Box Cox Transformation

- The Box-Cox transforms nonnormally distributed data to a set of data that has approximately normal distribution.

Chapter 1

Review Questions (Theory Questions)

1. Differentiate common (or chance) causes of variation in the quality of process output from assignable (or special) causes.
2. Differentiate a stable process from an unstable process.
3. Other than applying the 3-sigma rule for detecting the presence of an assignable cause, what else do we look for when studying a control chart?
4. Describe how the output of a stable process can be improved. What actions do not improve a stable process, but rather, make the output more variable?
5. What is the purpose of maintaining control charts?
6. What is tampering in the context of process control?

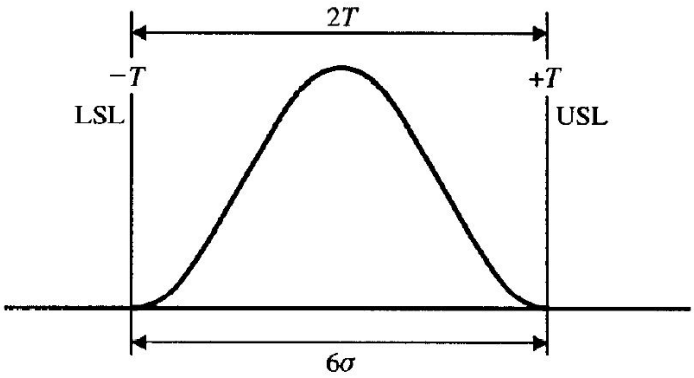
Chapter 2

Process Capability Indices

2.1 Process Capability

- In managing variables the usual aim is not to achieve exactly the same diameter for every piston, the same weight for every tablet, sales figures exactly as forecast, etc but to reduce the variation of products and process parameters around a **target value**.
- No adjustment of a process is called for as long as there has been no identified change in its accuracy or precision. This means that, in controlling a process, it is necessary to establish first that it is in statistical control, and then to compare its centering and spread with the specified target value and specification tolerance.
- We have seen previously that, if a process is not in statistical control, special causes of variation may be identified with the aid of ***control charts***.
- Only when all the special causes have been accounted for, or eliminated, can process capability be sensibly assessed. The variation due to common causes may then be examined and the "*natural specification*" compared with any imposed specification or tolerance zone.

- The relationship between process variability and tolerances may be formalized by consideration of the standard deviation, σ, of the process.
- In order to manufacture within the specification, the distance between the **upper specification limit** (USL) or upper tolerance ($+T$) and **lower specification limit** (LSL) or lower tolerance ($-T$), i.e. ($USL - LSL$) or 2T must be equal to or greater than the width of the base of the process bell, i.e. 6σ.

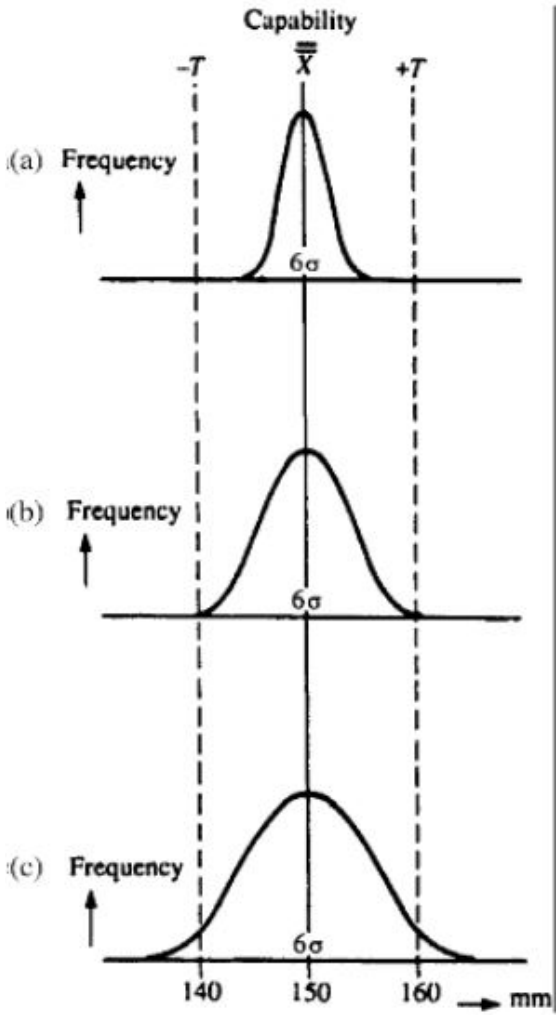


The relationship between $USL - LSL$ (i.e. 2T) and 6σ gives rise to three levels of precision of the process (Figure below):

(a) **High Relative Precision**, where the tolerance band is very much greater than 6σ ($2T >> 6\sigma$)

(b) **Medium Relative Precision**, where the tolerance band is just greater than 6σ ($2T > 6\sigma$)

(c) **Low Relative Precision**, where the tolerance band is less than 6σ ($2T < 6\sigma$)


Capability
−T
$\bar{\bar{X}}$
+T
(a) Frequency
6σ
(b) Frequency
6σ
(c) Frequency
6σ
140
150
160
mm

2.1.1 Process Capability Indices

- A process capability index is a measure relating the actual performance of a process to its specified performance, where processes are considered to be a combination of the plant or equipment, the method itself, the people, the materials and the environment.
- These indices assumes process output is approximately normally distributed.
- The absolute minimum requirement is that three process standard deviations each side of the process mean are contained within the specification limits.
- This means that approximately 99.7 per cent of output will be within the tolerances. A more stringent requirement is often stipulated to ensure that produce of the correct quality is consistently obtained over the long term.
- When a process is under statistical control (i.e. only random or common causes of variation are present), a process capability index may be calculated.
- Process capability indices are simply a means of indicating the variability of a process relative to the product specification tolerance.

C_p index

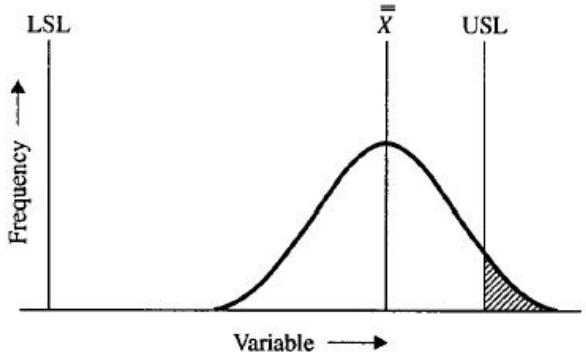
- In order to manufacture within a specification, the difference between the USL and the LSL must be less than the total process variation.
- A comparison of 6σ with (USL-LSL) or 2T gives an obvious process capability index, known as the C_p of the process:

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}}$$

- This estimates what the process is capable of producing if the process mean were to be centered between the specification limits. Clearly, any value of C_p below 1 means that the process variation is greater than the specified tolerance band so the process is incapable.
- For increasing values of C_p the process becomes increasingly capable. The C_p index makes no comment about the centring of the process, it is a simple comparison of total variation with tolerances.

C_{pk} index

- It is possible to envisage a relatively wide tolerance band with a relatively small process variation, but in which a significant proportion of the process output lies outside the tolerance band (Figure below).



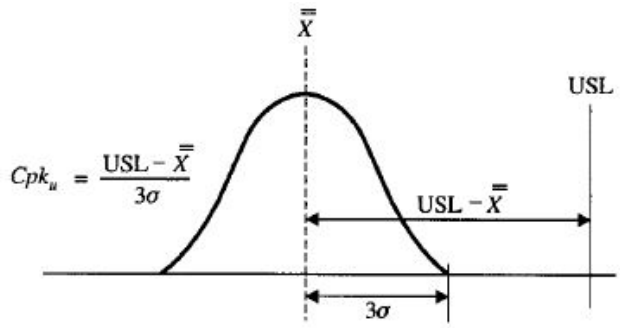
- This does not invalidate the use of Cp as an index to measure the potential capability of a process when centred, but suggests the need for another index which takes account of both the process variation and the centring. Such an index is the C_{pk}, which is widely accepted as a means of communicating process capability.

$$Cpk_{\mathrm{u}} = \frac{\mathrm{USL} - \overline{X}}{3\sigma} \qquad Cpk_{\mathrm{l}} = \frac{\overline{X} - \mathrm{LSL}}{3\sigma}.$$

- For upper and lower specification limits, there are two C_{pk} values, C_{pku} and C_{pkl}. These relate the difference between the process mean and the upper and the lower specification limits respectively, to 3σ (half the total process variation).
- The overall process C_{pk} is the lower value of C_{pku} and C_{pkl}. A C_{pk} of 1 or less means that the process variation and its centring is such that at least one of the tolerance limits will be exceeded

and the process is incapable. As in the case of C_p, increasing values of C_{pk} correspond to increasing capability.

- It may be possible to increase the C_{pk} value by centring the process so that its mean value and the mid-specification or target, coincide. A comparison of the C_p and the C_{pk} will show zero difference if the process is centred on the target value.



- The C_{pk} can be used when there is only one specification limit, upper or lower a one-sided specification. This occurs quite frequently and the C_p index cannot be used in this situation.

Example 1

Example 1.

In tablet manufacture, the process parameters from 20 samples of size $n=4$ are:

Mean Range ($\overline{R}$) = 91 mg, Process mean ($\overline{\overline{X}}$) = 2500 mg

Specified requirements USL = 2650 mg, LSL = 2350 mg

Standard Deviation (s) = 44.2

$$Cp = \frac{\text{USL} - \text{LSL}}{6\sigma} \quad \text{or} \quad \frac{2T}{6\sigma} = \frac{2650 - 2350}{6 \times 44.2} = \frac{300}{265.2} = 1.13$$

$$Cpk = \text{lesser of } \frac{\text{USL} - \overline{\overline{X}}}{3\sigma} \quad \text{or} \quad \frac{\overline{\overline{X}} - \text{LSL}}{3\sigma}$$

$$= \frac{2650 - 2500}{3 \times 44.2} \quad \text{or} \quad \frac{2500 - 2350}{3 \times 44.2} = 1.13.$$

Example 2

Example 2.

If the process parameters from 20 samples of size $n=4$ are:

Mean range $(\overline{R})$ = 91 mg, Process mean $(\overline{\overline{X}})$ = 2650 mg

Specified requirements USL = 2750 mg, LSL = 2250 mg

Standard Deviation (s) = 44.2

$$Cp = \frac{\text{USL} - \text{LSL}}{6\sigma} \quad \text{or} \quad \frac{2T}{6\sigma} = \frac{2750 - 2250}{6 \times 44.2} = \frac{500}{265.2} = 1.89$$

$$Cpk = \text{lesser of } \frac{2750 - 2650}{3 \times 44.2} \quad \text{or} \quad \frac{2650 - 2250}{3 \times 44.2}$$

$$= \text{lesser of } 0.75 \text{ or } 3.02 = 0.75.$$

Conclusion – the *Cp* at 1.89 indicates a potential for higher capability than in example (i), but the low *Cpk* shows that this potential is not being realized because the process is not centred.

It is important to emphasize that in the calculation of all process capability indices, no matter how precise they may appear, the results are only ever approximations we never actually know anything, progress lies in obtaining successively closer approximations to the truth. In the case of the process capability this is true because:

- there is always some variation due to sampling;
- no process is ever fully in statistical control;
- no output exactly follows the normal distribution or indeed any other standard distribution.

Interpreting process capability indices without knowledge of the source of the data on which they are based can give rise to serious misinterpretation.

Interpreting capability indices - IMPORTANT

In the calculation of process capability indices so far, we have derived the standard deviation, σ, and recognized that this estimates the short-term variations within the process. This short term is the period over which the process remains relatively stable, but we know that processes do not remain stable for all time and so we need to allow within the specified tolerance limits for:

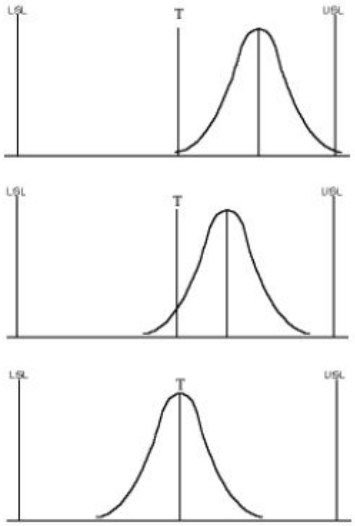
- some movement of the mean;
- the detection of changes of the mean;
- possible changes in the scatter (range);
- the detection of changes in the scatter;
- the possible complications of non-normal distributions.

Taking these into account, the following values of the C_{pk} index represent the given level of confidence in the process capability:

- $C_{pk} < 1$ A situation in which the production system is **not capable** and there will inevitably be non-conforming output from the process.
- $C_{pk} = 1$ A situation in which the production system is not really capable, since any change within the process will result in some undetected non-conforming output.
- $C_{pk} = 1.33$ A still far from acceptable situation since non-conformance is not likely to be detected by the process control charts.
- $C_{pk} = 1.5$ Not yet satisfactory since non-conforming output will occur and the chances of detecting it are still not good enough.
- $C_{pk} = 1.67$ Promising, non-conforming output will occur but there is a very good chance that it will be detected.
- $C_{pk} = 2$ High level of confidence in the production system, provided that control charts are in regular use.

The C_{pm} Index

- Another Index C_{pm} incorporates the target when calculating the standard deviation. The standard error, denoted $\hat{\sigma}_{Cpm}$ compares each observation to a reference value.
- However, instead of comparing the data to the mean, the data is compared to the target. These differences are squared. Thus any observation that is different from the target observation will increase the $\hat{\sigma}_{Cpm}$ standard deviation.
- As this difference increases, so does the Cpm. And as this index becomes larger, the C_{pm} gets smaller.
- If the difference between the data and the target is small, so too is the sigma. And as this sigma gets smaller, the Cpm index becomes larger. The higher the C_{pm} index, the better the process.
- In the following charts the process is the same, but as the process becomes more centred, the C_{pm} gets better.

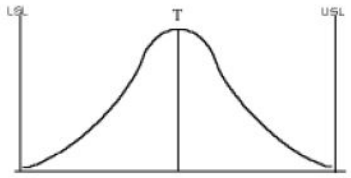


This Cpm is good.

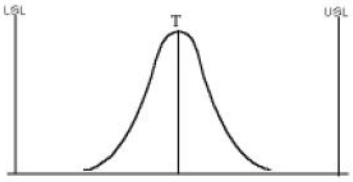
This Cpm is better.

This Cpm is best.

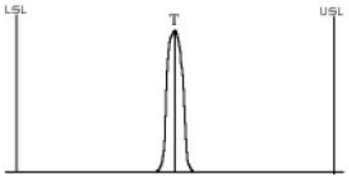
In these 3 charts, the process stays centred about the target, but as the variation is reduced, the Cpm gets better.



This Cpm is reasonably good.



This Cpm is better.



This Cpm is best.

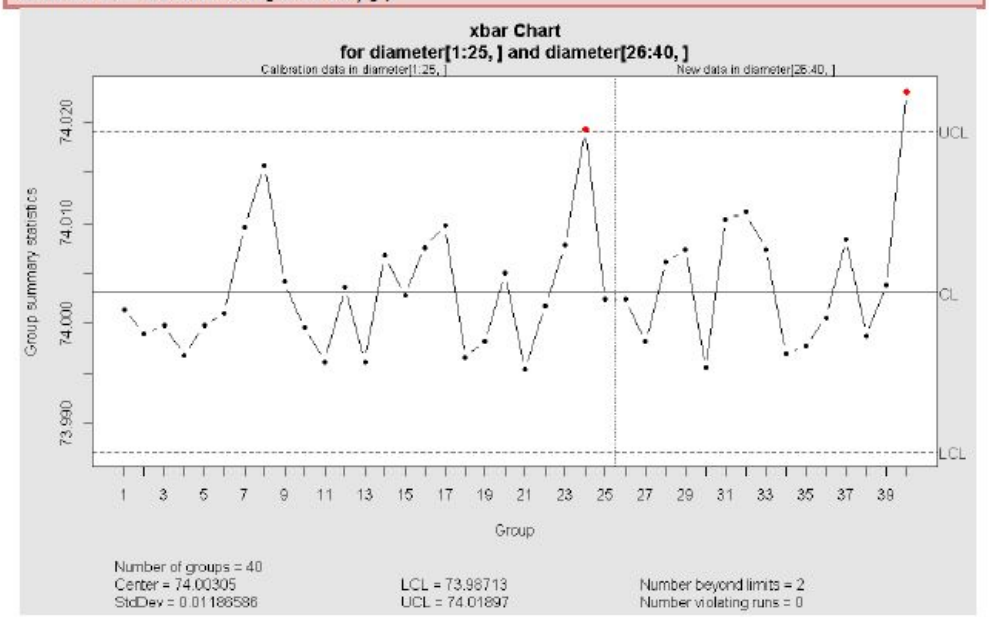
	Population Known	Population Unknown
Cp	$C_p = \frac{USL - LSL}{6\sigma}$	$\hat{C}_p = \frac{USL - LSL}{6s}$
Cpk	$C_{pk} = \min[\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}]$	$\hat{C}_{pk} = \min[\frac{USL - \bar{x}}{3s}, \frac{\bar{x} - LSL}{3s}]$
Cpm	$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}}$	$\hat{C}_{pm} = \frac{USL - LSL}{6\sqrt{s^2 + (\bar{x} - T)^2}}$

2.1.2 Worked Example with R

Data set used diameter (piston rings data set)

R code used previously, reminding ourselves about the data set.

```
data(pistonrings)
attach(pistonrings)
dim(pistonrings)
diameter <- qcc.groups(diameter, sample)
obj <- qcc(diameter[1:25,], type="xbar",
newdata=diameter[26:40,])
```



2.1.3 Implementation of Process Capability Analysis

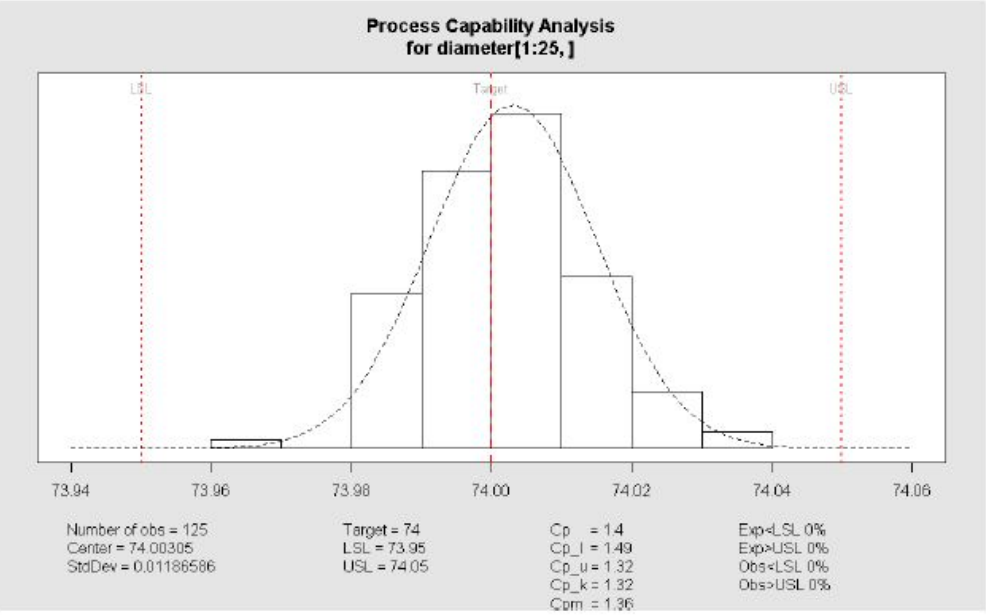
Indices and Confidence intervals for those indices.

```
> process.capability(obj, spec.limits=c(73.95,74.05))

Process Capability Analysis

Call:
process.capability(object = obj, spec.limits = c(73.95,
74.05))

Number of obs = 125          Target = 74
       Center = 74.00305        LSL = 73.95
       StdDev = 0.01186586      USL = 74.05

Capability indices:

      Value   2.5%  97.5%
Cp    1.405  1.230  1.579
Cp_l  1.490  1.327  1.653
Cp_u  1.319  1.173  1.465
Cp_k  1.319  1.145  1.493
Cpm   1.360  1.187  1.534

Exp<LSL 0%   Obs<LSL 0%
Exp>USL 0%   Obs>USL 0%
```

2.2 Quality Control Charts (MA4605 Only)

- In the Lab classes, we will look at using the R package qcc, which stands for quality control charts.
- This package can be used to create the following. Shewhart quality control charts for continuous, attribute and count data.
  1. CUSUM and EWMA charts.
  2. Operating characteristic curves.
  3. Process capability analysis.
  4. Pareto chart and cause-and-effect chart.
  5. Multivariate control charts

  (We wont have time to get through all of this material)



- There are two inbuilt data sets, available from the qcc package, that we will use

  1. pistonrings
  2. orangejuice

  The pistonrings data set (diameter) The first data set describes piston rings for an automotive engine are produced by a forging process.

- The inside diameter of the rings manufactured by the process is measured on 25 samples, each of size 5, drawn from a process being considered 'in control'. There are further 15 sets of observation (i.e. 40 batches in total) to demonstrate alternate outcomes. The dataset is restructured as a data set called diameter. The orangejuice data set The second data set describes frozen orange juice concentrate is packed in 6-oz cardboard cans. These cans are formed on a machine by spinning them from cardboard stock and attaching a metal bottom panel.

2.3 More on Control Charts

2.3.1 Control Chart Selection

- Correct control chart selection is a critical part of creating a control chart. If the wrong control chart is selected, the control limits will not be correct for the data.

- The type of control chart required is determined by the type of data to be plotted and the format in which it is collected.

- Data collected is either in variables or attributes format, and the amount of data contained in each sample (subgroup) collected is specified.
- **Variables data** is defined as a measurement such as height, weight, time, or length. Monetary values are also variables data.
    - Generally, a measuring device such as a weighing scale, vernier, or clock produces this data.
    - Another characteristic of variables data is that it can contain decimal places e.g. 3.4, 8.2.
- **Attributes data** is defined as a count such as the number of employees, the number of errors, the number of defective products, or the number of phone calls. A standard is set, and then an assessment is made to establish if the standard has been met.
    - The number of times the standard is either met or not is the count. Attributes data never contains decimal places when it is collected, it is always whole numbers, e.g. 2, 15.

2.3.2 Attribute Control Charts

- The Shewhart control chart plots quality characteristics that can be measured and expressed numerically. We measure weight, height, position, thickness, etc. If we cannot represent a particular quality characteristic numerically, or if it is impractical to do so, we then often resort to using a quality characteristic to sort or classify an item that is inspected into one of two "buckets".
- An example of a common quality characteristic classification would be designating units as "conforming units" or "nonconforming units".
- Another quality characteristic criteria would be sorting units into "non defective" and "defective" categories. Quality characteristics of that type are called ***attributes***.
- *Note that there is a difference between "nonconforming to an engineering specification" and "defective" – a nonconforming unit may function just fine and be, in fact, not defective at all, while a part can be "in spec" and not fucntion as desired (i.e., be defective).*
- Examples of quality characteristics that are attributes are the number of failures in a production run, the proportion of malfunctioning wafers in a lot, the number of people eating in the cafeteria on a given day, etc.

2.3.3 Types of Attributes Control Charts

- Control charts dealing with the proportion or fraction of defective product are called ***p-charts*** (for proportion).
- Control charts dealing with the number of defective product are called ***np-charts***.
- Control charts dealing with the number of defects or nonconformities are called ***c-charts*** (for count).
- There is another chart which handles defects per unit, called the ***u-chart*** (for unit). This applies when we wish to work with the average number of nonconformities per unit of product.

2.3.4 p-charts

- The p-chart is a type of control chart used to monitor the **proportion of nonconforming units** in a sample, where the sample proportion nonconforming is defined as the ratio of the number of nonconforming units to the sample size, n.
- The p-chart only accommodates dichotomous PASS/FAIL-type inspection as determined by a series of tests, effectively applying the specifications to the data before they are plotted on the chart.
- Other types of control charts display the magnitude of the quality characteristic under study, making troubleshooting possible directly from those charts.
- A p-chart is an attributes control chart used with data collected in subgroups of varying sizes. Because the subgroup size can vary, it shows a proportion on nonconforming items rather than the actual count.
- *p-charts show how the process changes over time. The process attribute (or characteristic) is always described in a yes/no, pass/fail, go/no go form.*
- Example: use a p-chart to plot the proportion of incomplete insurance claim forms received weekly. The subgroup would vary, depending on the total number of claims each week.

2.3.5 np-charts

The np-chart is a type of control chart used to monitor the number of nonconforming units in a sample. An np-chart is an *attributes* control chart used with data collected in subgroups that are the **same size**.

It is an adaptation of the p-chart and used in situations where personnel find it easier to interpret process performance in terms of concrete numbers of units rather than the somewhat more abstract proportion.

The np-chart differs from the p-chart in only the three following aspects:

- The control limits are
$$n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})}$$
, where n is the sample size and $\bar{p}$ is the estimate of the long-term process mean established during control-chart setup.
- The number nonconforming (np), rather than the fraction nonconforming (p), is plotted against the control limits.
- The sample size, n, is constant.

2.3.6 The c-chart

- In this chart, we plot the number of defectives (per batch, per day, per machine, per 100 feet of pipe, etc.).
- This chart assumes that defects of the quality attribute are rare, and the control limits in this chart are computed based on the Poisson distribution (distribution of rare events).
- The c-chart is a type of control chart used to monitor "count"-type data, typically total number of nonconformities per unit.It is also occasionally used to monitor the total number of events occurring in a given unit of time.
- The c-chart differs from the p-chart in that it accounts for the possibility of more than one nonconformity per inspection unit, and that (unlike the p-chart and u-chart) it requires a fixed sample size.
- The p-chart models "pass"/"fail"-type inspection only, while the c-chart (and u-chart) give the ability to distinguish between (for example) 2 items which fail inspection because of one fault each and the same two items failing inspection with 5 faults each; in the former case, the p-chart will show two non-conformant items, while the c-chart will show 10 faults.
- The Poisson distribution is the basis for the chart and requires the following assumptions:
  - The number of opportunities or potential locations for nonconformities is very large
  - The probability of nonconformity at any location is small and constant
  - The inspection procedure is same for each sample and is carried out consistently from sample to sample

2.3.7 The u-chart

- In this chart we plot the rate of defectives, that is, the number of defectives divided by the number of units inspected (the n; e.g., feet of pipe, number of batches).
- Unlike the C chart, this chart does not require a constant number of units, and it can be used, for example, when the batches (samples) are of different sizes.
- The u-chart is a type of control chart used to monitor "count"-type data where the sample size is greater than one, typically the average number of nonconformities per unit.
- The u-chart differs from the c-chart in that it accounts for the possibility that the number or size of inspection units for which nonconformities are to be counted may vary. Larger samples may be an economic necessity or may be necessary to increase the area of opportunity in order to track very low nonconformity levels.

2.4 Control Charts with the qcc Package

```
# install.package(qcc)
library(qcc)

# series of value w/ mean of 10 with a little random noise added in
x <- rep(10, 100) + rnorm(100)

# a test series w/ a mean of 11
new.x <- rep(11, 15) + rnorm(15)

# qcc will flag the new points
qcc(x, newdata=new.x, type="xbar.one")
```

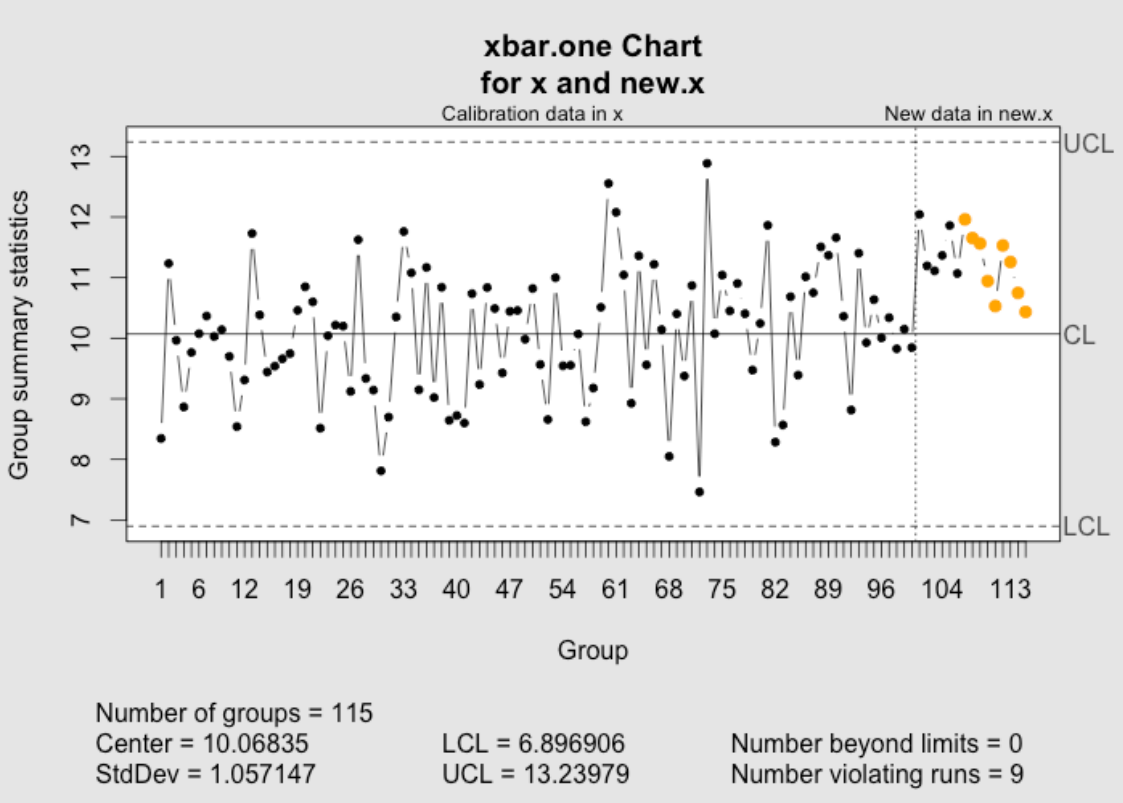
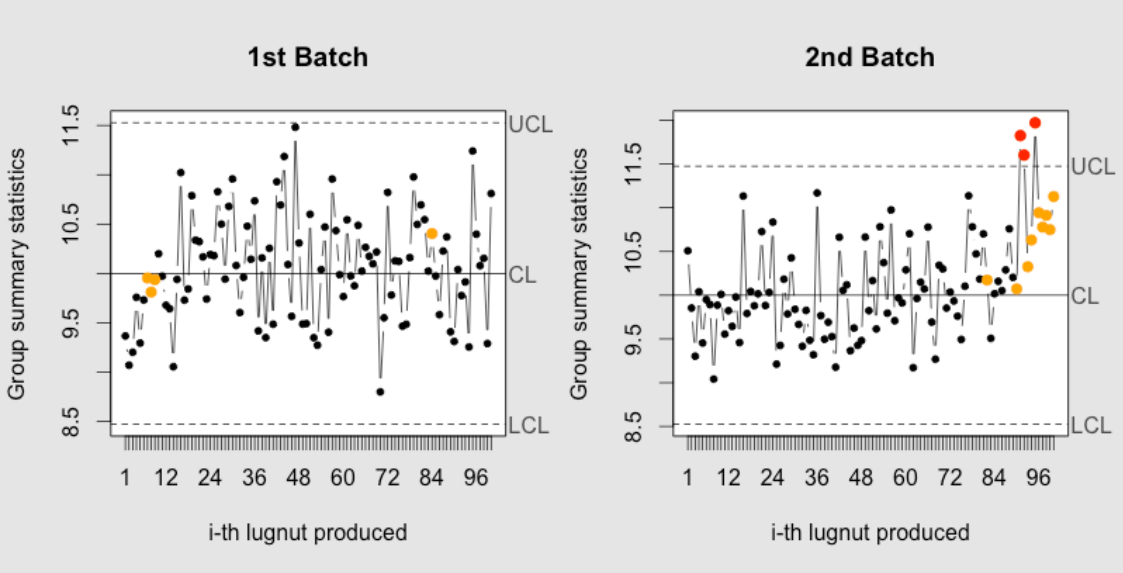


Figure 2.1:




```
library(qcc)
#make 2 plots in 1 figure
par(mfrow=c(1,2))

#points have base value of 10 w/ normally distributed error
lugnuts <- rep(10, 100) + rnorm(100, mean=0, sd=0.5)
qcc(lugnuts, type="xbar.one", center=10, add.stats=FALSE,
title="1st Batch",
xlab="i-th lugnut produced")
```

Second Batch

- First 90 points have mean value of 10 with normally distributed error,
- Last 10 points have mean value of 11 with normally distributed error

```
lugnuts <- c(rep(10, 90), rep(11, 10)) + rnorm(100, mean=0, sd=0.5)
qcc(lugnuts, type="xbar.one", center=10, add.stats=FALSE,
title="2nd Batch",
xlab="i-th lugnut produced")
```

```
> set.seed(1234)
> lugnuts <- rep(10, 100) + rnorm(100, mean=0, sd=0.5)
> summary(lugnuts)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
8.827   9.552   9.808   9.922  10.240  11.270
> length(lugnuts)
[1] 100
> newLugnuts <- rep(11, 10) + rnorm(10, mean=0, sd=0.5)
> summary(newLugnuts)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
10.55   10.75   11.00   10.92   11.08   11.21
> length(newLugnuts)
[1] 10
```

```
qcc1 <- qcc(lugnuts, type="xbar.one", center=10, add.stats=TRUE,
title="1st Batch of 100",
xlab="i-th lugnut produced")



qcc2 <- qcc(lugnuts, newdata=newLugnuts,
type="xbar.one", center=10,
add.stats=TRUE, title="All Lugnuts",
xlab="i-th lugnut produced")

mode(qcc1)
class(qcc1)
names(qcc1)
```

```
> mode(qcc1)
[1] "list"
> class(qcc1)
[1] "qcc"
> names(qcc1)
[1] "call"       "type"       "data.name"  "data"       "statistics"
[6] "sizes"      "center"     "std.dev"    "nsigmas"    "limits"
[11] "violations"
> qcc1$violations
$beyond.limits
integer(0)

$violating.runs
[1] 13 38 39 40 48 49 50 51 52 53 54 55
```

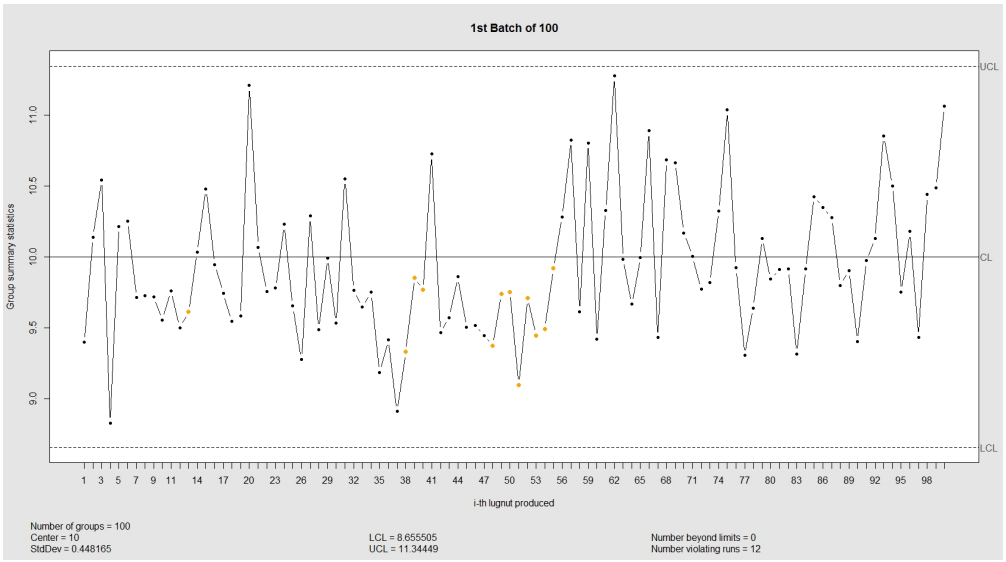


Figure 2.2:

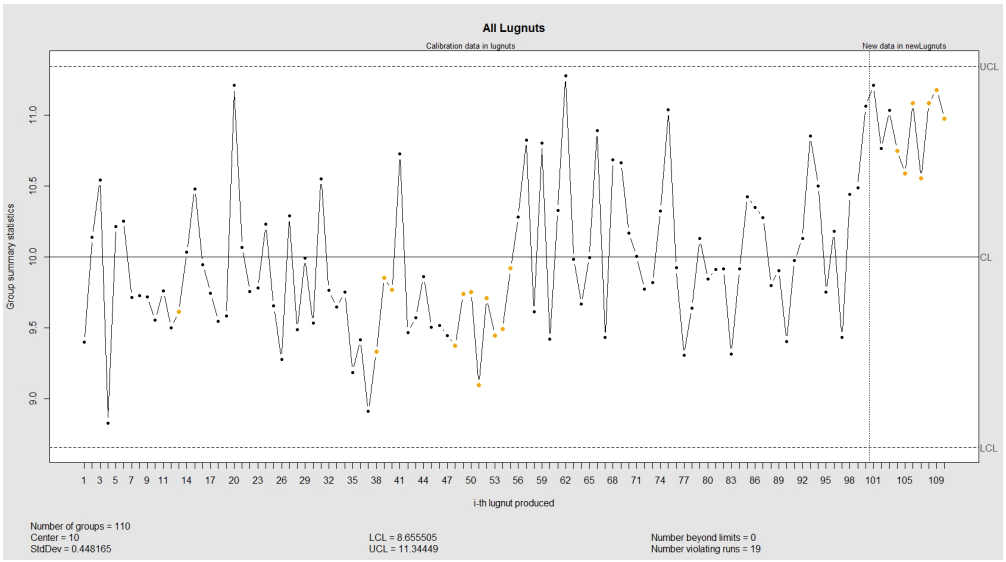


Figure 2.3:

2.4.1 Using the summary command

```
Call:
qcc(data = lugnuts, type = "xbar.one", center = 10, add.stats = TRUE,
title = "1st Batch of 100", xlab = "i-th lugnut produced")

xbar.one chart for lugnuts

Summary of group statistics:
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
8.827   9.552   9.808   9.922  10.240  11.270

Group sample size:  1
Number of groups:  100
Center of group statistics:  10
Standard deviation:  0.448165

Control limits:
LCL      UCL
8.655505 11.34449
```

```
Call:
qcc(data = lugnuts, type = "xbar.one", center = 10, newdata = newLugnuts,
add.stats = TRUE, title = "All Lugnuts", xlab = "i-th lugnut produced")

xbar.one chart for lugnuts
.....

Summary of group statistics in newLugnuts:
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
10.55   10.75   11.00   10.92   11.08   11.21

Group sample size:  1
Number of groups:  10

Control limits:
LCL      UCL
8.655505 11.34449
```

Example used by Drew Conway

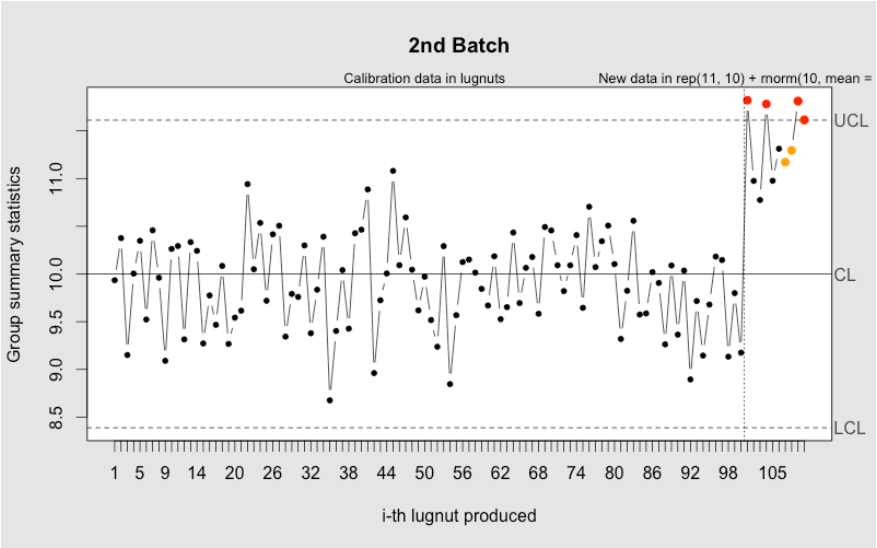


Figure 2.4:

```
lugnuts <- rep(10, 100) + rnorm(100, mean=0, sd=0.5)
qcc(lugnuts, newdata=rep(11, 10) + rnorm(10, mean=0, sd=0.5),
type="xbar.one", center=10,
add.stats=FALSE, title="2nd Batch",
xlab="i-th lugnut produced")
```

Remarks

- `newdata`
- `add.stats`

2.4.2 Using `R`

Advantages of using R statistical program along with the **qcc** package:

- There are several packages for interfacing with databases, RODBC being a common and useful one on MS windows.

- Allows of automation: you can program a regular event loop to check for new data and run a new set of charts and notifications if there is new data.
- The **mail** and **sendmailR** packages were designed to automatically send e-mails with regular reports and warning messages.
- It produces the standard SPC charts, these can go to the screen or a file to be sent out.
- Bespokes tests can be program for out of control signals
- Full programming language with common (and uncommon) statistics so you can preproccess you data in many ways to reduce dimension.
- You can have multiple instances running on multiple or a single computer, each processing for a single department, or you can combine it all into one script to run for all the departments.

Chapter 3

Multivariate SPC

3.1 Multivariate Control Charts

- With the enhancements in data acquisition systems it is usual to deal with processes with more than one correlated quality characteristic to be monitored.
- A common practice is to control the stability of the process using univariate control charts.
- This practice increases the probability of false alarm of special cause of variation.
- Therefore, the analysis should be performed through a multivariate approach; that is, the variables must be analyzed together, not independently.

3.1.1 Multivariate Control Charts

- Multivariate control charts monitor multiple process characteristics. Independent variables can be charted individually, but if the variables are correlated, a multivariate chart is needed to determine whether the process is in control.
- Multivariate control charts can detect shifts in the mean or the relationship between several related variables.
- The multivariate control chart plots Hotellings T2 statistic. The calculation for the control limit differs based on whether targets have been specified.

3.1.2 The MSQC package

In his book, Edgar Santos-Fernandez present the multivariate normal distribution, the data structure of the multivariate problems dealt in this book, the mult.chart function that allows the computation in R, and the most used multivariate control charts:

- The control ellipsoid or w2 control chart
- The T2 or Hotelling chart
- The Multivariate Exponentially Weighted Moving Average (MEWMA) chart
- The Multivariate Cumulative Sum (MCUSUM) chart
- The chart based on Principal Components Analysis (PCA)

3.1.3 The `mult.chart` Function

The performing of the multivariate control chart in R can be carried out with the function mult.chart which is a general function that allows to compute the most accepted and diversified continuous multivariate chart such as

- χ^2
- Hotelling T^2
- MEWMA
- MCUSUM according to Crosier (1988)
- MCUSUM by Pignatiello and Runger (1990)

Finally the function `mult.chart` returns:

- The T2 statistics
- The Upper Control Limit (UCL)
- The sample covariance matrix (S)
- The mean vector (Xmv)
- And if any point falls outside of the UCL and its decomposition

```
mult.chart(dowel1, type = "chi", alpha = 0.05)
```

3.1.4 T2 control chart

The origin of the T2 control chart dates back to the pioneer works of Harold Hotelling who applied this method to the bombsight problem in Second World War. The Hotelling (1947) procedure has become without doubt the most applied in multivariate process control and it is the multivariate analogous of the Shewhart control chart. For that reason, it is also known as multivariate Shewhart control chart.

```
data("carbon1")
mult.chart(type = "t2", carbon1)
mult.chart(type = "t2", carbon1)$t2
```

3.1.5 mqcc Example

```
# library(mqcc)
# Ryan (2000, Table 9.2) data with p = 2 variables,
#  m = 20 samples, n = 4 sample size:

X1 = matrix(c(72, 56, 55, 44, 97, 83, 47, 88, 57, 26, 46,
49, 71, 71, 67, 55, 49, 72, 61, 35, 84, 87, 73, 80, 26, 89, 66,
50, 47, 39, 27, 62, 63, 58, 69, 63, 51, 80, 74, 38, 79, 33, 22,
54, 48, 91, 53, 84, 41, 52, 63, 78, 82, 69, 70, 72, 55, 61, 62,
41, 49, 42, 60, 74, 58, 62, 58, 69, 46, 48, 34, 87, 55, 70, 94,
49, 76, 59, 57, 46), ncol = 4)

X2 = matrix(c(23, 14, 13, 9, 36, 30, 12, 31, 14, 7, 10,
11, 22, 21, 18, 15, 13, 22, 19, 10, 30, 31, 22, 28, 10, 35, 18,
11, 10, 11, 8, 20, 16, 19, 19, 16, 14, 28, 20, 11, 28, 8, 6,
15, 14, 36, 14, 30, 8, 35, 19, 27, 31, 17, 18, 20, 16, 18, 16,
13, 10, 9, 16, 25, 15, 18, 16, 19, 10, 30, 9, 31, 15, 20, 35,
12, 26, 17, 14, 16), ncol = 4)

X = list(X1 = X1, X2 = X2)
q = mqcc(X, type = "T2")
summary(q)
```

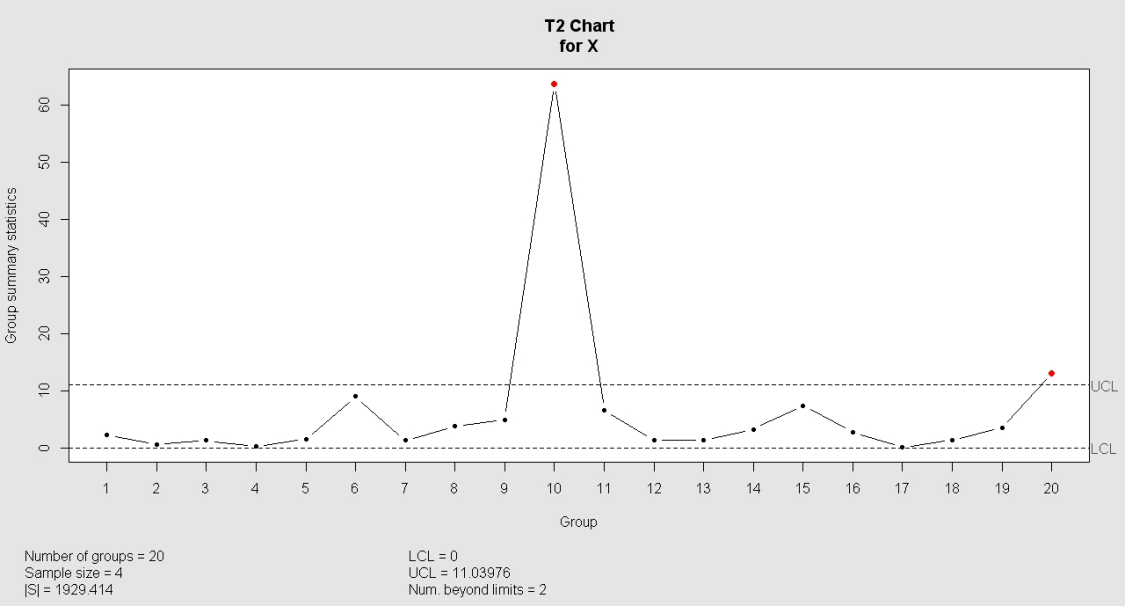


Figure 3.1:

```
Call:
mqcc(data = X, type = "T2")

T2 chart for X

Summary of group statistics:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1243  1.3250  2.5030  6.4700  5.3490 63.7600

Number of variables:  2
Number of groups:  20
Group sample size:  4

Center:
     X1      X2
60.3750 18.4875

Covariance matrix:
         X1        X2
X1 222.0333 103.11667
```

```
X2 103.1167  56.57917
|S|:  1929.414

Control limits:
 LCL      UCL
   0 11.03976
```

3.2 The qcc R package - The 7QC tools revisited

- The **qcc** package was built by Luca Scrucca for nothing but statistical quality control.
- It's extremely easy to use. You provide it with data and it tells you which points are considered to be outliers based on the Shewart Rules.
- It even color codes them based on how irregular each point is.
- Even though statistical quality control an old topic, statistical quality control is still highly relevant. There are probably have lots of jobs, processes, logs, or databse metric tha could be monitored using control charts.

3.2.1 qcc : Quality Control Charts

Some Remarks

- Shewhart quality control charts for continuous, attribute and count data.
- Cusum and EWMA charts.
- Operating characteristic curves.
- Process capability analysis.
- Pareto chart and cause-and-effect chart.
- Multivariate control charts.

3.2.2 Types of Control Chart supported by qcc

"xbar" mean - means of a continuous process variable

"R" range ranges of a continuous process variable

"S" standard deviation standard deviations of a continuous variable

"xbar.one" mean one-at-time data of a continuous process variable

"p" proportion proportion of nonconforming units

"np" count number of nonconforming units

"c" count nonconformities per unit

"u" count average nonconformities per unit

"g" count number of non-events between events

3.2.3 Pareto Chart Analysis.

- A Pareto chart is a barplot where the categories are ordered in non increasing order, and a line is also added to show the cumulative sum.
- Quality problems are rarely spread evenly across the different aspects of the production process or different plants. Rather, a few "bad apples" often account for the majority of problems.
- This principle has come to be known as the Pareto principle, which basically states that quality losses are mal-distributed in such a way that a small percentage of possible causes are responsible for the majority of the quality problems.
- For example, a relatively small number of "dirty" cars are probably responsible for the majority of air pollution; the majority of losses in most companies result from the failure of only one or two products. To illustrate the "bad apples", one plots the Pareto chart,

3.2.4 Pareto Analysis (Implementation with qcc package)

```
defect <- c(80, 27, 66, 94, 33)

names(defect) <- c("price code", "schedule date",
 "supplier code", "contact num.", "part num.")

# 1
pareto.chart(defect, ylab = "Error frequency")

#2
pareto.chart(defect, ylab = "Error frequency", xlab = "Error causes", las=1)

#3
pareto.chart(defect, ylab = "Error frequency", col=rainbow(length(defect)))

#4
pareto.chart(defect, cumperc = seq(0, 100, by = 5),
    ylab2 = "A finer tickmarks grid")
```

Output to accompany graphs

Pareto chart analysis for defect

	Frequency	Cum.Freq.	Percentage	Cum.Percent.
contact num.	94	94	31.33333	31.33333
price code	80	174	26.66667	58.00000
supplier code	66	240	22.00000	80.00000
part num.	33	273	11.00000	91.00000
schedule date	27	300	9.00000	100.00000

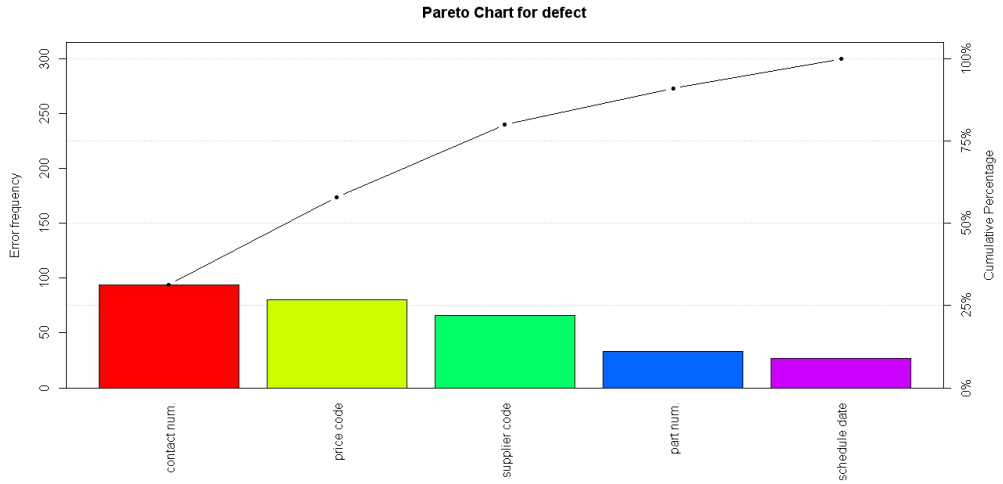


Figure 3.2: Third Implementation

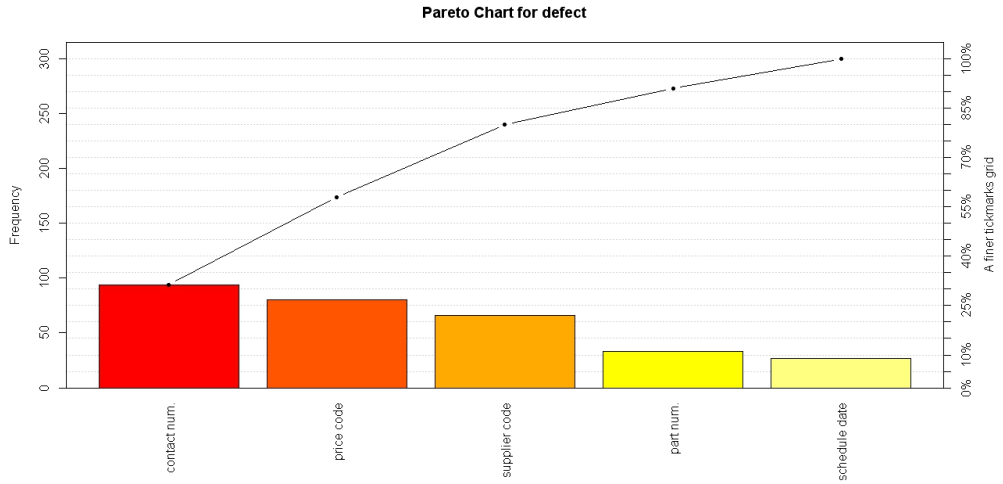


Figure 3.3: Fourth Implementation

3.2.5 Cause and Effect Diagrams

The cause and effect diagram is also known as "Ishikawa diagram", and has been widely used in Quality Management. It is one of the Seven Basic Tools of Quality.

```
cause.and.effect(cause=list(
  Measurements=c("Micrometers", "Microscopes", "Inspectors"),
  Materials=c("Alloys", "Lubricants", "Suppliers"),
  Personnel=c("Shofts", "Supervisors", "Training", "Operators"),
  Environment=c("Condensation", "Moisture"),
  Methods=c("Brake", "Engager", "Angle"),
  Machines=c("Speed", "Lathes", "Bits", "Sockets")),
effect="Surface Flaws")
```

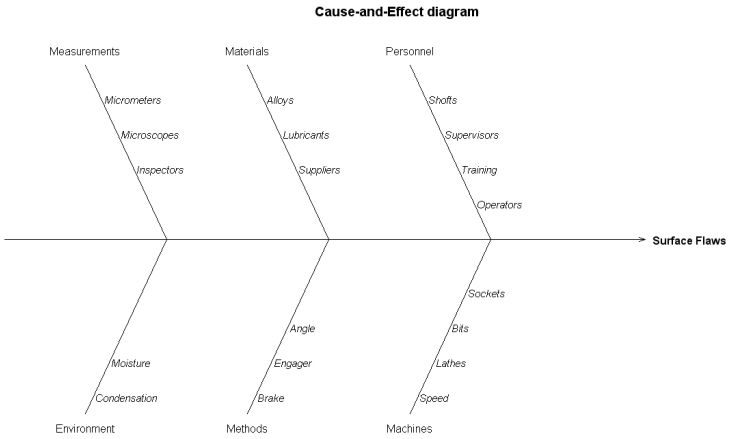


Figure 3.4:

Implementation with Six Sigma Package

```
effect <- "Flight Time"
causes.gr <- c("Operator", "Environment", "Tools", "Design",
"Raw.Material", "Measure.Tool")
causes <- vector(mode = "list", length = length(causes.gr))
causes[1] <- list(c("operator #1", "operator #2", "operator #3"))
causes[2] <- list(c("height", "cleaning"))
causes[3] <- list(c("scissors", "tape"))
causes[4] <- list(c("rotor.length", "rotor.width2", "paperclip"))
causes[5] <- list(c("thickness", "marks"))
causes[6] <- list(c("calibrate", "model"))
ss.ceDiag(effect, causes.gr, causes, sub = "Paper Helicopter Project")
```

3.2.6 Constructing Process Maps

```
inputs.overall<-c("operators", "tools", "raw material", "facilities")
outputs.overall<-c("helicopter")
steps<-c("INSPECTION", "ASSEMBLY", "TEST", "LABELING")
```

```
#Inputs of process "i" are inputs of process "i+1"
input.output<-vector(mode="list",length=length(steps))
input.output[1]<-list(c("sheets", "..."))
input.output[2]<-list(c("sheets"))
input.output[3]<-list(c("helicopter"))
input.output[4]<-list(c("helicopter"))
```

Parameters of each process

```
x.parameters<-vector(mode="list",length=length(steps))
```

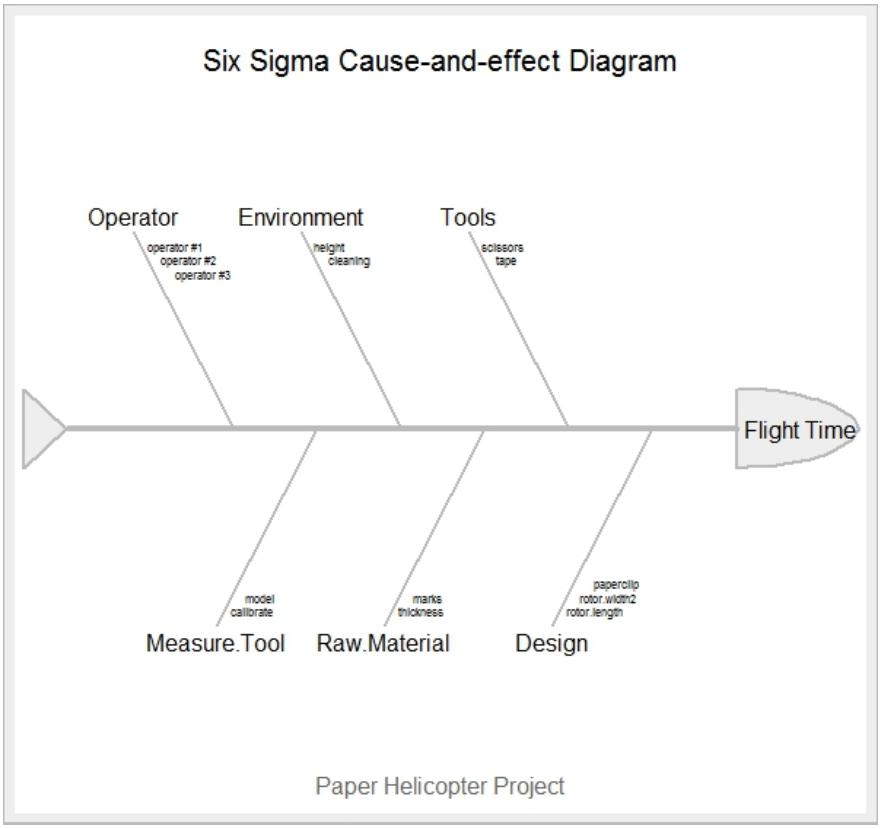


Figure 3.5:

```
x.parameters[1]<-list(c(list(c("width", "NC")),list(c("operator", "C")),
                        list(c("Measure pattern", "P")), list(c("discard", "P"))))
x.parameters[2]<-list(c(list(c("operator", "C")),list(c("cut", "P")),
                        list(c("fix", "P")), list(c("rotor.width", "C")),
                        list(c("rotor.length",
                                                              list(c("paperclip", "C"))
x.parameters[3]<-list(c(list(c("operator", "C")),
list(c("throw", "P")),
                        list(c("discard", "P")),
                        list(c("environment", "N"))))
x.parameters[4]<-list(c(list(c("operator", "C")),
list(c("label", "P"))))
```

```
x.parameters
```

```
#Features of each process
y.features<-vector(mode="list",length=length(steps))
y.features[1]<-list(c(list(c("ok", "Cr"))))
y.features[2]<-list(c(list(c("weight", "Cr"))))
y.features[3]<-list(c(list(c("time", "Cr"))))
y.features[4]<-list(c(list(c("label", "Cr"))))
y.features
ss.pMap(steps, inputs.overall, outputs.overall,
        input.output, x.parameters, y.features,
        sub="Paper Helicopter Project")
```

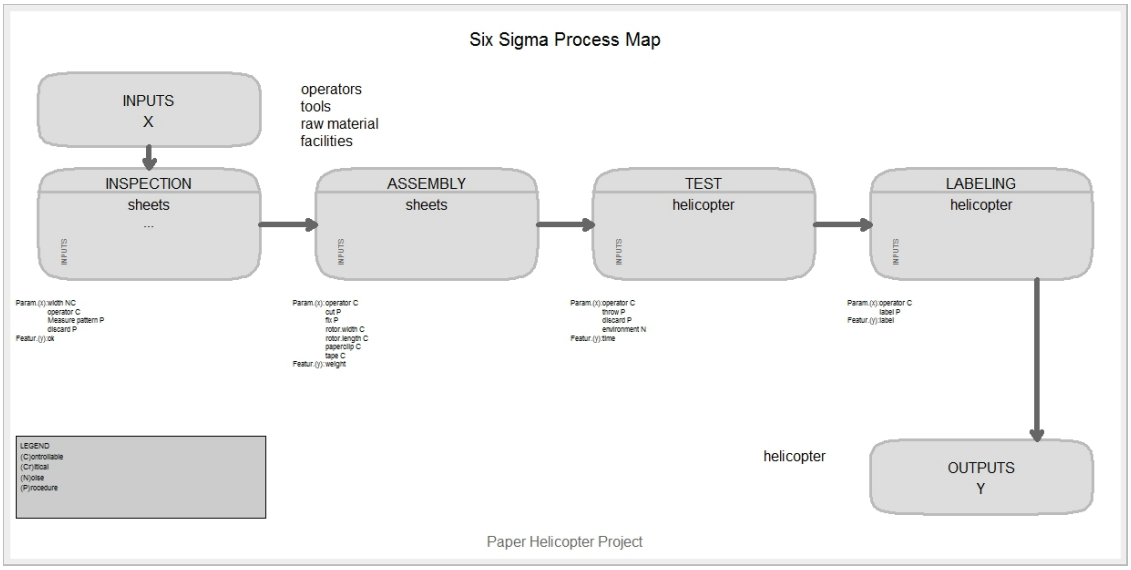


Figure 3.6:

3.3 The qcc R package - Other Types of Graph

3.3.1 Operating Characteristic (OC) Curves

- The OC Curve is used in sampling inspection. It plots the probability of accepting a batch of items against the quality level of the batch.
- The figure below shows an 'OC' (Operating Characteristic) Curve for a sample of 50 items taken from a batch of 2000 and using a critical acceptance number 'c' of 2 (the batch will be accepted if there are two or less defectives in the sample).
- From the curve you can see that there is about a 23% probability of accepting a batch that contains 8% of defective items.
- When designing a sampling plan it is usual to decide on two points, the AQL and LQL and the associated Producer's Risk and Consumer's Risk. The necessary sample size and acceptance number for the curve to pass through these points is then calculated and hence the shape of the curve.
- OC Curves are mainly associated with sampling inspection but they are also used to find the Average Run Length in control charts.
- A common supplementary plot to standard quality control charts is the so-called operating characteristic or OC curve (see example below). One question that comes to mind when using standard variable or attribute charts is how sensitive is the current quality control procedure? Put in more specific terms, how likely is it that you will not find a sample (e.g., mean in an X-bar chart) outside the control limits (i.e., accept the production process as "in control"), when, in fact, it has shifted by a certain amount?
- This probability is usually referred to as the (beta) error probability, that is, the probability of erroneously accepting a process (mean, mean proportion, mean rate defectives, etc.) as being "in control."
- Note that operating characteristic curves pertain to the false-acceptance probability using the sample-outside-of- control-limits criterion only, and not the runs tests described earlier.
- Operating characteristic curves are extremely useful for exploring the power of our quality control procedure. The actual decision concerning sample sizes should depend not only on the cost of implementing the plan (e.g., cost per item sampled), but also on the costs resulting from not detecting quality problems. The OC curve allows the engineer to estimate the probabilities of not detecting shifts of certain sizes in the production quality.

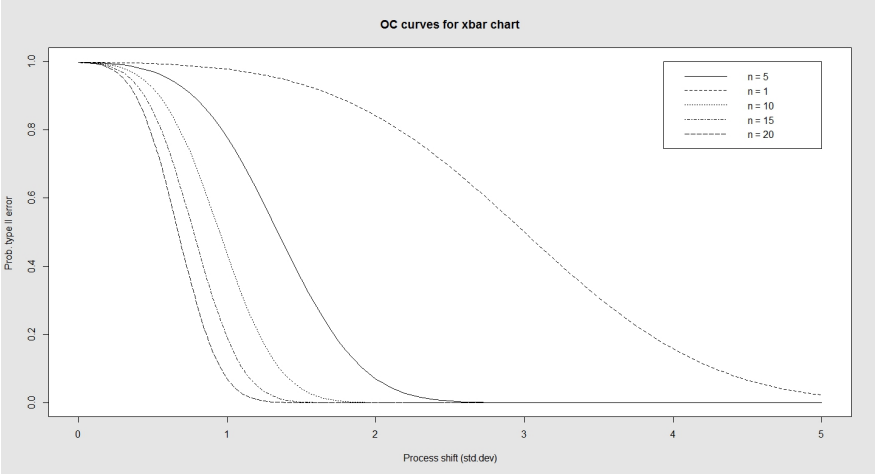


Figure 3.7:

pistonrings Data

```
data(pistonrings); attach(pistonrings);

diameter <- qcc.groups(diameter, sample)
beta <- oc.curves.xbar(qcc(diameter, type="xbar", nsigmas=3, plot=FALSE))
print(round(beta, digits=4))

# or to identify points on the plot use
## Not run: oc.curves.xbar(qcc(diameter,
    type="xbar", nsigmas=3, plot=FALSE), identify=TRUE)

detach(pistonrings)
```

3.3.2 Moving Average (MA) Chart

- To return to the piston ring example, suppose we are mostly interested in detecting small trends across successive sample means.
- For example, we may be particularly concerned about machine wear, leading to a slow but constant deterioration of quality (i.e., deviation from specification).
- Another way is to use some weighting scheme that summarizes the means of several successive samples; moving such a weighted mean across the samples will produce a moving average chart (as shown in the following graph).

GExponentially-weighted MA (EWMA) Chart

```
data(pistonrings)
attach(pistonrings)
diameter <- qcc.groups(diameter, sample)
q <- ewma(diameter[1:25,], lambda=0.2, nsigmas=3)
summary(q)

q <- ewma(diameter[1:25,], lambda=0.2, nsigmas=2.7,
newdata=diameter[26:40,], plot = FALSE)
summary(q)

plot(q)
```

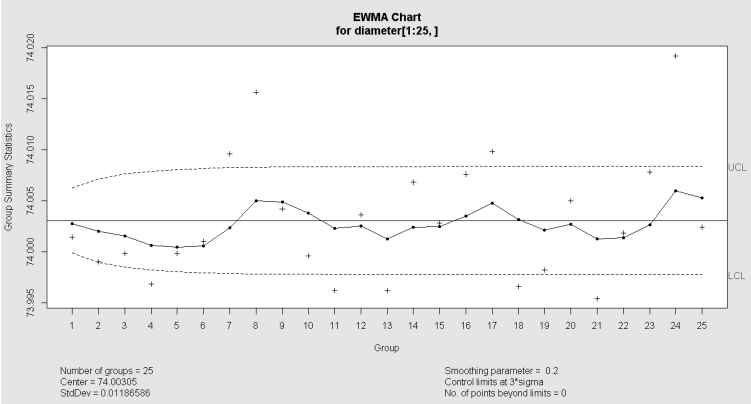


Figure 3.8:

```
> summary(q)

Call:
ewma(data = diameter[1:25, ], lambda = 0.2, nsigmas = 2.7,
newdata = diameter[26:40,     ], plot = FALSE)

ewma chart for diameter[1:25, ]
```

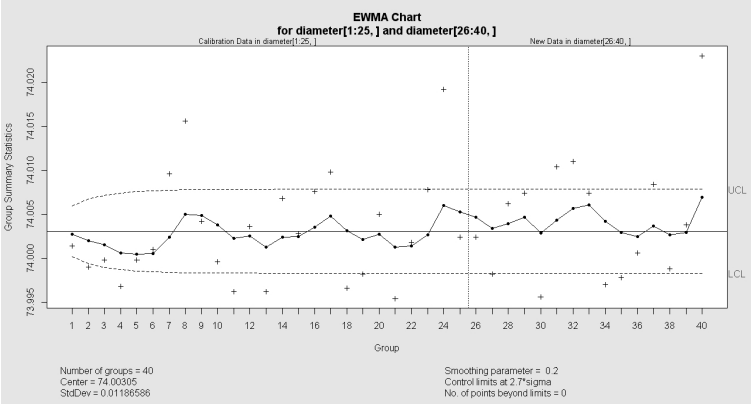


Figure 3.9:

```
Summary of group statistics:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  74.00   74.00   74.00   74.00   74.01   74.02

Group sample size:  5
Number of groups:  25
Center of group statistics:  74.00305
Standard deviation:  0.01186586

Summary of group statistics in diameter[26:40, ]:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  74.00   74.00   74.00   74.00   74.01   74.02

Group sample size:  5
Number of groups:  15

Smoothing parameter: 0.2
Control limits:
         LCL      UCL
1   74.00018 74.00591
2   73.99938 74.00672
...
40  73.99827 74.00782
```

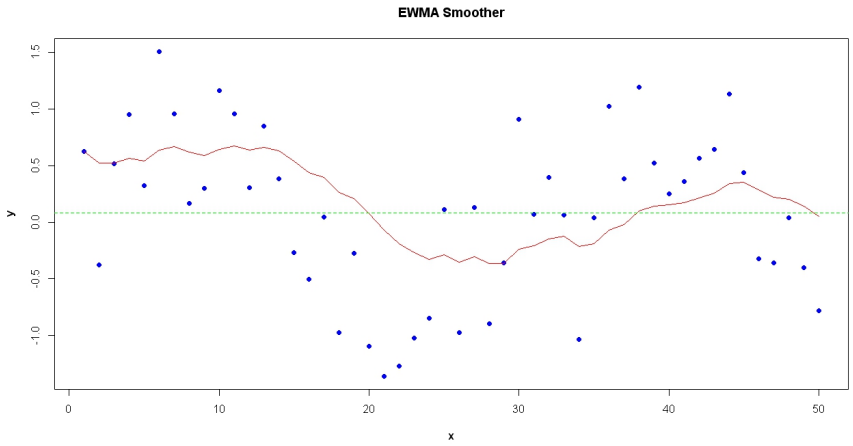


Figure 3.10:

EWMA : Individual observations

```
x <- c(33.75, 33.05, 34, 33.81, 33.46, 34.02, 33.68, 33.27, 33.49, 33.20,
33.62, 33.00, 33.54, 33.12, 33.84) # viscosity data (Montgomery, pag. 242)
q <- ewma(x, lambda=0.2, nsigmas=2.7)
summary(q)
```

```
x <- 1:50
y <- rnorm(50, sin(x/5), 0.5)
plot(x,y,pch=16,col="blue",font.lab=2)
lines(ewmaSmooth(x,y,lambda=0.1), col="red")
abline(h=mean(y),col="green",lty=2)
title("EWMA Smoother")
```

3.3.3 CUSUM charts

- CUSUM charts, while not as intuitive and simple to operate as Shewhart charts, have been shown to be more efficient in detecting small shifts in the mean of a process.
- In particular, analyzing ARL's for CUSUM control charts shows that they are better than Shewhart control charts when it is desired to detect shifts in the mean that are 2 sigma or less.

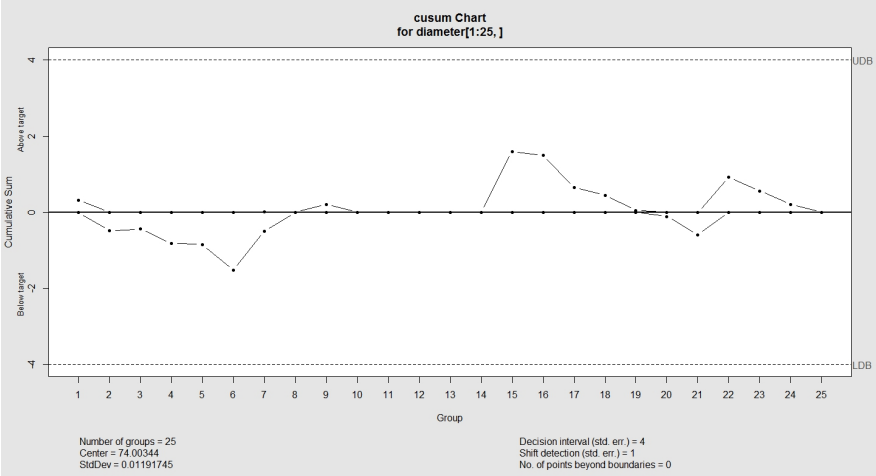


Figure 3.11:

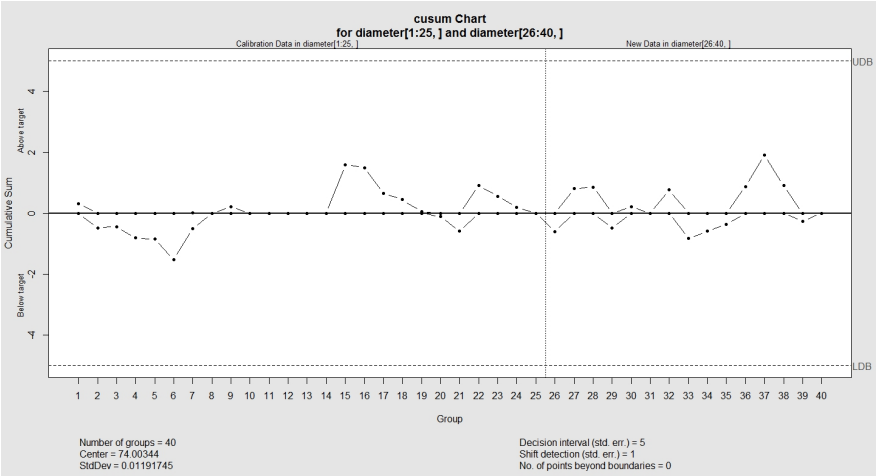


Figure 3.12:

3.3.4 Individual/Moving-Range chart

- The individual/moving-range chart is a type of control chart used to monitor variables data from a business or industrial process for which it is impractical to use rational subgroups.
- The chart is necessary in the following situations:
  - ∗ Where automation allows inspection of each unit, so rational subgrouping has less benefit.
  - ∗ Where production is slow so that waiting for enough samples to make a rational subgroup unacceptably delays monitoring
  - ∗ For processes that produce homogeneous batches (e.g., chemical) where repeat measurements vary primarily because of measurement error
- The "chart" actually consists of a pair of charts: one, the individuals chart, displays the individual measured values; the other, the moving range chart, displays the difference from one point to the next.
- As with other control charts, these two charts enable the user to monitor a process for shifts in the process that alter the mean or variance of the measured statistic.

Chapter 4

Process Capability

4.1 Process Capability

Process capability is the measure of process performance. Capability refers to the ability of a process to make parts that are well within engineering specifications. A capability study is done to answer the questions, *"Does the process need to be improved?"* and *"How much does the process need to be improved?"*

To define the study of process capability from another perspective, a capability study is a technique for analyzing the random variability found in a production process. In every manufacturing process there is variability. This variability may be large or small, but it is always present. It can be divided into two types:

- Variability due to common (random) causes
- Variability due to assignable (special) causes

The first type of variability can be expected to occur naturally within a process. It is attributed to common causes that behave like a constant system of chances. These chances form a unique and describable distribution. This variability can never be completely eliminated from a process. Variability due to assignable causes, on the other hand, refers to the variation that can be linked to specific or special causes. If these causes, or factors, are modified or controlled properly, the process variability associated with them can be eliminated. Assignable causes cannot be described by a single distribution.

4.1.1 Capability Study

A capability study measures the performance potential of a process when no assignable causes are present (when it is in statistical control). Since the inherent variability of the process can be described by a unique distribution, usually a normal distribution, capability can be

evaluated by utilizing this distributions properties. Simply put, capability is expressed as the proportion of in-specification process output to total process input.

Capability calculations allow predictions to be made regarding quality, enabling manufacturers to take a preventive approach to defects. This statistical approach contrasts to the traditional approach to manufacturing, which is a two-step process: production personnel make the product, and quality control personnel inspect and eliminate those products that do not meet specifications. This is wasteful and expensive, since it allows time and materials to be invested in products that are not always usable. It is also unreliable, since even 100% inspection would fail to catch all defective products. Control Limits are Not an Indication of Capability

Those new to SPC often believe they dont need capability indices. They think they can compare the control limits to the specification limits instead. This is not true, because control limits look at the distribution of averages and capability indices look at the distribution of individuals. The distribution of individuals will always spread out further than the distribution of averages (see figure below).

4.1.2 What is Process Capability?

Distribution of averages compared to distribution of individuals, for the same sample data. Control limits (based on averages) would probably be inside specification limits, even though many parts are out of specification. This shows why you should not compare control limits to specification limits.

Therefore, the control limits are often within the specification limits, but the ±3 Sigma distribution of parts is not. Subgroup averages follow more closely a normal distribution. This is why we can create control charts for processes that are not normally distributed. But averages cannot be used for capability calculations, because capability concerns itself with individual parts, or samples from a process. After all, parts, not averages, get shipped.

4.1.3 Capability Indices

Capability The uniformity of product which a process is capable of producing. Can be expressed numerically using CP, CR, CpK, and Zmax/3 when the data is normally distributed.

CP For process capability studies: CP is a capability index defined by the formula. CP shows the process capability potential but does not consider how centered the process is. CP may range in value from 0 to infinity, with a large value indicating greater potential capability. A value of 1.33 or greater is usually desired.

CR For process capability studies: the inverse of CP, CR can range from 0 to infinity in value, with a smaller value indicating a more capable process.

CpK For process capability studies: an index combining CP and K to indicate whether the process will produce units within the tolerance limits. CpK has a value equal to CP if the process is centered on the nominal; if CpK is negative, the process mean is outside the specification limits; if CpK is between 0 and 1, then some of the 6 sigma spread falls outside the tolerance limits. If CpK is larger than 1, the 6 sigma spread is completely within the tolerance limits. A value of 1.33 or greater is usually desired.

4.1.4 Interpreting Capability Indices

- The greater the CpK value, the better. A CpK greater than 1.0 means that the $6\sigma(\pm 3\sigma)$ spread of the data falls completely within the specification limits. A CpK of 1.0 means that one end of the 6σ spread falls on a specification limit. A CpK between 0 and 1 means that part of the 6σ spread falls outside the specification limits. A negative CpK indicates that the mean of the data is not between the specification limits.
- Since a CpK of 1.0 indicates that 99.73% of the parts produced are within specification limits, in this process it is likely that only about 3 out of 1,000 need to be scrapped or rejected. Why bother to improve the process beyond this point, since it will produce no reduction in scrap or reject costs? Improvement beyond just meeting specification may greatly improve product performance, cut warranty costs, or avoid assembly problems.
- Many companies are demanding CpK indexes of 1.33 or 2.0 of their suppliers products. A CpK of 1.33 means that the difference between the mean and specification limit is 4σ (since 1.33 is 4/3). With a CpK of 1.33, 99.994% of the product is within specification. Similarly a CpK of 2.0 is 6σ between the mean and specification limit (since 2.0 is 6/3).
- This improvement from 1.33 to 2.0 or better is sometimes justified to produce more product near the optimal target. Depending on the process or part, this may improve product performance, product life, customer satisfaction, or reduce warranty costs or assembly problems.
- Continually higher CpK indexes for every part or process is not the goal, since that is almost never economically justifiable. A cost/benefit analysis that includes customer satisfaction and other true costs of quality is recommended to determine which processes should be improved and how much improvement is economically attractive.

4.2 Process Capability

Process capability is the measure of process performance. Capability refers to the ability of a process to make parts that are well within engineering specifications. A capability study is done to answer the questions, *"Does the process need to be improved?"* and *"How much does the process need to be improved?"*

To define the study of process capability from another perspective, a capability study is a technique for analyzing the random variability found in a production process. In every manufacturing process there is variability. This variability may be large or small, but it is always present. It can be divided into two types:

- Variability due to common (random) causes
- Variability due to assignable (special) causes

The first type of variability can be expected to occur naturally within a process. It is attributed to common causes that behave like a constant system of chances. These chances form a unique and describable distribution. This variability can never be completely eliminated from a process. Variability due to assignable causes, on the other hand, refers to the variation that can be linked to specific or special causes. If these causes, or factors, are modified or controlled properly, the process variability associated with them can be eliminated. Assignable causes cannot be described by a single distribution.

4.2.1 Capability Study

- A capability study measures the performance potential of a process when no assignable causes are present (when it is in statistical control). Since the inherent variability of the process can be described by a unique distribution, usually a normal distribution, capability can be evaluated by utilizing this distributions properties.
- Simply put, capability is expressed as the proportion of in-specification process output to total process input.
- Capability calculations allow predictions to be made regarding quality, enabling manufacturers to take a preventive approach to defects. This statistical approach contrasts to the traditional approach to manufacturing, which is a two-step process: production personnel make the product, and quality control personnel inspect and eliminate those products that do not meet specifications.
- This is wasteful and expensive, since it allows time and materials to be invested in products that are not always usable. It is also unreliable, since even 100% inspection would fail to catch all defective products.
- Control Limits are Not an Indication of Capability
- Those new to SPC often believe they dont need capability indices. They think they can compare the control limits to the specification limits instead.
- This is not true, because control limits look at the distribution of averages and capability indices look at the distribution of individuals. The distribution of individuals will always spread out further than the distribution of averages.

4.2.2 What is Process Capability?

Distribution of averages compared to distribution of individuals, for the same sample data. Control limits (based on averages) would probably be inside specification limits, even though many parts are out of specification. This shows why you should not compare control limits to specification limits.

Therefore, the control limits are often within the specification limits, but the ± 3 Sigma distribution of parts is not. Subgroup averages follow more closely a normal distribution. This is why we can create control charts for processes that are not normally distributed. But averages cannot be used for capability calculations, because capability concerns itself with individual parts, or samples from a process. After all, parts, not averages, get shipped.

4.2.3 Capability Indices

Capability The uniformity of product which a process is capable of producing. Can be expressed numerically using CP, CR, CpK, and Zmax/3 when the data is normally distributed.

CP For process capability studies: CP is a capability index defined by the formula. CP shows the process capability potential but does not consider how centered the process is. CP may range in value from 0 to infinity, with a large value indicating greater potential capability. A value of 1.33 or greater is usually desired.

CR For process capability studies: the inverse of CP, CR can range from 0 to infinity in value, with a smaller value indicating a more capable process.

CpK For process capability studies: an index combining CP and K to indicate whether the process will produce units within the tolerance limits. CpK has a value equal to CP if the process is centered on the nominal; if CpK is negative, the process mean is outside the specification limits; if CpK is between 0 and 1, then some of the 6 sigma spread falls outside the tolerance limits. If CpK is larger than 1, the 6 sigma spread is completely within the tolerance limits. A value of 1.33 or greater is usually desired.

4.2.4 Interpreting Capability Indices

- The greater the CpK value, the better. A CpK greater than 1.0 means that the $6\sigma(\pm 3\sigma)$ spread of the data falls completely within the specification limits. A CpK of 1.0 means that one end of the 6σ spread falls on a specification limit. A CpK between 0 and 1 means that part of the 6σ spread falls outside the specification limits. A negative CpK indicates that the mean of the data is not between the specification limits.
- Since a CpK of 1.0 indicates that 99.73% of the parts produced are within specification limits, in this process it is likely that only about 3 out of 1,000 need to be scrapped or rejected. Why bother to improve the process beyond this point, since it will produce no reduction in scrap or reject costs? Improvement beyond just meeting specification may greatly improve product performance, cut warranty costs, or avoid assembly problems.
- Many companies are demanding CpK indexes of 1.33 or 2.0 of their suppliers products. A CpK of 1.33 means that the difference between the mean and specification limit is 4σ (since 1.33 is 4/3). With a CpK of 1.33, 99.994% of the product is within specification. Similarly a CpK of 2.0 is 6σ between the mean and specification limit (since 2.0 is 6/3).
- This improvement from 1.33 to 2.0 or better is sometimes justified to produce more product near the optimal target. Depending on the process or part, this may improve product performance, product life, customer satisfaction, or reduce warranty costs or assembly problems.
- Continually higher CpK indexes for every part or process is not the goal, since that is almost never economically justifiable. A cost/benefit analysis that includes customer satisfaction and other true costs of quality is recommended to determine which processes should be improved and how much improvement is economically attractive.

4.2.5 Process Capability Analysis

- Process capability compares the output of an in-control process to the specification limits by using capability indices.
- The comparison is made by forming the ratio of the spread between the process specifications (the specification "width") to the spread of the process values, as measured by 6 process standard deviation units (the process "width").

4.2.6 Intepreting Process Capability Indices

- **CP**
  Historically, this is one of the first capability indexes used. The "natural tolerance" of the process is computed as 6s . The index simply makes a direct comparison of the process natural tolerance to the engineering requirements. Assuming the process distribution is normal and the process average is exactly centered between the engineering requirements, a CP index of 1 would give a "capable process." However, to allow a bit of room for process drift, the generally accepted minimum value for CP is 1.33. In general, the larger CP is, the better. The CP index has two major shortcomings. First, it cannot be used unless there are both upper and lower specifications. Second, it does not account for process centering. If the process average is not exactly centered relative to the engineering requirements, the CP index will give misleading results. In recent years, the CP index has largely been replaced by CPK (see below).

- **CPM**
  A CPM of at least 1 is required, and 1.33 is preferred. CPM is closely related to CP. The difference represents the potential gain to be obtained by moving the process mean closer to the target. Unlike CPK, the target need not be the center of the specification range.

4.3 Process Capability Analysis

Process Capability Indices for a characteristic of interest from a continuous process can be obtained using the command process.capability.

Lower and Upper Specification Limits must be specified.

$$c_p = \frac{\text{USL} - \text{LSL}}{6 \times s} = \frac{12}{6 \times 1.956} = 1.02$$

$$c_{pk}(\text{upper}) = \frac{\text{USL} - \bar{x}}{3 \times s} = \frac{506 - 500.38}{3 \times 1.956} = 0.96$$

$$c_{pk}(\text{lower}) = \frac{\bar{x} - \text{USL}}{3 \times s} = \frac{500.38 - 494}{3 \times 1.956} = 104$$

4.4 Worked Examples

4.4.1 MA6001 Autumn 2007/2008 Question 4

Over several weeks of normal, or in-control, operation, 20 samples of 150 packages each of synthetic-gut tennis strings were tested for breaking strength. A total of 141 packages of the 3000 tested failed to conform to the manufacturer's specifications.

(i.) What is an estimate of the process proportion defective when the system is in control?

(ii.) Compute the upper and lower control limits for an np chart.

(iii.) Using the results of part (ii) above, what conclusion should be made about the process if tests on a new sample of 150 packages find 12 defective?

A quality control process monitors the weight per carton of laundry detergent. Control limits for the mean are set at $UCL = 20.12$ ounces and $LCL = 19.90$ ounces. Samples of size 5 are used for the sampling and inspection process.

(i.) What are the process mean and process standard deviation for the manufacturing operation?

(ii.) If the process mean shifts to 20.25 ounces, what is the probability of a type II error?

For a single sampling plan $50(0, 1)$ what is the probability of accepting the batch when

(i.) $p = 0.01$

(ii.) $p = 0.02$

4.4.2 MA6001 Autumn 2008/2009 Question 4

Lear Seating of Kitchener, Ontario, manufactures seats for Chrysler, Ford, and GeneralMotors cars. Several years ago, Lear instituted statistical process control, which has resulted in improved quality and lower costs. One of the components of a front-seat cushion is a wire spring, produced from 4-mm steel wire. A machine is employed to bend the wire so that the spring's length is 500mm. If the springs are longer than 500mm, they will loosen and eventually fall out.If they are too short, they won't easily fit into position.(In fact, in the past, when there was a relatively large number of short springs, workers incurred arm and hand injuries when attempting to install the springs.)

To determine if the process is under control, random samples of four springs are taken every two hours. The results of 25 samples taken over the last week gave the following overall results:

(i.) Calculate the centreline and control limits for the $\bar{\bar{x}}$ and s charts.

(ii.) Assuming the process is in control and given that the tolerances are set at $500 \pm 6mm$, calculate capability indices and comment.

(iii.) Calculate the probability of a type II error if the process mean shifts to 501mm. item[(iv.)] What is the average run length (ARL) for the situation outlined at (iii) above?

Chapter 5

SixSigma

5.1 Six Sigma - Background to Six Sigma

- Motorola, one of the worlds leading manufacturers and suppliers of semiconductors and electronic equipment systems for civil and military applications, introduced the concept of six-sigma process quality to enhance the reliability and quality of their products, and cut product cycle times and expenditure on test/repair. Motorola used the following statement to explain:
- Sigma is a statistical unit of measurement that describes the distribution about the mean of any process or procedure. A process or procedure that can achieve plus or minus six-sigma capability can be expected to have a defect rate of no more than a few parts per million, even allowing for some shift in the mean. In statistical terms, this approaches zero defects.
- The approach was championed by Motorolas chief executive officer at the time, Bob Galvin, to help improve competitiveness. The six-sigma approach became widely publicized when Motorola won the US Baldrige National Quality Award in 1988.
- Six-sigma is a disciplined approach for improving performance by focusing on producing better products and services faster and cheaper. The emphasis is on improving the capability of processes through rigorous data gathering, analysis and action, and:
  - enhancing value for the customer;
  - eliminating costs which add no value (waste).

  Unlike simple cost-cutting programmes six-sigma delivers cost savings whilst retaining or even improving value to the customers.

5.1.1 Why six-sigma?

In a process in which the characteristic of interest is a variable, defects are usually defined as the values which fall outside the specification limits (LSLUSL). Assuming and using a normal distribution of the variable, the percentage and/or parts per million defects can be found.

For example, in a centred process with a specification set at $x \pm 3\sigma$ there will be 0.27 per cent or 2700 ppm defects. This may be referred to as an unshifted $\pm 3\sigma$ process and the quality called 3 sigma quality. In an unshifted 6 sigma process, the specification range is x 6 and it produces only 0.002 ppm defects.

It is difficult in the real world, however, to control a process so that the mean is always set at the nominal target value in the centre of the specification. Some shift in the process mean is expected.

5.2 Six Sigma R Package

5.2.1 DMAIC Process

- DMAIC (an abbreviation for Define, Measure, Analyze, Improve and Control) refers to a data-driven improvement cycle used for improving, optimizing and stabilizing business processes and designs.
- The DMAIC improvement cycle is the core tool used to drive Six Sigma projects.
- However, DMAIC is not exclusive to Six Sigma and can be used as the framework for other improvement applications.

5.2.2 Data Set - `ss.data.ca`

This data set is the volume measured in 20 bottles for a filling process in a winery

```
ss.data.ca
```

```
> ss.data.ca
   Volume
1  755.81
2  750.54
```

```
........
........
19 750.26
20 751.29
```

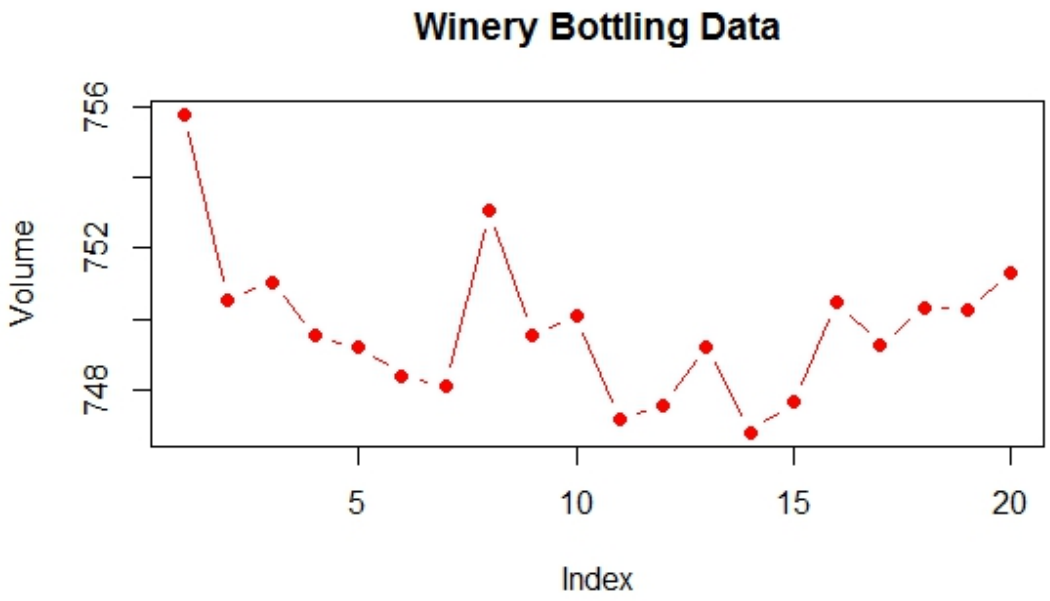


Figure 5.1:

5.2.3 `ss.ca.z` Capability Indices

Compute the Capability Indices of a process, Z (Sigma Score), C_p and C_{pk}.

```
> ss.ca.cp(ss.data.ca$Volume,740, 760)
[1] 1.584136
> ss.ca.cpk(ss.data.ca$Volume,740, 760)
[1] 1.546513
> ss.ca.z(ss.data.ca$Volume,740,760)
[1] 3.139539
```

5.2.4 Loss Fucntion Aanalysis

The Taguchi Loss Function is graphical depiction of loss developed by the Japanese business statistician Genichi Taguchi to describe a phenomenon affecting the value of products produced by a company. Praised by Dr. W. Edwards Deming (the business guru of the 1980s American quality movement),[1] it made clear the concept that quality does not suddenly plummet when, for instance, a machinist exceeds a rigid blueprint tolerance. Instead "loss" in value progressively increases as variation increases from the intended condition. This was considered a breakthrough in describing quality, and helped fuel the continuous improvement movement that since has become known as lean manufacturing.

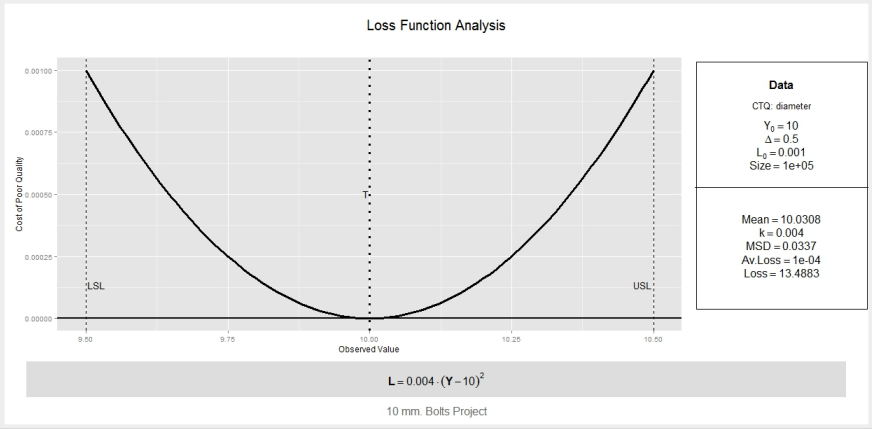


Figure 5.2:

5.3 Other R Packages

1. mpci
2. mspc
3. qualityTools
4. spc
5. SixSigma with R (Emilio Cano)

5.3.1 spc: Statistical Process Control

- Evaluation of control charts by means of the zero-state, steady-state ARL (Average Run Length) and RL quantiles.
- Setting up control charts for given in-control ARL.
- The control charts under consideration are one- and two-sided EWMA, CUSUM, and Shiryaev-Roberts schemes for monitoring the mean of normally distributed independent data.
- ARL calculation of the same set of schemes under drift are added.
- Other charts and parameters are in preparation.
- Further SPC areas will be covered as well (sampling plans, capability indices ...)

R package spc provides

```
xcusum.ad steady-state ARLs of CUSUM charts
xcusum.arl (zero-state) ARLs of CUSUM charts
xcusum.crit decision intervals of CUSUM charts
xewma.ad steady-state ARLs of EWMA charts
xewma.arl (zero-state) ARLs of EWMA charts
xewma.crit critical values of EWMA charts
```