Quality Engineering - Class Notes (experimental)

Jonathan Rosenblatt

November 5, 2015

Preface

This text accompanies my Quality Engineering course at the Dept. of Industrial Engineering at the Ben-Gurion University of the Negev. It has several purposes:

- Help me organize and document the course material.
- Help students during class so that they may focus on listening and not writing.
- Help students after class, so that they may self-study.

At its current state it is experimental. It can thus be expected to change from time to time, and include mistakes. I will be enormously grateful to whoever decides to share with me any mistakes found.

I also ask for the readers' forgiveness for my Wikipedia quoting style. It is highly unorthodox to cite Wikipedia as one would cite a peer reviewed publication. I do so, in this text, merely for technical convenience.

I hope the reader will find this text interesting and useful.

Contents

List of Figures

List of Definitions

Chapter 1

Introduction

Quality Engineering is the study and design of practices aimed improving the "quality" of production. Production is understood in a wide sense, and includes services as well. Quality is understood in many senses. Here are several definitions compiled verbatim from Montgomery (2007) and Wikipedia (2015d):

1. Montgomery: "The reciprocal of variability".
2. American Society for Quality: A combination of quantitative and qualitative perspectives for which each person has his or her own definition; examples of which include, "Meeting the requirements and expectations in service or product that were committed to" and "Pursuit of optimal solutions contributing to confirmed successes, fulfilling accountabilities. In technical usage, quality can have two meanings: (a) The characteristics of a product or service that bear on its ability to satisfy stated or implied needs. (b) A product or service free of deficiencies."
3. Subir Chowdhury: "Quality combines people power and process power".
4. Philip B. Crosby: "Conformance to requirements."
5. W. Edwards Deming: "The efficient production of the quality that the market expects".
6. W. Edwards Deming: "Costs go down and productivity goes up as improvement of quality is accomplished by better management of design, engineering, testing and by improvement of processes."
7. Peter Drucker: "Quality in a product or service is not what the supplier puts in. It is what the customer gets out and is willing to pay for."

8. Victor A. Elias: "Quality is the ability of performance, in each Theme of Performance, to enact a strategy."

9. ISO 9000: "Degree to which a set of inherent characteristics fulfills requirements."

10. Joseph M. Juran: "Fitness for use.".

11. Noriaki Kano and others, present a two-dimensional model of quality: "must-be quality" and "attractive quality." The former is near to "fitness for use" and the latter is what the customer would love, but has not yet thought about. Supporters characterize this model more succinctly as: "Products and services that meet or exceed customers' expectations."

12. Robert Pirsig: "The result of care."

13. Six Sigma: "Number of defects per million opportunities."

14. Genichi Taguchi: "Uniformity around a target value."

15. Genichi Taguchi: "The loss a product imposes on society after it is shipped."

16. Gerald M. Weinberg: "Value to some person".

17. Jonathan D. Rosenblatt: "The efficient fulfilment of a promise".

Collecting ideas

1. Quality is not only about production.
2. Quality is the means, not the end.
3. Quality may deal with the **design** or with **conformance** to a given design.

Almost all of the above definitions, may apply to different characteristics, we call *dimensions of quality*. Following Wikipedia (2015b) :

Dimensions of Quality

Performance Performance refers to a product's primary operating characteristics. This dimension of quality involves measurable attributes; brands can usually be ranked objectively on individual aspects of performance.

Features Features are additional characteristics that enhance the appeal of the product or service to the user.

Reliability Reliability is the likelihood that a product will not fail within a specific time period. This is a key element for users who need the product to work without fail.

Conformance Conformance is the precision with which the product or service meets the specified standards.

Durability Durability measures the length of a product's life. When the product can be repaired, estimating durability is more complicated. The item will be used until it is no longer economical to operate it. This happens when the repair rate and the associated costs increase significantly.

Serviceability Serviceability is the speed with which the product can be put into service when it breaks down, as well as the competence and the behavior of the service person.

Aesthetics Aesthetics is the subjective dimension indicating the kind of response a user has to a product. It represents the individual's personal preference.

Perceived Quality Perceived Quality is the quality attributed to a good or service based on indirect measures.

1.1 Terminology and Concepts

1.1.1 Basic Terminology

Quality Characteristics A.k.a. *Critical to Quality Characteristics* (CTQs). May be physical, sensory, or temporal properties of a process/product. Obviously related to the dimensions of quality.

Quality Engineering "The set of operational, managerial, and engineering activities that a company uses to ensure that the quality characteristics of a product are at the nominal or required levels and that the variability around these desired levels is minimum." (Montgomery, 2007)

Variables Continuous measurements of some CTQ.

Attributes Discrete measurements of some CTQ.

Target Value The desired level of a particular CTQ. A.k.a. *nominal* value.

USL & LSL Largest and smallest allowable values of a CTQ.

Specifications The set of permissible values for all CTQs. Either a set of target values, or USL-LSL intervals.

Non-conformity A non conforming product is one that fails to meet the specification.

Fallout The same as non-conformity.

Defect A non-conformity that is serious enough to affect the use of the product.

DPMO Defect per million opportunities.

PPM Parts per million. Interchangeable with DPMO.

1.1.2 Statistical Terminology

Exploratory Data Analysis (EDA) An assumption free quantitative inspection of data; "Story telling"; no inference.

Inference Data analysis with the intention of generalizing from a sample to a population.

Causal Inference Inference, with the intention of claiming causal relations between quantities under study.

Predictive Analytics Data analysis with the intention of making predictions with future data. Can be seen as inference, without aiming at causality.

Design of experiments (DOE) By far the best and most established way for causal inference. The *random assignment* of units to groups allows to interpret statistical correlations as causal.

Statistical Process Control (SPC) Data analysis with the intention of identifying anomalous behaviour with respect to history[1].

Computer Simulation Well, just what the name implies.

[1] Akin to *anomaly detection*, or *novelty detection*, in the machine learning literature.

Control Chart A graphic visualization of the historical evolution of one (or several) CTQs.

(Un)Controllable Inputs Each process has inputs that affect the behaviour of the CTQ. Some are controllable, and some are not.

Factorial Design In the language of DOE, controllable inputs are *factors*. A factorial design, is an experiment where factors are varied in order to study their effect on the CTQ.

Off/On-line process control SPC can be performed on or off line. On-line, a.k.a. *in-process control*, meaning control happens as the process evolves, and off-line meaning before it starts or after it has finished.

Engineering control A.k.a. *automatic control*, or *feedback control*. SPC that triggers an intervention that keeps the process in control

Outgoing/Ingoing Inspection Refers to the stage at which SPC is performed. As inputs come in (ingoing), or as outputs come out (outgoing).

1.2 Some History

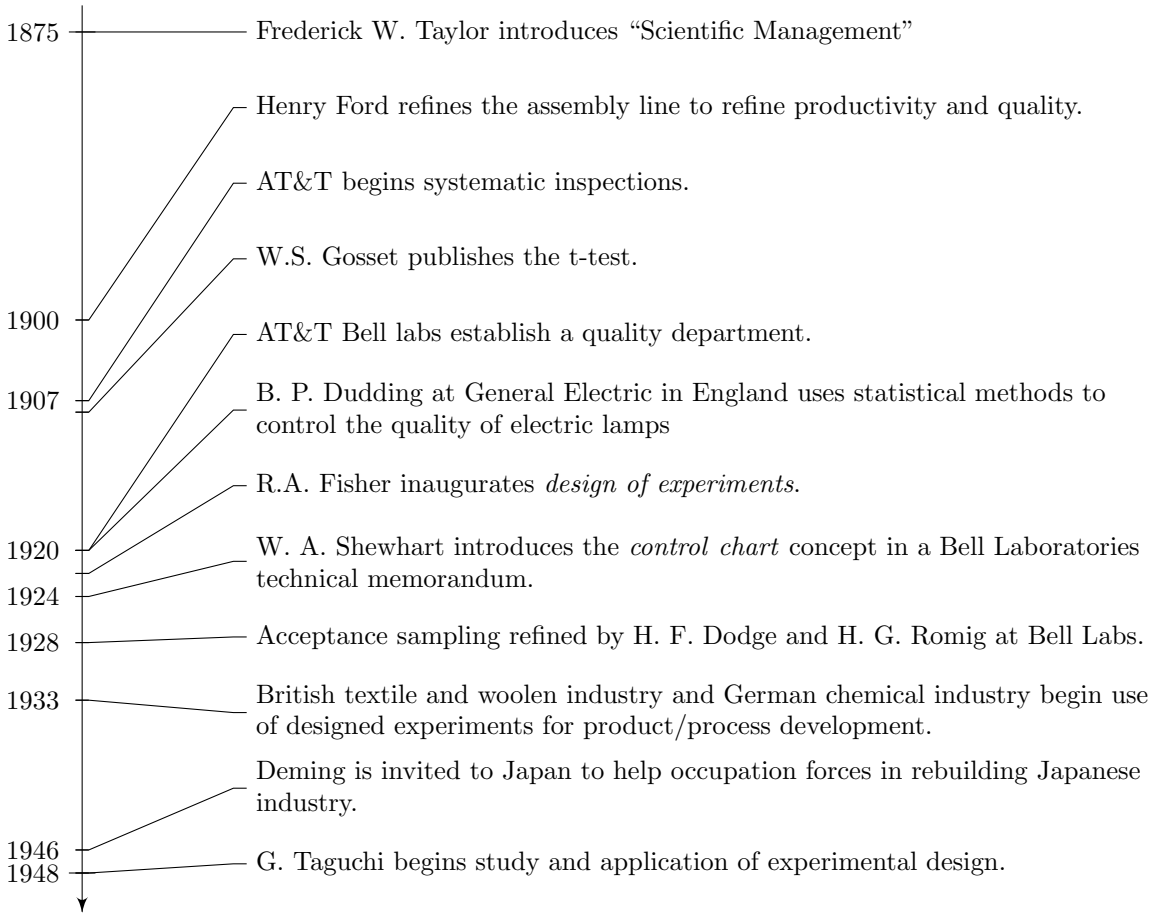

Table 1.1: Adapted from (Montgomery, 2007, Table 1.1).

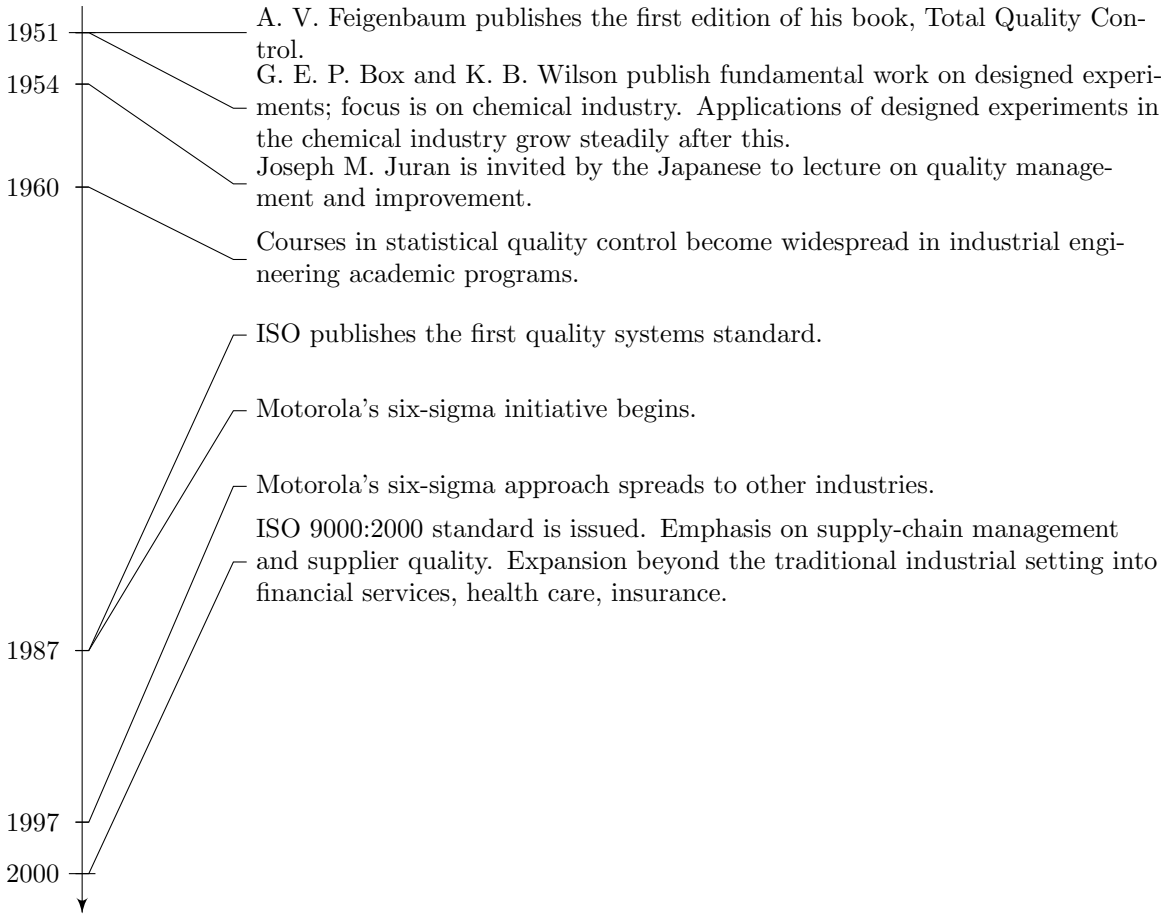

Table 1.2: Adapted from (Montgomery, 2007, Table 1.1).

1.3 Management Aspects of Improving Quality

The founding fathers of QC have many do's-and-don'ts for managers. See Montgomery (2007, Sec 1.4) for details. As usual, we collect recurring ideas:

1. The responsibility for quality rests with management.
2. QC is not a one-time project, but an on-going process. It may advance continuously, or incrementally.
3. QC is (or should be) manifested in organizational structure, training, recruitment, incentives, knowledge management, to name a few.

1.4 Programs and Initiatives

1.4.1 Zero Defects Program (ZD)

Quoting Wikipedia (2015f):

> ... a management-led program to eliminate defects in industrial production that enjoyed brief popularity in American industry from 1964 to the early 1970's. Quality expert Philip Crosby later incorporated it into his "Absolutes of Quality Management" and it enjoyed a renaissance in the American automobile industry, as a performance goal more than as a program, in the 1990s. Although applicable to any type of enterprise, it has been primarily adopted within supply chains wherever large volumes of components are being purchased (common items such as nuts and bolts are good examples).

1.4.2 Quality is Free Initiative

Quoting Montgomery (2007):

> ... in which management worked on identifying the cost of quality (or the cost of *nonquality*, as the Quality is Free devotees so cleverly put it). Indeed, identification of quality costs can be very useful, but the Quality is Free practitioners often had no idea about what to do to actually improve many types of complex industrial processes.

1.4.3 Value Engineering Program (VE)

Quoting Wikipedia (2015e):

> Value engineering (VE) is systematic method to improve the "value" of goods or products and services by using an examination of function. Value, as defined, is the ratio of function to cost. Value can therefore be increased by either improving the function or reducing the cost. It is a primary tenet of value engineering that basic functions be preserved and not be reduced as a consequence of pursuing value improvements.

1.4.4 Total Quality Management (TQM)

TQM originates in the 1980's with the ideas of Deming and Juran. It is a very wide framework that attempts at capturing the company-wide efforts required for QC. According to Montgomery (2007, p.23):

> TQM has only had **moderate success** for a variety of reasons, but frequently because there is insufficient effort devoted to widespread utilization of the technical tools of variability reduction. Many organizations saw the mission of TQM as one of training. Consequently, many TQM efforts engaged in widespread training of the workforce in the philosophy of quality improvement and a few basic methods. This training was usually placed in the hands of human resources departments, and much of it was ineffective. The **trainers often had no real idea about what methods should be taught**, and success was usually measured by the percentage of the workforce that had been "trained," not by whether any measurable impact on business results had been achieved.
>
> ... Another reason for the erratic success of TQM is that many managers and executives have regarded it as **just another "program" to improve quality**. During the 1950's and 1960's, programs such as Zero Defects and Value Engineering abounded, but they had little real impact on quality and productivity improvement.

1.4.5 Six-Sigma

Quoting Montgomery (2007):

> Products with many components typically have many opportunities for failure or defects to occur. Motorola developed the Six-Sigma program in the late 1980s as a response to the demand

for their products. The focus of six-sigma is reducing variability in key product quality characteristics to the level at which failure or defects are extremely unlikely.

Assume a device has m components. The failure probability of component $j \in 1, \ldots, m$ is α_j. What is the probability of the device failing, when assuming independent failures?

$$\begin{aligned} P(\text{failure}) &= P(\text{at least one failure}) \\ &= 1 - P(\text{no failure}) \\ &= 1 - \prod_{j=1}^{m}(1-\alpha_j) \end{aligned} \tag{1.1}$$

Assuming all components have the same fallout rate, we omit the index j in α_j.

The failure probability α is implied by the CTQs, and its specification limits (USL, LSL). Denoting the target value of the CTQ by T, then $USL = T + \delta$ and $LSL = T - \delta$. Three-sigma means that the production variability, σ, is small enough so that

$$3\sigma = \delta.$$

Assuming

$$CTQ \sim \mathcal{N}(T, \sigma),$$

we can compute:

$$\alpha = 1 - P(LSL < CTQ < USL) \tag{1.2}$$
$$= 1 - P(|CTQ| < \delta) \tag{1.3}$$
$$= 1 - P(|CTQ| < 3\sigma) = 0.0027. \tag{1.4}$$

The 3-sigma quality guarantee is also known as $2,700$ defective parts per million (ppm) for now obvious reasons. Plugging the 3-sigma performance in Eq.(1.1) returns PPM

$$P(\text{3-sigma failure}) < 1 - (1 - 0.0027)^m$$

Figure 1.1 illustrates the probability of failure against the number of components. For simple devices, the 3-sigma criterion may suffice. But now imagine the number of components in a car, a cellular phone, ... The 3-sigma rule is just not good enough. This is where 6-sigma requirement comes along. It implies that the production is process is so accurate that

$$6\sigma = \delta.$$

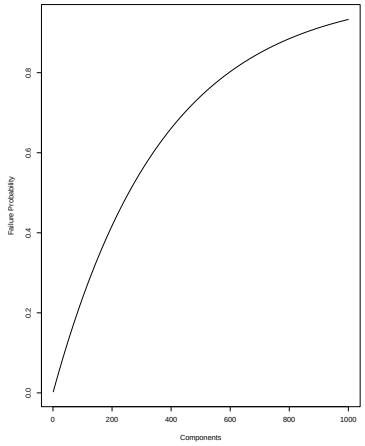

Figure 1.1: The probability of failure as a function of components under the 3-sigma standard.

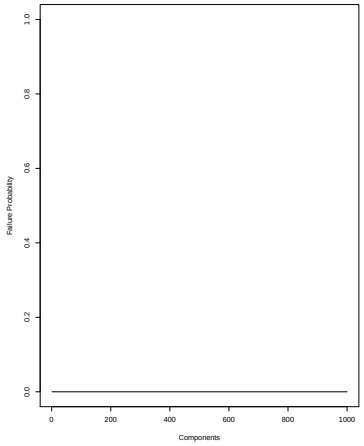

Figure 1.2: The probability of failure as a function of components under the 6-sigma standard.

Updating Eq.(1.2) we get that the defective *ppm* of 6-sigma is 0.002. This is obviously excellent news, except for the typically tremendous effort involved in achieving this level of quality.

According to Montgomery (2007), the 6-sigma methodology has gained more success than it predecessors:

> The reason for the success of six-sigma in organizations outside the traditional manufacturing sphere is that variability is everywhere, and where there is variability, there is an opportunity to improve business results.

1.4.6 Lean Systems

Quoting Wikipedia (2015c) (my own emphasis in bold):

> Essentially, lean is centered on making obvious what **adds value** by **reducing everything else**. Lean manufacturing is a management philosophy derived mostly from the Toyota Production System (TPS) (hence the term Toyotism is also prevalent) and identified as "lean" only in the 1990s.

1.4.7 Design for Six-Sigma (DFSS)

Quoting Wikipedia (2015a) (my own emphasis in bold):

It is based on the use of **statistical tools** like linear regression and enables empirical research similar to that performed in other fields, such as social science. While the tools and order used in Six Sigma require a process to be in place and functioning, DFSS has the objective of **determining the needs of customers** and the business, and driving those needs into the product solution so created. DFSS is relevant for relatively simple items / systems. It is used for product or process design in contrast with process improvement.

1.4.8 Quality Systems and Standards

The first quality standard was issued by the International Standards Organization (ISO) in 1987. Current quality standards are known as the *ISO9000 series*. These include: ISO9000

ISO9000:2000 Quality Management System-Fundamentals and Vocabulary.

ISO9001:2000 Quality Management System-Requirements.

ISO9004:2000 Quality Management System-Guidelines for Performance Improvement.

In Israel, it is the Standards Institute of Israel[2] that may give ISO9000 (like any ISO) certifications upon inspecting the candidate organization. As emphasized by Montgomery (2007, p.24), ISO9000 is a set of rules and best practices, mostly oriented at *knowledge management*. It may help to *preserve* quality, but it does not, nor does it aim to, *improve* quality. As such, it will not be the focus of our course, which will focus on *statistical tools*.

Extra Info 1. [TODO: Just-in-Time, Poka-Yoke]

1.5 DMAIC

There are many names for the process of quantitative re-evaluations of performance against a given target: *data driven decision making* (DDD), *Shewart cycle*, etc. We will focus on one such framework, illustrated in Figure 1.3 known as DMAIC: Define, Measure, Analyze, Improve, Control.

Here are some general observations on DMAIC:

[2] `https://portal.sii.org.il/heb/qualityauth/certificationtypes/qualitylinks/iso9001/`

Figure 1.3: The DMAIC cycle.
`http://www.sapartners.com/sigma-academy/`

1. It is aimed at promoting improvement and creative thinking.
2. It is not part of the six-sigma methodology, but will typically take part in its implementation.

What do the stages of DMAIC mean [3]?

Define the problem, improvement activity, opportunity for improvement, the project goals, and customer (internal and external) requirements.

Measure process performance.

Analyze the process to determine root causes of variation, poor performance (defects).

Improve process performance by addressing and eliminating the root causes.

Control the improved process and future process performance.

In the following chapter we give a set of statistical tools required for *measuring*,*analyzing* and *controlling* a process.

[3] `http://asq.org/learn-about-quality/six-sigma/overview/dmaic.html`

Chapter 2

Exploratory Data Analysis- EDA

In this chapter, we give a short review of methods for *exploratory data analysis* (EDA), a.k.a. *descriptive statistics*. Recall that our goal is an assumptions-free description of our data. EDA thus consist of computing interpretable summaries of the data, called *summary statistics*, and visualizations.

Descriptive Statistics

2.1 Summary Statistics

We now distinguish between summary statistics that apply to attributes, categorical by definition, and variables, continuous by definition.

2.1.1 Summarizing Categorical Data

Univariate

Summarizing a vector of categorical data can naturally be done by tabulating it, i.e., computing the frequency and relative frequency of each category. Clearly averages, medians, and the likes are incomputable, since categorical data has no ordering, nor does it admit simple operations such as summation.

Extra Info 2. Variability of categorical data can clearly not be measured by its variance, since it does not admit a summation operation. It is, however, possible to define different measures of variability that do apply. The *entropy* is such an example.

Entropy

Bivariate

Generalizing the univariate case to bivariate, or multivariate, one can keep tabulating. I.e., compute the frequency, and relative frequency, of combinations of categories.

2.1.2 Summarizing Continuous Data

Continuous variables admit many more mathematical manipulations than categorical attributes.

Univariate

We start by presenting the most natural summaries of the data. Without going into the formal definition, we refer to them as *summary of location*. These include:

Location Summaries

Definition 1 (The Mean)**.** The *mean*, or *average*, is defined as

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

Definition 2 (The Median)**.** The median is the observation that is smaller than half of the sample and larger than half of the sample.

Definition 3 (α-Trimmed Mean)**.** The α-trimmed mean is the average of the observations left after ignoring the largest and the smallest $(100\alpha)\%$ of them.

The naïve average is the 0-trimmed mean, and the median is the 0.5-trimmed mean.

From summaries of location, we move to summaries of *scale*.

Summary of Scale

Definition 4 (The Standard Deviation)**.**

$$s(x) := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.2}$$

For the following, we require the definition of the sample quantiles, themselves **not** a scale summary.

Definition 5 (α Quantile)**.** The α-quantile of the sample is the observation that is larger than $(100\alpha)\%$, and smaller then $(100(1-\alpha))\%$ of the sample.

The empirical maximum and minimum are then $x_{1.0}$ and $x_{0.0}$, respectively.

Definition 6 (The Range)**.**

$$Range(x) := \max_i\{x_i\} - \min_i\{x_i\} = x_{1.0} - x_{0.0} \tag{2.3}$$

Definition 7 (The Inter Quantile Range- IQR)**.**

$$IQR(x) := x_{0.75} - x_{0.25} \tag{2.4}$$

Definition 8 (The Median Absolute Deviation- MAD)**.**

$$MAD(x) := \{|x_i - x_{0.5}|\}_{0.5} \tag{2.5}$$

Note that the MAD may be sometimes scaled by 1.4826, so that it estimates σ. Such is the behaviour of the **R** function `mad()`.

After summaries of scale, we move to summaries of *skewness*, or *asymmetry*.

Definition 9 (Yule Skewness Measure)**.**

$$YULE(x) := \frac{\frac{1}{2}(x_{0.75} + x_{0.25}) - x_{0.5}}{IQR(x)} \tag{2.6}$$

Bivariate

From univariate data x, we move to bivariate x, y. Clearly we can apply univariate summaries component-wise. We want, however, to summarize the *joint* behaviour of the data. For this purpose, we assume that data comes in pairs, implying that x and y are of same length.

Definition 10 (Covariance)**.** The sample covariance, or *empirical* covariance is defined as

$$Cov(x, y) := \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n - 1} \tag{2.7}$$

Definition 11 (Pearson's Correlation Coefficient)**.** *Pearson's Correlation Coefficient*, or *Pearson's Moment Product Correlation Coefficient*, is defined as

$$r(x, y) := \frac{(n-1)Cov(x, y)}{S(x)S(y)} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{S(x)S(y)} \tag{2.8}$$

We can dwell into the meaning and intuition underlying Pearson's correlation coefficient, but we will not. The curious reader is reffered to Rodgers and Nicewander (1988).

The next measure of association captures a more general association.

Definition 12 (Spearman's Correlation Coefficient)**.** Spearman's correlation coefficient is merely Pearson's correlation coefficient computed on the *ranks* of x and y.

We conclude by noting that *regression coefficients* are also a measure of association.

Multivariate Data

Multivariate data, both continuous (variables), and discrete (attributes), admits a vast realm of method for summary and visualization. Clearly, associations between several variables can be very complicated so that the more we try to summarize, the more information we give up. On the other hand, and unlike the univariate and bivariate case, our minds will need some type of simplification since they cannot grasp the raw data (did you ever try to imagine how $\mathbb{R}^4$ looks like?). As usual, we emphasize that our purpose is to summarize the joint association in the data. For component-wise summaries, we can always apply the univariate summaries one variable at a time.

By far the most popular measures of joint association are the covariance matrix and correlation matrix.

Definition 13 (Covariance Matrix)**.** For a multivariate data consisting of $x_1, \ldots, x_p$ vectors, each with n entries: $x_{i,1}, \ldots, x_{i,n}$, we define the (sample) covariance matrix to be a $p \times p$ matrix whose elements are the (sample) covariances between corresponding vectors:

$$\hat{\Sigma}_{i,j} := Cov(x_i, x_j) \tag{2.9}$$

Extra Info 3 (Sample Covariance Matrix)**.** The matrix $\hat{\Sigma}$ has many useful properties. The curious reader is referred to Petersen and Pedersen (2006), and references therein, for more details.

Definition 14 (Correlation Matrix)**.** For a multivariate data consisting of $x_1, \ldots, x_p$ vectors, each with n entries: $x_{j,1}, \ldots, x_{j,n}$, we define the (sample) correlation matrix to be a $p \times p$ matrix whose elements are the (Pearson) correlations between corresponding vectors:

$$\hat{R}_{i,j} := r(x_i, x_j) \tag{2.10}$$

Extra Info 4 (Multivariate Data Analysis)**.** Multivariate analysis is an important, and very actively studied field in statistics and machine learning. A non-comprehensive list of methods that belong to this realm include Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Factor Analysis (FA), Independent Component Analysis (ICA), Dimensionality Reduction, Manifold Learning, Self Organizing Maps, etc. Ask me for reference books or courses if this topic interests you.

PCA, SVD,ICA

2.2 Visualization

2.2.1 Visualizing Categorical Data

Univariate

Much like computing summaries, there is not much to be said about visualizing univariate categorical variables. The most natural, and perhaps only visualization, is the *bar plot*, illustrated in Figure 2.1.

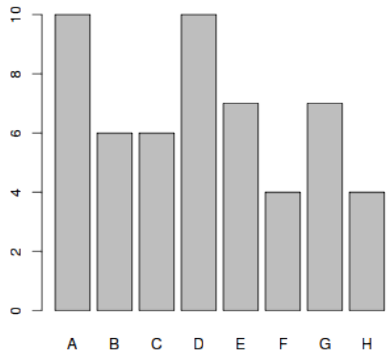

Figure 2.1: The Bar-Plot.
`http://www.r-tutor.com/elementary-statistics/qualitative-data/bar-graph`

Bivariate

Visualizing a two-way cross-table can be done using an extension of the bar-plot. Several extensions exist. By far, the most informative and recommended figure, in this author's view, is the *mosaic plot*, illustrated in Figure 2.2.

Mosaic Plot

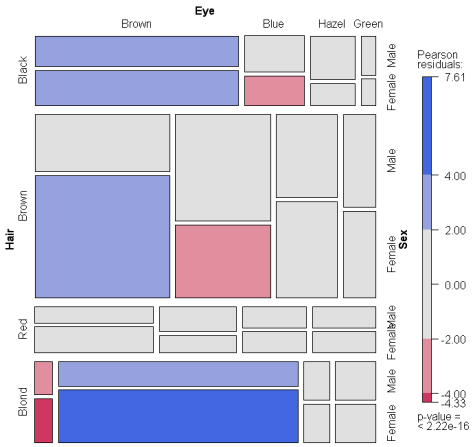

Figure 2.2: Mosaic Plot.
`http://www.statmethods.net/advgraphs/mosaic.html`

2.2.2 Visualizing Continuous Data

Univariate

Visualization of univariate continuous vectors can present the raw data, or it distribution (i.e.- discarding the indexes). The most basic visualizations are the *dotchart*, *histogram*, *boxplot*, *stem-and-leaf plot*. These are illustrated in figures 2.3, 2.4, 2.5, 2.6 respectively.

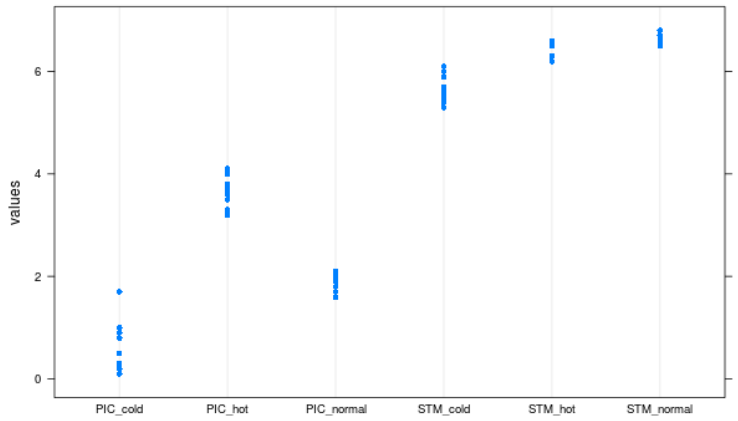

Figure 2.3: Dot Plot.
`http://stackoverflow.com/questions/15109822/r-creating-scatter-plot-from-data-frame`

Bivariate

The simultaneous visualization of two continuous variables, can naturally be done with a *scatter plot*. More sophisticated visualization, which generalizes the histogram into two dimensions, is the *hexbin plot*. These are illustrated in figures 2.7, and 2.8, respectively.

Multivariate Data

Since we cannot possibly visualize data in more than 3-dimensions, and we clearly prefer data in 1 or 2 dimensions, the visualization of multivariate data will typically consist of summarizing the data into $1D$ or $2D$, and then applying the above mentioned visualization techniques.

An important exception is due to the observation that a computer image, is essentially a matrix. We can thus visualize matrices, with a simple

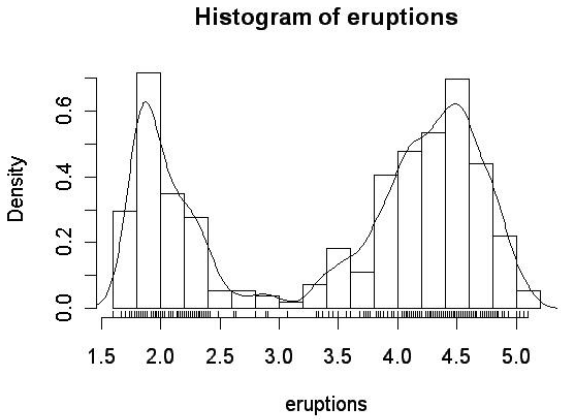

Figure 2.4: Histogram.
`http://compbio.pbworks.com/w/page/16252882/Basic`

image, and in particular, covariance and correlation matrices, as illustrated in Figure 2.9.

A second exception is when the data has both continous variables and discrete attributes. Endlessly many combinations are then possible. The author strongly recommends to visit Hans Rosling's *Gap Minder* at `http://www.gapminder.org/world` for an excellent interactive visualization.

Gap Minder

2.2.3 On-Line Visualization

For the purpose of quality control, we may often want an *on-line* visualization, and not *off-line*, as the ones previously discussed. This is the purpose of *dashboards*, illustrated in Figure 2.10.

Dash-board

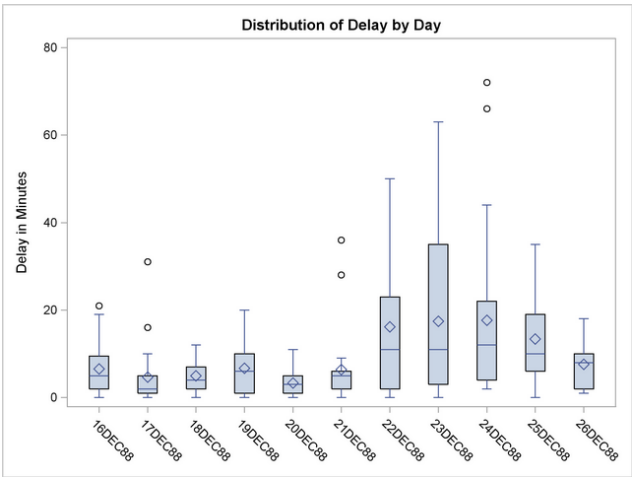

Figure 2.5: Boxplot.
`http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm`

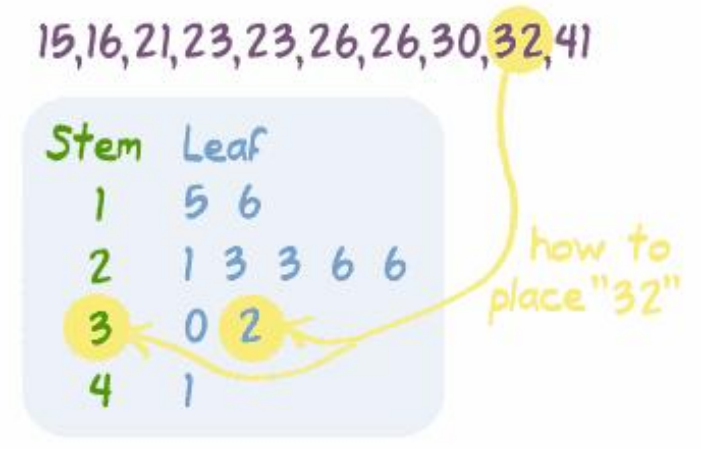

Figure 2.6: Stem-and-leaf plot.
`https://www.mathsisfun.com/data/stem-leaf-plots.html`

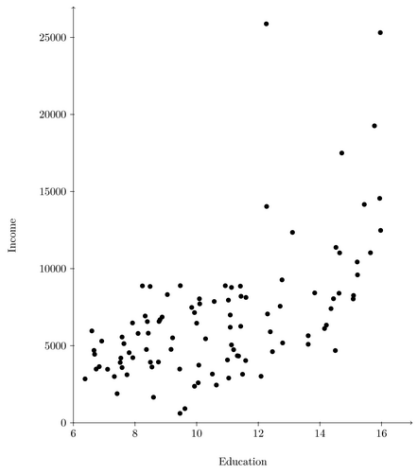

Figure 2.7: Scatter Plot.
`http://texample.net/tikz/examples/scatterplot/`

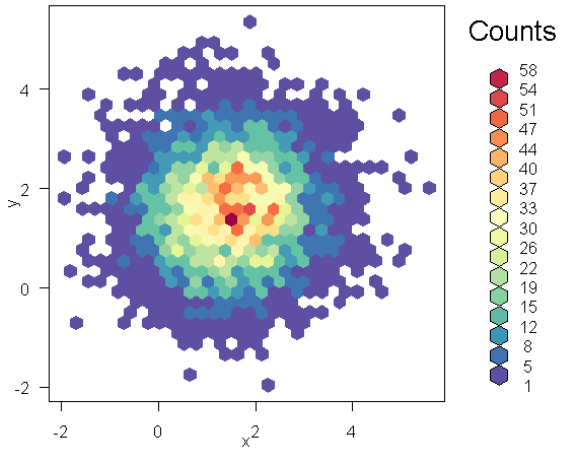

Figure 2.8: HexBin plot. Sorce: `http://www.r-bloggers.com/5-ways-to-do-2d-histograms-in-r/`

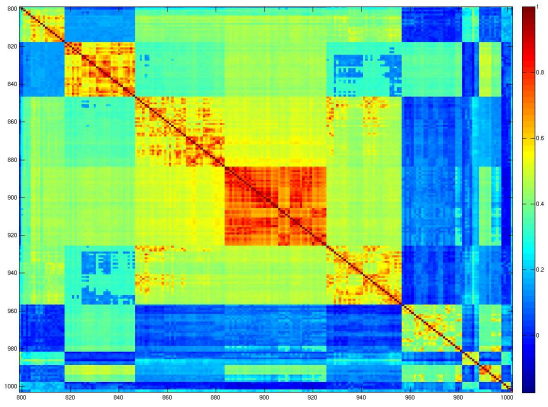

Figure 2.9: Image of covariance matrix.
`http://cs.brown.edu/courses/csci1950-g/results/final/sghosh/`

Figure 2.10: Dashboard.
`http://www.iconics.com/Home/Products/AnalytiX/Quality-AnalytiX.aspx`

Chapter 3

Statistical Inference

The idea of extrapolating knowledge from a *sample* to a population is known as *statistical inference*. It encompasses the ideas of *parameter estimation*, *confidence intervals*, and *hypothesis testing*. We will assume the reader is familiar with these ideas, but recall some required terminology. The QC and SPC terminology are not always consistent with statisticians' terminology. When new names are given to old ideas, we will emphasize this in the text.

Null/Alternative Hypothesis Some statement about the world we wish to test with data. The frequentist argument follows a Popperian philosophy: "to show the alternative hypothesis is true, I will show that the null hypothesis is not true".

Statistical Test The procedure of inferring from data on the truthfulness of the alternative hypothesis.

Assumptions As the name suggests, these are assumptions. We stress that unlike hypothesis, assumptions are not being tested in a statistical test.

Test Statistic The function of the data to be computed for the purpose of inference. As such, it is a random variable.

Null/Alternative Distribution The distribution of the test statistic under the null/alternative hypothesis.

Type I/II error See Figure 3.1.

False/True Positive/Negative See Figure 3.1.

Rejection Region The collection of event that will lead us to reject the null hypothesis, and believe in the alternative hypothesis.

	Null hypothesis (H_0) is true	Null hypothesis (H_0) is false
Reject null hypothesis	Type I error **False positive**	Correct outcome True positive
Fail to reject null hypothesis	Correct outcome True negative	Type II error **False negative**

Figure 3.1: Type I/II error confusion table.
`https://infocus.emc.com/william_schmarzo/beware-of-false-positives/`

p-value A.k.a. *observed significance*. The null probability of the observed, or "more extreme" events.

Significance Level A.k.a. α. The probability of a false positive.

Power The probability of a true positive.

i.i.d. "Independent and identically distributed" (i.i.d.) is an assumption made on the sampling distribution, meaning that samples are statistically independent, and all originating from the same distribution.

The following sections of this chapter present particular statistical inference methods we will be using in the following chapters.

3.1 Goodness of Fit- GOF

Goodness of fit (GOF) deals with the inference on the sampling distribution, a.k.a., the generative process. It can be approached via rigorous hypothesis testing, or by visualizations.

3.1.1 QQplot and QQnorm

The fundamental idea of the *quantile-quantile plot* (QQplot) is to compare the empirical quantiles in the sample, to the theoretical quantiles implied by the assumed distribution. If the theory and observations agree, we conclude our assumptions are plausible. For the particular case of testing the normality of the data, the corresponding QQplot is known as a *QQnorm plot*.

Figure 3.2 illustrates a QQnorm plot of normal distributed data, while Figure 3.3 is the same for non-normal data.

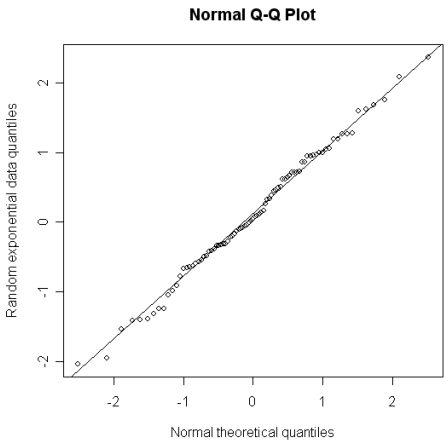

Figure 3.2: A QQplot of Gaussian distributed data.

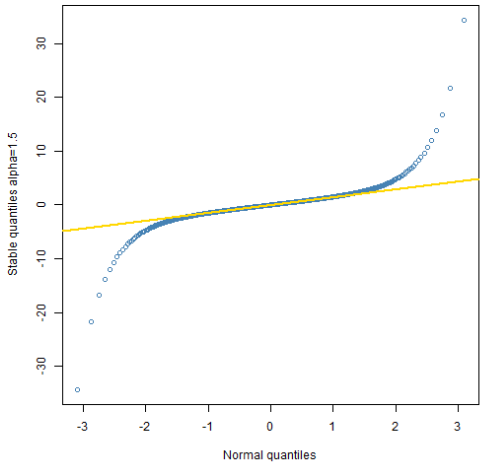

Figure 3.3: A QQnorm plot of non Gaussian distributed data.

3.1.2 Chi-Square GOF Test

The Chi-Square GOF test (not to be confused with the Chi-Square independence test), tests a hypothesis on the sampling distribution of discrete (attributes) data. Note that it is very general, since all continuous variables may be discretized, simply by binning.

Definition 15 (Chi-Square GOF Test)**.** Assume an i.i.d. sample $x_1, \ldots, x_n$. The Chi-Square GOF is a tests $H_0 : \mathbf{x}_i \sim P$ versus $H_1 : \mathbf{x}_i \not\sim P$. P is assumed to be discrete with K categories and $p_k := P(\mathbf{x}_i \in k)$. The test statistic, X^2, is defined as

$$X^2 := \sum_{k=1}^{K} \frac{(obs_k - exp_k)^2}{exp_k}, \tag{3.1}$$

where $obs_k := \#\{x_i \in k\}$, and $exp_k := p_k n$. The approximate null distribution of X^2 is χ^2_{K-1}.

3.1.3 Kolmogorov–Smirnov GOF Test

The Kolmogorov-Smirnof GOF test, tests a hypothesis on the sampling distribution of continuous (variable) data.

Definition 16 (Kolmogorov–Smirnov GOF Test)**.** Assume an i.i.d. sample $x_1, \ldots, x_n$. The Chi-Square GOF is a tests $H_0 : \mathbf{x}_i \sim P$ versus $H_1 : \mathbf{x}_i \not\sim P$. P is assumed to be continuous. The test statistic, D, is depicted in Figure 3.4.

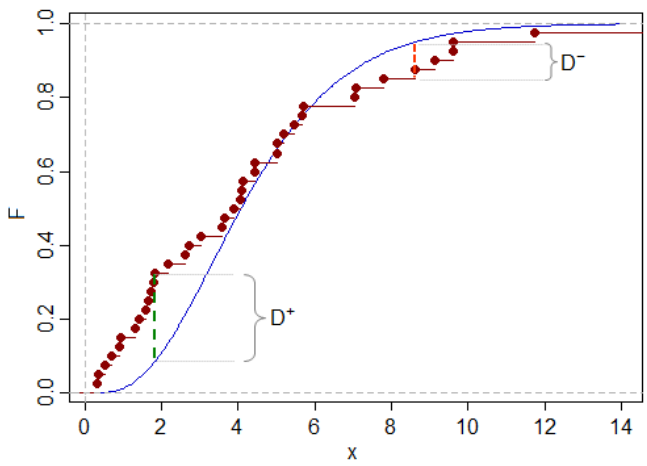

Figure 3.4: Kolmogorov-Smirnov Test. The D statistic is D_+ in the figure.

The null distribution of D is the *Kolmogorov distribution* obtained from tables.

Kolmogorov Distribution

Extra Info 5 (GOF tests)**.** There are endlessly many more GOF tests, such as Anderson-Darling, Jarque–Bera, Shapiro–Wilk, Kuiper's test, etc. Wikipedia is a good place for further reading.

Chapter 4

System Capability Analysis

In a system capability analysis, we essentially use statistical tools to measure the variability in a production process. This analysis can answer questions that are raised at the measuring, analyzing, and improving stages of the DMAIC cycle. The methods we will discus compare between the processes variability and specification. Note, however, that a simple statistical analysis of the production process, without relating to specification, may also be qualify as a system capability analysis.

We naturally want production processes that adhere to specifications, and want to quantify the level of adherence. The quantification is performed by comparing the variability in a CTQ to the specification. In this chapter we assume the process's capability is fixed over time. In Chapter 9 we revisit the same problems, when allowing the process's capability to vary over time.

The particular setup discusses in this chapter is also known as *product specification*. When we use the actual time, and ordering of data samples, as in a control chart, we will no longer regard it as product specification but rather as a bona-fide capability analysis. This is the subject of Chapter 9.

Product Specification

Because capability analysis, or product specification, is essentially the study of the CTQ's distribution, it can be approached with the aforementioned statistical tools such as univariate summary statistics and visualizations presented in Sections 2.1 and 2.2. If particular hypothesis want to be tested on the distribution of the CTQ, we may call upon the inference tools from Chapter 3.

4.1 Process Capacity Indexes

Classical statistical devices are not aimed at introducing the designed process's capabilities. We will thus introduce several measures that do so, col-

lectively called *process capability ratios* (PCR), or *process capability index*. The first, and most basic PCR is the C_p of a particular CTQ.

Process Capability Index

Definition 17 (C_p)**.**

$$C_p := \frac{USL - LSL}{6\sigma}, \tag{4.1}$$

where σ is the standard deviation of the CTQ.

Clearly C_p is a process parameter, that needs to be estimated.

Definition 18 ($\hat{C}_p$)**.**

$$\hat{C}_p := \frac{USL - LSL}{6\hat{\sigma}}, \tag{4.2}$$

where $\hat{\sigma}$ is some estimate of the standard deviation of the CTQ.

The most natural $\hat{\sigma}$ is the sample standard deviation s, but we will explore other options in the following.

There is a relation between C_p and the probability of non-conformance. To explore this relation we introduce the following notation, under the assumption of a symmetric specification about its target value.

Collecting Notation
$T := (USL + LSL)/2$, the target value.
$\delta := (USL - LSL)/2$.
$\mu := \mathbf{E}[CTQ]$, the expected CTQ.
$1 - p_{NC} := P(CTQ \in [LSL, USL])$, the probability of compliance, thus p_{NC} is the probability of non compliance.

With our new notation Eq.(4.1) is now $C_p = \frac{\delta}{3\sigma}$. Assuming $CTQ \sim \mathcal{N}(\mu, \sigma^2)$, and that the process is centred so that $\mu = T$, then C_p is related to p_{NC} via

$$p_{NC} = 2\,\Phi(-3C_p) \tag{4.3}$$

As a sanity check, we check this relation for a 3-sigma process. A 3-sigma process implies that $C_p = 1$, and Eq.(4.3), implies that $p_{NC} = 0.0027$, as we have already seen in six-sigma introduction in Section 1.4.5.

Montgomery (2007) recommends the following $\hat{C}_p$ values:

	$\hat{C}_p$ Value	Implied ppm
Existing processes	1.33	66
New processes	1.50	6.8
Safety, strength, or critical parameter, existing process	1.50	6.8
Safety, strength, or critical parameter, new process	1.67	0.5
Six Sigma quality process	2.00	0.002

To derive Eq.(4.3), and thus the ppm column in the table, we had to call upon several assumptions. Namely:

1. The process is in statistical control, i.e. μ is fixed over time
2. The CTQ has a normal distribution.
3. The process is centred, i.e. $\mu = T$.
4. T is mid-way between LSL and USL.

4.1.1 Non-Conformance for a Non-Gaussian CTQ

The first assumption we will now relax is the Gaussianity of CTQ.

Theorem 4.1.1 (Chebyshev's inequality)**.** *For any random variable* $\mathbf{x}$*, with* $\mu := \mathbf{E}[\mathbf{x}]$*, and* $\sigma^2 := \mathbf{E}[(\mathbf{x}-\mu)^2]$*, then*

$$P(|\mathbf{x}-\mu| > k\sigma) = \frac{1}{k^2} \tag{4.4}$$

An immediate application of Chebyshev's inequality yields that for *any* distribution of the CTQ, then $p_{NC} < 0.11$. This is nice, but far from amazing, since it means that a 3-sigma process, assumingly with $2,700ppm$, may actually have $111,111ppm$, if we were wrong about the Gaussianity assumption. We now present a similar bound on p_{NC} that applies to all *unimodal* CTQ distributions.

Theorem 4.1.2 (Vysochanskij–Petunin inequality)**.** *For any random variable* $\mathbf{x}$*, with* $\mu := \mathbf{E}[\mathbf{x}]$*,* $\sigma^2 := \mathbf{E}[(\mathbf{x}-\mu)^2]$*, with unimodal distribution, and for* $k > \sqrt{8/3} = 1.63299$*, then*

$$P(|\mathbf{x}-\mu| > k\sigma) = \frac{4}{9k^2} \tag{4.5}$$

Since 3 is obviously larger than 1.63299, then an immediate application of Vysochanskij–Petunin inequality yields that for a unimodal CTQ, then $p_{NC} < 0.0493$, or $49,383ppm$.

Again we see that if we wrongly assumed normality, we may have many more non-compliances than we thought. Then again, both Chebyshev and Vysochanskij–Petunin are very loose bounds. They should thus be seen as a worst-case, while reality is not as bad.

Another important approach to dealing with non-Gaussianity of the CTQ is by a non-linear change of scale. As a rule of thumb, you may try scale

transformation such as a log-transform and a square-root, and then inspect the normality of the data with the tools discusses in Chapters 2 and 3.

We now consider $\hat{C}_p$, i.e., the estimation of C_p. For a Gaussian CTQ, then σ can naturally be estimated by s. If, however, the Gaussianity assumption does not hold, even if the process is under control, then s may be a poor estimator of σ. We may now recall that in Chapter 2 we encountered several measure of scale such as the IQR (Definition 7), and MAD (Definition 8). Since s is essentially a measure of scale, we may replace s by the IQR, or the MAD, for example, and still retain a valid estimator of C_p.

4.1.2 Process Capability of a non-centred Process

We will now relax the assumption of $\mu = T$, while still in statistical control.

Definition 19 (C_{pk})**.**

$$C_{pk} := \min\{C_{pu}, C_{pl}\} \tag{4.6}$$

where $C_{pu} := \frac{USL-\mu}{3\sigma}$ and $C_{pl} := \frac{\mu-LSL}{3\sigma}$.

For a non-centred process, this definition is more informative on the probability of non-compliance. Indeed, Eq.(4.3) will not hold, but we can derive an updated version:

$$p_{NC} \approx \Phi(-3C_{pk})[TODO : fix] \tag{4.7}$$

Generally, $C_{pk} \leq C_p$, with equality holding for centred processes ($\mu = T$). An illustration is given in Figure 4.1.

The C_{pk} index is motivated by conserving the relation between the index and the fallout rate p_{NC}. Another index, known as C_{pm}, or the *Taguchi* capability index, deals with the non centring slightly differently. It is motivated by the observation that for $\mu = T$, then $\sigma = \sqrt{\mathbf{E}\left[(CTQ - T)^2\right]}$. This relation does not hold if $\mu \neq T$, which leads us to the following definition:

Definition 20 (C_{pm})**.**

$$C_{pm} := \frac{USL - LSL}{6\sqrt{\mathbf{E}\left[(CTQ - T)^2\right]}} \tag{4.8}$$

$$= \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \tag{4.9}$$

$$= \frac{C_p}{\sqrt{1 + (\frac{\mu-T}{\sigma})^2}}. \tag{4.10}$$

Eq.(4.10) readily shows that just like the C_{pk}, then $C_{pm} < C_p$.

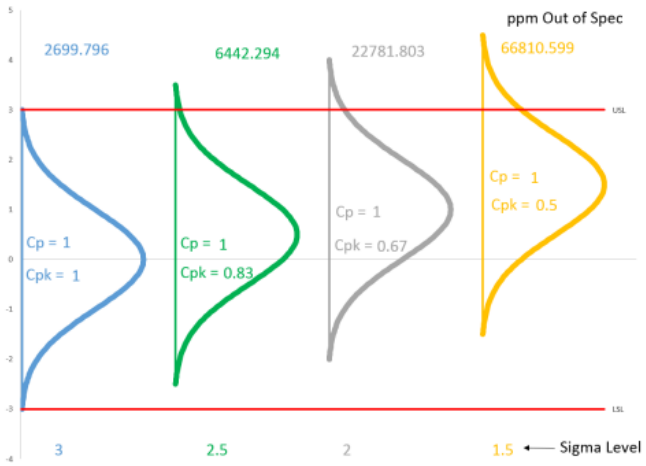

Figure 4.1: C_{pk} and C_p.
`https://www.spcforexcel.com/knowledge/process-capability/interactive-look-process-capability`.

4.1.3 Interval Estimation for Capability Indexes

Since the various process capability indexes are merely population parameters that depend on specifications, we can also construct confidence intervals (CIs) for them, which may be very informative if sample sizes are small. The simplest case is that of C_p. Being a monotone transformation of σ, we can call upon confidence intervals for the variance of a normal population, so that with probability $1-\alpha$:

$$C_p \in \left[\hat{C}_p\sqrt{\frac{\chi^2_{\alpha/2,n-1}}{n-1}}, \hat{C}_p\sqrt{\frac{\chi^2_{1-\alpha/2,n-1}}{n-1}}\right] \tag{4.11}$$

Intervals for the other capability indexes, are available in Montgomery (2007) and references therein.

4.1.4 Testing Hypotheses on Capability

Consider a supply contract, which requires production to have $C_p > 1.5$. It may easily be the case, that the $C_p > 1.5$, even if $\hat{C}_p < 1.5$, especially if the sample size is small. It thus makes a lot of sense, to design hypothesis tests on process capabilities. We observe that for an i.i.d. sample from a Gaussian

population, where $\hat{C}_p$ is estimated with s, then

$$(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 \sim \chi^2_{n-1} \tag{4.12}$$

so that $(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2$ may serve as test statistic.

Example 1 (C_p test for 6-sigma compliance)**.**

$$H_0 : C_p \leq 2$$
$$H_1 : C_p > 2$$
$$(n-1)\left(\frac{2}{\hat{C}_p}\right)^2 \overset{H_0}{\sim} \chi^2_{n-1}$$

so that the $1-\alpha$ rejection region in $\hat{C}_p$ scale is

$$\hat{C}_p > \sqrt{\frac{4(n-1)}{\chi^2_{n-1,1-\alpha}}}.$$

Note that we should not be testing this hypothesis with the confidence interval in Eq.(4.11) because this particular hypothesis is directional.

4.1.5 Process Performance Indices

Process *performance* indices measure compliance to specification of a process out of statistical control. These include the P_p and P_{pk} indices. Besides mentioning their existence, we will not give them further attention, since we adopt Montgomery (2007)'s view that their use is strongly discouraged.

Chapter 5

Statistical Process Control-SPC

Statistical process control, a.k.a. *change detection algorithms* deals with the quantitative analysis of a "process", which may be a production line, a service, or any other repeated operation. As such, it may be found in the Analyze, Improve, and Control stages of the DMAIC cycle. The purpose of the SPC, in the terms coined by Shewart, is to seperate the variability in the process into *assignable* causes of variation and *chance* causes of variation. These are also known as *special* and *common* causes of variation, respectively. A process is said to be in *statistical control* if all its variation is attributable to chance causes. If this is not the case, we call it *out of control* and we will seek the assignable causes, remove then, and re-analyze.

Change Detection

Causes of variation

All the previously mentioned statistical tools may be called upon for this analysis. In the context of process control, a subset of tools has gained the nick-name "The Magnificent Seven". These include:

Histogram and stem-and-leaf plot. As described in Chapter 2.

Check Sheet. [TODO: add figure]

Pareto chart. An ordered bar plot of the events due to the various assignable variability causes. [TODO: add figure]

Cause-and-effect diagram. A visualization of candidate assignable variability causes. [TODO: add figure]

Defect concentration diagram. A visual inspection of the location of defects on the product.

Scatter plot. As described in Chapter 2.

Control chart. A powerful analysis tool to which we devote the rest of this chapter.

Extra Info 6 (A more rigorous treatment)**.** The contents of this chapter is mostly derived from Montgomery (2007). For a more mathematically rigorous treatment of the topic see Basseville et al. (1993). For an **R** oriented exposition of the topic, see Qiu (2013).

5.1 A soft start. The $\bar{x}$-chart

We demonstrate the concepts and utility of Control Charts with the simplest, yet most popular of them all, the $\bar{x}$-chart . The chart borrows its name from the fact that it is essentially a visualization of the time evolution of the average ($\bar{x}$) of the CTQ of a sample of products. The chart is also augmented with visual aids that help in determining if the process is *in control*, i.e., if it consistent with its own history. Process capability analysis may benefit from the ideas of control charts. We emphasize however, that control charts have no information on the specifications of the process, merely on its own history.

An illustration of a $\bar{x}$-chart is given in Figure 5.1. The ingredients of this chart is the centerline, the control limits, and $\bar{x}$ evolving in time. If at each period $t = 1, \ldots, \tau$, we compute the average of n samples, we denote $\bar{x}_t := 1/n \sum_{i=1}^{n} x_{it}$.

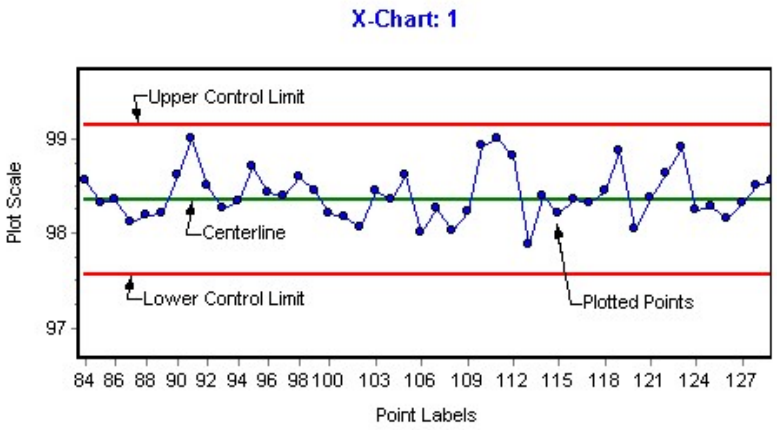

Figure 5.1: $\bar{x}$-chart .
`https://mvpprograms.com/help/P-mvpstats/spc/WhatAreControlCharts`

Figure 5.1 makes it evident $\bar{x}$-chart requires several design decisions. A standard design decision is setting the center line as the grand average of the process:

$$1/\tau \sum_{t=1}^{\tau} \bar{x}_t \tag{5.1}$$

If it is unclear to you, how may we compute the grand average of a process that is still evolving and has not finished, you are right! We thus introduce the ideas of *Phase I* and *Phase II*. Initially we assume the process it out of control, we identify and remove assignable causes of variation, until we are left with a "well-behaved" subset of data points. We call this Phase I, and we use it to initialize required quantities such as the centre line. Eq.(5.1) thus implies that in Phase I we were left with τ samples assumingly in statistical control. After the chart has been calibrated, and major assignable sources of variability removed, we can finally start monitoring the process, known as Phase II.

Phase I/II

Other design decisions to be made are:

1. UCL and LCL [1].
2. Sample size in each sample.
3. Rational groupings (within-period sampling scheme).
4. Frequency of samples (between-period sampling scheme).
5. Other stopping rules.

These deign decisions ultimately govern the error rate (false positive rate) and power (false negative rate) of the chart, which in turn, incur some financial costs. For now we will restrict attention to type I/II error rates, until Section 5.4 where we consider these choices as economical optimization problems.

For ease of exposition, control chart design is demonstrated for the $\bar{x}$-chart , but equally apply to other control charts. We start by a type I error rate analysis. Denote α_t the false alarm probability at period t. How do our

[1] Do not confuse with USL and LSL!

design choices affect α_t?

$$\begin{aligned}\alpha_t &:= 1 - P_{H_0}(\bar{x}_t \in [UCL, LCL]) && (5.2)\\ &= 2P_{H_0}(\bar{x}_t < UCL) && (5.3)\\ &= 2P_{H_0}(Z < \frac{UCL - \mu}{\sigma_{\bar{x}}}) && (5.4)\\ &= 2P_{H_0}(Z < -k) && (5.5)\\ &= 2\Phi(-k) && (5.6)\end{aligned}$$

The above follows from assuming that $UCL := \mu + k\sigma_{\bar{x}}, LCL := \mu - k\sigma_{\bar{x}}$, $\mathbf{x}_{it} \sim \mathcal{N}(\mu, \sigma^2)$, and denoting $\sigma_{\bar{x}} := \frac{\sigma}{n}$. A typical design choice is $k = 3$, known as *3-sigma control limits*, implying a false alarm rate of $\alpha_t = 0.0027$. Since we assumed the process is fixed over time, then so is α_t and we can simply write $\alpha_t = \alpha$.

3-Sigma Control Limits

A power analysis for our design choices follows the same lines. Denote β_t, and $\pi_t = 1 - \beta_t$ the type II error rate, and power, at period t. We then have

$$\pi_t := 1 - P_{H_1}(\bar{x}_t \in [UCL, LCL]) \quad (5.7)$$

and the rest follow from the distribution of $\bar{x}$ when the process is out of control. Assuming the out-of-control process is a shift of magnitude $k\sigma$, i.e.: $\mathbf{x} \sim \mathcal{N}(\mu_1, \sigma^2); \mu_1 = k\sigma$, we plot in Figure 5.2, the detection power of a 3-sigma chart, as a function of k. This is known as statistical literature as a *power function*, and in the engineering literature as the *true positive rate* operator characteristic (TPR-OR.)

ARL Another important related quantity is the *average run length* (ARL), which is the expected number of periods between two crossings of control limits. We denote by ARL_0 the average run length when the process is under statistical control, and ARL_1 otherwise[2]. If $\bar{x}_t$ are statistically independent, then clearly the number of periods until a crossing is geometrically distributed. Using the expectation of a geometric random variable we can conclude that

ARL

$$\begin{aligned}ARL_0 &= 1/\alpha && (5.8)\\ ARL_1 &= 1/\pi && (5.9)\end{aligned}$$

Clearly we can convert to time units by multiplying the ARL by the duration of sampling interval. This is known as the *average time to signal* (ATS).

ATS

[2]Note that it is implied that the process has a *stable* distribution, even though it is out if control.

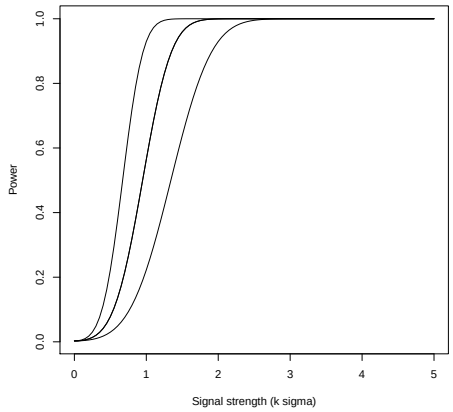

Figure 5.2: Power function of the 3-sigma $\bar{x}$-chart with $n = 5, 10, 20$ and $\mu_1 = k\sigma$.

Now assume that we are unhappy with some control chart. It simply makes too many false alarms, or takes too long to detect loss of statistical control. What can we do about it? Well, this is exactly the same question as when increasing the power or lowering the type I error of a statistical hypothesis test. This is obviously no coincidence, since control charts are nothing but a statistical test! Here are some action courses:

1. Increase k. This is the same as shrinking a rejection region: it will decrease the false alarm rate, at the cost of statistical sensitivity.

2. Increase n. Brilliant! Statistically, there is nothing to lose. It may, however, cost time and money.

3. Increase the sampling frequency. Brilliant again! Nothing to lose, except time and money...

4. Change the sampling scheme within period. We elaborate on this in Section 5.1.2.

5. Add other stopping rules: this acts just like growing a rejection region. It will increase power, at the cost of type I error. We elaborate in Section 5.1.3.

6. Pool together more time points or more variables. We elaborate on this in sections 5.2 and 5.3, respectively.

5.1.1 Control Limits and the Alarm Rate

As previously discussed, k governs the tradeoff between type I and type II errors, or sensitivity versus specificity. It is very common to set $k = 3$. For a Normally distributed CTQ, this implies $2,700$ false alarms per million periods. This also implies an ARL of $1/\alpha = 370$ periods. We may however, discard this convention, and directly set UCL and LCL so they guarantee some desirable false alarm rate.

If normality of $\bar{x}$ can be assumed, then one may estimate σ from phase I, and set LCL and UCL by finding the k that solves $2\Phi(-k) = \alpha$. If normality cannot be assumed, there are many ways to go about. Here are some options:

1. Increase n: even if $\mathbf{x}_i$ is non normal, for large enough n, then $\bar{x}$ will be via the central limit theorem.

2. Use empirical quantiles: If phase I was returned enough data, then we may estimate $\mathbf{x}_{\alpha/2}$ and $\mathbf{x}_{1-\alpha/2}$ using the empirical quantiles of phase I. The false alarm rate will be α since $P(\mathbf{x} \notin [\mathbf{x}_{\alpha/2}, \mathbf{x}_{1-\alpha/2}]) = \alpha$, if $[\mathbf{x}_{\alpha/2}, \mathbf{x}_{1-\alpha/2}]$ is well estimated.

3. TODO:simulation

5.1.2 Rational Groupings

Recall that at each period we compute the average of n samples. How should we draw this samples? All at the same time from the same machine? At different times from the same machine? Many configurations are possible, and the correct approach depends on the type of out-of-control behaviour one seeks. *Rational groupings* merely reminds us to sample "rationally" in each period. Quoting Montgomery (2007)'s words of caution:

> ... we can often make any process appear to be in statistical control just by stretching out the interval between observations in the sample.

5.1.3 Other Stopping Rules

The assumption that we may only create alarms if $\bar{x}$ exceeds some control limits is needle sly restrictive. A first relaxation is by allowing multiple regions. It is quite common to define *warning limits* which only call for inspection, and *action limits*. Each may have its own alarm rate. We may even change the sapling scheme if limits are breached. Increasing the sampling rate once

the warning limits have been breached is known as *adaptive sampling*, or *variable sampling*.

Adaptive Sampling

Another, more strict approach, is to define multiple sets of stopping rules Here an example:

1. One or more points outside of the control limits.
2. Two of three consecutive points outside the 2-sigma warning limits but still inside the control limits.
3. Four of five consecutive points beyond the 1-sigma limits.
4. A run of 8 consecutive points on one side of the center line.

The above set of rules is known as the Western Electric Rules, a.k.a. the *WECO* rules. Augmenting the set of rules is the same as increasing a rejection region. It adds more sensitivity, at the cost of false alarms. If the rules are properly selected, the gain in sensitivity is worth the increase in false alarms.

WECO

As a quick exercise, we may compute α for m independent rules, each with α^* type I error itself:

$$\alpha = 1 - (1 - \alpha^*)^m. \tag{5.10}$$

Having 4 rules, like WECO, each at $\alpha^*0.0027$ implies that we actually have $\alpha = 0.01$.

Extra Info 7 (Stopping Rules)**.** There are many sets of stopping rules. These include WECO, Nelson, AIAG, Juran, Hughes, Duncan, Gitlow, Westgard, and more. See `http://www.quinn-curtis.com/spcnamedrulesets.htm` for a quick introduction.

5.2 Pooling Information Over Periods

Assume an out-of-control process is simply a mild shift of the controlled-process. This shift may be hard to detect in Shewart chart, especially if n is not too large (as seen in Figure 5.2). If the shift persists, we may gain power, i.e., sensitivity, by pooling several periods together. We now present several ways to pool information from history. These are typically applied in Phase II, where out-of-control processes are expected to have only mild shifts, and not major ones as in Phase I.

5.2.1 Moving Average Chart- MA

One way to pool information from different periods is by a *moving average*.

Definition 21 (Moving Average- MA)**.** The moving average at period t is defined as

$$M_t := \frac{\bar{x}_t + \cdots + \bar{x}_{t-w+1}}{w} \tag{5.11}$$

Assuming $\bar{x}_t \sim \mathcal{N}(\mu_0, \sigma^2/n)$ then clearly

$$M_t \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{nw}\right). \tag{5.12}$$

It is quite advantageous, and common, to take one observation at a time, so that $n = 1$.

The control limits on M_t are typically

$$UCL := \mu_0 + 3\sigma_{M_t} = \mu_0 + 3\frac{\sigma_x}{\sqrt{nw}} \tag{5.13}$$

$$UCL := \mu_0 - 3\sigma_{M_t} = \mu_0 - 3\frac{\sigma_x}{\sqrt{nw}}. \tag{5.14}$$

The fallout rate of this criterion is trivially $\alpha = 0.0027$. The ARL, however, is no longer so simple to compute. This is because the pooling of periods has violated the independence between periods, and Eqs.(5.8,5.9) are no longer valid. Do not despair, however, as the ARL may still be computed. You can always use simulation to compute it, or try using the `spc` **R** package.

We are free to choose the magnitude of w. If w is too small, there is no real pooling from history. At the limit, where $w = 1$, we are back to the classical Shewart chart. If w is too large, then each new observation has very small importance, and it may take a long time to detect a shift. Which is the right intermediate value, is left for you to decide.

5.2.2 Exponentially Weighted Moving Average Chart- EWMA

The moving average gives all observations the same importance. We want to change this, so that we may capture drifts quickly when they occur, while enjoy the power benefits that come with pooling information over periods. The *Exponentially Weighted Moving Average* (EWMA), a.k.a. the *geometric moving average* (GMA), does just that.

Definition 22 (Exponentially Weighted Moving Average- EWMA)**.**

$$z_t := \lambda \bar{x}_t + (1-\lambda) z_{t-1} \tag{5.15}$$

By recursive substitution we have that

$$\sigma^2_{z_t} = \frac{\sigma^2_x}{n}\left(\frac{\lambda}{2-\lambda}\right)(1-(1-\lambda)^{2t}), \tag{5.16}$$

and

$$z_t \sim \mathcal{N}\left(\mu_0, \sigma^2_{z_t}\right). \tag{5.17}$$

Eq.(5.16) may be used to construct control limits for EWMA: It is however, more economic to observe that for large λ and t: $(1-(1-\lambda)^{2t}) \approx 1$ so that we may use

$$\begin{aligned} UCL &:= \mu_0 + 3\sigma_{z_t} = \mu_0 + 3\sqrt{\frac{\sigma^2_x}{n}\left(\frac{\lambda}{2-\lambda}\right)}, \\ UCL &:= \mu_0 - 3\sigma_{z_t} = \mu_0 - 3\sqrt{\frac{\sigma^2_x}{n}\left(\frac{\lambda}{2-\lambda}\right)}. \end{aligned} \tag{5.18}$$

By now, you should immediately know what is the fallout rate of these limits. By now, you should also know that because of the dependence between z_t's, computing the ARL is not as simple as for Shewart charts. The `xewma.arl()` **R** function, in package `spc`, permits doing so easily. Its output for various λ and k is illustrated in Figure 5.3.

Note that if we do not opt for the approximation in Eq.(5.18), and use the exact Eq.(5.16) then the alarm limits change over time.

In the MA chart, we used the choice of w to balance between quick response (small w) and sensitivity (large w). EWMA has no window-width parameter, since it looks on all of the history. On the other hand, we can control it by choosing λ. Large λ gives more importance to the present. At the limit, $\lambda = 1$, EWMA collapses to a standard Shewart chart.

5.2.3 Filtered Derivative Chart

[TODO]

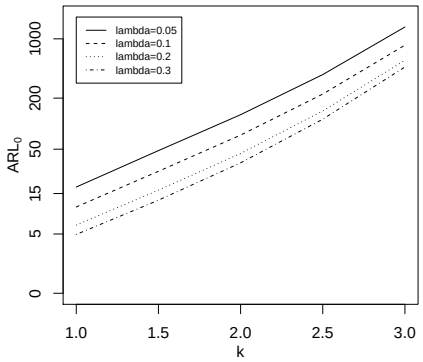

Figure 5.3: ARL_0 for EWMA.
Code from `http://users.phhp.ufl.edu/pqiu/research/book/spc/r-codes/fig53.r`

5.2.4 CUSUM Chart

The *cumulative sum* chart is similar to the EWMA in that it pools information from the history. It does differ, in that it gives all the history an equal weight. The CUSUM simply sums deviations from the centre line. If the process is in control, deviation will cancel each other, and their sum will vary around 0. If the process is ot of control, some drift will appear. The statistic to be plotted is

$$C_t := \sum_{j=0}^{t} (\bar{x}_j - \mu_0) = C_{t-1} + (\bar{x}_t - \mu_0) \tag{5.19}$$

Observing that when under control then $C_t \sim \mathcal{N}\left(\mu_0, \frac{t}{n}\sigma_x^2\right)$, we could set

$$\begin{aligned} UCL &:= \mu_0 + 3\sigma_{C_t} = \mu_0 + 3\sqrt{\frac{t}{n}\sigma_x^2}, \\ UCL &:= \mu_0 - 3\sigma_{C_t} = \mu_0 - 3\sqrt{\frac{t}{n}\sigma_x^2}. \end{aligned} \tag{5.20}$$

This approach, is rarely seen in practice, because it is suboptimal. Indeed,

5.2.5 Combined Shewart and Running Window Charts

Extra Info 8 (Local Methods)**.**

5.3 Multivariate Control Charts

5.4 Economical Design of Control Charts

5.5 Non-Statistical Target Functions

5.6 Other Control Statistics

5.6.1 R Chart

5.6.2 s Chart

5.6.3 s^2 Chart

5.6.4 Shewhart Individuals Control Chart

5.6.5 Three-way Chart

5.6.6 p Chart

5.6.7 np Chart

5.6.8 c Chart

5.6.9 u Chart

5.6.10 Time Series Model

5.6.11 Regression Control Chart

5.6.12 Running Window Versions

5.7 Notes

Chapter 6

Design of Experiments - DOE

6.1 Terminology

Alias

Balanced Design

Blocking

Factor Encodings

Confounding

Control Group

Design

Design Matrix

Effect

Experimental Unit

Factors

Fixed Effect

Random Effect

Interaction

Lack of fit error

Pure Error

Orthogonality

Replication

Resolution

Response

Screening Design

Test Plan

Treatment

Treatment Group

Variance Components

6.2 Randomization

6.3 Full Factorial Designs

6.4 Fractional Factorial Designs

6.5 Robust Parameter Design (RPD)

6.6 Latin-Square Design

6.7 Plackett–Burman Design

6.8 Computer Experiments

Chapter 7

Acceptance Sampling

Chapter 8

Reliability

Chapter 9

Revisiting System Capability Analysis

[TODO]

9.1 System Capability with Control Charts

9.2 System Capability with Designed Experiments

Appendix A

Notation

In this text we use the following notation conventions:

x A column vector, or scalar, as implied by the text.

$:=$ An assignment, or definition. $A := a$ means that A is defined to be a.

$\prod_{i=1}^{n}$ The product operator: $\prod_{i=1}^{n} x_i := x_1 \times \cdots \times x_n$

$\#\{A\}$ The count operator. Returns the number of elements of the set A. Also known as the *cardinality*.

$\Phi(t)$ The standard Gaussian CDF at t: $\Phi(t) := P(Z < t)$.

$\phi(t)$ The standard Gaussian density at t: $\phi(t) := \frac{\partial}{\partial t}\Phi(t)$.

Appendix B

SPC in R

Here is a list of packages worth knowing for SPC in R. The list is taken from Qiu (2013).

cpm [TODO]

mnspc [TODO]

msqc [TODO]

qcc [TODO]

spc [TODO]

Bibliography

M. Basseville, I. V. Nikiforov, and others. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

D. C. Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007.

K. B. Petersen and M. S. Pedersen. *The matrix cookbook*. Citeseer, 2006.

P. Qiu. *Introduction to Statistical Process Control*. Chapman and Hall/CRC, Boca Raton, 2013.

J. L. Rodgers and W. A. Nicewander. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1):59–66, 1988.

Wikipedia. *Design for Six Sigma — Wikipedia, The Free Encyclopedia*. 2015a. [Online; accessed 29-October-2015].

Wikipedia. *Eight dimensions of quality — Wikipedia, The Free Encyclopedia*. 2015b. [Online; accessed 28-October-2015].

Wikipedia. *Lean manufacturing — Wikipedia, The Free Encyclopedia*. 2015c. [Online; accessed 29-October-2015].

Wikipedia. *Quality (business) — Wikipedia, The Free Encyclopedia*. 2015d. [Online; accessed 28-October-2015].

Wikipedia. *Value engineering — Wikipedia, The Free Encyclopedia*. 2015e. [Online; accessed 29-October-2015].

Wikipedia. *Zero Defects — Wikipedia, The Free Encyclopedia*. 2015f. [Online; accessed 29-October-2015].