Quality Management

Lecture 5, 23/01/23

Objectives

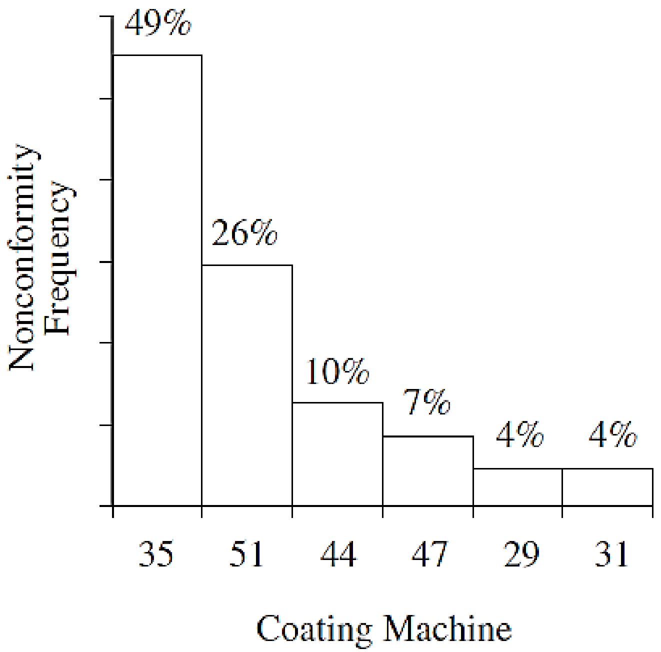
- Understanding the seven QC tools of continuous improvement and solving problems
- Understanding basic statistical concepts like measures of central tendency and dispersion, population, sample and normal distribution
- Overview of application of data-based approach for basic statistical tools for continuous improvement and solving problems
- Studying statistical control charts, types and some application examples

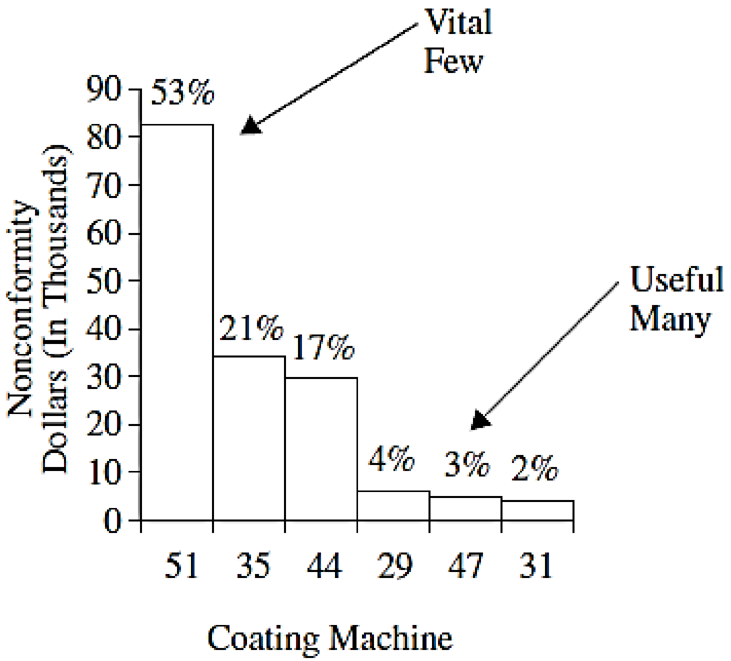
Statistical Process Control

- Statistical process control (SPC) is defined as **the use of statistical techniques to control a process or production method**. SPC tools and procedures can help you monitor process behavior, discover issues in internal systems, and find solutions for production issues.
- There are seven basic techniques. Since the first four techniques are not really statistical, the word *statistical* is somewhat of a misnomer.
- Furthermore, this technical tool not only *controls* the process but has the capability to improve it as well

Pareto Diagram

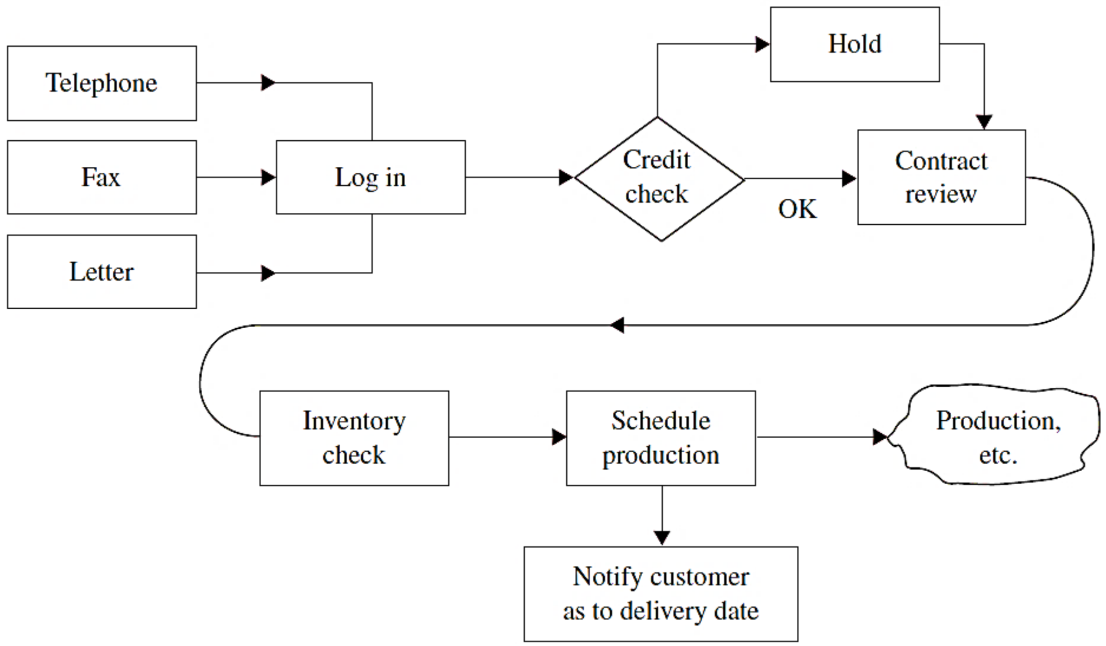
- ✓ Alfredo Pareto (1848–1923) conducted extensive studies of the distribution of wealth in Europe. He found that there were a few people with a lot of money and many people with little money.
- ✓ This unequal distribution of wealth became an integral part of economic theory. Dr. Joseph Juran recognized this concept as a universal that could be applied to many fields. He coined the phrases *vital few* and *useful many*.
- ✓ A Pareto diagram is a graph that ranks data classifications in descending order from left to right.
- ✓ Quality improvement of the vital few, say, 50%, is a much greater return on investment than a 50% improvement of the useful many.





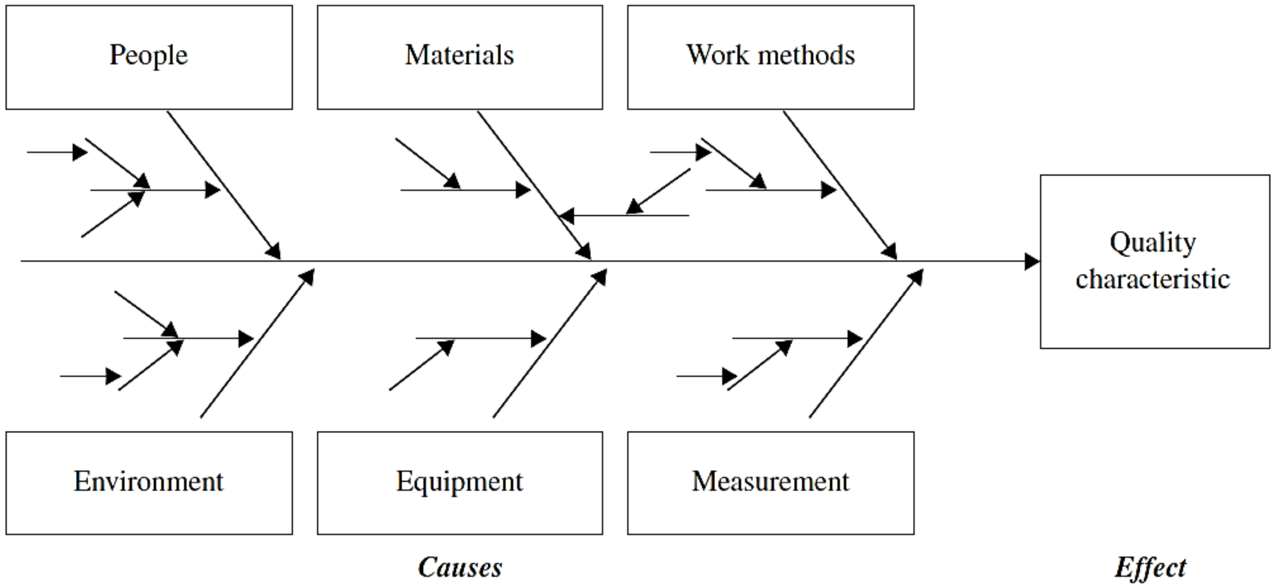
Process Flow Diagram/Virtual Platform

- ✓ For many products and services, it may be useful to construct a process flow diagram.
- ✓ Figure shows a flow diagram for the order entry activity of a make-to-order company that manufactures gasoline filling station hose nozzles.



Cause and Effect Diagram

1. Participation by every member of the team
2. Quantity of ideas, rather than quality, is encouraged
3. Criticism of an idea is not allowed
4. Visibility of the diagram is a primary factor of participation
5. Create a solution-oriented atmosphere and not a gripe session
6. Let the ideas incubate for a period of time (at least overnight) and then have another brainstorming session.



Check Sheets

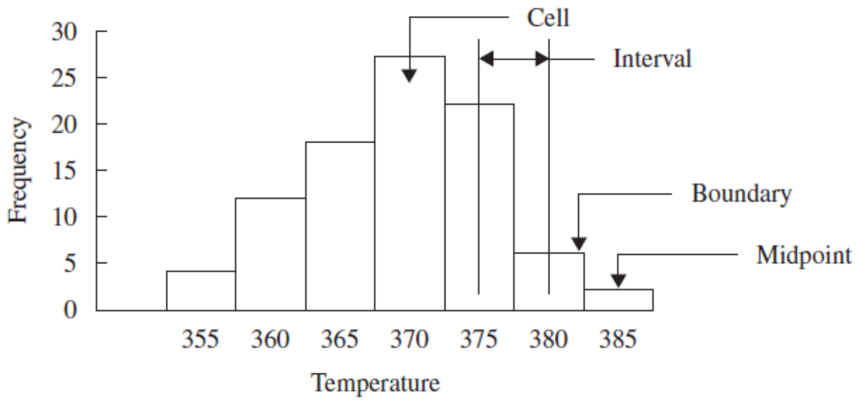
CHECK SHEET

Product: Bicycle 32 — Number inspected: 2217

Nonconformity type	Check	Total
Blister	𝍸 𝍸 𝍸 𝍸 𝍷	21
Light spray	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷𝍷	38
Drips	𝍸 𝍸 𝍸 𝍸 𝍷𝍷	22
Overspray	𝍸 𝍸 𝍷	11
Runs	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷	47
Others	𝍸	5
	Total	144
Number Nonconforming	𝍸 𝍷𝍷𝍷	113

Histogram

- The first "statistical" SPC technique is the histogram. It describes the variation in the process, as illustrated by Figure
- The histogram graphically estimates the process capability and, if desired, the relationship to the specifications and the nominal (target). It also suggests the shape of the population and indicates if there are any gaps in the data.
- Histograms can give sufficient information about a quality problem to provide a basis for decision making without further analysis. They can also be compared in regard to location, spread, and shape.



Big Data? Statistical Process Control Can Help!

"Big data" is a buzzword these days due to an enormous amount of data-rich applications in different industries and research projects.

In practice, big data often take the form of data streams in the sense that new batches of data keep being collected over time.

One fundamental research problem when analyzing big data in a given application is to monitor the underlying sequential process of the observed data to see whether it is longitudinally stable, or how its distribution changes over time.

To monitor a sequential process, one major statistical tool is the SPC charts, which have been developed and used mainly for monitoring production lines in the manufacturing industries.

With many new and versatile SPC methods developed in the recent research, it is believed that SPC can become a powerful tool for handling many big data applications that are beyond the production line monitoring.

https://www.tandfonline.com/doi/epdf/10.1080/00031305.2019.1700163?needAccess=true&role=button

Big Data? Statistical Process Control Can Help!

The longitudinal process is **one in which no particular trend over time is expected**; the interest is in assessing how much responses vary over time for a typical individual.

Big Data? Statistical Process Control Can Help!

- ✓ To sequentially monitor a longitudinal process, a major statistical tool is SPC. Traditional SPC concepts and methods are developed mainly for monitoring production lines in the manufacturing industry, to detect any special cause variation(e.g., process mean shift or drift) in the observed data.
- ✓ In case when the data variation is mainly due to random noise, it is often called common cause variation, and the process under monitoring is considered to be in statistical control, or simply in-control(IC).When a process has a special cause variation present, the process is considered to be out-of-control(OC).

Big Data? Statistical Process Control Can Help!

- ✓The four types of basic SPC charts are based on the assumptions that IC process observations at different time points are iid and follow a parametric distribution (e.g., normal). In practice, these assumptions would hardly be valid, especially in data-rich applications.
- ✓So, much recent SPC research focuses on developing new control charts that are appropriate to use without these assumptions.

Recent SPC for Monitoring Processes With Complicated Data

Serially Correlated Data Monitoring

In practice, process observations at different time points are usually correlated with each other. Many existing methods are based on parametric time series modeling of the observed data, and on sequential monitoring of the residuals obtained from the time series modeling.

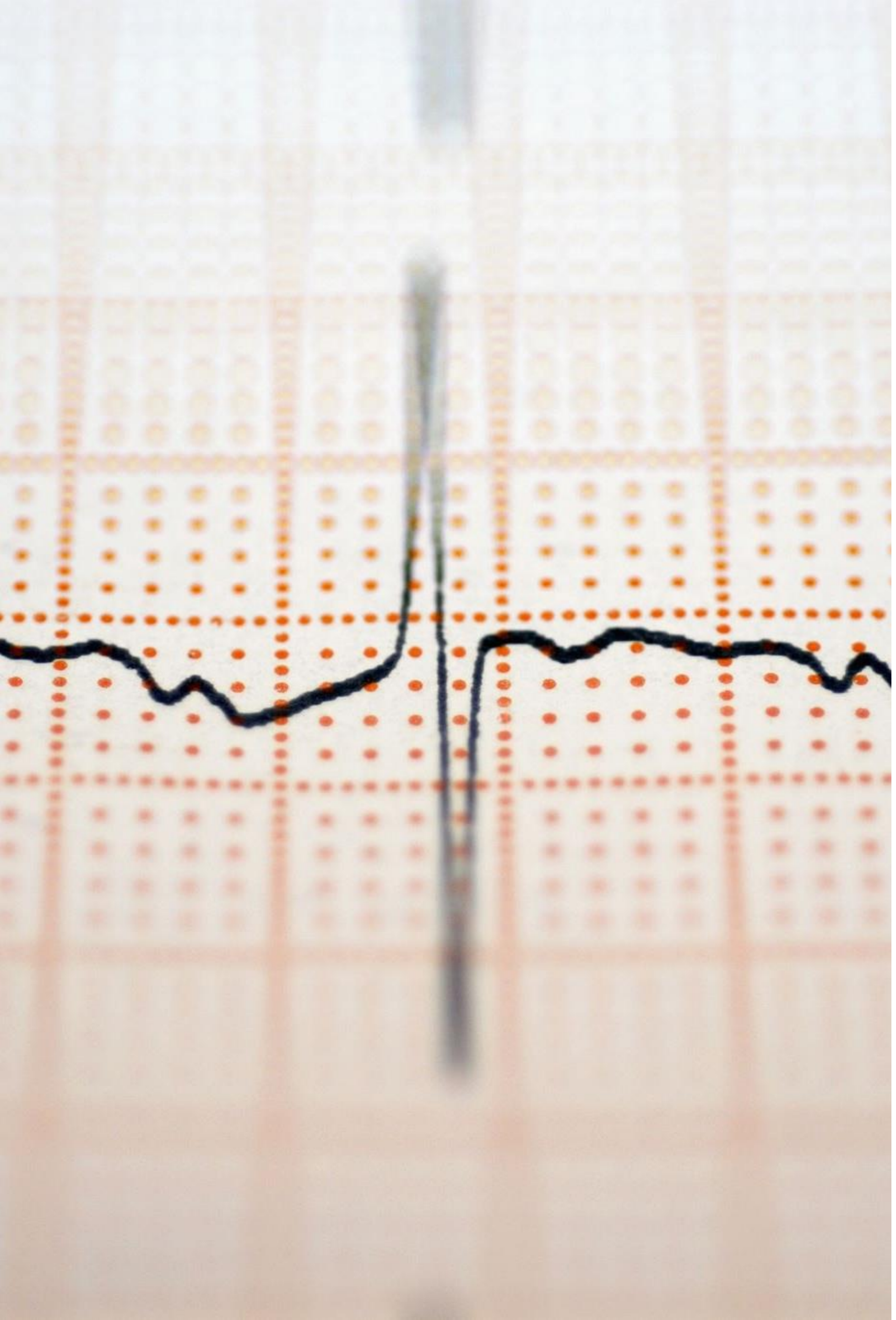
Big Data? Statistical Process Control Can Help!

For instance, it is assumed that IC process observations $\{Xn, n \geq 1\}$ follow the ARMA (p,q) model

Serially Correlated Data Monitoring

$$X_n = \mu_0 + \sum_{j=1}^{p} \phi_j (X_{n-j} - \mu_0) + \varepsilon_n - \sum_{j=0}^{q} \theta_j \varepsilon_{n-j},$$

where μ_0 is the IC mean, $\{\varepsilon_n\}$ are iid random errors with the distribution $N(0, \sigma^2)$, and $\{\phi_j\}$ and $\{\theta_j\}$ are coefficients.



Dynamic Process Monitoring

- For many longitudinal processes, their distributions could change over time, even when their performance is considered IC. One example is about our medical indices, such as blood pressure readings and cholesterol levels.

- For many longitudinal processes, their distributions could change over time, even when their performance is considered IC. One example is about our medical indices, such as blood pressure readings and cholesterol levels.

Dynamic Process Monitoring

- In the past several years, a new method, called dynamic screening system (DySS), for sequential monitoring of dynamic processes has been developed.
- The DySS method will be introduced using the example to early detect stroke by sequentially monitoring a person's total cholesterol level readings. It consists of the following three main steps:

(i) Estimation of the regular longitudinal pattern: We first estimate the IC longitudinal pattern of the total cholesterol level from the observed total cholesterol level data of a set of non stroke people.

(ii) Cross-sectional comparison: For a specific person to monitor, we standardize her/his total cholesterol level observations using the estimated regular longitudinal pattern obtained in Step (i).

(iii) Sequential monitoring: Apply a conventional control chart to the standardized data obtained in Step (ii) for sequential process monitoring

Nonparametric SPC

Classic control charts, are based on the assumption that process observations follow a normal or another parametric distribution, which is rarely valid in practice. It has been well demonstrated that the classic control charts are unreliable to use when their distributional assumption is violated.

So, nonparametric has been in rapid development in recent years, and many nonparametric control charts have been developed.

Nonparametric SPC

Nonparametric statistics is the branch of statistics that is not based solely on parametrized families of probability distributions (common examples of parameters are the mean and variance).

Nonparametric statistics is based on either being distribution-free or having a specified distribution but with the distribution's parameters unspecified. Nonparametric statistics includes both descriptive statistics and statistical inference. Nonparametric tests are often used when the assumptions of parametric tests are violated.

Recent SPC Research for Handling Data-Rich Applications

- Sequential Monitoring of High-Dimensional Multivariate Processes
- Sequential Monitoring of Univariate and Multivariate Profiles
- Sequential Monitoring of Spatial Data
- Sequential Monitoring of Network Data

The Enron E-mail corpus is a well-known dataset in social net-work research. After the bankruptcy of the Enron Corporation in October 2001, all E-mails to and from the Enron employees during the period from 1998 and 2002 were made public by the ruling of the Federal Energy Regulatory Commission.

Recent SPC Research for Handling Data-Rich Applications

There have been some existing SPC methods for monitoring network process. These methods usually apply SPC charts to aggregated measures of the topological characteristics (e.g., density, degree, clustering coefficient, and scan statistic) of the entire network or relevant subnetworks. Some of them are based on parametric modeling while the others are nonparametric.

Interface Between Big Data and SPC :Feature-Based Process Monitoring

- When data get larger, they often have a more complicated structure. Consequently, proper analysis of them becomes more challenging. some challenging issues that are quite common to various big data monitoring problems:

Because of the big data size in certain process monitoring problems, people tend to simplify the observed data, by first extracting certain features of the observed data and then monitoring the extracted features. One example is image monitoring.

Interface Between Big Data and SPC: Accommodation of Complicated Data Structure

- Big data often have complicated data structures. One such structure concerns data correlation. Usually, quality variables are correlated with each other, and their observations at different time points are also serially correlated. Such data correlation is often a reflection of the impact of certain confounding variables, such as weather and geographical conditions in the Landsat image

Interface Between Big Data and SPC: Accommodation of Covariate

- In practice, performance of a process is often affected by various covariates. Observations of some covariates could be available to us. Therefore, we should make use of the helpful information in covariates when monitoring the related process. As an example, the sequence of Landsat images of a given region may depend on the weather and geographical conditions of the region, and observations of these conditions can be obtained from the databases

Interface Between Big Data and SPC: Dynamic Process Monitoring

- As discussed, many processes to monitor in practice are dynamic processes in the sense that their IC distributions would change over time. Although dynamic process monitoring by DySS provides a reasonable solution to this important process monitoring task, there are still some fundamentally important issues to address. First, we need to determine a time period from the history data of the process under monitoring, in which the performance of the process is believed to be IC.

Many big data in practice are in the form of data streams, and they are sequential observations of certain underlying longitudinal processes. To study the patterns of these processes over time, SPC is a relevant statistical tool.

Statistical Fundamentals

Measures of Central Tendency

A measure of central tendency of a distribution is a numerical value that describes the central position of the data or how the data tend to build up in the center. There are three measures in common use in quality: (1) the average, (2) the median, and (3) the mode. The average is the sum of the observations divided by the number of observations. It is the most common measure of central tendency and is represented by the equation

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

where $\overline{X}$ = average and is read as "X bar"
n = number of observed values
X_i = observed value
Σ = sum of

References:

- https://www.sciencedirect.com/science/article/abs/pii/S0098135414002804
- https://www.theijes.com/papers/vol6-issue11/Version-1/H0611013948.pdf
- https://www.tandfonline.com/doi/epdf/10.1080/00031305.2019.1700163?needAccess=true&role=button
- https://www.researchgate.net/profile/Federico-Barravecchia-2/publication/362230598_Statistical_Process_Control_techniques_to_monitor_quality_determinants_in_digital_Voice-of-Customer/links/62de495b82bb472992a1b9ae/Statistical-Process-Control-techniques-to-monitor-quality-determinants-in-digital-Voice-of-Customer.pdf
- https://www.sciencedirect.com/science/article/pii/S1877050915037011
- https://www.proquest.com/openview/ba990fce1c70387e9c3c7ce97dbf723c/1?pq-origsite=gscholar&cbl=25222