cause.

The number of defects in an item X follows a Poisson distribution:

$$X \sim pois(D)$$

$$P(X = x) = \frac{e^{-D} D^x}{x!}$$

where D is the defect density. The value of D is calculated form fallout data and used to monitor the defect level for a lot or unit. Recall that, for Poisson:

$$E(X) = D,$$

$$Var(X) = D.$$

Refer to the SQC section for an example of a control chart based on count data.

The number of defects X is then serially plotted, with 3-sigma control limits, to create a control chart for the number of defects found in each production unit. If these charts and other monitoring methods demonstrate sustained decreases in defect density, the ESS plan can be gradually relaxed. Eventually, HASS will be replaced with Product Reliability Verification Tests (PVRT). This is a much lighter version of HASS that aims to simply confirm that incoming units meet reliability requirements; the more active goal of precipitating and fixing defects takes a back seat.

5. INTRODUCTION TO MISTAKE PROOFING

In the previous discussions of defect detection and quality control, there is one thing in common: these methods are reactive, not proactive. First, some manufacturing issue caused defects, then the defects are detected, and finally the affected units are either discarded or go through costly rework.

SQC can help reduce the frequency and impact of manufacturing issues, but it is not foolproof. Strict distributional assumptions – and the fact that reality cannot be exactly modeled with a smooth probability curve in the first place – means that SQC underestimates the frequency of fringe or "tail" events that disrupt production (Hinckley 2001).

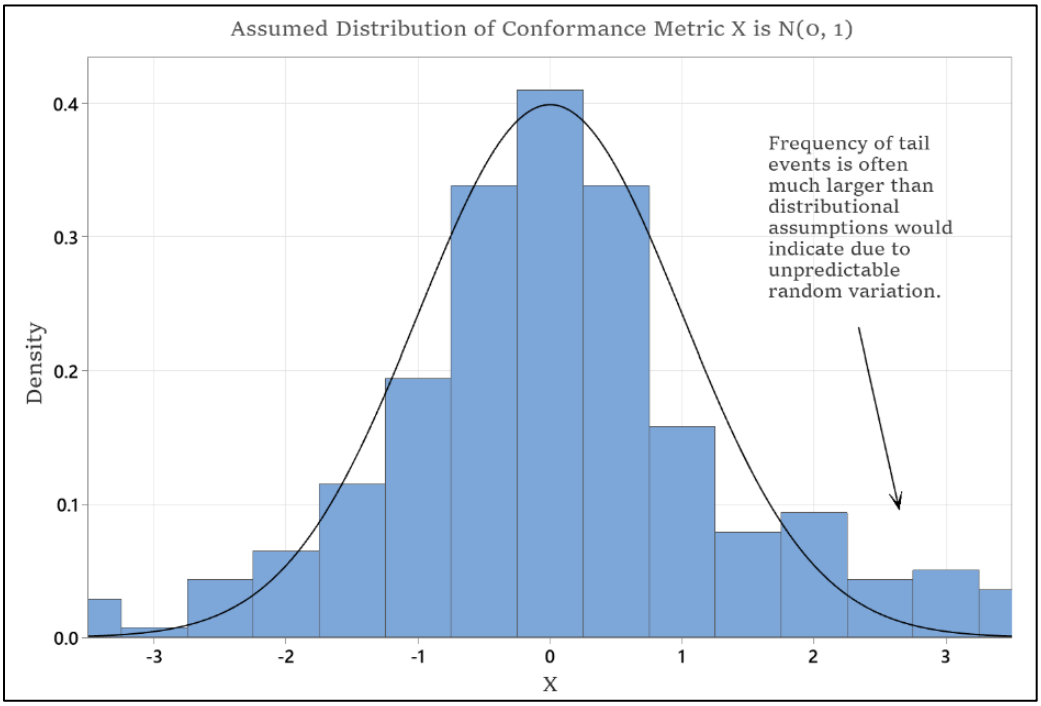

Figure 27. Probability Distribution Function demonstrating the difference in expected vs actual tail events that cause large shifts away from requirements

A wise woman named Hannah Montana once said: "Everybody makes mistakes. Everybody has those days." And it's true – even if the root causes of all defects were identified and eliminated from the manufacturing process, the number of mistakes will never be zero because human error is inevitable.

Mistake Proofing, also known by its Japanese name Poka-Yoke, is a manufacturing controls method pioneered by engineer Shigeo Shingo. It addresses the fundamental limitations of SQC by promoting an awareness of human nature and a deep understanding of the manufacturing process at hand.

Defects can be reduced if mistakes, which will inevitably happen, never manifest as defects.

Shingo believed that mistake proofing methods could reduce the defect rate to zero (Shingo, 1986). This assertion may or may not be realistic, but regardless, no manufacturing control plan is complete without proper mistake proofing.

5.1. Characterizing Mistakes

Some examples of manufacturing mistakes given by Hinkley (2001) are:

- Not performing a required action
- Performing a prohibited action
- Misinterpreting instructions
- Setup mistakes
- Missing parts
- Faulty operation or adjustments

Mistakes should be traced to their source in the production process, which is often called "root cause analysis." A popular method within quality control circles is called the "five whys method." (Institute of Industrial and Systems Engineers 2005)

This involves asking "Why did *that* happen?" multiple times (not always exactly five times) in succession until the root cause of a mistake is identified. For example, leg pieces for Luke Skywalker action figures are created with a series of molds on a conveyor belt, which are filled with hot plastic and then solidified with a cooling process. The recent batch of leg pieces were shorter than expected. (Aren't you a little short for a Stormtrooper?)

"Why is this recent batch of Luke Skywalker legs so short?"
"Because the leg production machine isn't filling up the leg molds all the way."

"Why isn't the production machine filling up each mold all the way?"
"Because it moves on to the next mold too quickly, before the current leg is finished."

"Why is the machine moving on to the next mold too quickly?"
"Because the timing of the machine has been mis-calibrated."

"Why is the timing of the machine mis-calibrated?"
"Because the instructions are unclear, and new employees are easily confused."

The too-short Luke Skywalker legs have been traced to a human calibration mistake, which was caused by unclear instructions. Now, we know the mistake-proofing efforts should be focused on improving instructions so miscalibration doesn't happen again.

5.2. Reducing Complexity

Sometimes, the best way to reduce error in a manufacturing process is by reducing the complexity of the process itself. There is a clear link between manufacturing complexity, manufacturing time, and manufacturing defects (Hinkley 2001).
Further, complexity not only causes mistakes, but makes it more difficult to detect defects once they occur. Before adding new Poka-Yoke devices or instructions, it's worth taking a look at the production process or product design and make it simpler, if possible.
Hinckley (2001) lays out the following steps for complexity reduction:

1. Choose a product to simplify and create an assembly diagram for it. This includes all parts and subcomponents.
2. Create a tree diagram for the product, which includes all assembly operations that must be performed to create the product.
3. Use something called a Quality by Control of Complexity (QCC) worksheet, which crosses each part with each operation and records the associated assembly time between them, if it exists.

For a QCC sheet example, consider this excerpt of a lightsaber manufacturing process for Luke Skywalker action figures at Kenner Production Plant #66:

Table 12. Example QCC sheet for a lightsaber

	Snap in place	**Use Screwdriver**	**Twist-Lock**	**Apply adhesive**
Hilt			2 sec	
Kyber beam emitter			2 sec	
Kyber Crystal	2 sec			3 sec
On/off switch	1 sec			4 sec
Screw x10		5 sec (x10)		
Spring x5	1 sec (x5)			

Since assembly time is so closely related to process complexity and the mistake rate, these QCC sheets can be used to inform mathematical comparisons between different designs and manufacturing processes. The following formulas are given in Hinckley (2001).

Let T_i be the total assembly time and let O_i be the total number of assembly operations for the i^{th} product design. The complexity factor for the product design is then:

$$CF_i = T_i - t_0 * O_i$$

Where t_0 is the threshold assembly time, which has historically been shown to have a value of 2.4 seconds per operation.
The defect rates of two product designs, as a result of complexity, can be compared with the relationship:

$$\frac{D_i}{D_j} = \left(\frac{CF_i}{CF_j}\right)^k$$

Where the constant $k = 1.3$, the slope of D over CF on a log-log plot.

Kenner analysts have reported very high popularity of the first batch of lightsaber-equipped Luke Skywalker action figures. They believe it has counteracted the negative effects of the earlier loose head debacle. However, Kenner Production Plant #66 has seen high rework and scrap costs because of the complicated lightsaber design. Operating personnel are taking longer to get the hang of this process than expected. Engineers have proposed a second, less complex lightsaber design. They hope it can reduce the defect rate enough to increase output and meet customers' frenzy for lightsabers.

The original design has a total assembly time of 90 seconds and 12 total assembly operations. The new design has a total assembly time of 65 seconds and just 10 total assembly operations. Let CF_1 and CF_2 be the complexity factors of the first and second designs, respectively.

$$CF_1 = 90 - 2.4 * 12 = 61.2$$
$$CF_2 = 60 - 2.4 * 10 = 36$$

This leads to a defect rate comparison:

$$\frac{D_1}{D_2} = \left(\frac{36}{61.2}\right)^{1.3} = 0.50167$$

Which means the new design will lead to a more than 50% reduction in complexity-driven defects. There are other, more intuitive methods of reducing complexity in a process covered in Hinckley (2001). One option is the division of labor into repetitive tasks that take 1 or 2 minutes, known as "takt time." This allows operating personnel to fall into an easy routine.

Another is ensuring a good organizational flow the factory floor. Machines should be arranged so that the product travels a short distance from one manufacturing step to the next. Unnecessary inventory, broken machinery, and other clutter should be clearly marked to prevent confusion. This is known as the "red tag tactic," after the red-colored tags that are often used for this purpose.

Complexity reduction is key to mistake proofing, and a simple design will reduce the amount of mistake proofing procedures needed.

5.3. Poka-Yoke Concepts and Methods

Poka-Yoke is the leading mistake proofing ideology, developed by Shingo. It involves the implementation of simple methods, called "devices," that make it impossible to make a specific mistake during the manufacturing process.

One of the earliest examples of a Poka-Yoke device is described by Grout and Downs (2024). A product includes two push buttons, each with a spring that must be inserted underneath. Sometimes workers forget to add one of the springs, which is clearly a defect. To solve this problem, a dish is placed in front of the assembly station. Workers must place two springs in the dish before they begin assembly. When the unit is finished, the worker knows they have forgotten a spring if one is left in the dish. In this case, they fix the unit before passing it on.

A small alteration in the process leads to a significant reduction in defects caused by natural human forgetfulness. Other common devices include checklists, counters, and guiding machinery, or warning infrastructure.

Poka-Yoke devices should be simple and inexpensive, so the benefits of mistake proofing are not offset by complexity or high costs. They should also be human-centered and placed in areas where mistakes are commonly made. This way, the problem is eliminated at its root, before a true defect is generated.

These ideas are also closely related to another Japanese concept called "jidoka," or autonomation (Tommelein 2008). The devices should themselves be immune to human error, and the best way to achieve this is to make the devices autonomous. Then, operators can step in and solve the problem. Devices can be organized into groups; the number of groups and what they entail varies in different sources. The following division into three groups is a rough consensus of ideas, with examples from Kenner Production Plant #66.

The first group is warning devices. This may include lights, bells, or other attention-grabbing techniques to alert the operator that something has gone wrong or is about to go wrong. Operators at Kenner Production Plant #66 must insert the Kyber crystal into each lightsaber, but they sometimes forget. Lightsabers are weighed after moving on to the next step of assembly. If the

weight is too low, a red light flashes, signaling that the crystal has been forgotten. This is a warning device.

The second group is control devices. These devices automatically fix mistakes or make it impossible for the operator to proceed if a mistake has been made. One of the final steps of lightsaber production is attaching the Kyber beam emitter to the hilt with a twist-lock operation. If the emitter is not fully locked in place, the total height of the lightsaber is slightly higher. The lightsaber must pass through a small conveyor belt to reach the next step of assembly. The door to the conveyor belt is just the right size, so if the emitter isn't locked down, the operator cannot fit the lightsaber on the conveyor belt. This is a control device.

The third group is shutdown devices. These devices completely halt the machine or process if a mistake is detected. All lightsabers must be equipped with a safety switch to prevent accidents in transit. Sometimes, operators forget to attach the safety switch below the on-off button. Kenner engineers, anticipating this mistake, designed the lightsaber so the on-off button is stuck in "off" unless the safety switch is included. This is a shutdown device.

The general methodology behind Poka-Yoke can be broken down into five steps (Six Sigma US 2024):

1. Identify critical defects and root causes (five whys)
2. Redesign process to avoid identified errors (i.e. complexity reduction)
3. Incorporate controls and alerts (i.e. Poka-Yoke devices)
4. Validate proof of concept (Are defects actually being prevented?)
5. Expand implementation

Once a new Poka-Yoke device is developed and determined to be effective, it should be documented. Then, it can be used for future products with similar manufacturing operations and mistakes.

Work instructions (WIs), also known as operating instructions, are a central component of mistake proofing at SNL. Every minute detail of a manufacturing or operating process is clearly documented and illustrated in a WI. This includes an outline of the process, complete descriptions of materials, tools, equipment, safety requirements, quality standards, and more.

Jackson (2020) notes that if a WI is written properly, an employee should be able to follow it independently after training. He also emphasizes that a section outlining clear definitions – abbreviations, acronyms, and more – is essential. Humans may even have different names for the same component or process. A WI should remove as much ambiguity as possible to minimize human error.

5.4. Mistake-Proofing Culture

As with all elements of manufacturing controls, mistake proofing isn't just a set of methodologies or steps to follow – it requires a mindset shift throughout the organization. Employees at all levels should adopt the Zero Quality Control (ZQC) philosophy and believe that a process with zero defects is possible (Shingo 1986).

A mistake-proofing culture should be constructive and collaborative. Hinkley (2001) notes that imposing consequences on workers for making mistakes can actually have a negative effect on production. The number of mistakes or defects doesn't decrease – instead, workers feel threatened and try to conceal mistakes from management.

This prevents mistakes from being identified and resolved. Production issues will persist. An alternative is to reward people for reporting mistakes and implementing successful mistake-proofing strategies.

Mistake proofing is especially important in high-consequence industries, like aviation, health care, nuclear power, and weapons manufacturing.

6. CASE STUDY

Many of the manufacturing controls ideas covered in this chapter can be demonstrated with a case study of measurement data from SNL.

SNL creates many piece parts that are included in higher-level assemblies. One such piece is a cylinder.

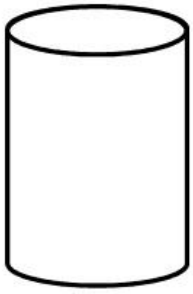

Figure 28. The cylinder a sub-component produced by SNL.

Below is a subset of a dataset measuring the z-dimension – sort of like the height – of a stream of cylinders produced by the labs. The z-dimension was measured in twelve different locations, each labeled X1, X2, X3, … X12. Each cylinder is identified with a serial number, and a total of 157 cylinders were measured.

Table 13. Excerpt of the cylinder data.

Serial Number	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1913695	1.00314	1.00314	1.00345	1.00386	1.00447	1.00375	1.00365	1.00386	1.00324	1.00355	1.00314	1.00283
1917692	0.99873	0.99924	1.00006	0.99965	0.99904	0.99894	0.99822	0.99812	0.99720	0.99740	0.99781	0.99843
1917694	0.99976	1.00017	0.99883	0.99883	0.99843	0.99832	0.99822	0.99843	0.99883	0.99945	0.99965	1.00006
1917695	0.99832	0.99812	0.99761	0.99812	0.99873	0.99843	0.99843	0.99914	0.99904	0.99914	0.99904	0.99873
1918687	1.00006	1.00058	1.00027	0.99996	0.99955	0.99914	0.99873	0.99863	0.99853	0.99894	0.99996	1.00037
…	…	…	…	…	…	…	…	…	…	…	…	…

Also included in the data are columns for the mean, range, minimum, and maximum of the twelve measurements for each row:

Table 14. Excerpt of the cylinder data.

Mean in Row ($\bar{x}$)	Max in Row (x_{max})	Min in Row (x_{min})	Range in Row (R)
1.00351	1.00447	1.00283	0.0016395
0.99857	1.00006	0.99720	0.0028690
0.99908	1.00017	0.99822	0.0019468
0.99857	0.99914	0.99761	0.0015370
0.99956	1.00058	0.99853	0.0020493

Since this data is recording z-dimensions, the range represents the parallelism of the cylinder. It should, ideally, be close to zero. Specification limits for the z-dimension of the cylinder were given; the USL is 1.0045 and the LSL is 0.9955.

Both the z-dimension and parallelism of the cylinder are quantities of interest that we can investigate with our manufacturing controls toolbox. We will start with a brief exploratory data analysis (EDA) and distributional analysis, followed by a demonstration SQC techniques. The analysis will be carried out in Minitab, with accompanying instructions.

6.1. EDA and Distributional Analysis

EDA and identifying rough distributions for our variables will help us better understand our data and prepare for appropriate statistical quality analysis.

The most obvious summary statistic is the grand average, or the mean of "mean in row," which we will denote with $\bar{\bar{x}}$. In Minitab, we can go under the "Calc" menu, select "column statistics," and then choose the mean of "mean in row."

This gives us $\bar{\bar{x}} = 1$.

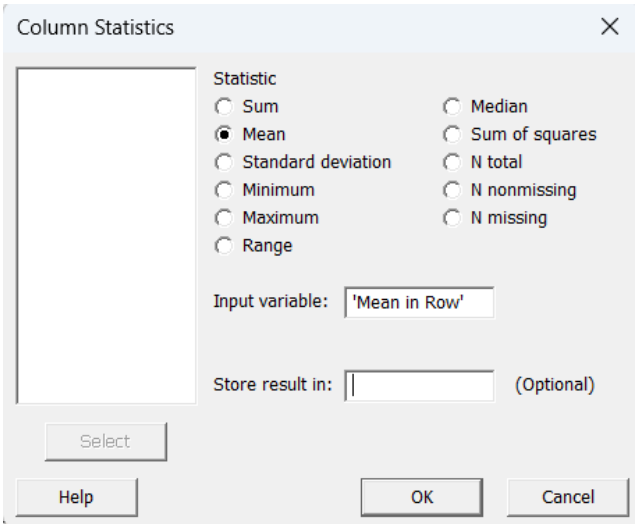

Figure 29. Menu selections for the grand average.

A similar process can be used to find the mean of "range in row," or $\bar{R}$, and other important summary statistics. We have that $\bar{R} = 0.003324$.

The complete set of summary statistics for all columns can be calculated at once by selecting the "Stat" menu, then "Basic Statistics," then "Display Descriptive Statistics." All columns can be selected in the following manner:

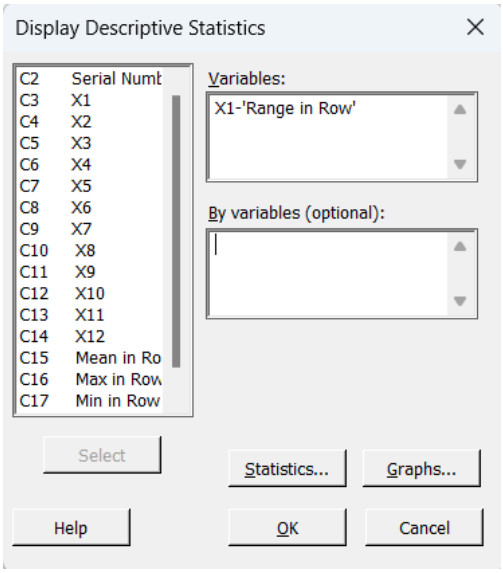

Figure 30. Menu selections for finding all descriptive statistics.

This produces many key summary statistics for each column, including mean, standard deviation, and quantiles:

Statistics

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
X1	0.999867	0.0021925	0.996273	0.998220	0.999655	1.00147	1.00478
X2	1.00020	0.0025436	0.996376	0.998220	0.999860	1.00173	1.01164
X3	1.00001	0.0022800	0.995761	0.998220	0.999655	1.00147	1.00970
X4	1.00007	0.0021956	0.995761	0.998323	0.999552	1.00173	1.00724
X5	1.00023	0.0024841	0.995966	0.998630	0.999706	1.00168	1.00990
X6	1.00021	0.0023601	0.995351	0.998528	0.999552	1.00163	1.00959
X7	1.00000	0.0022333	0.995044	0.998425	0.999552	1.00145	1.00498
X8	0.999877	0.0022746	0.995556	0.998246	0.999450	1.00150	1.00488
X9	0.999904	0.0024080	0.995556	0.998015	0.999603	1.00193	1.00806
X10	1.00001	0.0025867	0.995863	0.998015	0.999603	1.00181	1.00929
X11	0.999695	0.0021476	0.995761	0.997810	0.999603	1.00129	1.00447
X12	0.99999	0.0022062	0.995761	0.998246	0.999808	1.00170	1.00529
Mean in Row	1.00000	0.0020454	0.996820	0.998333	0.999488	1.00161	1.00464
Max in Row	1.00200	0.0027636	0.998015	0.999885	1.00124	1.00386	1.01164
Min in Row	0.998673	0.0023095	0.995044	0.996786	0.998015	1.00042	1.00396
Range in Row	0.0033242	0.0021274	0.0008197	0.0019468	0.0027153	0.0036631	0.0098367

Figure 31. Descriptive statistics for all the data.

We see here that each of the 12 z-dimension measurements, X1-X12, have very similar averages; all fall between 0.999 and 1.001. Recall that the specification limits are 0.9955 and 1.0045. This means that, on average, the z-dimension at all 12 locations is on target.

A series plot of "mean in row" with specification limits supports this statement and demonstrates strong process capability. Note that the following is not a control chart, because it plots specification limits, not control limits!

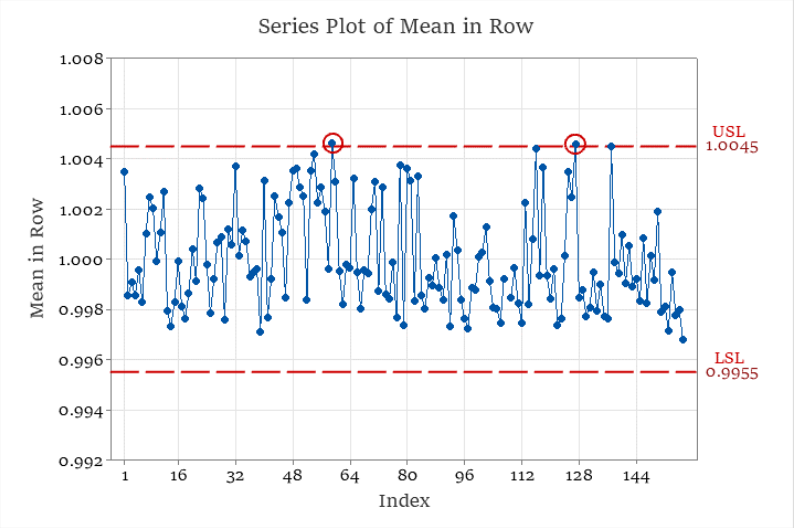

Figure 32. Series chart with specification limits.

However, note that the maximum "mean in row" is 1.00464. Indeed, we see a few points on the series plot that are at or above the upper specification limit. A control chart might be a good idea to ensure that this process is meeting standards.

There's something else we should check before creating a control chart. Suppose we want to track the "mean in row" for each cylinder. Recall that for $\bar{x}$ and R charts, the range is used to estimate the standard deviation, which is used to calculate the control limits on the $\bar{x}$ chart.

This worked fine when each sample consisted of independent units and independent measurements – like the radii of 5 different Luke Skywalker heads. In this dataset, however, we are averaging across 12 measurements made on the *same* unit. As a result, X1-X12 are not independent and will be highly correlated. Indeed, by selecting the "Stat" menu, "Basic statistics," and then "Correlation," we see the following results:

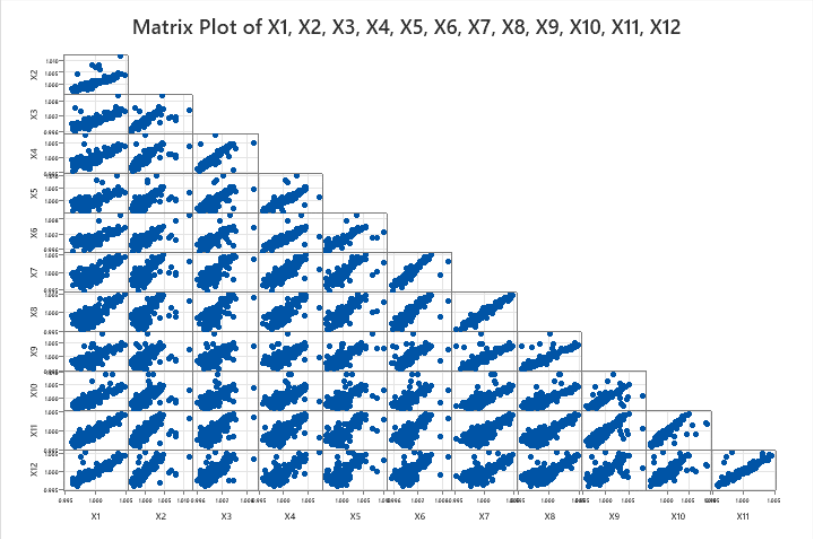

Figure 33. Visual correlation matrix for X1-X12.

Correlations

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
X2	0.809										
X3	0.821	0.776									
X4	0.788	0.742	0.821								
X5	0.750	0.668	0.753	0.820							
X6	0.754	0.785	0.766	0.829	0.815						
X7	0.760	0.660	0.752	0.808	0.817	0.898					
X8	0.771	0.626	0.716	0.774	0.782	0.854	0.964				
X9	0.740	0.539	0.612	0.633	0.745	0.681	0.773	0.834			
X10	0.795	0.589	0.631	0.645	0.640	0.670	0.779	0.805	0.756		
X11	0.915	0.690	0.731	0.712	0.694	0.720	0.805	0.841	0.827	0.857	
X12	0.880	0.707	0.721	0.764	0.682	0.679	0.726	0.745	0.718	0.747	0.883

Figure 34. Correlation coefficient matrix for X1-X12.

Correlation between all 12 locations is high; none of the coefficients drop below 0.50. This is especially true for "neighbors." For example, the correlation between X7 and X8 is 0.964. Intuitively, this makes sense. On the same cylinder, 12 different measurements of the height in different locations should tend in the same direction.

For our analysis, this correlation will cause the range between X1-X12 for one cylinder to be small, which means our standard deviation will be underestimated, and our control limits will be too narrow.

This means a standard $\bar{x}$-R control chart will likely show an out-of-control process – even if that's not necessarily the case. Under the "Stat" menu, then "Control charts," then "Variables charts for subgroups," then "Xbar-R," we can create a naïve first pass at a control chart:

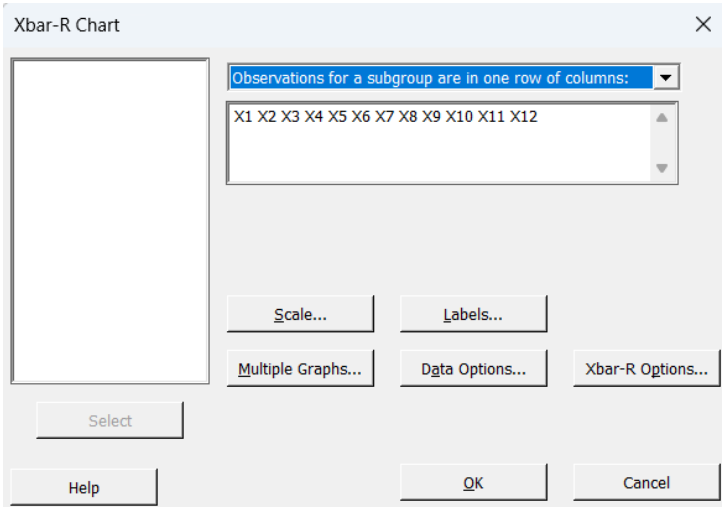

Figure 35. Menu settings for naïve control chart.

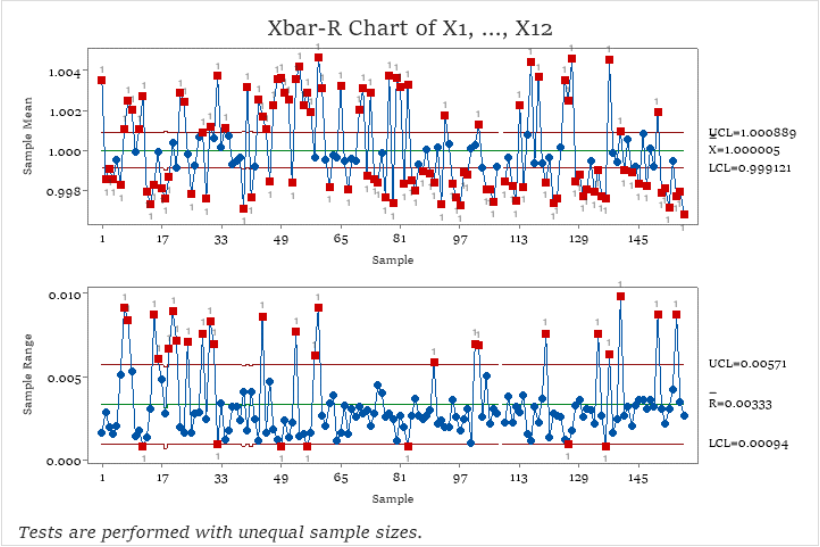

Figure 36. Naïve control chart.

As we predicted, many points are plotting out of control here. But this doesn't mean there's an assignable cause. Instead, our approach is incorrect because of the strong dependence structure between the variables.

A correct approach that mitigates the effect of the correlation is outlined later in this case study. It's also necessary to determine, roughly, the distribution of variables in the dataset before creating control charts.

Minitab can generate probability plots for a number of common distributions. These plots include a hypothesis test for fit. If the p-value is less than the significance level, this means there is significant evidence against the fit. In other words, higher p-values mean a better fit.

These plots can be created simply by selecting "Probability Plot" under the "Graph" menu. The pop-up box gives options for changing the featured variable and the distribution.
Several probability plots were created for "mean in row." The following p-values for fit were obtained:

Table 15. p-values for goodness of fit tests.

Distribution	p-value
Normal	<0.005
Lognormal	<0.005
Gamma	<0.005
Weibull	<0.010
Logistic	<0.005

In summary, the mean in row isn't defined particularly well by any of the most common distributions.

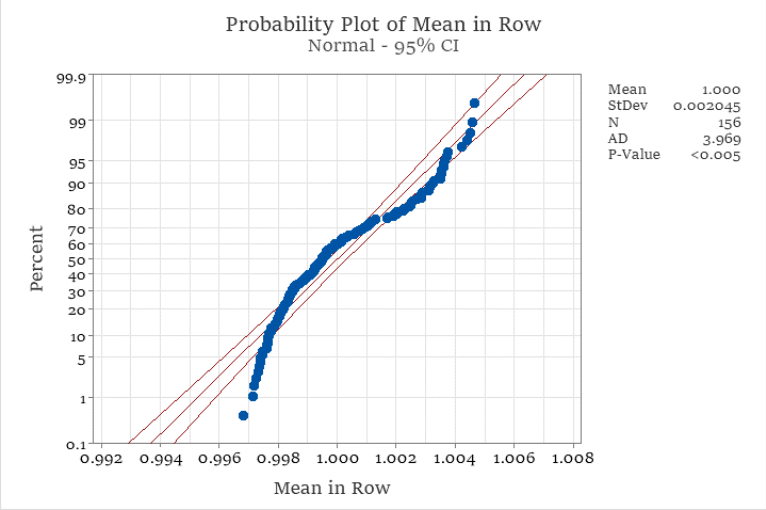

Figure 37. A normal probability plot for "mean in row," showing poor fit.

Perhaps a histogram will give us a better idea of what's going on with this variable. Histograms can be found under the "Graph" menu. Here, we want to select a "Simple" histogram, and make sure "mean in row" is the selected variable. To change the number of bins, double-click the bars and go to the "Binning" tab.

By setting the bin number to 15, we reveal an interesting pattern:

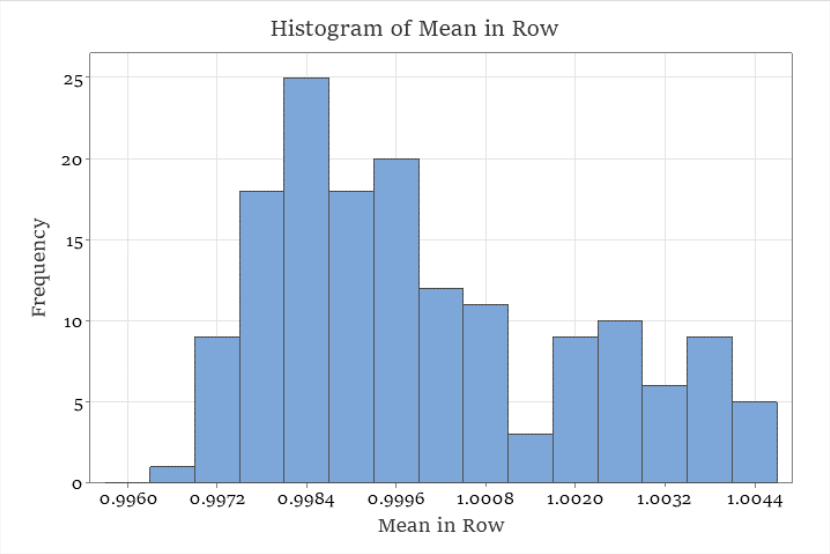

Figure 38. Histogram with 15 bins for "mean in row."

This histogram suggests that our cylinder population actually consists of two subpopulations with two different distributions. A subset of the cylinders might come from a process different enough to change the distribution of the mean z-dimension.

To proceed, we will try to isolate each population and see if there's an appropriate fit for each. Using the histogram as a guide, we will split the dataset into cylinders with an average z-dimension less than or equal to 1.0014 – group 1 – and greater than 1.0014 – group 2.

6.1.1. Group 1

Several probability plots were created for "mean in row" of the group 1 data, with the 3-parameter Weibull distribution presenting a strong fit.

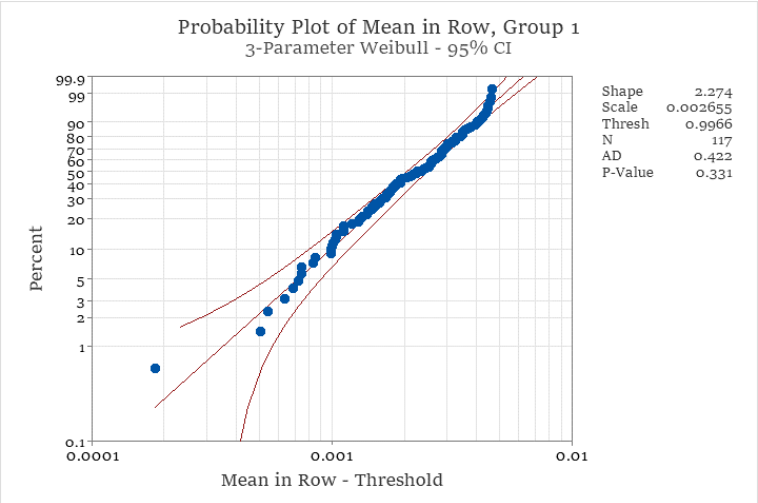

Figure 39. 3-parameter Weibull probability plot of "mean in row" for Group 1.

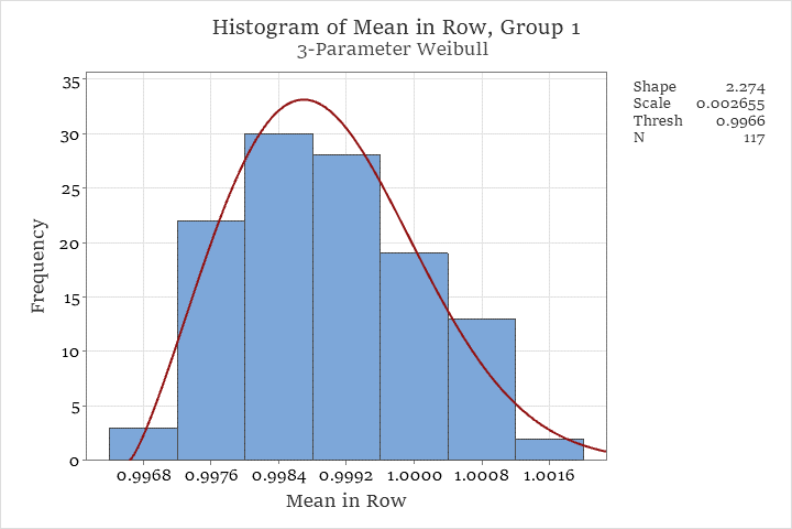

Figure 40. Histogram with 3-parameter Weibull fit for Group 1.

The above histogram can be created by selecting "With Fit" instead of "Simple" under the histogram pop-up menu. The default distribution will be normal, but this can be changed by double-clicking the red distribution line and selecting the "Options" tab:

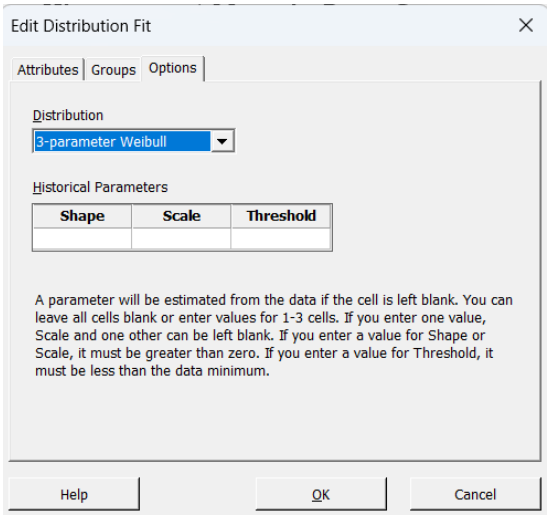

Figure 41. Menu for changing the estimated distribution on a histogram.

Parameters for the distribution are featured on the right side of the histogram plot. The distribution of group 1 is estimated as:

$$W \sim Weibull(\beta = 2.274, \eta = 0.002655, \gamma = 0.9966)$$

Where β is the shape parameter, η is the scale parameter, and γ is the location parameter. The probability distribution function for 3-parameter Weibull distributed variable is:

$$f(w) = \frac{\beta}{\eta}\left(\frac{w-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{w-\gamma}{\eta}\right)^{\beta}}$$

Let's also investigate the distribution of the parallelism, or "range in row," of the group 1 cylinders. Initially, none of the probability plots looked promising. A histogram gave some hints, and the data seemed to reasonably match with a 3-parameter lognormal fit.

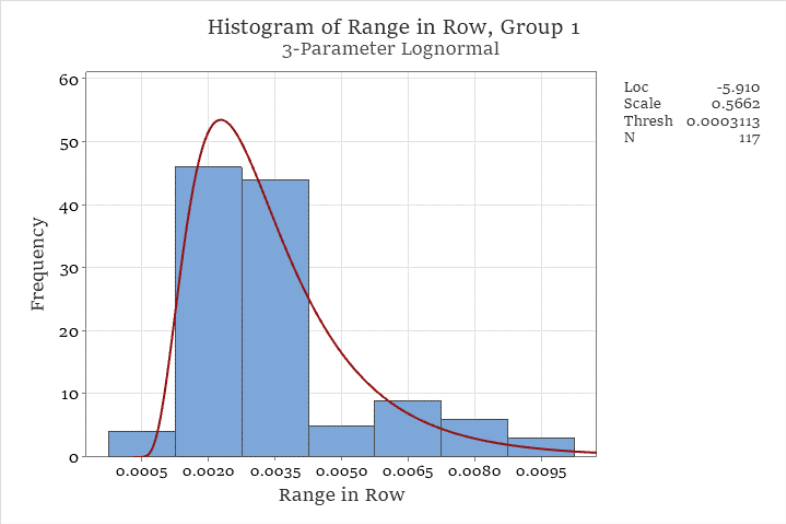

Figure 42. Histogram with 3-parameter lognormal fit for range in row of Group 1

A probability plot of "range in row" with the 3-parameter lognormal distribution did not report a p-value for the goodness of fit test, but visually, the plot was acceptable.

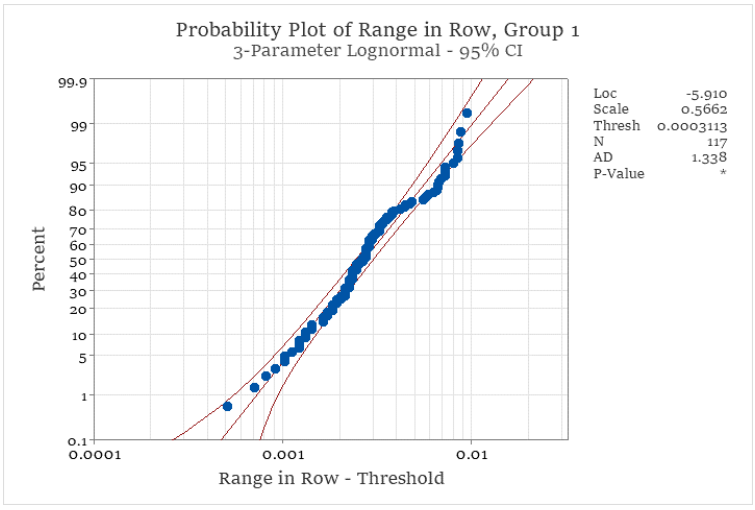

Figure 43. Probability plot for range in row of Group 1

Using the estimated parameters, the parallelism of group 1 cylinders has the following distribution:

$$R_W \sim lognormal(\mu = -5.910, \sigma = 0.5662, \theta = 0.0003113)$$

6.1.2. Group 2

For the group 2 data, several probability plots showed high p-values. Since the p-value for the 3-parameter Weibull distribution slightly beat out the others, we will select it for Group 2 as well.

Table 15. A selection of p-values for testing goodness of fit of Group 2 data.

Distribution	p-value
Normal	0.478
Lognormal	0.480
3-Parameter Weibull	>0.500
Largest Extreme Value	>0.250
Logistic	>0.250

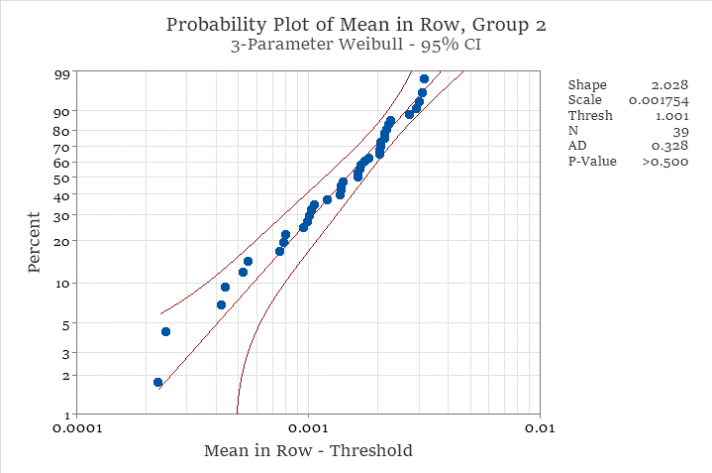

Figure 44. 3-parameter Weibull probability plot of "mean in row" for Group 2.

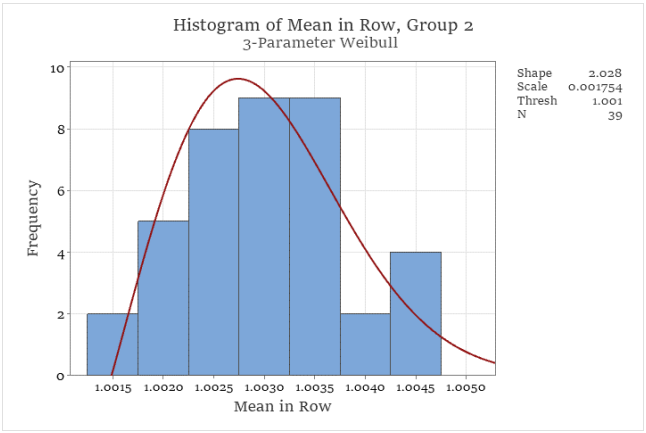

Figure 45. Histogram with 3-parameter Weibull fit for Group 2.

The distribution of Group 2 is estimated as:

$$V \sim Weibull(\beta = 2.028, \eta = 0.001754, \gamma = 1.001)$$

There were challenges in fitting a distribution to the "range in row" of group 2. As with group 1, a histogram and a probability plot showed a good-enough match with the 3-parameter lognormal fit.

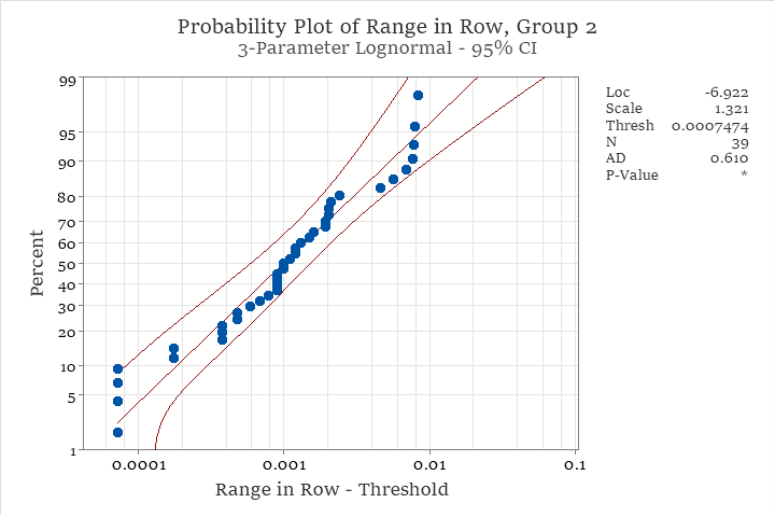

Figure 46. 3-parameter lognormal probability plot for "range in row" of Group 2.

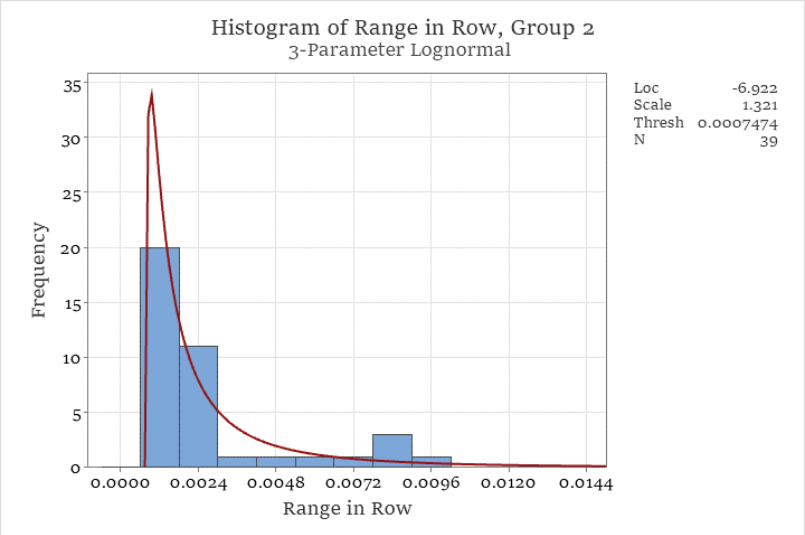

Figure 47. Histogram with 3-parameter lognormal fit for range in row of Group 2.

Using the estimated parameters, the parallelism of group 2 cylinders has the following distribution:

$$R_V \sim lognormal(\mu = -6.922, \sigma = 1.321, \theta = 0.0007474)$$

In summary, our data is split into two groups. Each group has a 3-parameter Weibull distribution for the mean in row and a 3-parameter lognormal distribution for the range in row, but with different values for the parameters. As part of the investigation, it is best practice to discuss this two-subpopulation phenomenon with subject matter experts (SMEs). They might be able to identify the cause of the split and provide information to help with analysis.

6.2. Control Charts Demonstration

Now that we have a better idea of what we can – and can't – do with our dataset, we can move into more meaningful analysis. This data is historical. If we can determine that the process that created these cylinders was in control, we can set up control charts to help monitor the production of future cylinders.

As noted previously, the high correlation between X1-X12 make the basic $\bar{x}$-R and $\bar{x}$-s charts inaccurate. To remove the dependence structure, we can treat the "mean in row," or the average z-dimension across X1-X12, as an individuals measurement.

The best control chart to use in this situation is the Shewhart control chart for individuals, also known as the individuals-moving range (I-MR) control chart. Refer to section 8.1.7 for more information about I-MR charts.

Recall that the standard procedure for making I-MR charts assumes normally distributed data; the charts are highly sensitive to deviations from this assumption.

"Mean in row" is highly non-normal – in fact, we showed it can actually be separated into two different 3-parameter Weibull distributions. In cases like this, it's best practice to use probability limits instead of limits based on multiples of the standard deviation. The rest of this section should be read with a British accent.

Since we broke up our data into two populations, we will create separate control charts for each. Let us begin with Group 1.

6.2.1. Group 1

To set probability limits, we select a desired type I error probability, α, and then find the corresponding percentiles. The most popular choice for probability limits is 0.002, or 0.2%. Using Minitab's Inverse Cumulative Distribution function, we find for Group 1:

$$LCL = w_{\alpha/2} = w_{0.001} = 0.996727$$
$$UCL = w_{1-\alpha/2} = w_{0.999} = 1.00281$$

To create a control chart with custom control limits, we will make a time series chart instead of using Minitab's built-in control chart function.

Under the "Graph" menu, select "Time Series Chart," then in the pop-up menu, choose "Simple." The y-axis scale will probably have to be widened to make room for the control limits; this can be done by double-clicking one of the y-labels.

Control limits will be added as horizontal reference lines, under the "Add Item" menu of the chart editor.

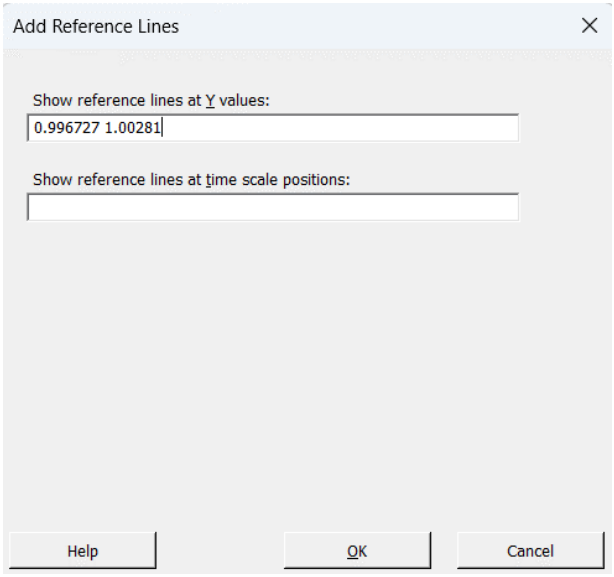

Figure 48. Adding control limits to a time series plot.

We'll use the average "mean in row" of group 1 as the center line. This can be added as a reference line in the same way. The plot can be customized and annotated to the user's preferences. The final individuals chart for "mean in row" of the group 1 data is:

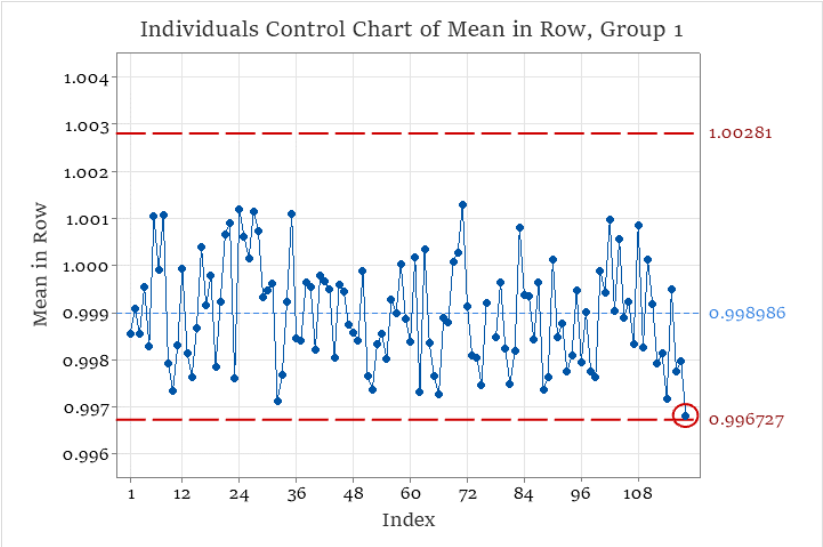

Figure 49. Individuals chart for "mean in row" of Group 1.

This chart shows strong evidence of an in-control process. The last point seems to cross the LCL, but just one point is usually not a cause for alarm. It's expected that we will occasionally see a random out-of-control point.

Since we just made an Individuals Chart, it might seem like the next step is to make a moving range chart for the same data. However, since we split the data into two groups, the moving range is no longer a meaningful statistic. Neighboring observations in the Group 1 data might not actually appear next to each other in the full dataset.

We were also interested in monitoring the parallelism of each cylinder, defined by the "range in row" variable. A natural replacement for the traditional moving range chart is a control chart for the parallelism of each cylinder.

Using the 3-paramter lognormal distribution found for the range in row of Group 1, we can calculate probability limits:

$$LCL = R_{W_{\alpha/2}} = R_{W_{0.001}} = 0.0007828$$
$$UCL = R_{W_{1-\alpha/2}} = R_{W_{0.999}} = 0.0159140$$

With this information, we can the time series plot to create and customize our control chart.

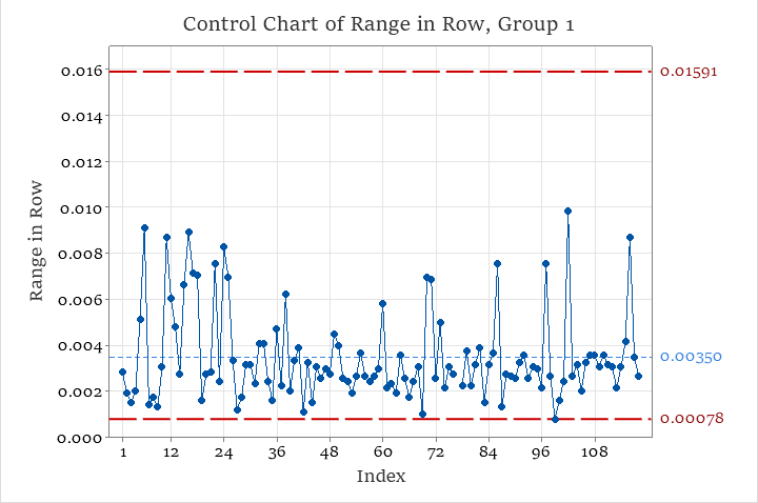

Figure 50. Control chart for parallelism of Group 1.

This chart also shows evidence of an in-control process. Recall that a highly left-skewed lognormal distribution was fit to the data. This causes the UCL to be much higher than most of the data, and the center line to be closer to the LCL.

Note that since our goal for the parallelism is 0, we shouldn't be too concerned about points going below the LCL.

6.2.2. Group 2

We will follow the same process to make control charts for the second group.

Our probability limits for the individuals control chart of "mean in row" for Group 2, based on the 3-parameter Weibull distribution, are:

$$LCL = v_{\alpha/2} = v_{0.001} = 1.00106$$
$$UCL = v_{1-\alpha/2} = v_{0.999} = 1.00555$$

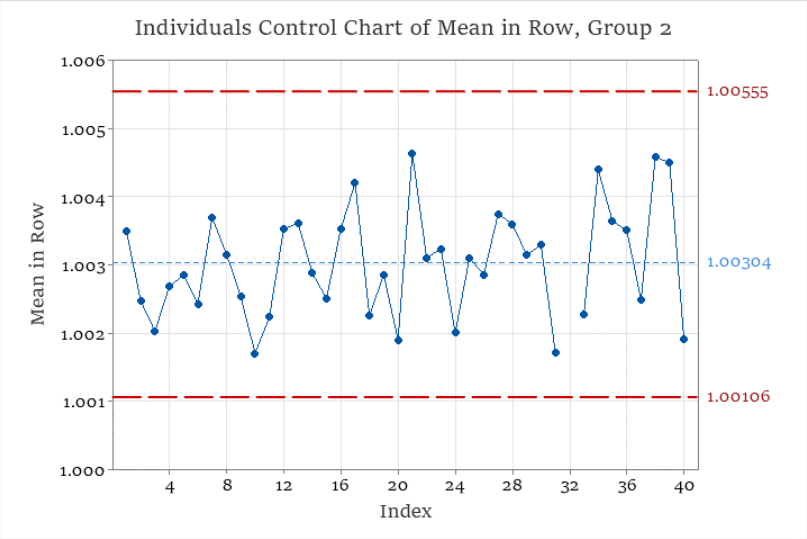

Figure 51. Individuals control chart for "mean in row" of Group 2.

No points have plotted out of control in the second group. The UCL is higher than the USL here, which should be kept in mind if this chart is used to monitor the production of future cylinders. Generally, we want our control limits to be within, or narrower than, the specification limits. This way, the natural variation of the process has almost no chance of violating requirements.
Finally, we make the control chart for the parallelism of Group 2.

The probability limits for "range in row" of Group 2, calculated with the 3-parameter lognormal distribution, are:

$$LCL = R_{v_{\alpha/2}} = R_{v_{0.001}} = 0.0007640$$
$$UCL = R_{v_{1-\alpha/2}} = R_{v_{0.999}} = 0.0591846$$

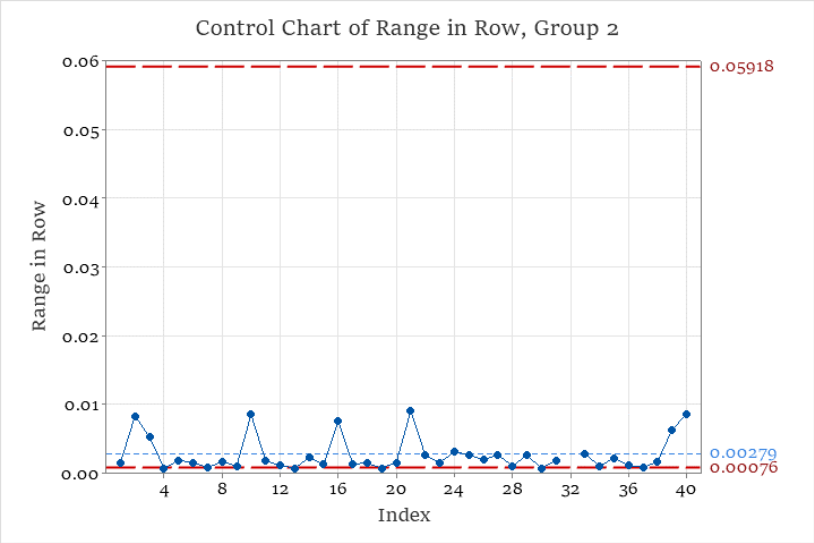

Figure 52. Control chart for parallelism of Group 2.

Again, the highly skewed lognormal distribution causes the UCL to seem skewed. This chart is not ideal, but without a clear distribution to build from, it's the best we can set up. It shows no evidence of an out-of-control process.

6.3. Conclusion

In the case study above, we demonstrated the preparation and decision making that goes into creating control charts with real manufacturing data. It's not all as cut-and-dry as Kenner's normally distributed Luke Skywalker data. Sometimes, setting up effective control charts require a little creativity – which is why we should always be open to revising our design as more data arrives. For now, we can reasonably say that our process is in control. We should investigate the reason for the data being split into two different distributions, because ideally, all output should have the same distribution. There must have been some disruption to the process that wasn't severe enough to knock the z-dimension out of specification but changed its distribution enough to create a new population.

7. SUMMARY

We have covered four of the most essential manufacturing controls methods: statistical quality control, acceptance sampling, highly accelerated stress screening, and mistake proofing.

Statistical quality control leverages the idea of variance in a production process to statistically detect changes in the process that negatively affect the output. Core SQC ideas include chance causes of variation versus assignable causes of variation; the Six Sigma philosophy; variables control charts like $\bar{x}$ and R charts, individuals charts, and EWMA charts; attributes charts like p charts and c charts; and chart diagnostics like the average run length and the OC curve.

Acceptance sampling is the creation of statistics-based inspection plans to safeguard against lots with high rates of poor-quality units, whether that be incoming raw material or outgoing assemblies. Core acceptance sampling ideas include single sampling plans; double sampling plans; sequential sampling plans; rectifying inspection and average outgoing quality; diagnostics like Type A and Type B OC curves, average quality level, lot tolerance percent defective, and average total inspection; and sampling schemes and systems.

Highly accelerated stress screening is a special case of environmental stress screening that puts a product through higher-than-normal environmental stresses, mainly temperature and vibration, in order to precipitate built-in, hidden defects. Core HASS ideas include patent versus latent defects; defect density and screening strength; temperature and vibration cycles; operational limits; and safety of screen.

Mistake proofing takes a proactive, human-centered approach to recognize that mistakes are inevitable, but defects can be prevented by simplifying the production process and creating simple devices to reduce human error and catch defective product before it escapes. Core mistake proofing concepts include design and process complexity; the five whys; Poka-Yoke devices, including warning, control, and shutdown devices; work instructions; and jidoka, or autonomation.

Manufacturing controls will help extended testing at SNL become easier and more successful. The methods in this chapter should be viewed as an essential supplement to the testing procedures outlined in the rest of this report. If we ensure the quality of our product at each step in the manufacturing process and stop defects before the final level of assembly, our high-consequence single use systems will be safer, more effective, and more reliable.

8. RELATED READINGS

In this report, we assume constant sample sizes during the construction of control charts. However, due to the nature of certain manufacturing processes, this may not always be possible. Montgomery (2020) describes, for several of the control charts discussed above, different ways to deal with variable sample sizes, one of which is to simply use the average sample size. He also discusses impacts this may have on control chart interpretation.

Another type of control chart, called the cumulative sum (CUSUM) control chart, was not discussed in this report. First developed by E.S. Page (1961), CUSUM control charts are better at detecting small changes in a process than the standard Shewhart control chats. For this reason, it is often compared to the EWMA control chart. Hawkins and Wu (2014) note that neither has a strong advantage over the other, but Montgomery (2020) suggests that EWMA charts are easier to interpret. Crowder (1987) investigates the distribution of run lengths of EWMA charts, essential for chart design and comparison. This includes simplified methods to find ARLs and the standard deviation of the run length for EWMA charts.

There are other types of attribute control charts that are covered by Montgomery (2020) but not in this report. Most of these involve similar ideas, but data that is collected or counted differently, and thus yield slightly different control limits. This includes two charts called the *u* chart and the *np* chart.

A notable modification to the p chart was created by Laney (2002) and is aptly known as the Laney *p'* chart. Because of strict distributional assumptions, standard p charts perform poorly when dealing with under or overdispersion in the data, especially when sample sizes are very large. Under or overdispersion often results from uncontrollable environmental factors that cause excess variance but aren't really "assignable causes" that can be fixed. This can cause overly narrow control limits, which leads to false alarms. Laney's process calculates control limits that account for both within-group variation and between-group variation; the control limits are wider and depend on sample size from point to point. The best description of these charts is found in Laney's original paper, cited below. Minitab has a built-in function to create Laney p' charts.

This report mentions MIL-STD-105E, or its civilian equivalent, as the primary source of standard acceptance sampling plans. However, this is not the only guideline. Another popular plan design is called the Dodge-Romig plans, which can either focus on a target lot tolerance percent defective or average outgoing quality level. Dodge-Romig tables were originally published in the 1940s and are now widely available online.

A type of acceptance sampling plan not described in this report are skip-lot plans. Not all lots are inspected in this plan, and it is typically employed when there is already a substantial history record of high quality. More about these plans can be read in Schilling and Neubauer (2009).

Often paired with Highly Accelerated Stress Screening in literature is a similar concept called Highly Accelerated Life Testing (HALT). HALT was not covered here to avoid confusion, since it is a type of product testing, but is not an acceptance screen. We draw a careful distinction between testing, which has the goal of discovery and design improvement, and environmental screening, which precipitates manufacturing defects. HALT is performed in the design phase and can be used to inform the environmental parameters of HASS or ESS screens. Further information on HALT can be found by Collins, Huzurbazar, and Warr (2024) or Thomas (2015).

Zero Quality Control (ZQC) is the philosophy promoted first by Shigeo Shingo that a zero-defect production process is possible. ZCQ incorporates ideas beyond Poka-Yoke and mistake proofing methods; it also promotes 100% source inspection and organizational improvements. The best source on ZQC is Shigeo's book (1986).

REFERENCES

Bahret, Adam. 2016. "ESS and HASS." Accendo Reliability. November 23, 2016. https://accendoreliability.com/ess-hass/.

Best, M, and D Neuhauser. 2006. "Walter a Shewhart, 1924, and the Hawthorne Factory." *Quality and Safety in Health Care* 15 (2): 142–43. https://doi.org/10.1136/qshc.2006.018093.

Collins, David H., Aparna V. Huzurbazar, and Richard L. Warr. 2024. "Highly Accelerated Life Testing (HALT): A Review from a Statistical Perspective." *Wiley Interdisciplinary Reviews Computational Statistics* 16 (4). Wiley. doi:https://doi.org/10.1002/wics.70000.

Crowder, Stephen V. 1987. "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts." *Technometrics* 29 (4): 401. doi:https://doi.org/10.2307/1269450.

Crowder, Stephen V. 1989. "Design of Exponentially Weighted Moving Average Schemes." *Journal of Quality Technology* 21 (3): 155–62. https://doi.org/10.1080/00224065.1989.11979164.

———. 2013. "HASS Screening: Estimating Screen Efficiency and Fraction Defective in Fielded Product." Sandia National Laboratories.

———. 2017. "A Note on Screen Efficiency." Sandia National Laboratories.

Crowder, Stephen V., and Marc D. Hamilton. 1992. "An EWMA for Monitoring a Process Standard Deviation." *Journal of Quality Technology* 24 (1): 12–21. https://doi.org/10.1080/00224065.1992.11979369.

Crowder, S., Collins, E., Wilson, L. Ghanbari, Z. Butler, N., Duffy, T., Frerking, A. (2025). "Minimum Sample Size Test Plans to Demonstrate Reliability of War Reserve Components." Internal Sandia National Laboratories SAND2025-12600R.

Department of Defense. 1993. *Military Handbook: Environmental Stress Screening (ESS) of Electronic Equipment*.

Downs, Brian T., and John R. Grout. 2024. "A Brief Tutorial on Mistake-Proofing, Poka-Yoke, and ZQC." MistakeProofing.com. March 4, 2024. https://mistakeproofing.com/a-brief-tutorial-on-mistake-proofing-poka-yoke-and-zqc/.

Duncan, Acheson. 1965. *Quality Control and Industrial Statistics*. Richard D. Irwin, Inc.

Hawkins, Douglas M., and Qifan Wu. 2014. "The CUSUM and the EWMA Head-To-Head." *Quality Engineering* 26 (2): 215–22. doi:https://doi.org/10.1080/08982112.2013.817014.

Hinckley, C. Martin. 2001. *Make No Mistake!* Productivity Press.

Institute of Industrial and Systems Engineers. 2005. "What You Need to Know about Mistake-Proofing (Poka Yoke)." https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwje-pqm48aOAxVnIUQIHWMOB1kQFnoECBoQAQ&url=https%3A%2F%2Fwww.iise.org%2Fuploadedfiles%2FIIE%2FEducation%2FSix_Sigma_Green_Belt_Transition%2FPokeYoke.pdf&usg=AOvVaw1gJb6UNqEOOfgHTQ2aQJLX&opi=89978449.

Jackson, Robert P. 2020. "Writing Effective Work Instructions." PDH Center. https://pdhonline.com/courses/g373/g373content.pdf.

Kececioglu, Dimitri, and Feng-Bin Sun. 1995. *Environmental Stress Screening: Its Quantification, Optimization, and Management*. Prentice Hall PTR.

Laney, David B. 2002. "Improved Control Charts for Attributes." *Quality Engineering* 14 (4): 531–37. doi:https://doi.org/10.1081/qen-120003555.

Montgomery, Douglas. 2020. *Introduction to Statistical Quality Control*. Eighth. John Wiley & Sons.

National Institute of Standards and Technology. n.d. "What Is Acceptance Sampling?" Engineering Statistics Handbook. https://www.itl.nist.gov/div898/handbook/pmc/section2/pmc21.htm.

Noskievičová, Darja. 2013. "Complex Control Chart Interpretation." *International Journal of Engineering Business Management* 5 (January): 13. https://doi.org/10.5772/56441.

Nowak, R. 2011. "Lecture 9: Sequential Testing (ECE 830 Statistical Signal Processing)." Https://nowak.ece.wisc.edu/ece830/index.html, 2011.

Page, E. S. 1961. "Cumulative Sum Charts." *Technometrics* 3 (1). Taylor & Francis: 1–1. doi:https://doi.org/10.2307/1266472.

Peterson, Bruce. 2004. "Environmental Stress Screening Tutorial." Accolade Engineering Solutions.

Rice, William B. 1947. *Control Charts in Factory Management*. John Wiley & Sons.

Safety and Reliability Society. 2012. *Applied R&M Manual for Defence Systems*. https://www.sars.org.uk/BOK/Applied%20R&M%20Manual%20for%20Defence%20Systems%20(GR-77)/P3c45.pdf.

Schilling, Edward, and Dean Neubauer. 2009. *Acceptance Sampling in Quality Control: Second Edition*. Taylor & Francis Group, LLC.

Shewhart, W. A. 1931. *Economic Control of Quality Manufactured Product*. D. Van Nostrand Company, Inc.

Shingo, Shigeo. 1986. *Zero Quality Control: Source Inspection and the Poka-Yoke System*. CRC Press.

Six Sigma US. n.d. "What Is Six Sigma?" Six Sigma. https://www.6sigma.us/six-sigma/.

Six Sigma US Editorial Team. 2024. "A Complete Guide to Poka-Yoke in Six Sigma."

SixSigma.us. 2024. https://www.6sigma.us/lean-tools/poka-yoke-six-sigma/.

Thomas, Edward V. 2015. "A Statistical Perspective on Highly Accelerated Testing." OSTI.gov: Sandia National Laboratories. https://www.osti.gov/servlets/purl/1170512.

Tommelein, I. D. 2008. "'Poka Yoke' or Quality by Mistake Proofing Design and Construction Systems." *Proceedings of the 16th Annual Conference of the International Group for Lean Construction* IGLC 16: 195–205.

Vardeman, Stephen, and J. Marcus Jobe. 2016. *Statistical Methods for Quality Assurance: Basics, Measurement, Control, Capability, and Improvement*. Springer-Verlag.

Vining, Geoff. 2009. "Technical Advice: Phase I and Phase II Control Charts." *Quality Engineering* 21 (4): 478–79. https://doi.org/10.1080/08982110903185736.

Wald, Abraham. 1947. *Sequential Analysis*. Dover Publications, Inc.

Wetherill, G. Barrie, and Don Brown. 1991. *Statistical Process Control: Theory and Practice*. Chapman and Hall.

Zwetsloot, Inez M., L. Allison Jones-Farmer, and William H. Woodall. 2023. "Monitoring Univariate Processes Using Control Charts: Some Practical Issues and Advice." *Quality Engineering* 36 (3): 1–13. https://doi.org/10.1080/08982112.2023.2238049.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Stephen V. Crowder	05574	svcrowd@sandia.gov
Elmer W. Collins	00533	ewcolli@sandia.gov
Lauren Wilson	05574	lcwilso@sandia.gov
Clark Snow	07585	cssnow@sandia.gov
Technical Library	01911	sanddocs@sandia.gov

This page left blank

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.