Chapter 3
Uncovering Underlying Pattern
Blurred by Sampling Instability

Section 3.3
Pattern and Process

© 2020 by David W. Gerbing

School of Business Administration
Portland State University

- Pattern and Process
  - Stable Process Model
  - Control Chart – Individual Values
  - Control Chart of the Mean
  - Control Chart of the Standard Deviation
  - Tampering

3.3a
Stable Process Model

Variability

A key aspect of a process is the variability of its output

- Variation in the outcome of the product or service is inevitable
- Some variation results in marginally acceptable products or even completely unacceptable products
- The goal is to narrow the range of variation in order to achieve quality output virtually all of the time
- **Key Idea**: Improve the process by understanding the causes of variation, and then adjust these causes to reduce variation
- But before the reasons for the variability can be understood and corrected, first assess the variability

The Reality that Underlies Every Data Value

The concrete vs the abstract

- Every data value is generated by an underlying process with specific characteristics that include a mean at specific level and a specific variability about that mean
- The reality assessed with descriptive statistics is the concrete reality observed as the data, which reflects inherent random variation
- **Key Concept**: A more interesting reality than the randomly fluctuating data is the stable but underlying, abstract reality not directly observed, presumed to have generated the data
- The underlying reality refers to characteristics of the system that generated the data, the population values
- Population values include μ for population mean and σ for population standard deviation

Stable Process Model of Each Data Value

A more refined definition of a stable process

- The value of measurement Y_i follows from two components
  - **Random error**: Unique, random component of each data value Y_i
  - The process mean μ, which is shared by all of the data values the process generates
- **Stable Process Model** or system-in-control: Each data value, Y_i, is formed from a shared underlying stable component, the process mean, μ, plus a unique random error component, ϵ_i

$$Y_i = \mu + \epsilon_i \quad \text{where } \sigma_\epsilon \text{ is a constant for all } Y_i$$

- **Key Concept**: All variation of a stable system is due to common causes, which together contribute to the error variation, displacing each data value Y_i from the underlying mean, μ

Control Chart vs Run Chart

Control chart provides more information

- As seen in the previous chapter, a run chart is a plot of the values of process over time, traditionally with the median as the centerline for comparison of values
- The more refined version of the run chart is the control chart, both designed to detect any unwanted changes in a process
- **Control Chart**: A run chart with the mean as the center line with upper and lower control limits, which define the extent of a deviation from the mean that becomes an outlier
- Set the upper and lower limits at three times the standard deviation of the process, on either side of the mean
- For a normal distribution of output, 99.73% of the common cause variation fall within these limits

3.3b
Control Chart – Individual Values

R: Control Chart, Individual Values

qcc function in the qcc package

- Obtain the display of the control chart from the qcc function in the contributed package of functions, qcc
- The qcc package was accessed previously for the Pareto chart
- So the library should already have been downloaded with

```
> install.packages("qcc")
```

- Load the functions into the library, read the data

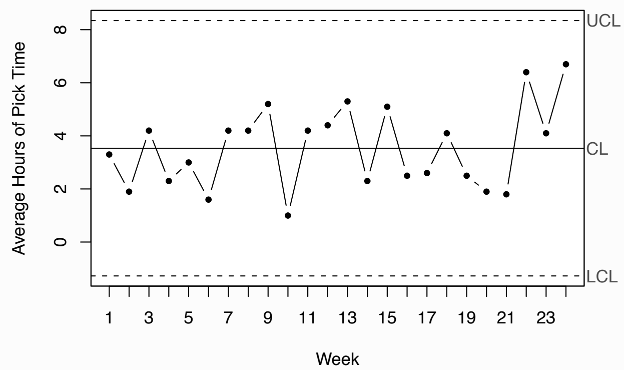
```
> library(qcc)
> d <- rd("http://lessRstats.com/data/pick.csv")
```

- Obtain the control chart with the qcc function
- Invoke the type="xbar.one" option for this chart to indicate that each displayed value is a *single* data value
- Later other types of control charts are introduced based on the analysis of *samples* of data values

Pattern: Stable or "Constant-Cause" System

System-in-control yields random deviations from center

```
> qcc(d$Hours, type="xbar.one", add.stats=FALSE,
  xlab="Week", ylab="Average Hours of Pick Time")
```



Analyze Out-of-Control or Unstable Processes

Look for a new process in addition to the reference process

- The existence of *any* pattern other than random deviation about the mean indicates an assignable cause
- The `qcc` function labels deviant points, which signal some potential assignable causes underling these data values
  - Red dot: Corresponding data value is beyond the upper or lower control limit
  - Yellow dot: Corresponding data value is an extension of a run beyond seven consecutive data values
- There are other possibilities as well, which include:
  - Two out of three successive points on the same side of the centerline and more than two standard deviations from the centerline
  - Six or more values all increasing or decreasing

Disentangle Stability from Randomness

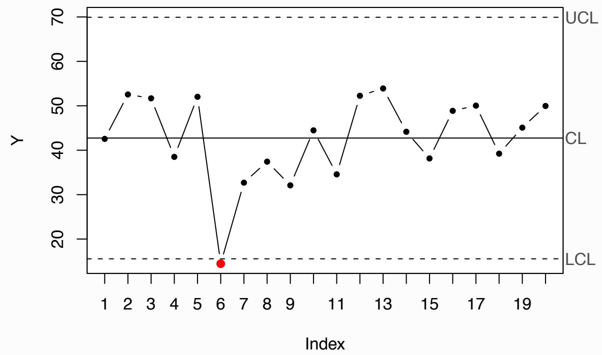
This separation is central to data analysis

- Regardless if the variation from a data value is due exclusively to common causes, or to common and assignable causes, random common cause variation is always present
- So there is a reasonable probability, but never certainty, if a deviant point is due to an assignable cause

Pattern: One Value from Another Process

Assignable cause yields a single outlier

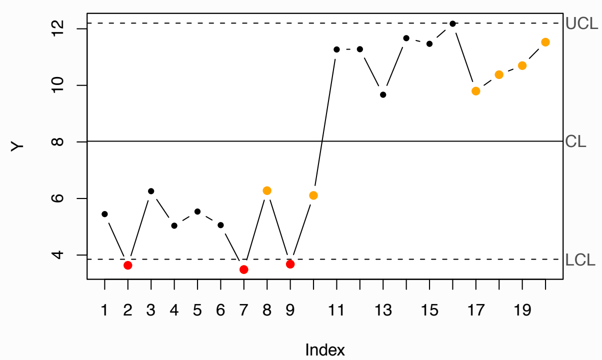
- Data value #6 is so small that it is *likely* from another process than that which generated the other data values



Pattern: Level Shift

Assignable cause due to formation of a new process

- Something happened to the process after Value 10, as the mean dramatically increased, *likely* indicating a new process



Need to Recalibrate

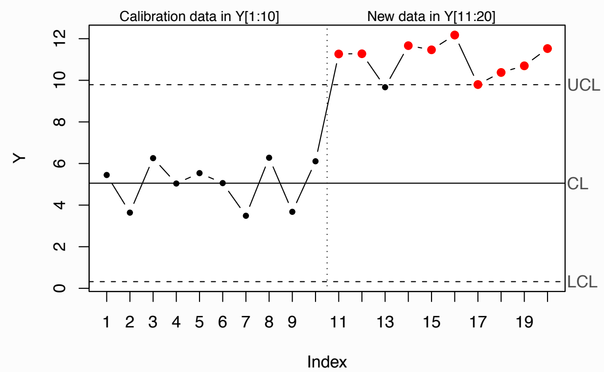
Control limits not meaningful for mixed processes

- In the preceding graph, many data values indicate a process out of control in reference to the meaningless control limits
- Instead, isolate the data influenced by an assignable cause and either remove, or, if software permits, base the analysis on just the output from the reference process of interest
- For the `qcc` function, invoke the `newdata` option, and specify both the original data the new data
- Split the data from the same file with the R index or sub-setting operation, indicated by the square brackets
  - d$Y[1:10] indicates just the first 10 values of Y
  - d$Y[11:20] indicates just the last 10 values of Y
- With re-calibration the comparison is explicit, the last 10 outcome values are mostly outliers in reference to the 1^{st} 10

Pattern: Level Shift, Recalibrate

Assignable cause due to formation of a new process

```
> qcc(d$Y[1:10], newdata=d$Y[11:20],
     type="xbar.one", xlab="Index", ylab="Y")
```



Pattern: Level Shift, Alternate Analysis

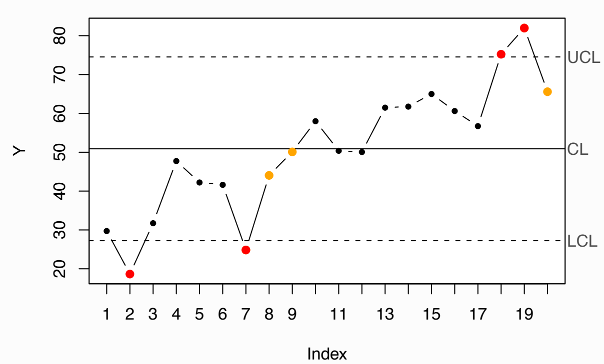
Can also formally compare group means with inference

- This R statement calculates the sample mean of the first ten values, `mean(Y[1:10])`, and this for the last 10 data values `mean(Y[11:20])`
- $m_1 = 5.055$ and $m_2 = 10.995$, the mean of the second group is more than twice the size of the first group mean
- The control chart demonstrates a differentiation of the means in these two sets of data
- To more formally demonstrate a distinction among the corresponding group means, μ_1 and μ_2, need the *t*-test of a mean difference from statistical inference from Sec 6.1
- Although not often done in practice, the stability of the process underlying each group is an assumption that should be verified before conducting the inferential test

Pattern: Trend

Assignable cause due to a new mean for every data value

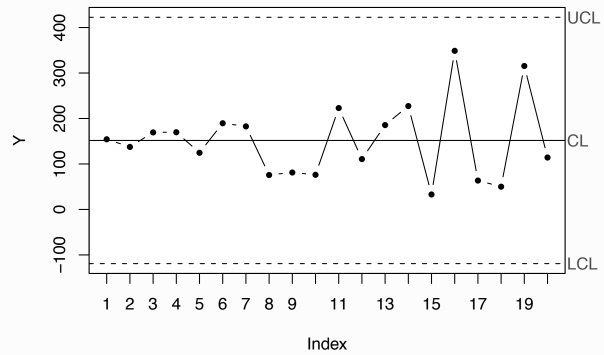
- The value of the underlying population mean continually increases over time, so a new process underlies every value



Pattern: Dispersion Shift

Assignable cause due to increasing variability

- As time goes on, the mean remains constant, but the process output becomes increasingly variable



Technical Note: Standard Deviation for the Control Chart

Cannot assume the process is stable

- To calculate the control limits, need an estimate of the process standard deviation
- If the process is not stable, the mean and/or standard deviation of the data may be changing, once, or many times
- So instead of an estimate of the usual standard deviation of all of the data about a common mean, use the moving range
- **Moving Range**: Mean of the successive ranges across all pairs of values, such as for $n = 20$
  - $|Y_2 - Y_1| + |Y_3 - Y_2| + \ldots + |Y_{20} - Y_{19}|$
  - Convert the sum to an average, divide by number of pairs
- It turns out that the corresponding estimate of the standard deviation of the data is this average moving range divided by the scaling constant of 1.128

3.3c
Control Chart of the Mean

The X-bar Chart

Processing sub-group means

- Previous examples of control charts were of individual data values, resulting in the I-chart
- The preference is for subgroups of data values, with, for example, the subgroup mean, *m*, replacing the individual data value in the I-chart
- **X-bar chart**: Control chart of the sample means from subgroups of data collected over time
- For time ordered data, have about 4 to 7 or so data values in each subgroup, with the data values in a subgroup assessed at approximately the same time so that only random variation separate the values within a group
- Multiple values per time period provides a more stable estimate of the system mean at that time, as well as provides an estimate of the standard deviation

The R X-bar Chart: Data

Data values for a specific time period appear in a single row

- The data values were collected over 23 days, three observations per day, so there are three columns and 24 rows of data in the data file, plus the column names in the first row

The first part of the data file in `.csv` format

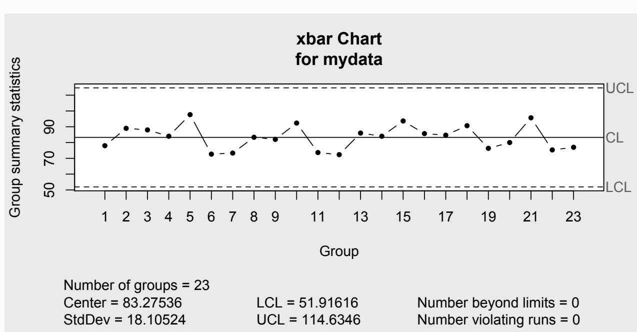
```
T1,T2,T3
86,73,75
90,82,95
101,74,89
```

- `d <- Read( "http://web.pdx.edu/~gerbing/data/Carey_p61.csv")`

The R X-bar Chart

- The qcc function `qcc`, here applied to all the variables in the data table `d` for an x-bar chart (sample mean, *m*)

```
> qcc(d, type="xbar")
```



3.3d
Control Chart of the Standard Deviation

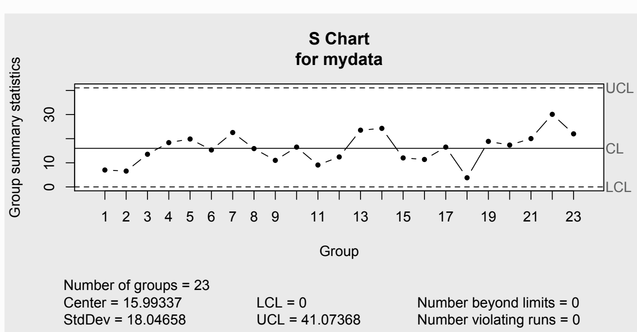
The Sigma-Chart

- A constant-cause system is characterized not only by a mean common to all data values, but also a constant level of variation
- **Sigma-Chart**: Evaluates the stability of the variability of the process by displaying the standard deviation of each group
- In-control variability is prerequisite for a meaningful X-bar chart because the *average* standard deviation is used to construct the X-bar chart, and so must be representative of variability across time
- The Sigma-chart uses the same data as the X-bar chart

The R Sigma-Chart

- The qcc function `qcc`, here applied to all the variables in the data table `d` for an S chart

```
> qcc(d, type="S")
```



3.3e
Tampering

Tampering I: Based on Random Fluctuation

Do not be coerced into action by random fluctuation

- A common theme of managing a process is to reduce the overall variation of the process, such as in manufacturing or minimizing queue times, such as the wait in the ER
- Tampering is modifying a process on the basis of data that was not generated by the process of interest
- **Tampering** of a stable process: Adjusting a process in reaction to moderate deviation due only to random fluctuation
- Instead of diminishing variability, tampering has the opposite effect of *increasing* variability
- If a moderate random deviation results in a data value far from the mean, then reducing the mean of the process to compensate suppresses the mean for all subsequent output
- After tampering with the process, the data become more variable than if no adjustment had been made

Ex: Tampering Based on Random Fluctuation

React according to reality, not chance

- This form of tampering mistakenly treats a chance occurrence as a systematic part of the process
- Consider the drilling of a hole for a machined part
  - Perhaps purely by chance the last three holes were drilled to the right of center
  - Adjusting the drilling set-up in response to these chance deviations leads to more erratic output than does leaving the process alone
- Evaluate the entire process, and adjust the level and/or minimize variability only when all the output is deemed to be generated by a stable process

Tampering II: Based on a Process Change

Do not adjust the process due to an assignable cause

- **Tampering** of an unstable process: Adjusting a process because of a deviant value due to an unique assignable cause
- The improper management response is to treat all the data values as a single system and adjust the system on the basis of the deviant value
- If the analysis reveals an out-of-control deviation, that is, an outlier, and, particularly, if an investigation reveals the reason for the deviation, then
  - discard the value from the current analysis
  - fix the problem responsible for the outlier

Ex: Tampering Based on an Assignable Cause

Eliminate assignable causes before adjusting process

- Within a manufacturing context, consider some assignable causes that lead to errors in terms of scrapped parts
  - A temporary employee drills an off-center hole
  - A substandard batch of metal was delivered for machining
- However, only when the usual employees are working or the proper metal is used should the process be analyzed for adjustment and modification
- To reduce variability due to the common causes, the process must have settled down so that only common causes are randomly operating
- **Key Concept**: If an outlier due to an assignable cause is present, the mean of all of the data is not the process mean because all of the data are not from the same process with only one common underlying mean and standard deviation

The Key to Process Management

Is there a single process that generated the data?

- **Key Concept**: The presence of non-random sources of variation implies that the measured outcomes are generated by *multiple* processes
- Before actively managing a process to improve performance, first identify the process that needs to be managed
- To improve the results of any process, directly manage the sources of variation underlying the outcomes, such as including an analysis of a control chart, instead of just reacting to the process results such as the mean of all the data values
- Adjusting the wrong aspect of a process with the goal of reducing variability can result in exactly the opposite result
- **Key Concept**: Manage the process, not the results
- A crucial tool for process management is the control chart

Index

Subtract 2 from each listed value to get the Slide #

- The End