Introduction to Statistical Quality Control

1st Hour

Based on the book “Introduction to Statistical Quality Control”, 7th Edition by Douglas C. Montgomery.

Speaker Introductions

- Dr. Bersimis Sotirios is an Assistant Professor at the Statistics and Insurance Science Department of the University of Piraeus, where he has been providing his academic services since 2010.
- He is a PhD holder in Statistics and Probability, from the Statistics and Insurance Science Department of the University of Piraeus. He also holds an MSc in Statistics from Athens University of Economics and Business and a BSc in Statistics and Insurance Science from the University of Piraeus.
- His doctoral research has been funded by a scholarship from the Hellenic General Secretary of Research and Technology. He has also been granted a post-doctoral fellow scholarship by the Hellenic State Scholarships Foundation.

STATISTICAL QUALITY CONTROL

QUALITY

QUALITY DIMENSIONS AND CHARACTERISTICS

Statistical Quality Control

- Which is the exact definition of Statistical Quality Control?
- With the term of Statistical Quality Control (SQC) we describe the toolbox of Statistical methods which are used in Industry (as a whole) in order to ensure the quality of the produced goods and services.

Definition of Quality

- What is Quality?
- There are a lot of definitions for the term Quality.
- According to Joseph Juran “Quality” can be defined as the adjustment of a product’s (or Sevice’s) characteristics so as to respond to consumers’ needs and expectations.
- The term “Quality” should not be confused with the term of “high standards”.

Definition

Quality means fitness for use.

- This is a traditional definition
- Quality of design
- Quality of conformance

DIMENSIONS OF QUALITY

What is a "Qualitative" Product/Service

The product/service is adjusted to consumers' expectations and needs in terms of:

- Performance
- Reliability
- Durability
- Serviceability
- Aesthetics
- Features
- Perceived Quality
- Conformance of Standards

DIMENSIONS OF QUALITY

- Most people “confuse” “Quality” with the dimensions a product or service must possess

DIMENSIONS OF QUALITY

- **Performance:** Does the product/service meet the purpose of which it is destined? Does it work better than similar products/services?
- **Reliability:** Does the product need frequent repairs?
- **Durability:** Which is the lifespan of the product?
- **Serviceability:** In case of a damage to the product, how fast this can be restored? At what cost?

DIMENSIONS OF QUALITY

- **Aesthetics:** Is the product satisfactory in terms of its external characteristics? (color, shape, wrapping, etc.)
- **Features:** Which are the additional capabilities of the product?
- **Perceived Quality:** Is the company's reputation good or bad?
- **Conformance of Standards:** Is the product manufactured according to specifications set by its designer?

QUALITY CHARACTERISTICS OF A PRODUCT

- Is a product’s characteristic (a variable) which can be controlled (monitored) and it is related to one or more of the products’ quality dimensions.

QUALITY CHARACTERISTICS OF A PRODUCT

Every product possesses a number of elements that jointly describe what the user or consumer thinks of as quality. These parameters are often called **quality characteristics.** Sometimes these are called **critical-to-quality (CTQ)** characteristics. Quality characteristics may be of several types:

1. **Physical:** length, weight, voltage, viscosity
2. **Sensory:** taste, appearance, color
3. **Time Orientation:** reliability, durability, serviceability

Since variability can only be described in statistical terms, **statistical methods** play a central role in quality improvement efforts. In the application of statistical methods to quality engineering, it is fairly typical to classify data on quality characteristics as either **attributes** or **variables** data. Variables data are usually continuous measurements, such as length, voltage, or viscosity. Attributes data, on the other hand, are usually discrete data, often taking the form of counts. We will describe statistical-based quality engineering tools for dealing with both types of data.

DIMENSIONS OF QUALITY & QUALITY CHARACTERISTICS (VARIABLES)

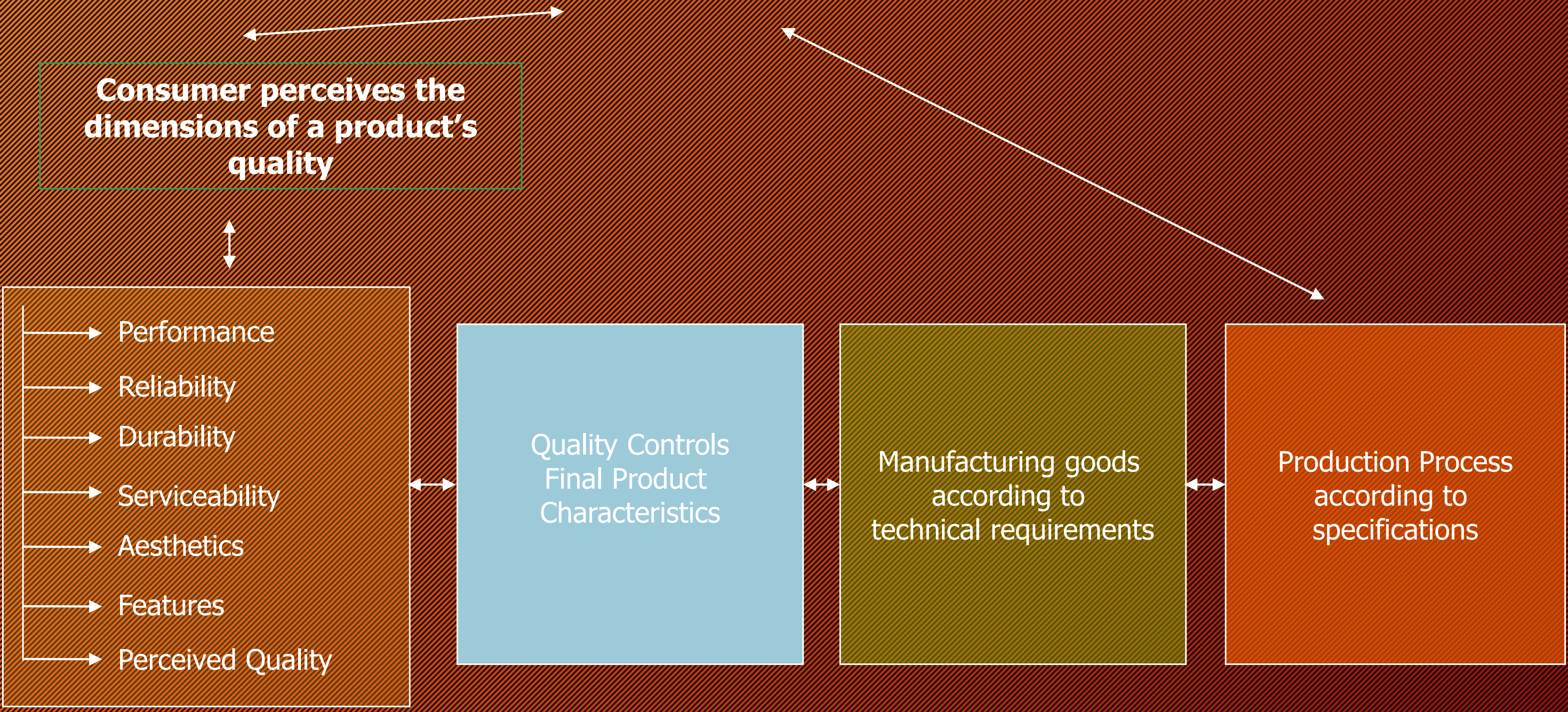
- Απόδοση
- Αξιοπιστία
- Διάρκεια ζωής
- Επισκευασιμότητα
- Αισθητική
- Δυνατότητες
- Φήμη εταιρείας
- Κατασκευή σύμφωνη με τις σχεδιαστικές προδιαγραφές

Quality Characteristics

- Length: X_1
- Strength : X_2
- Temp : X_3
- Length : X_4
- Color Factor $X5$
- Diameter: X_6

QUALITY - CONTROL - SPECIFICATIONS

Quality of the Final Product = adjustment of the product's characteristics to consumers' expectations and needs



QUALITY & QUALITY SPECIFICATIONS

- **Crucial Points:**
  - Quality of the Final Product = adjustment of product’s characteristics to the consumer’s needs!
  - Quality of the Final Product ≠ “High” Standards!
- **Example:**
  - A car can be characterized as a quality car whether it costs 10.000€ or 30.000€.
  - But, undoubtedly, a car costing 30.000€ has better Product Specifications (greater engine power, more safety characteristics etc.)

QUALITY OF A PRODUCT & PRODUCTION

- **Crucial Points:**
- Consumer recognizes “quality” through its basic dimensions.
- These dimensions are correlated to the qualitative characteristics of the product .
- Accordingly, the qualitative characteristics of the product are related to the quality of manufacturing and designing of the product as well as with the production process

QUALITY AND SPECIFICATIONS

- **Example:**
- A car is manufactured (specifications) in order to meet consumers' needs with specific characteristics.
- In order to achieve this goal, in the phase of manufacturing some technical specification s are set not only for the product itself but also for the production process.
- In order for this car to be a quality car:
  - the production process must be in accordance to technical specifications so as,
  - the product to be consistent to the manufacturing specifications, that is,
  - the qualitative characteristics (which are connected to the quality dimensions) to be in accordance with the specifications, and finally,
  - the consumer to perceive that the product satisfies his needs and expectations.

STATISTICAL QUALITY CONTROL

STATISTICS AND QUALITY

STATISTICS AND QUALITY

- How is Statistics related to Quality?
  - A more recent term defines quality as ***inversely proportional to the variability of the characteristics of the production process which define the quality of a product***.
  - We should therefore aim at reducing this variability of the qualitative characteristics which are related to the dimensions of quality and in this way, finally, the way in which we perceive as consumers the quality of the product itself.

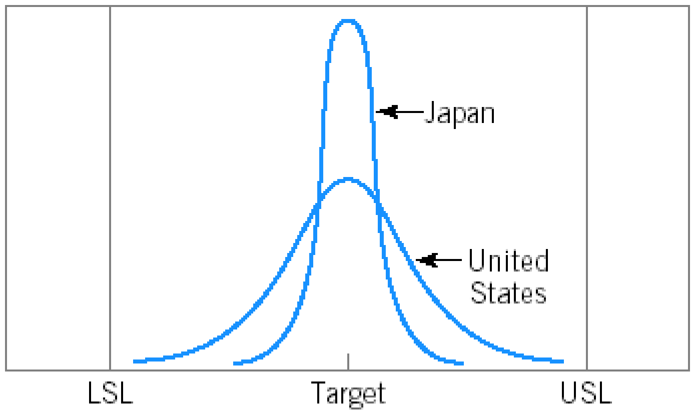


Figure 1-2 Distributions of critical dimensions for transmissions.

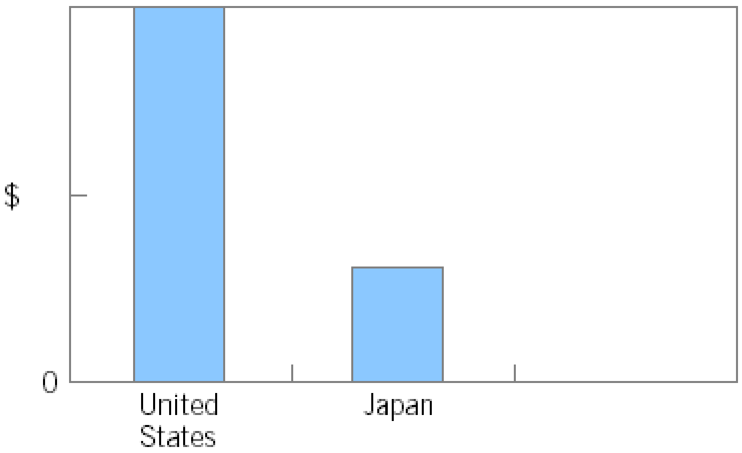


Figure 1-1 Warranty costs for transmissions.

Definition

Quality improvement is the reduction of variability in processes and products.

- The transmission example illustrates the utility of this definition
- An equivalent definition is that quality improvement is the **elimination of waste**. This is useful in service or transactional businesses.

STATISTICS AND QUALITY

Definition

Quality is inversely proportional to variability.

Definition

Quality improvement is the reduction of variability in processes and products.

This is a modern definition of quality

STATISTICAL QUALITY CONTROL

PRODUCTION PHASES &

TECHNICS OF STATISTICAL QUALITY CONTROL

PRODUCTION PHASES AND SQC

Production Phase	Product Design	Production Process	Final Product
SQC Technic	Design and Analysis of Experiments	Statistical Process Control	Acceptance Sampling

Acceptance Sampling

- Trend today is toward developing testing methods that are so quick, effective, and inexpensive that products are submitted to <u>100% inspection/testing</u>
- Every product shipped to customers is inspected and tested to determine if it meets customer expectations
- But there are situations where this is either impractical, impossible or uneconomical
  - Destructive tests, where no products survive test
- In these situations, acceptance plans are sensible

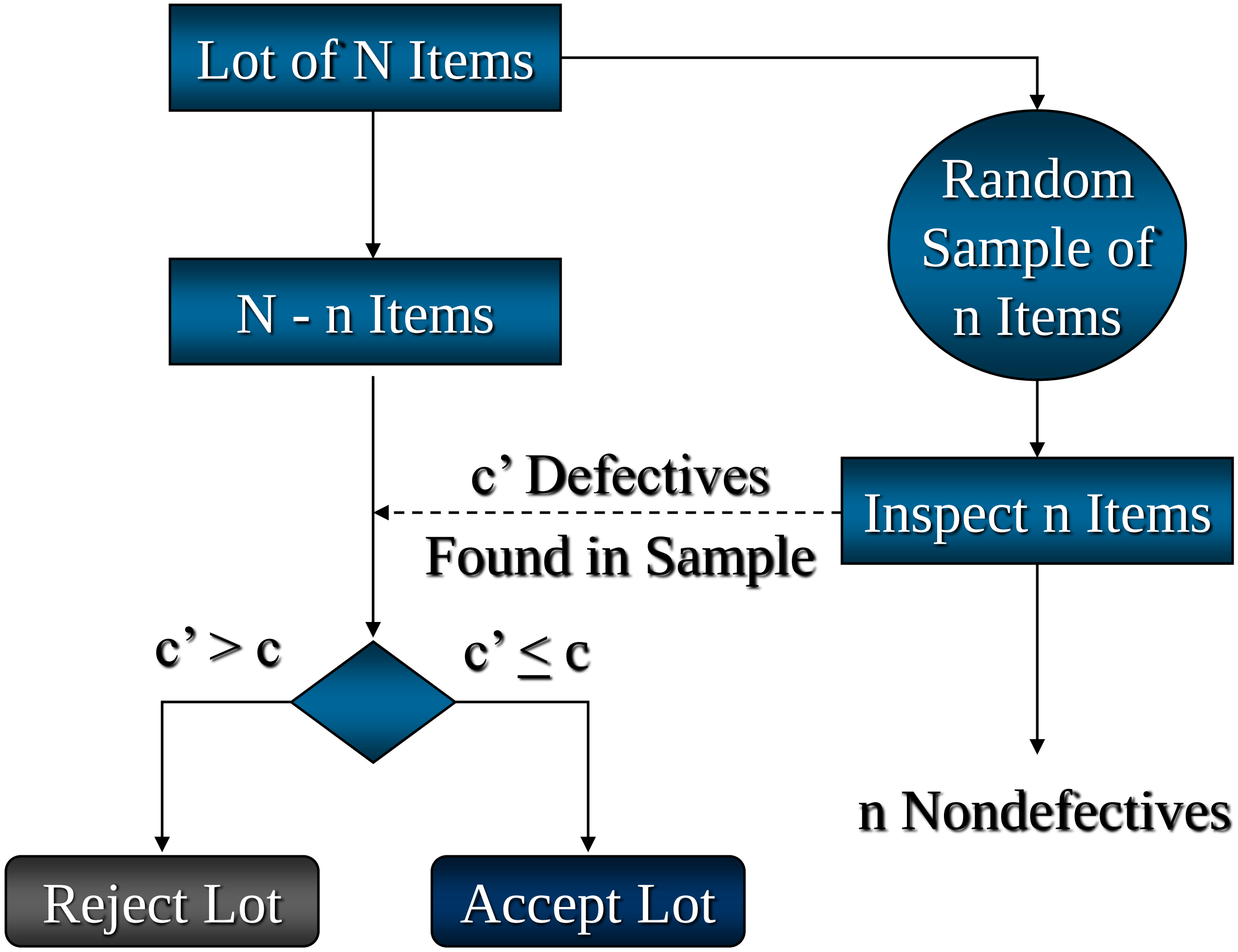
Acceptance Plans

- An acceptance plan is the overall scheme for either accepting or rejecting a lot based on information gained from samples.
- The acceptance plan identifies the:
  - Size of samples, n
  - Type of samples
  - Decision criterion, c, used to either accept or reject the lot
- Samples may be either single, double, or sequential.

Single-Sampling Plan

- Acceptance or rejection decision is made after drawing only one sample from the lot.
- If the number of defectives, c', does not exceed the acceptance criteria, c, the lot is accepted.

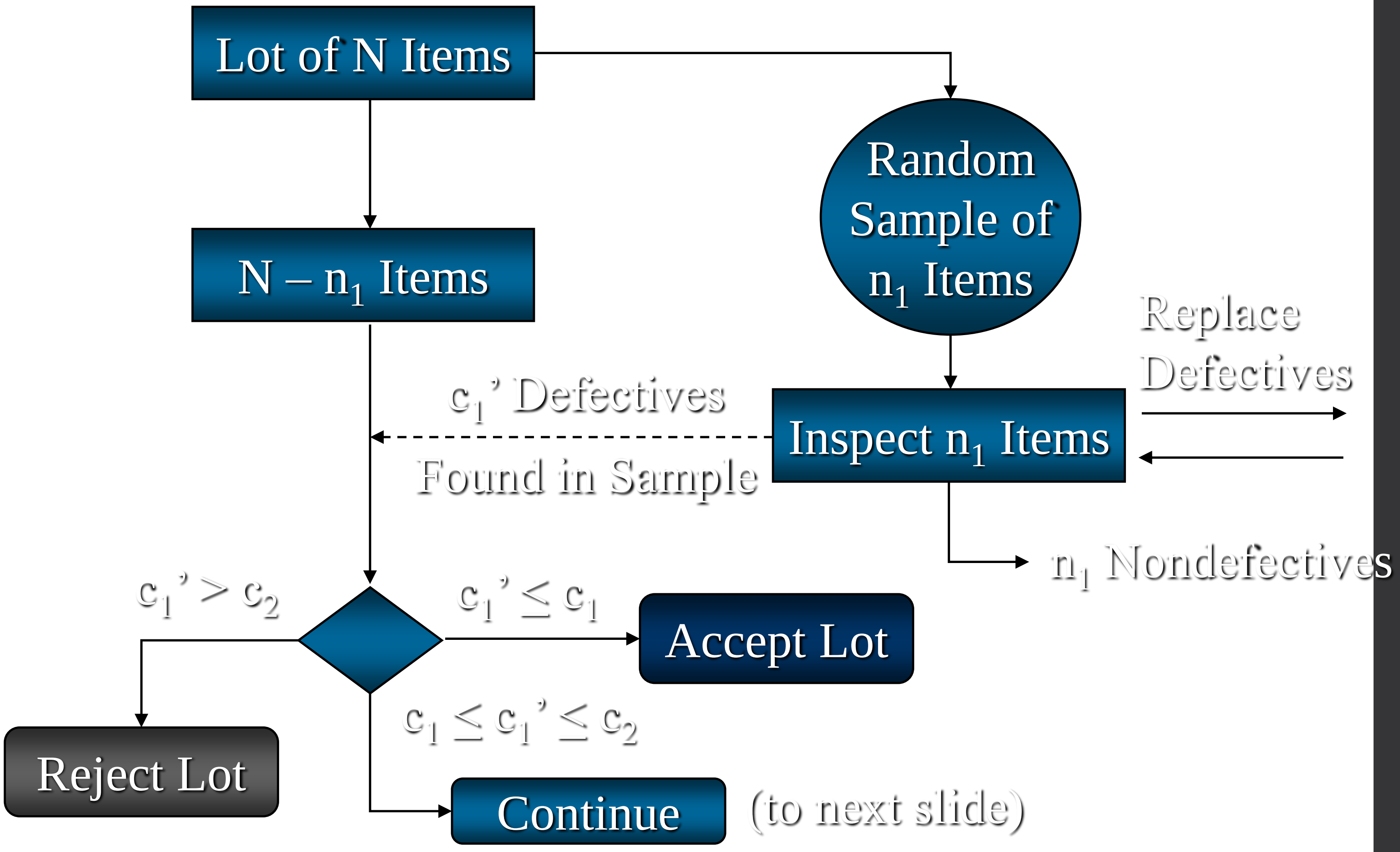
Single-Sampling Plan



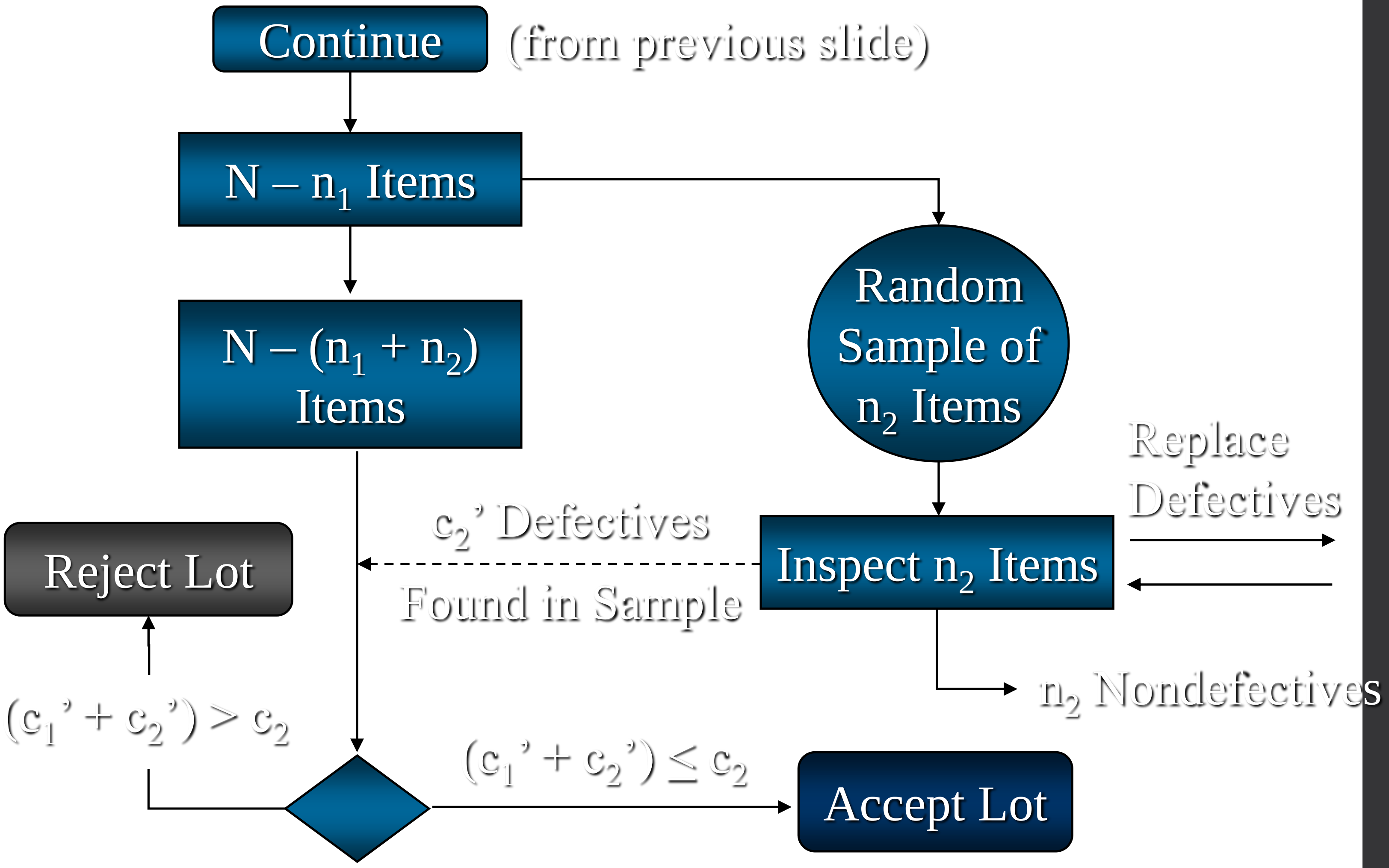
Double-Sampling Plan

- One small sample is drawn initially.
- If the number of defectives is less than or equal to some lower limit, the lot is accepted.
- If the number of defectives is greater than some upper limit, the lot is rejected.
- If the number of defectives is neither, a second larger sample is drawn.
- Lot is either accepted or rejected on the basis of the information from both of the samples.

Double-Sampling Plan



Double-Sampling Plan



NEGATIVE CONSEQUENCES IN USING ACCEPTANCE SAMPLING (SOLELY)

Cost of Labor and raw materials

- Lost manhours
- In the most likely event of a sole reprocess of the product, a loss of earnings for the company is present
- In case the product cannot be reprocessed at all, then is considered waste/scrap

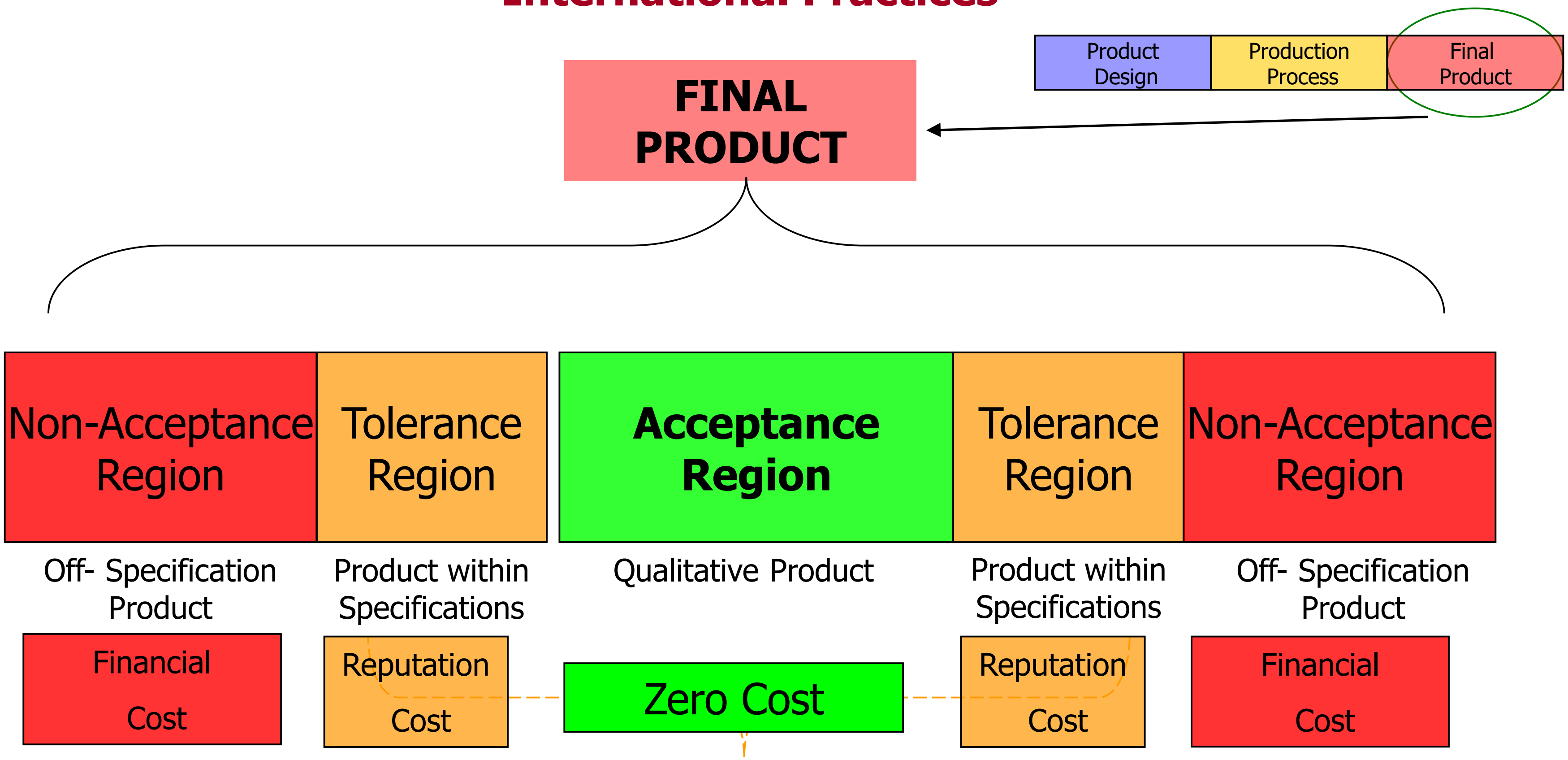
Cost in company's reputation

- This is the worst scenario – with UNPREDICTED consequences for the company

Quality Control of the Final Product related to a Qualitative Characteristic (screw, length of a screw))

Traditional Sampling Control of the Final Product

International Practices



The product can reach the end user only if it is up to the orange region

Red regions mean products inappropriate for the market or that the product must be reprocessed

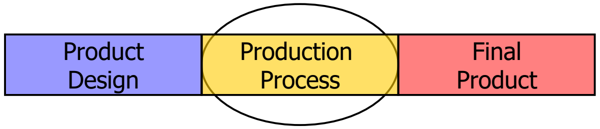
THE SOLUTION

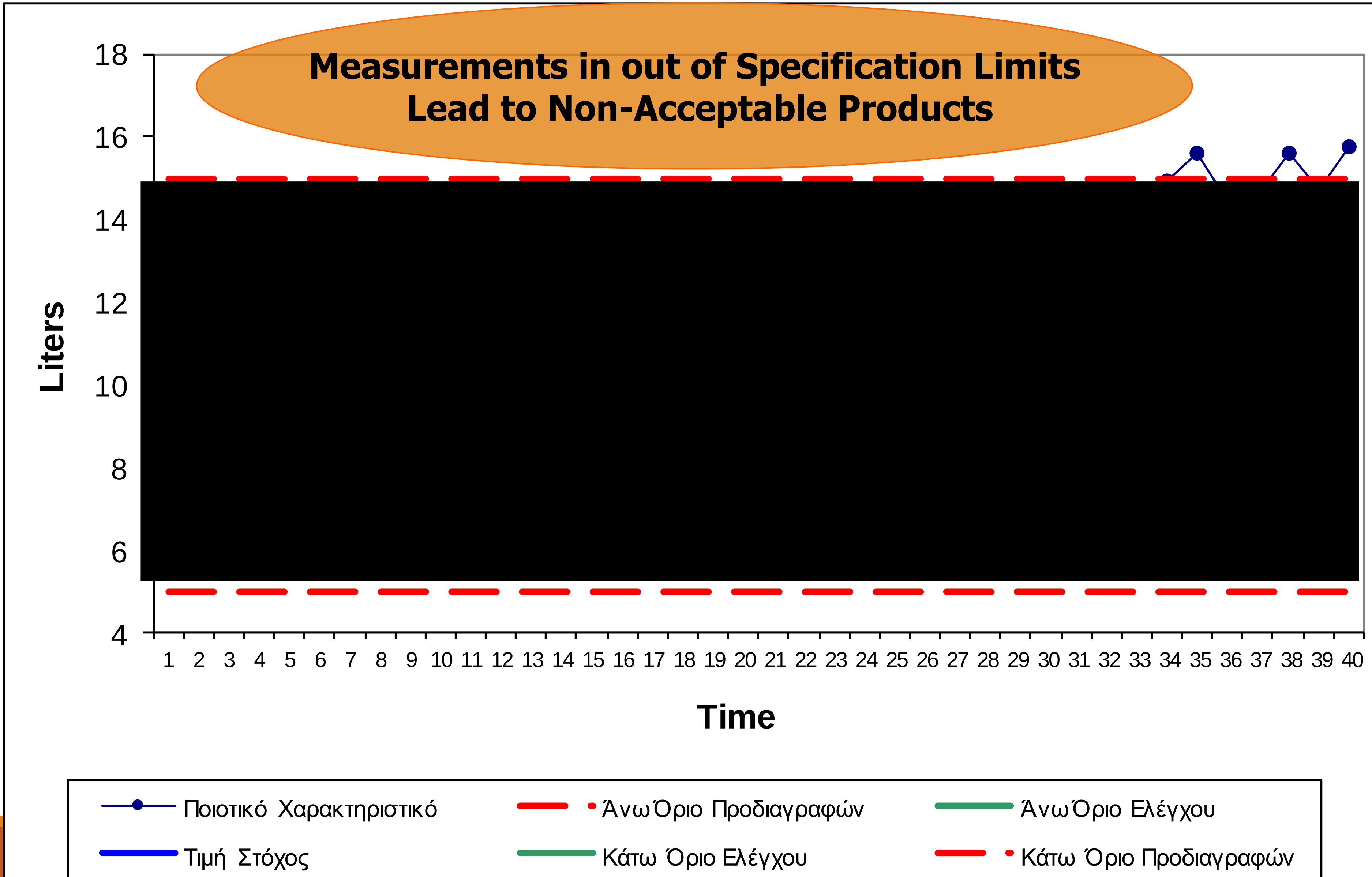
- Strict production control with the use of Statistical Process Control in all production phases.
- Statistical Process Control aims at diagnosing any potential violation of quality borders by real-time surveillance of the production processes, and alerting the responsible for the production personnel for the malfunctions long before the process goes out of control!
- Real-time production surveillance is feasible by monitoring a product variable (of qualitative characteristics) or by monitoring a process variable (which has a prompt influence at the product's qualitative characteristics).

Terminology

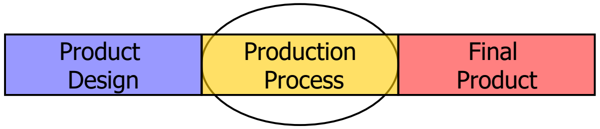
- Specifications
  - Lower specification limit
  - Upper specification limit
  - Target or nominal values
- Defective or nonconforming product
- Defect or nonconformity
- Not all products containing a defect are necessarily defective

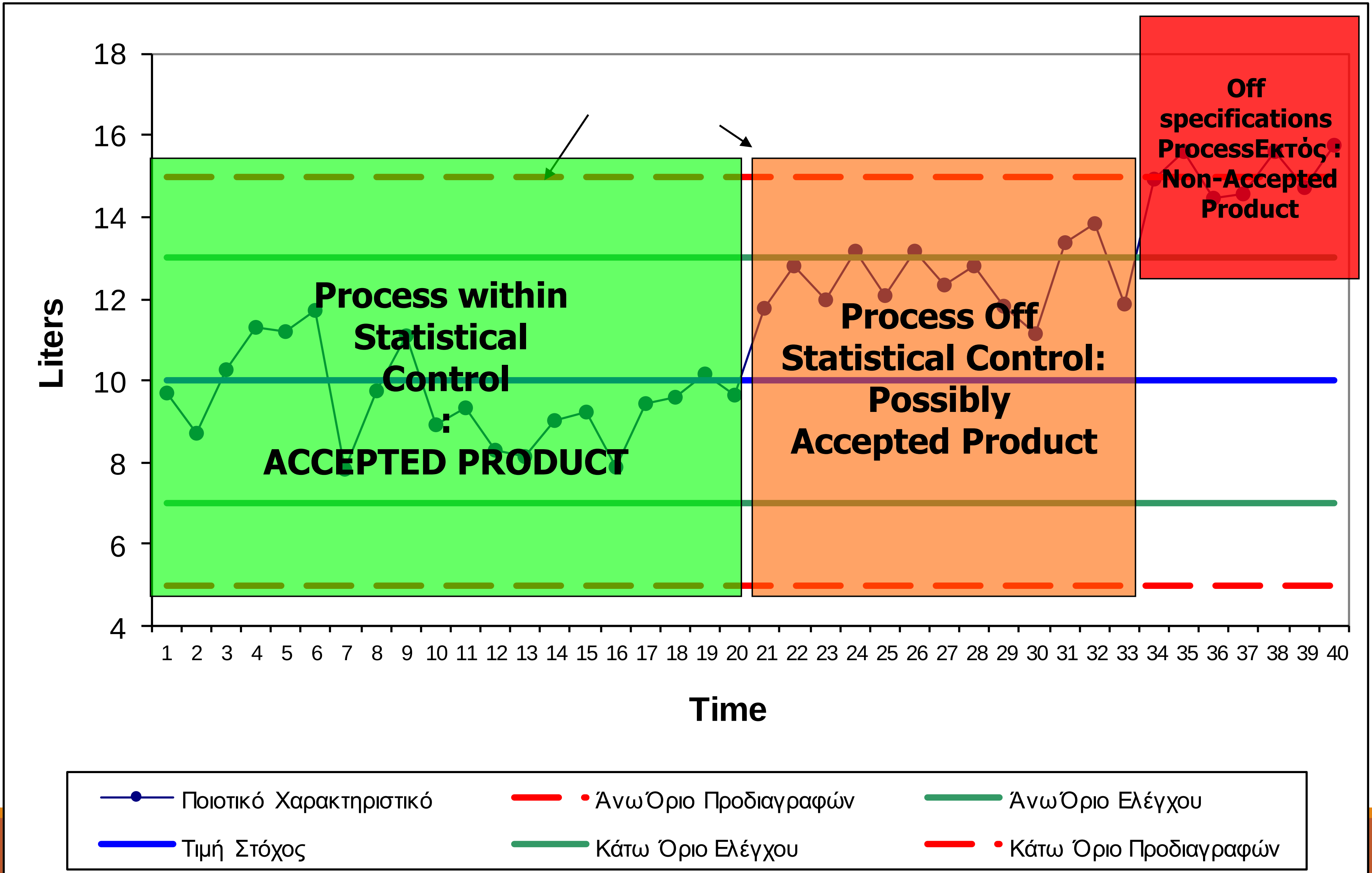
An example of a process flow without the use of Statistical Process Control (SPC)



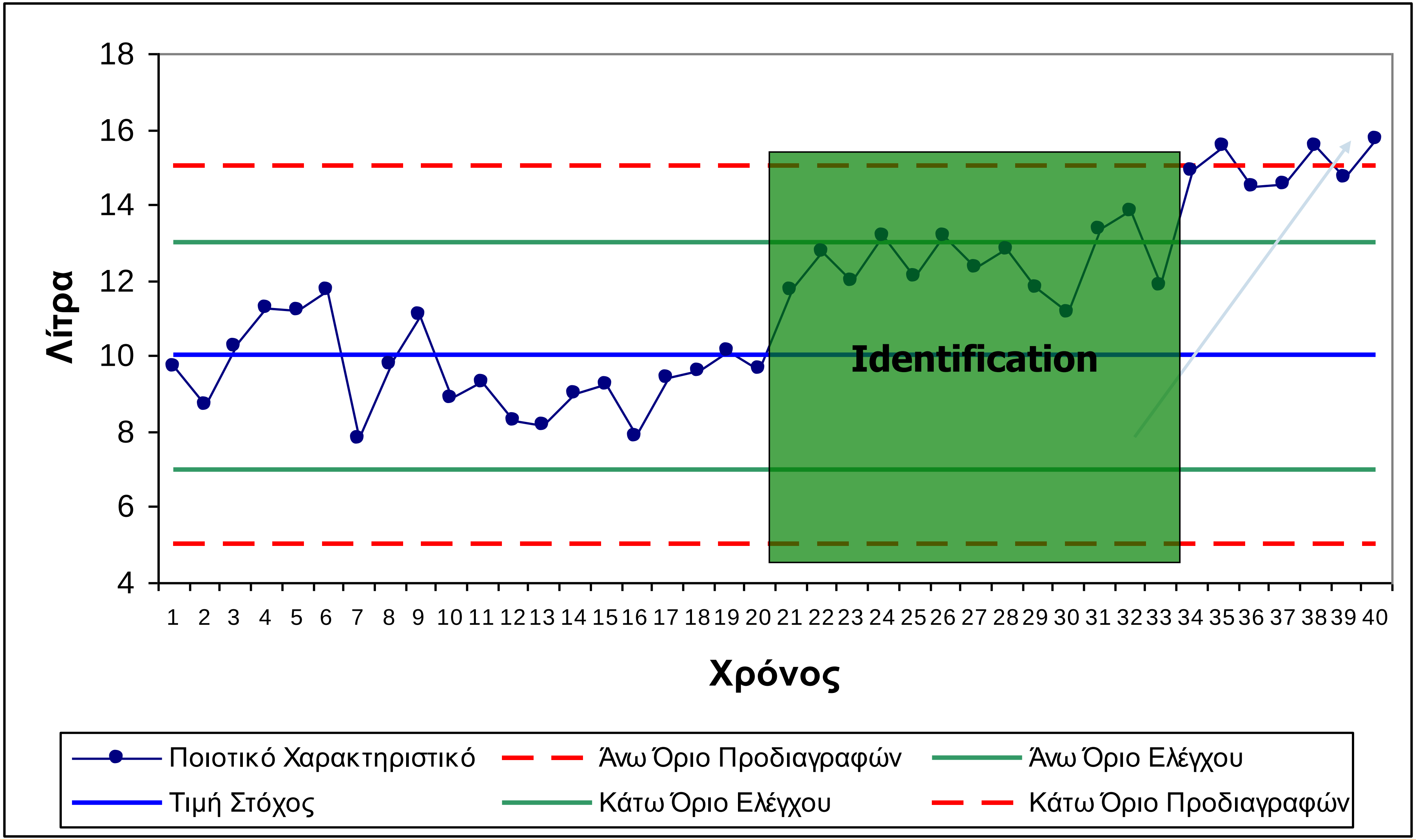


An example of a process flow with the use of Statistical Process Control (SPC)



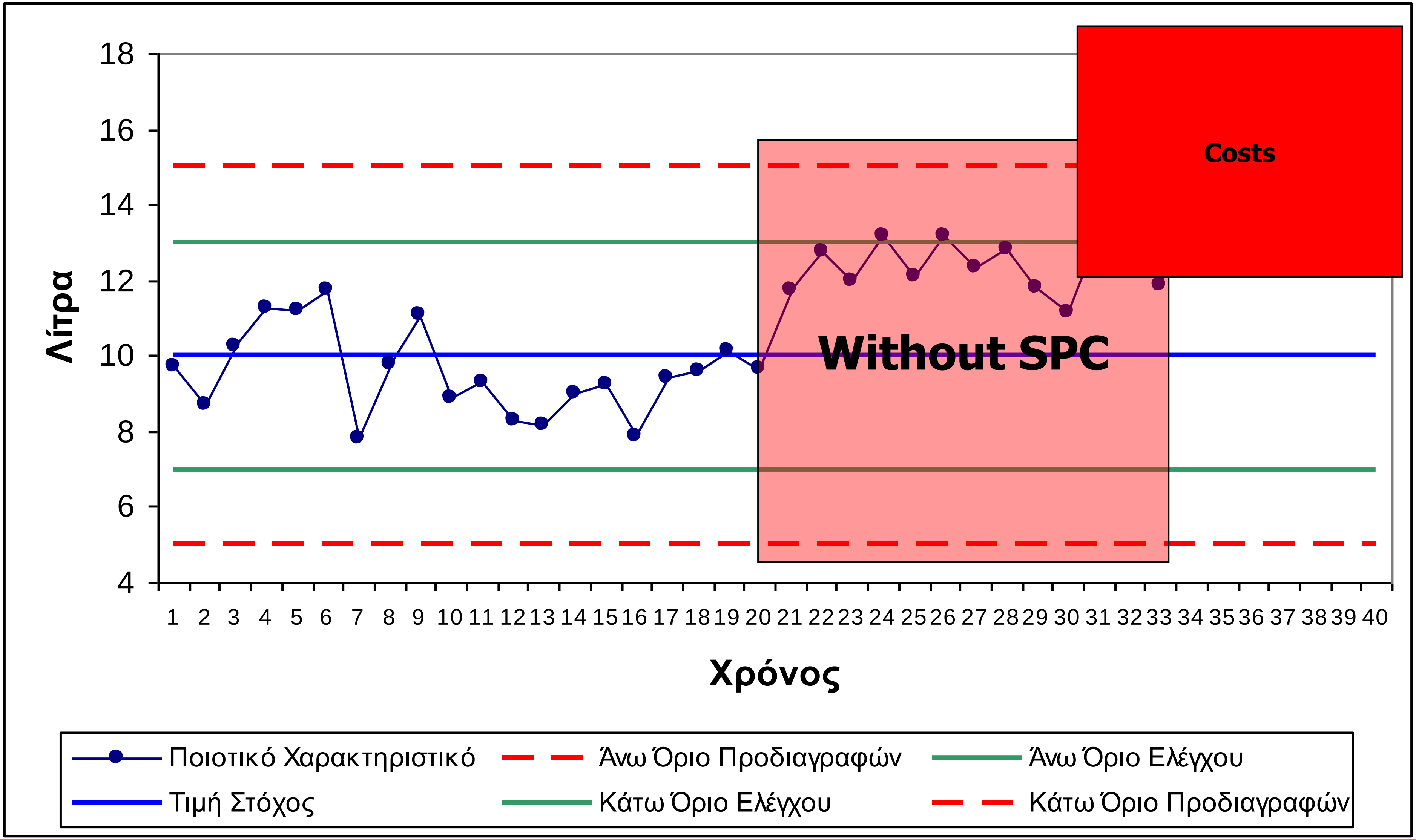


An example of a process flow with the use of Statistical Process Control (SPC)



Without SPC

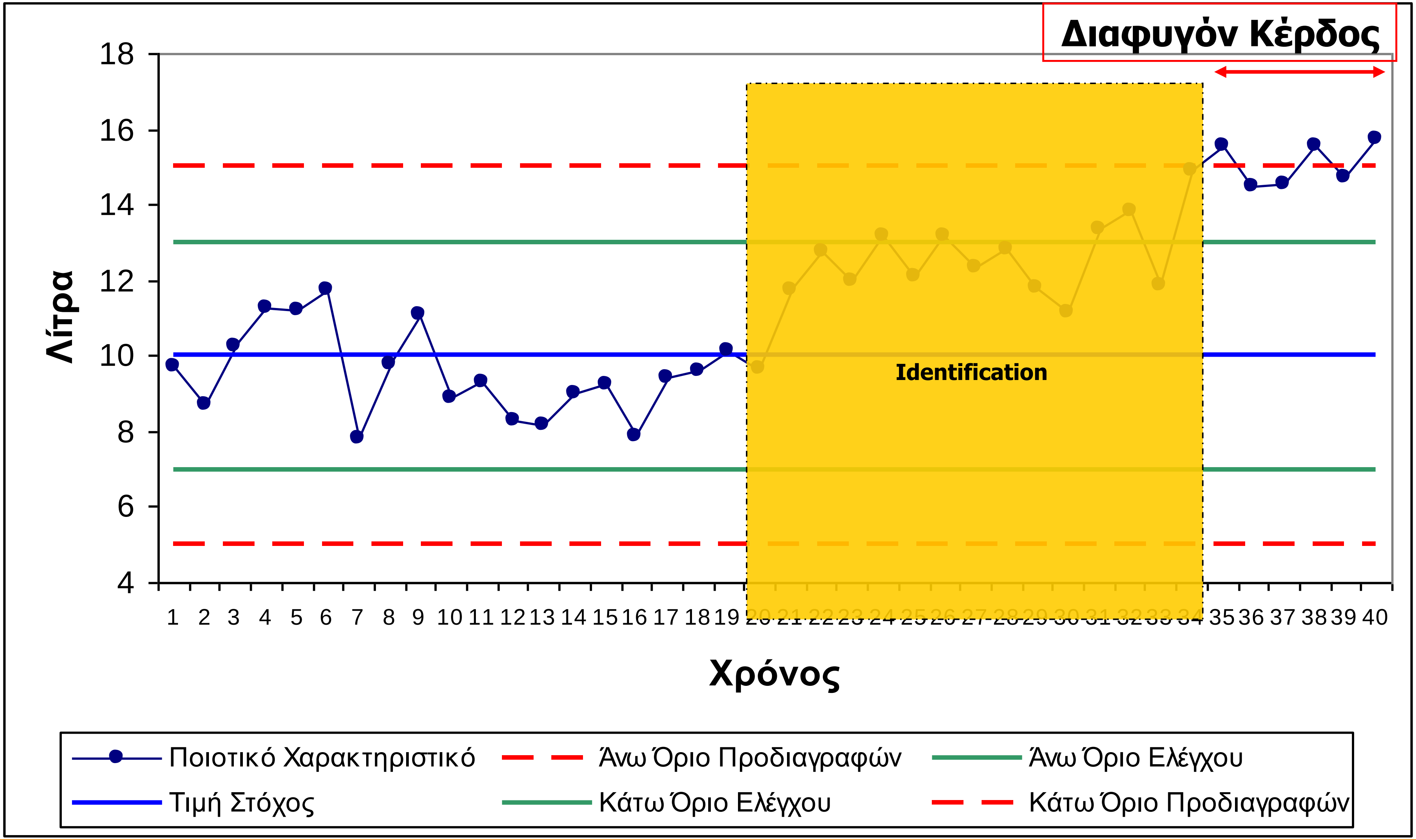
Σχεδιασμός Προϊόντος
Παραγωγική Διεργασία
Τελικό Προϊόν



Costs
Without SPC
Λίτρα
Χρόνος
18
16
14
12
10
8
6
4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Ποιοτικό Χαρακτηριστικό
Άνω Όριο Προδιαγραφών
Άνω Όριο Ελέγχου
Τιμή Στόχος
Κάτω Όριο Ελέγχου
Κάτω Όριο Προδιαγραφών

With SPC

Σχεδιασμός Προϊόντος	Παραγωγική Διεργασία	Τελικό Προϊόν



Walter A. Shewart (1891-1967)

- Trained in engineering and physics
- Long career at Bell Labs
- Developed the first control chart about 1924

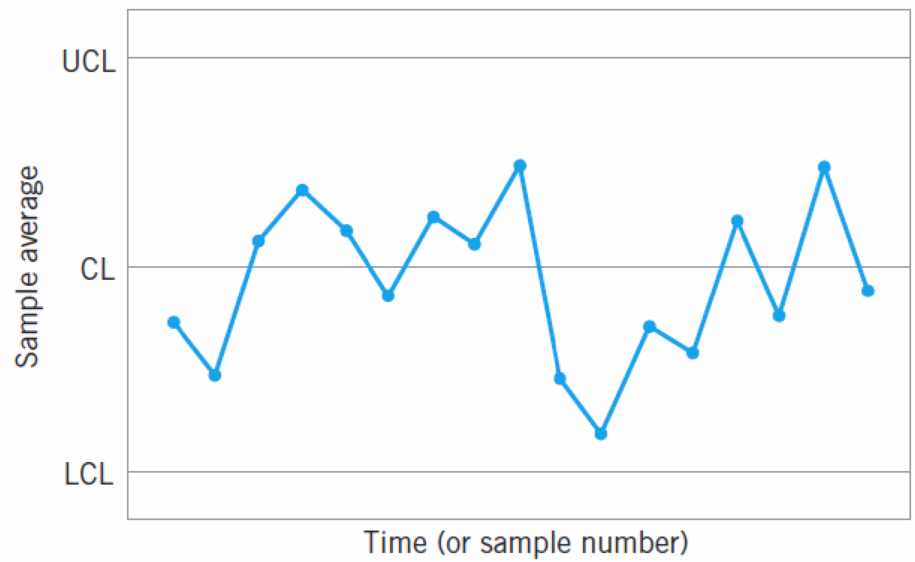


FIGURE 1.4 A typical control chart.

What is Experimental Design?

- The design of experiments (DOE, DOX, or experimental design) is the design of any task that aims to describe or explain the variation of information under conditions that are hypothesized to reflect the variation.
- In its simplest form, an experiment aims at predicting the outcome by introducing a change of the preconditions, which is represented by one or more independent variables, also referred to as "input variables" or "predictor variables."
- The change in one or more independent variables is generally hypothesized to result in a change in one or more dependent variables, also referred to as "output variables" or "response variables."
- The experimental design may also identify control variables that must be held constant to prevent external factors from affecting the results.
- Experimental design involves not only the selection of suitable independent, dependent, and control variables, but planning the delivery of the experiment under statistically optimal conditions given the constraints of available resources.
- Correctly designed experiments advance knowledge in the natural and social sciences and engineering. Other applications include marketing and policy making.

What is Experimental Design?

Experimental design includes both

- Strategies for organizing data collection
- Data analysis procedures *matched* to those data collection strategies

Classical treatments of design stress analysis procedures based on the analysis of variance (ANOVA)

Other analysis procedure such as those based on hierarchical linear models or analysis of aggregates (e.g., class or school means) are also appropriate

Why Do We Need Experimental Design?

Because of variability

We wouldn't need a science of experimental design if

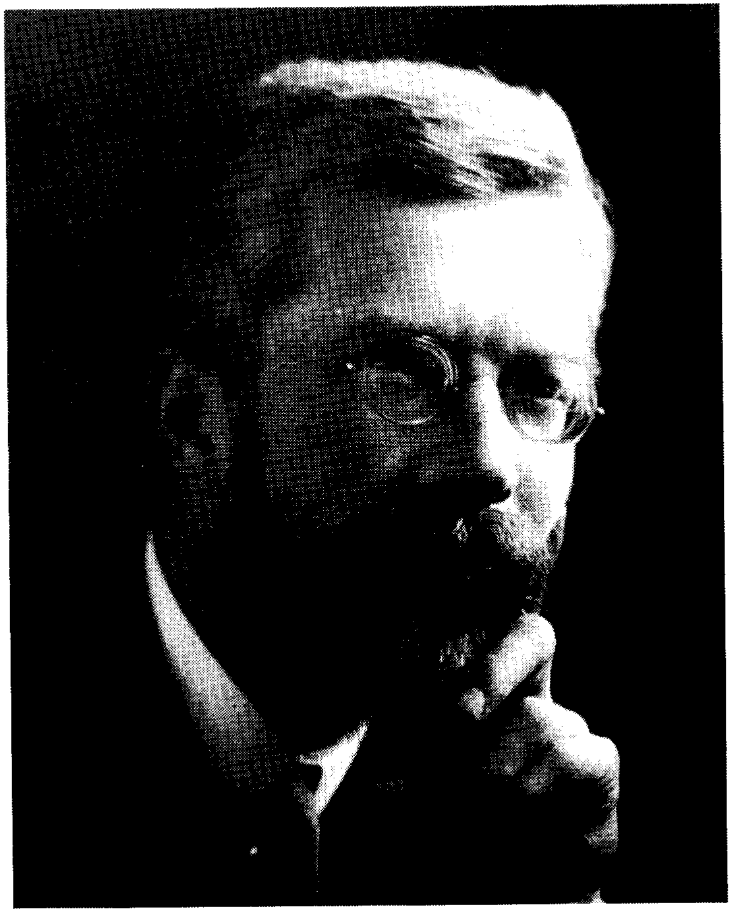
- If all units were identical

and

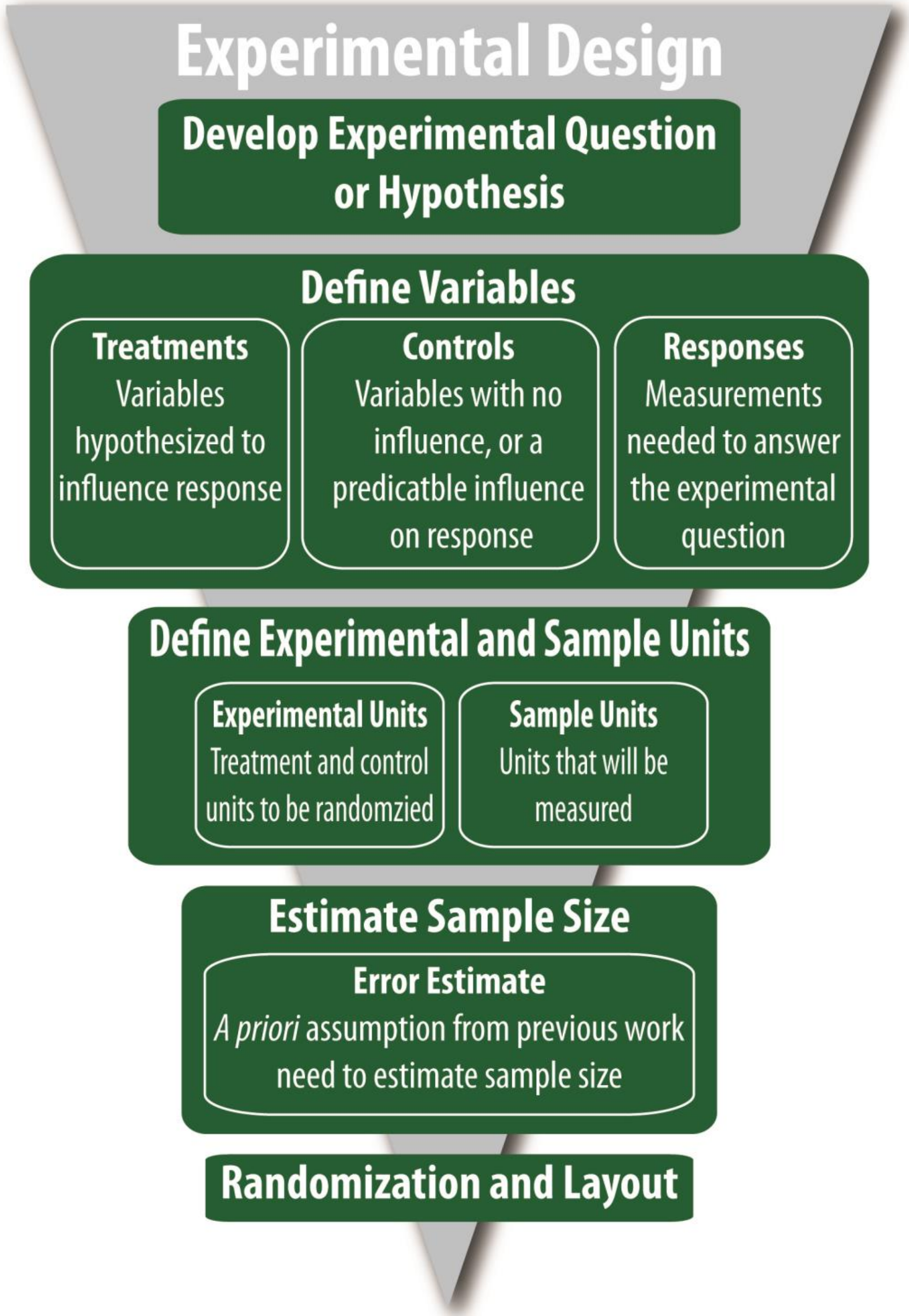
- If all units responded identically to treatments

We need experimental design to control variability so that treatment effects can be identified

The Father of DOE



R. A. FISHER 1929


Experimental Design
Develop Experimental Question or Hypothesis
Define Variables
Treatments
Variables hypothesized to influence response
Controls
Variables with no influence, or a predicatble influence on response
Responses
Measurements needed to answer the experimental question
Define Experimental and Sample Units
Experimental Units
Treatment and control units to be randomzied
Sample Units
Units that will be measured
Estimate Sample Size
Error Estimate
A priori assumption from previous work need to estimate sample size
Randomization and Layout

Introduction to Statistical Quality Control

2nd Hour

STATISTICAL QUALITY CONTROL

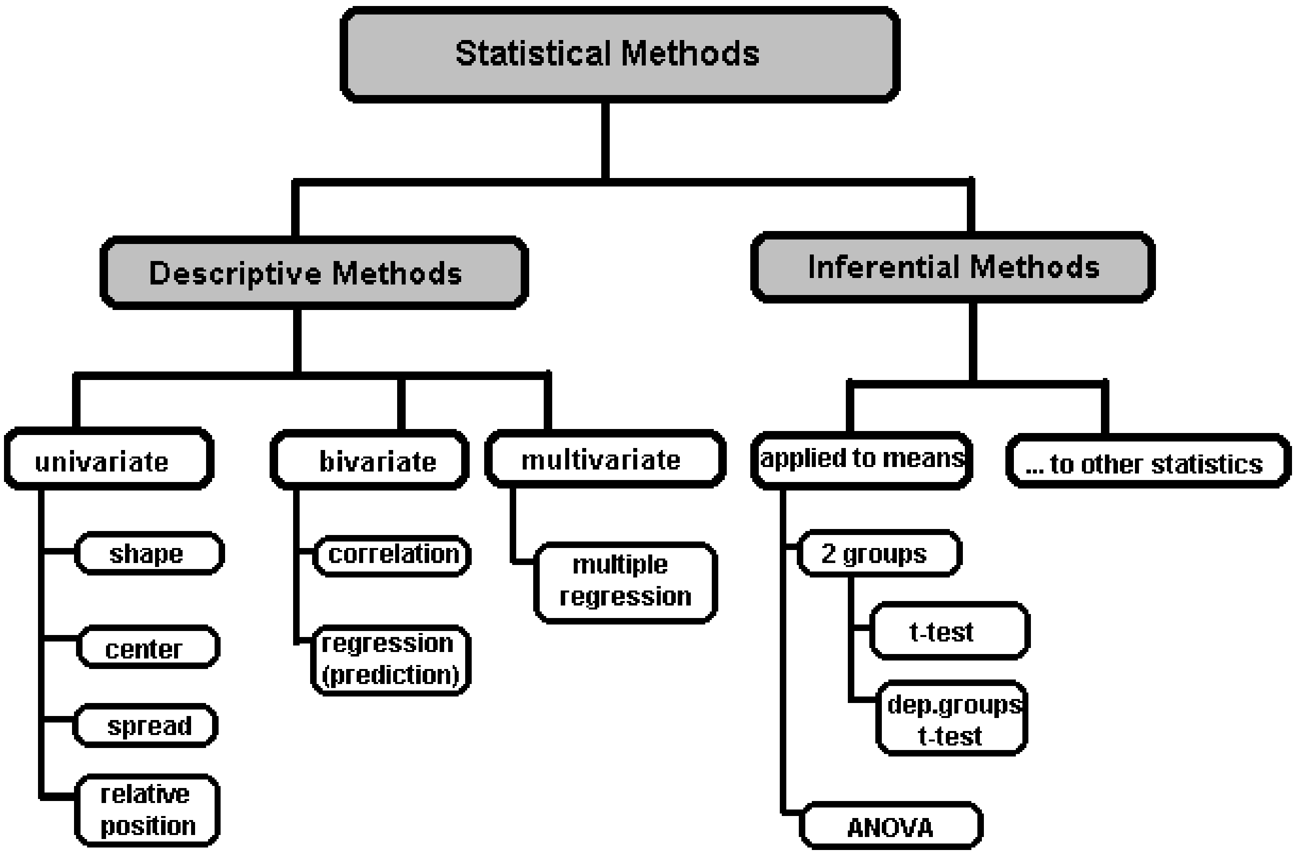
Basic Statistics

BASICS OF STATISTICS

Definition: Science of collection, presentation, analysis, and reasonable interpretation of data.

Statistics presents a rigorous scientific method for gaining insight into data. For example, suppose we measure the weight of 100 patients in a study. With so many measurements, simply looking at the data fails to provide an informative account. However statistics can give an instant overall picture of data based on graphical presentation or numerical summarization irrespective to the number of data points. Besides data summarization, another important task of statistics is to make inference and predict relations of variables.

A TAXONOMY OF STATISTICS



STATISTICAL DESCRIPTION OF DATA

- Statistics describes a numeric set of data by its
  - Center
  - Variability
  - Shape
- Statistics describes a categorical set of data by
  - Frequency, percentage or proportion of each category

SOME DEFINITIONS

Variable - any characteristic of an individual or entity. A variable can take different values for different individuals. Variables can be *categorical* or *quantitative*. Per S. S. Stevens…

- Nominal - Categorical variables with no inherent order or ranking sequence such as names or classes (e.g., gender). Value may be a numerical, but without numerical value (e.g., I, II, III). The only operation that can be applied to Nominal variables is enumeration.
- Ordinal - Variables with an inherent rank or order, e.g. mild, moderate, severe. Can be compared for equality, or greater or less, but not *how much* greater or less.
- Interval - Values of the variable are ordered as in Ordinal, and additionally, differences between values are meaningful, however, the scale is not absolutely anchored. Calendar dates and temperatures on the Fahrenheit scale are examples. Addition and subtraction, but not multiplication and division are meaningful operations.
- Ratio - Variables with all properties of Interval plus an absolute, non-arbitrary zero point, e.g. age, weight, temperature (Kelvin). Addition, subtraction, multiplication, and division are all meaningful operations.

SOME DEFINITIONS

Distribution - (of a variable) tells us what values the variable takes and how often it takes these values.

- Unimodal - having a single peak
- Bimodal - having two distinct peaks
- Symmetric - left and right half are mirror images.

FREQUENCY DISTRIBUTION

Consider a data set of 26 children of ages 1-6 years. Then the frequency distribution of variable ‘age’ can be tabulated as follows:

Frequency Distribution of Age

Age	1	2	3	4	5	6
Frequency	5	3	7	5	4	2

Grouped Frequency Distribution of Age:

Age Group	1-2	3-4	5-6
Frequency	8	12	6

CUMULATIVE FREQUENCY

Cumulative frequency of data in previous page

Age	1	2	3	4	5	6
Frequency	5	3	7	5	4	2
Cumulative Frequency	5	8	15	20	24	26

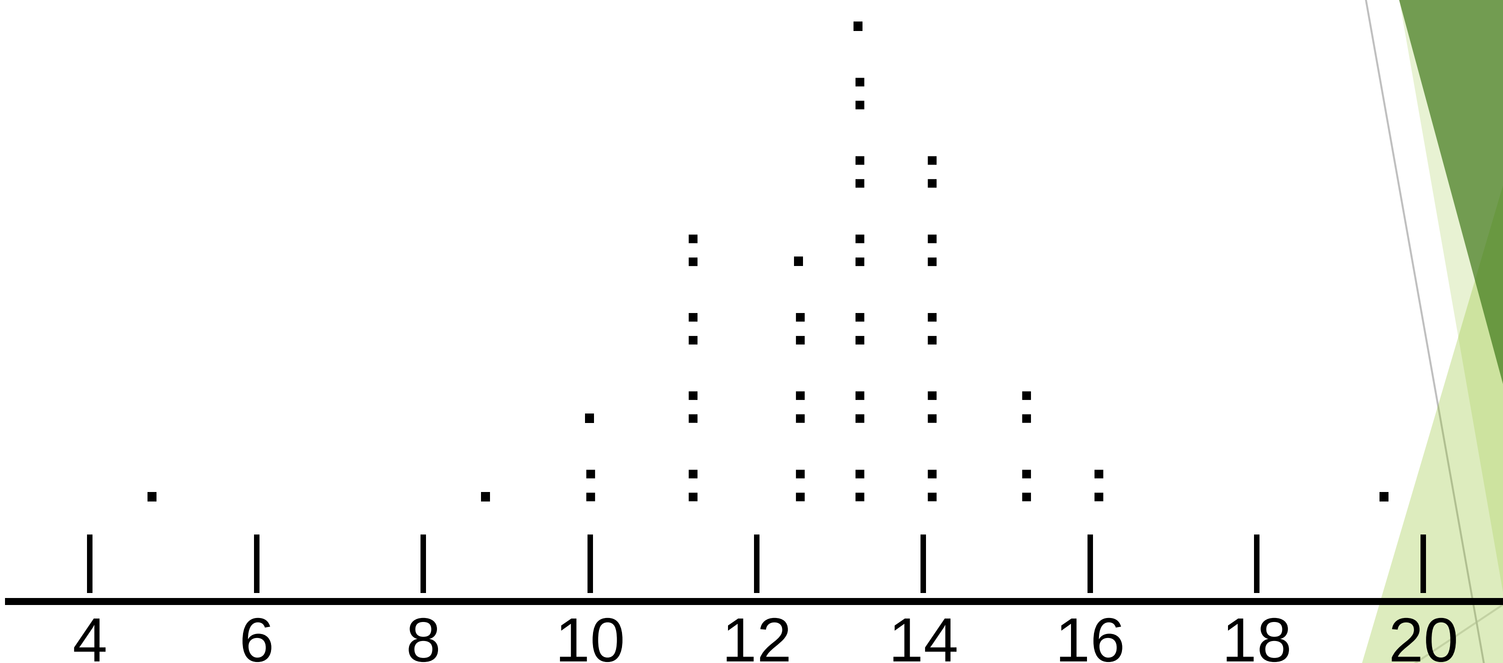
Age Group	1-2	3-4	5-6
Frequency	8	12	6
Cumulative Frequency	8	20	26

% of Population over 65—Data

Alabama	13	Louisiana	11	Ohio	13
Alaska	5	Maine	14	Oklahoma	14
Arizona	13	Maryland	11	Oregon	14
Arkansas	15	Mass	14	Penn	16
California	11	Michigan	12	R Island	16
Colorado	10	Minnesota	12	S Carolina	12
Connecticut	14	Mississippi	12	S Dakota	14
Delaware	13	Missouri	14	Tennesse	13
Florida	19	Montana	13	Texas	10
Georgia	10	Nebraska	14	Utah	9
Hawaii	13	Mevada	11	Vermont	12
Idaho	11	N Hampshire	12	Virginia	11
Illinois	13	N Jersey	14	Washington	12
Indiana	13	N Mexico	11	W Virginia	15
Iowa	15	N York	13	Wisconsin	13
Kansas	14	N Carolina	13	Wyoming	11
Kentucky	13	N Dakota	15		

% of Population over 65—Dot Plot

Percent of Population over 65 years of Age in the 50 States



Source: Statistical Abstract of the US

x-axis: Number in % to nearest integer

% of Population over 65—Data

	Class			Tally	Fre-quency	Rel Freq
A	4	≤ x <	6 %	______	______	______
B	6	≤ x <	8 %	______	______	______
C	8	≤ x <	10 %	______	______	______
D	10	≤ x <	12 %	______	______	______
E	12	≤ x <	14 %	______	______	______
F	14	≤ x <	16 %	______	______	______
G	16	≤ x <	18 %	______	______	______
H	18	≤ x <	20 %	______	______	______
				Totals		

% of Population over 65—Data

	Class				Tally	Fre-quency	Rel Freq
A	4	$\leq x <$	6	%	I	1	0.02
B	6	$\leq x <$	8	%	0	0	0.00
C	8	$\leq x <$	10	%	I	1	0.02
D	10	$\leq x <$	12	%	~~IIIII~~ ~~IIIII~~ I	11	0.22
E	12	$\leq x <$	14	%	~~IIIII~~ ~~IIIII~~ ~~IIIII~~ ~~IIII~~	20	0.40
F	14	$\leq x <$	16	%	~~IIIII~~ ~~IIIII~~ IIIII	14	0.28
G	16	$\leq x <$	18	%	II	2	0.04
H	18	$\leq x <$	20	%	I	1	0.02
				Totals	50	50	1.00

DATA PRESENTATION

Two types of statistical presentation of data - graphical and numerical.

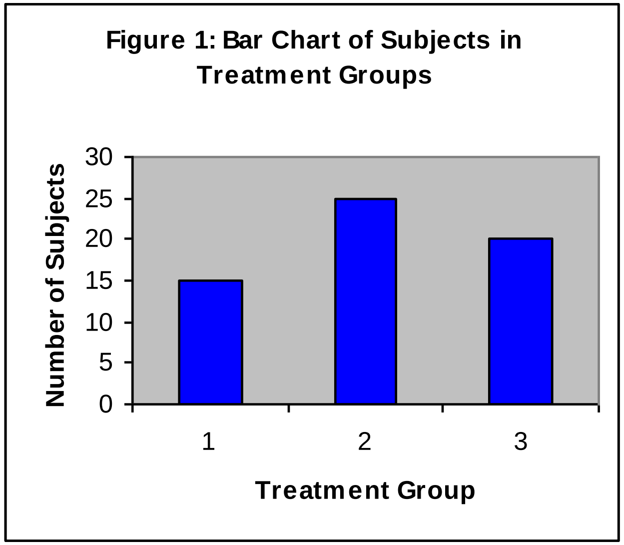
Graphical Presentation: We look for the overall pattern and for striking deviations from that pattern. Over all pattern usually described by shape, center, and spread of the data. An individual value that falls outside the overall pattern is called an *outlier*.

Bar diagram and Pie charts are used for categorical variables.

Histogram, stem and leaf and Box-plot are used for numerical variable.

Data Presentation –Categorical Variable

Bar Diagram: Lists the categories and presents the percent or count of individuals who fall in each category.

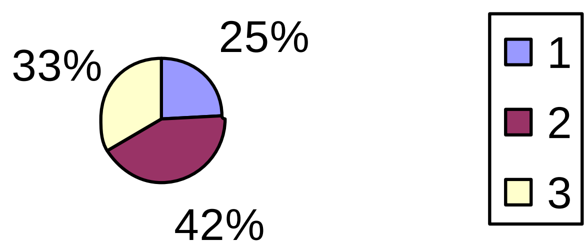


Treatment Group	Frequency	Proportion	Percent (%)
1	15	(15/60)=0.25	25.0
2	25	(25/60)=0.333	41.7
3	20	(20/60)=0.417	33.3
Total	60	1.00	100

Data Presentation –Categorical Variable

Pie Chart: Lists the categories and presents the percent or count of individuals who fall in each category.

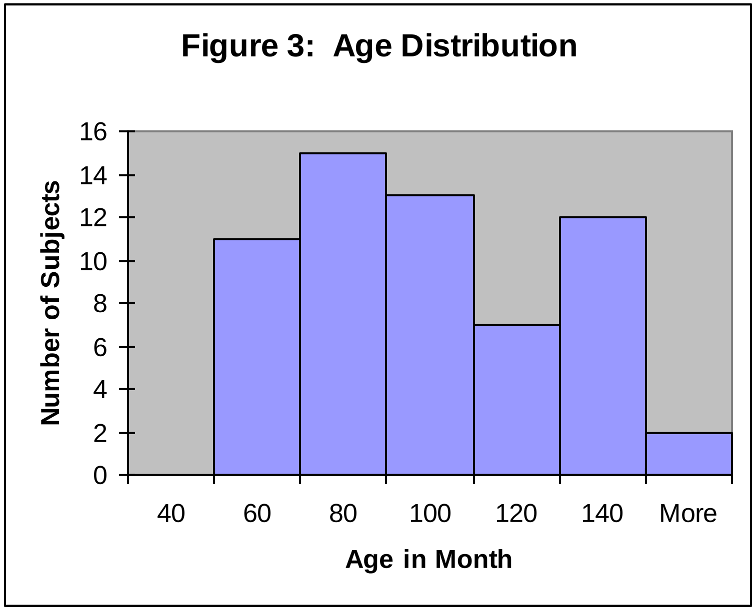
Figure 2: Pie Chart of Subjects in Treatment Groups



Treatment Group	Frequency	Proportion	Percent (%)
1	15	(15/60)=0.25	25.0
2	25	(25/60)=0.333	41.7
3	20	(20/60)=0.417	33.3
Total	60	1.00	100

GRAPHICAL PRESENTATION –NUMERICAL VARIABLE

Histogram: Overall pattern can be described by its shape, center, and spread. The following age distribution is right skewed. The center lies between 80 to 100. No outliers.



Mean	90.41666667
Standard Error	3.902649518
Median	84
Mode	84
Standard Deviation	30.22979318
Sample Variance	913.8403955
Kurtosis	-1.183899591
Skewness	0.389872725
Range	95
Minimum	48
Maximum	143
Sum	5425
Count	60

HISTOGRAMS – USEFUL FOR LARGE DATA

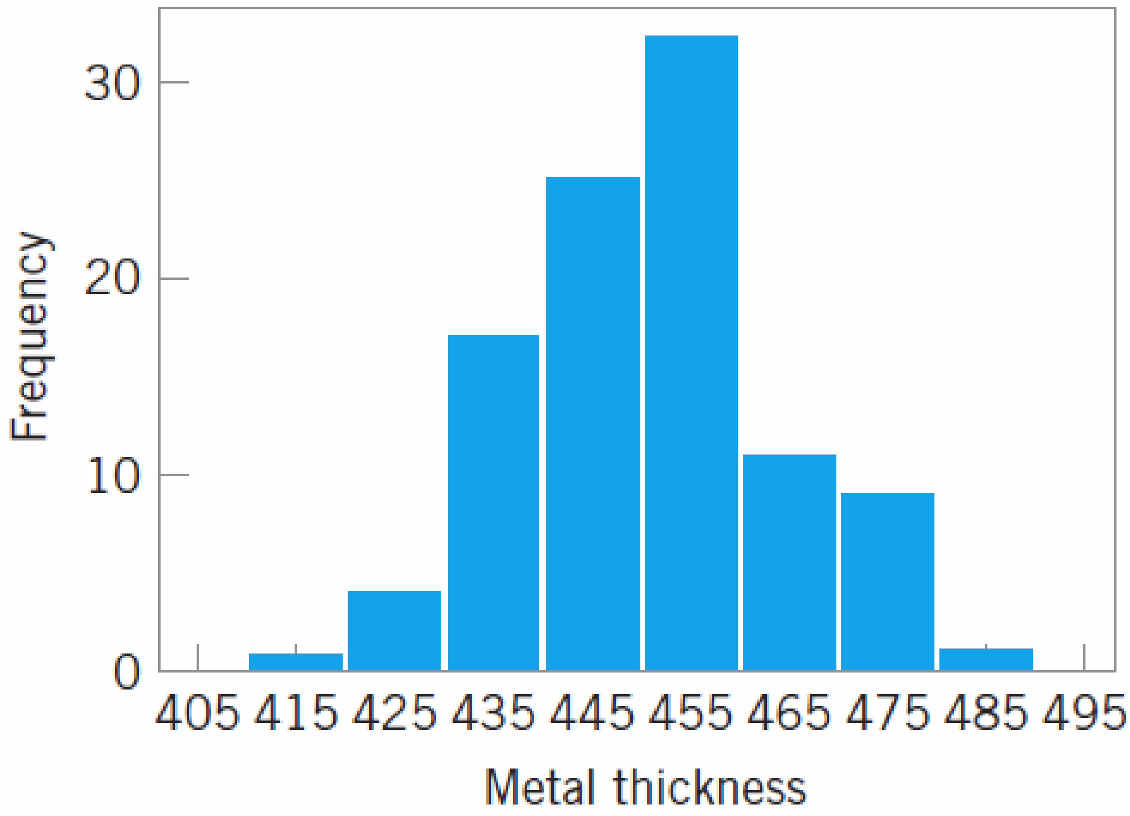
TABLE 3.2

Layer Thickness (Å) on Semiconductor Wafers

438	450	487	451	452	441	444	461	432	471
413	450	430	437	465	444	471	453	431	458
444	450	446	444	466	458	471	452	455	445
468	459	450	453	473	454	458	438	447	463
445	466	456	434	471	437	459	445	454	423
472	470	433	454	464	443	449	435	435	451
474	457	455	448	478	465	462	454	425	440
454	441	459	435	446	435	460	428	449	442
455	450	423	432	459	444	445	454	449	441
449	445	455	441	464	457	437	434	452	439

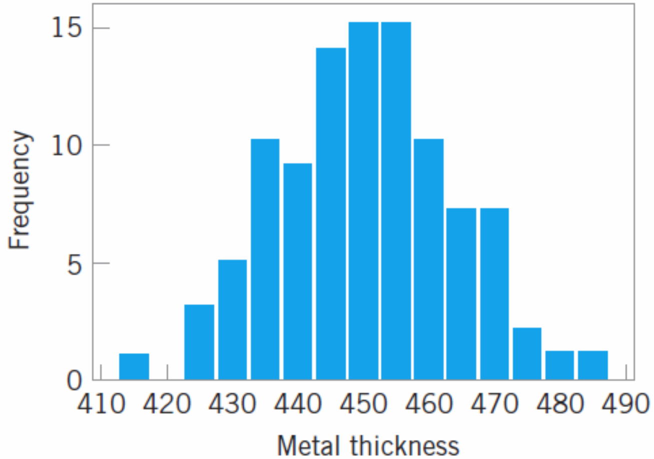
Group values of the variable into bins, then count the number of observations that fall into each bin

Plot frequency (or relative frequency) versus the values of the variable

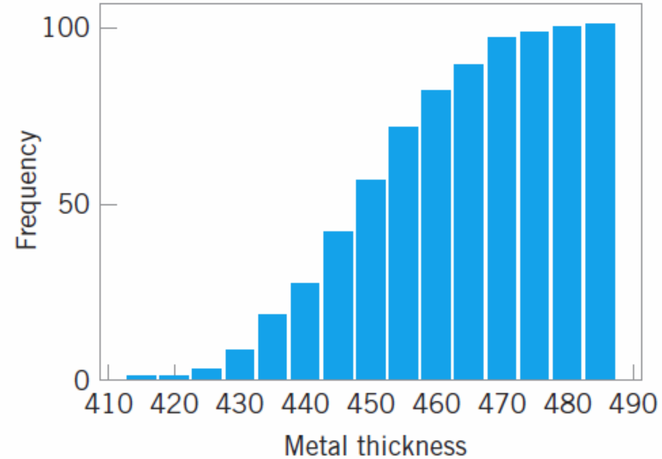


■ FIGURE 3.3 Minitab histogram for the metal layer thickness data in Table 3.2.

ADDITIONAL MINITAB GRAPHS



■ **FIGURE 3.4** Minitab histogram with 15 bins for the metal layer thickness data.



■ **FIGURE 3.5** A cumulative frequency plot of the metal thickness data from Minitab.

TABLE 3.3

Surface Finish Defects in Painted Automobile Hoods

6	1	5	7	8	6	0	2	4	2
5	2	4	4	1	4	1	7	2	3
4	3	3	3	6	3	2	3	4	5
5	2	3	4	4	4	2	3	5	7
5	4	5	5	4	5	3	3	3	12

Figure 3.6 is the histogram of the defects. Notice that the number of defects is a discrete variable. From either the histogram or the tabulated data we can determine

$$\text{Proportions of hoods with at least 3 defects} = \frac{39}{50} = 0.78$$

and

$$\text{Proportions of hoods with between 0 and 2 defects} = \frac{11}{50} = 0.22$$

These proportions are examples of relative frequencies.

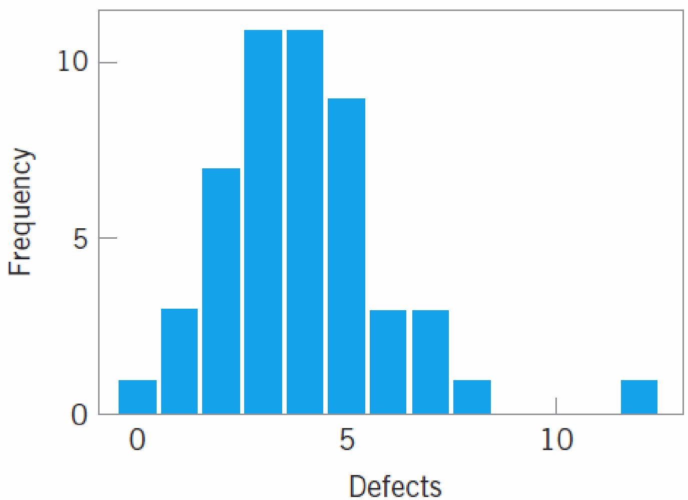


FIGURE 3.6 Histogram of the number of defects in painted automobile hoods (Table 3.3).

NUMERICAL PRESENTATION

A fundamental concept in summary statistics is that of a *central value* for a set of observations and the extent to which the central value characterizes the whole set of data. Measures of central value such as the mean or median must be coupled with measures of data dispersion (e.g., average distance from the mean) to indicate how well the central value characterizes the data as a whole.

To understand how well a central value characterizes a set of observations, let us consider the following two sets of data:

A: 30, 50, 70

B: 40, 50, 60

The mean of both two data sets is 50. But, the distance of the observations from the mean in data set A is larger than in the data set B. Thus, the mean of data set B is a better representation of the data set than is the case for set A.

METHODS OF CENTER MEASUREMENT

Center measurement is a summary measure of the overall level of a dataset

Commonly used methods are mean, median, mode, geometric mean etc.

Mean: Summing up all the observation and dividing by number of observations.
Mean of 20, 30, 40 is (20+30+40)/3 = 30.

Notation: Let $x_1, x_{2,} \dots x_n$ are n observations of a variable x. Then the mean of this variable,

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

METHODS OF CENTER MEASUREMENT

Median: The middle value in an ordered sequence of observations. That is, to find the median we need to order the data set and then find the middle value. In case of an even number of observations the average of the two middle most values is the median. For example, to find the median of {9, 3, 6, 7, 5}, we first sort the data giving {3, 5, 6, 7, 9}, then choose the middle value 6. If the number of observations is even, e.g., {9, 3, 6, 7, 5, 2}, then the median is the average of the two middle values from the sorted sequence, in this case, (5 + 6) / 2 = 5.5.

Mode: The value that is observed most frequently. The mode is undefined for sequences in which no observation is repeated.

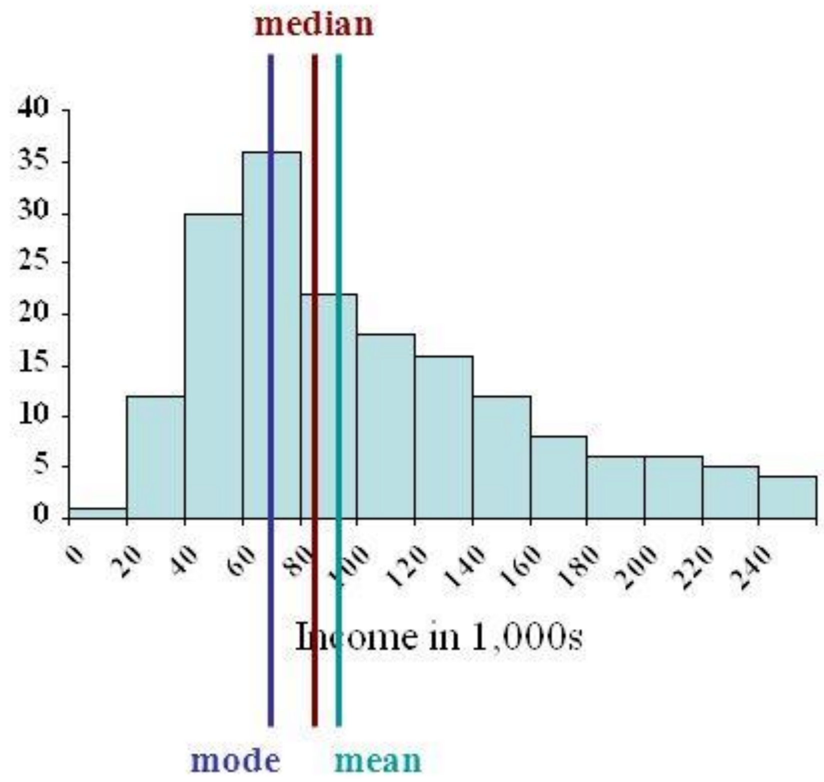
MEAN OR MEDIAN

The median is less sensitive to outliers (extreme scores) than the mean and thus a better measure than the mean for highly skewed distributions, e.g. family income. For example mean of 20, 30, 40, and 990 is (20+30+40+990)/4 =270. The median of these four observations is (30+40)/2 =35. Here 3 observations out of 4 lie between 20-40. So, the mean 270 really fails to give a realistic picture of the major part of the data. It is influenced by extreme value 990.

Skewed distributions

- Skewness refers to the asymmetry of the distribution
- A positively skewed distribution is asymmetrical and points in the positive direction.

Mode = 70,000$
Median = 88,700$
Mean = 93,600$



•**mode** < **median** < **mean**

METHODS OF VARIABILITY MEASUREMENT

Variability (or dispersion) measures the amount of scatter in a dataset.

Commonly used methods: *range*, *variance*, *standard deviation*, *interquartile range*, *coefficient of variation etc*.

Range: The difference between the largest and the smallest observations. The range of 10, 5, 2, 100 is (100-2)=98. It’s a crude measure of variability.

METHODS OF VARIABILITY MEASUREMENT

Variance: The variance of a set of observations is the average of the squares of the deviations of the observations from their mean. In symbols, the variance of the n observations x_1, x_2,…x_n is

$$S^2 = \frac{(x_1 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}$$

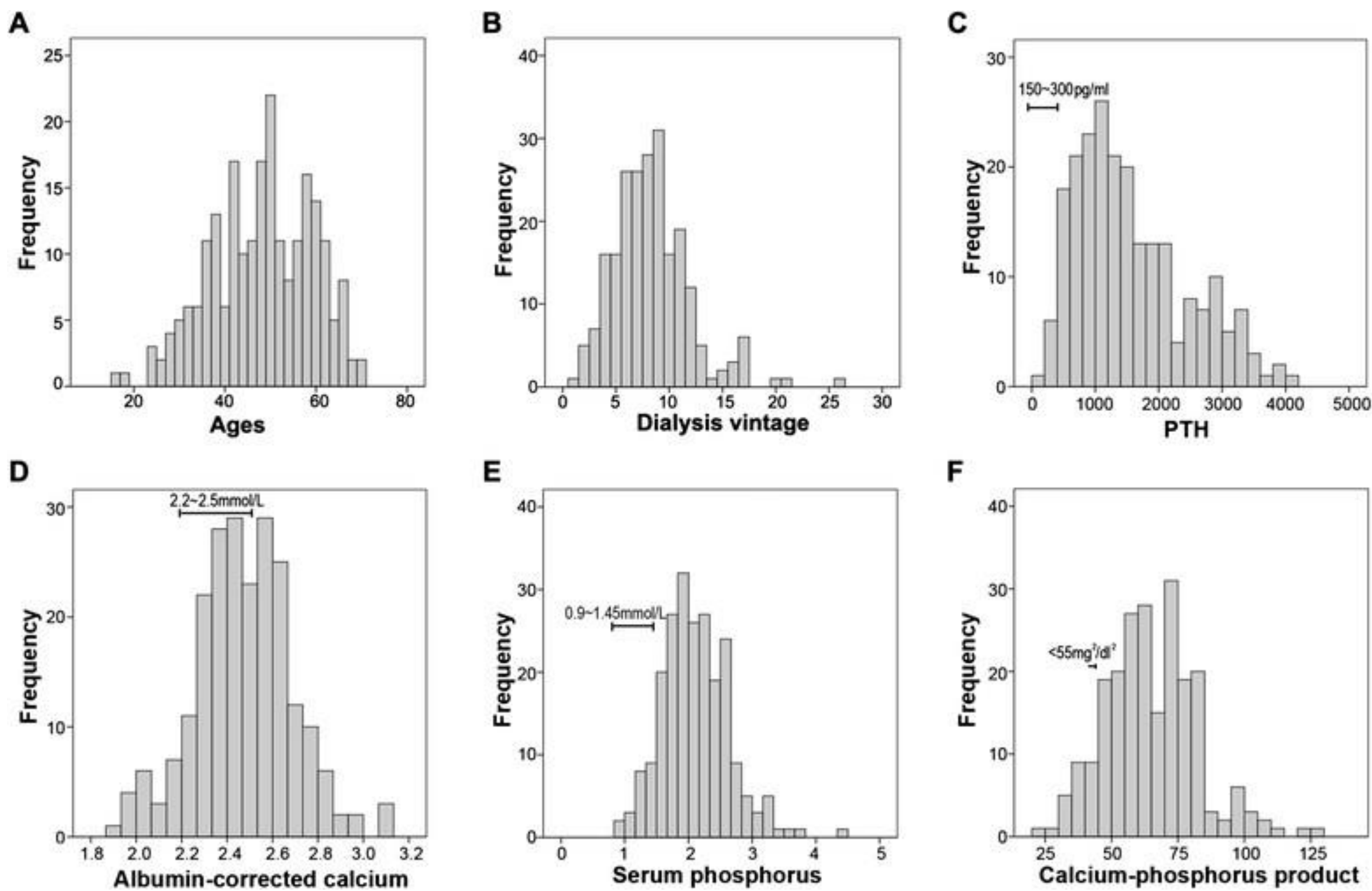
Variance of 5, 7, 3? Mean is (5+7+3)/3 = 5 and the variance is

$$\frac{(5-5)^2 + (3-5)^2 + (7-5)^2}{3-1} = 4$$

Standard Deviation: Square root of the variance. The standard deviation of the above example is 2.

SHAPE OF DATA

- Shape of data is measured by
  - Skewness
  - Kurtosis



SKEWNESS

- Measures asymmetry of data
  - Positive or right skewed: Longer right tail
  - Negative or left skewed: Longer left tail

Let $x_1, x_2, \ldots x_n$ be n observations. Then,

$$\text{Skewness} = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$$

KURTOSIS

- Measures peakedness of the distribution of data. The kurtosis of normal distribution is 0.

Let $x_1, x_2, \ldots x_n$ be n observations. Then,

$$\text{Kurtosis} = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} - 3$$

NUMERICAL SUMMARY OF DATA

Sample average:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.1}$$

Note that the sample average $\bar{x}$ is simply the arithmetic mean of the n observations. The sample average for the metal thickness data in Table 3.2 is

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100} = \frac{45.001}{100} = 450.01\,\text{Å}$$

Refer to Fig. 3.3 and note that the sample average is the point at which the histogram exactly "balances." Thus, the sample average represents the center of mass of the sample data.

The variability in the sample data is measured by the **sample variance:**

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1} \tag{3.2}$$

Note that the sample variance is simply the sum of the squared deviations of each observation from the sample average $\overline{x}$, divided by the sample size minus one. If there is no variability in the sample, then each sample observation $x_i = \overline{x}$, and the sample variance $s^2 = 0$. Generally, the larger is the sample variance s^2, the greater is the variability in the sample data.

THE STANDARD DEVIATION

The units of the sample variance s^2 are the square of the original units of the data. This is often inconvenient and awkward to interpret, and so we usually prefer to use the square root of s^2, called the **sample standard deviation** s**,** as a measure of variability.

It follows that

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{3.3}$$

The primary advantage of the sample standard deviation is that it is expressed in the original units of measurement. For the metal thickness data, we find that

$$s^2 = 180.2928\,\text{Å}^2$$

and

$$s = 13.43\,\text{Å}$$

PROBABILITY DISTRIBUTIONS

The histogram (or stem-and-leaf plot, or box plot) is used to describe *sample* data. A **sample** is a collection of measurements selected from some larger source or **population.** For example, the measurements on layer thickness in Table 3.2 are obtained from a sample of wafers selected from the manufacturing process. The population in this example is the collection of all layer thicknesses produced by that process. By using statistical methods, we may be able to analyze the sample layer thickness data and draw certain conclusions about the process that manufactures the wafers.

A **probability distribution** is a mathematical model that relates the value of the variable with the probability of occurrence of that value in the population. In other words, we might visualize layer thickness as a **random variable,** because it takes on different values in the population according to some random mechanism, and then the probability distribution of layer thickness describes the probability of occurrence of any value of layer thickness in the population. There are two types of probability distributions.

Definition

1. **Continuous distributions.** When the variable being measured is expressed on a continuous scale, its probability distribution is called a *continuous distribution*. The probability distribution of metal layer thickness is continuous.

2. **Discrete distributions.** When the parameter being measured can only take on certain values, such as the integers 0, 1, 2, . . . , the probability distribution is called a *discrete distribution*. For example, the distribution of the number of nonconformities or defects in printed circuit boards would be a discrete distribution.

Sometimes called a
probability mass function

Sometimes called a
probability density function

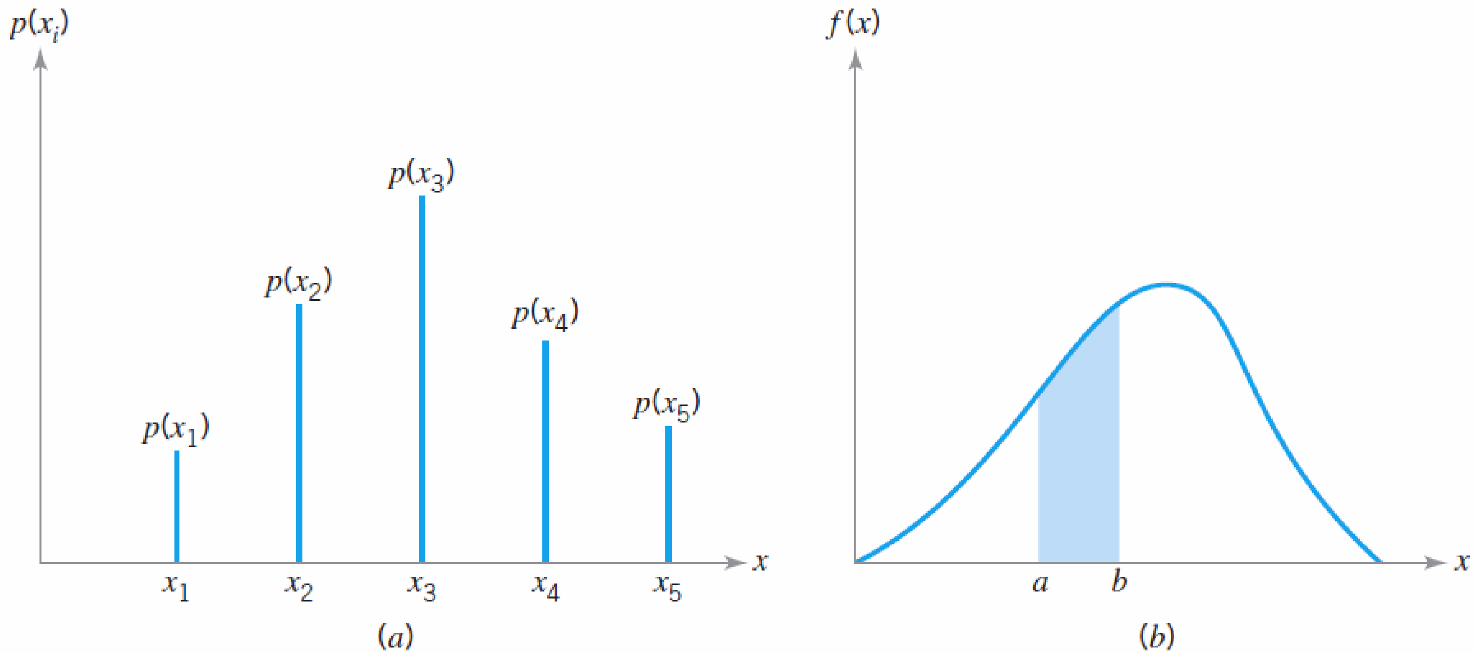


FIGURE 3.9 Probability distributions. (*a*) Discrete case. (*b*) Continuous case.

Will see many examples in the text

EXAMPLE 3.5 A Discrete Distribution

A manufacturing process produces thousands of semiconductor chips per day. On the average, 1% of these chips do not conform to specifications. Every hour, an inspector selects a random sample of 25 chips and classifies each chip in the sample as conforming or nonconforming. If we let x be the random variable representing the number of nonconforming chips in the sample, then the probability distribution of x is

$$p(x) = \binom{25}{x}(0.01)^x(0.99)^{25-x} \qquad x = 0,1,2,\ldots,25$$

where $\binom{25}{x} = 25!/[x!\,(25-x)!]$. This is a *discrete* distribution, since the observed number of nonconformances is $x = 0, 1, 2, \ldots, 25$, and is called the **binomial distribution.** We may calculate the probability of finding one or fewer nonconforming parts in the sample as

$$\begin{aligned}
P(x \le 1) &= P(x=0) + P(x=1) \\
&= p(0) + p(1) \\
&= \sum_{x=0}^{1} \binom{25}{x}(0.01)^x(0.99)^{25-x} \\
&= \frac{25!}{0!\,25!}(0.99)^{25}(0.01)^0 + \frac{25!}{1!\,24!}(0.99)^{24}(0.01)^1 \\
&= 0.7778 + 0.1964 = 0.9742
\end{aligned}$$

EXAMPLE 3.6 A Continuous Distribution

Suppose that x is a random variable that represents the actual contents in ounces of a 1-pound bag of coffee beans. The probability distribution of x is assumed to be

$$f(x) = \frac{1}{1.5} \qquad 15.5 \leq x \leq 17.0$$

This is a *continuous* distribution, since the range of x is the interval [15.5, 17.0]. This distribution is called the **uniform distribution,** and it is shown graphically in Figure 3.10. Note that the area under the function $f(x)$ corresponds to probability, so that the probability of a bag containing less than 16.0 oz is

$$P\{x \leq 16.0\} = \int_{15.5}^{16.0} f(x)dx = \int_{15.5}^{16.0} \frac{1}{1.5} dx$$

$$= \frac{x}{1.5}\bigg|_{15.5}^{16.0} = \frac{16.0 - 15.5}{1.5} = 0.3333$$

This follows intuitively from inspection of Figure 3.9.

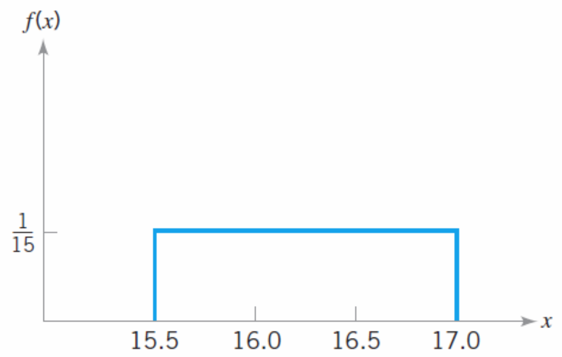


FIGURE 3.10 The uniform distribution for Example 3.6.

The **mean** μ of a probability distribution is a measure of the **central tendency** in the distribution, or its **location.** The mean is defined as

$$\mu = \begin{cases} \int_{-\infty}^{\infty} x f(x)\, dx, x \text{ continuous} & (3.5a) \\ \sum_{i=1}^{\infty} x_i p(x_i), x \text{ discrete} & (3.5b) \end{cases}$$

For the case of a discrete random variable with exactly N equally likely values [that is, $p(x_i) = 1/N$], then equation 3.5b reduces to

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

The mean is the point at which the distribution exactly “balances”.

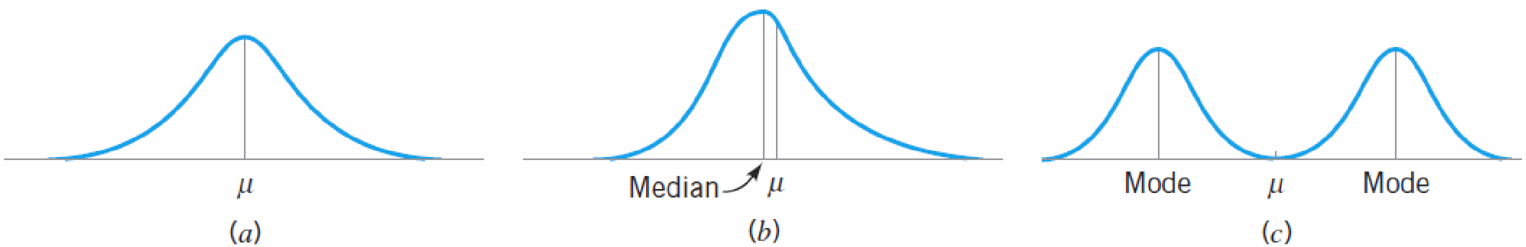


FIGURE 3.11 The mean of a distribution.

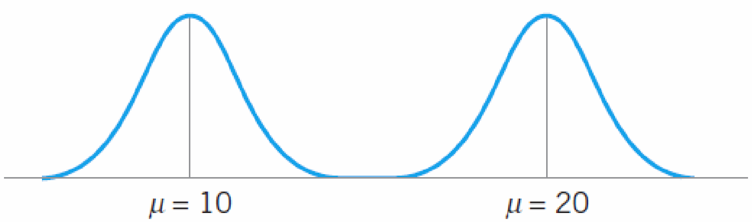


FIGURE 3.12 Two probability distributions with different means.

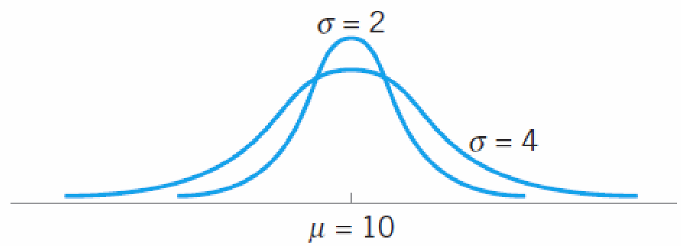


FIGURE 3.13 Two probability distributions with the same mean but different standard deviations.

The mean is not necessarily the 50th percentile of the distribution (that's the median)

The mean is not necessarily the most likely value of the random variable (that's the mode)

The scatter, spread, or variability in a distribution is expressed by the **variance** σ^2. The definition of the variance is

$$\sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, x \text{ continuous} & (3.6a) \\ \sum_{i=1}^{\infty} (x_i-\mu)^2 p(x_i), x \text{ discrete} & (3.6b) \end{cases}$$

when the random variable is discrete with N equally likely values, then equation 3.6b becomes

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}$$

and we observe that in this case the variance is the average squared distance of each member of the population from the mean. Note the similarity to the sample variance s^2, defined in equation 3.2. If $\sigma^2 = 0$, there is no variability in the population. As the variability increases, the variance σ^2 increases. The variance is expressed in the square of the units of the original variable. For example, if we are measuring voltages, the units of the variance are (volts)^2. Thus, it is customary to work with the square root of the variance, called the **standard deviation** σ. It follows that

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}} \tag{3.7}$$

The standard deviation is a measure of spread or scatter in the population expressed in the original units. Two distributions with the same mean but different standard deviations are shown in Figure 3.13.

3.3 Important Continuous Distributions

The Normal Distribution

Definition

The **normal distribution** is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty \qquad (3.21)$$

The mean of the normal distribution is μ $(-\infty < \mu < \infty)$ and the variance is $\sigma^2 > 0$.

The normal distribution is used so much that we frequently employ a special notation, $x - N(\mu, \sigma^2)$, to imply that x is normally distributed with mean μ and variance σ^2. The visual appearance of the normal distribution is a symmetric, unimodal or **bell-shaped** curve and is shown in Figure 3.16.

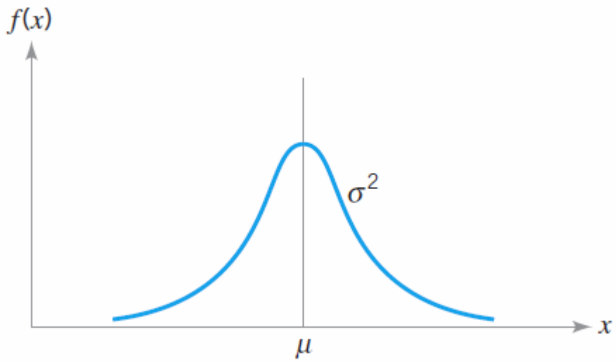


FIGURE 3.16 The normal distribution.

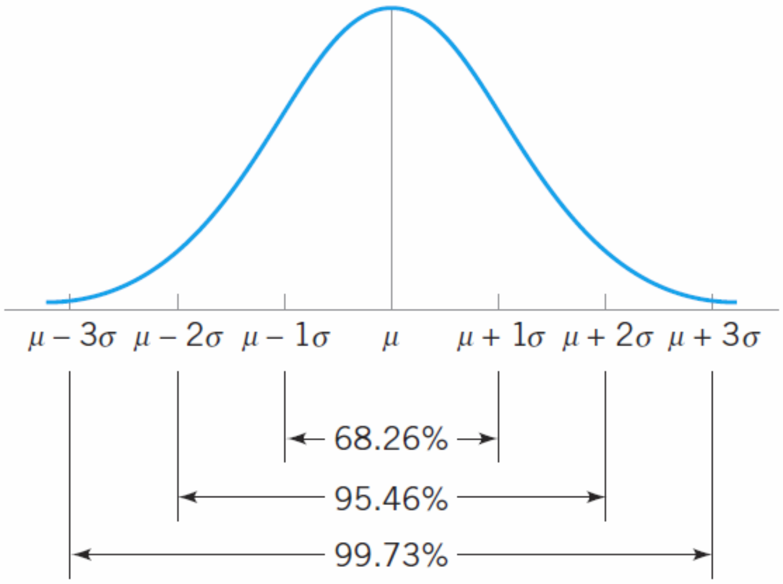


FIGURE 3.17 Areas under the normal distribution.

The cumulative normal distribution is defined as the probability that the normal random variable x is less than or equal to some value a, or

$$P\{x \le a\} = F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \qquad (3.22)$$

This integral cannot be evaluated in closed form. However, by using the change of variable

$$z = \frac{x-\mu}{\sigma} \qquad (3.23)$$

the evaluation can be made independent of μ and σ^2. That is,

$$P\{x \le a\} = P\left\{z \le \frac{a-\mu}{\sigma}\right\} \equiv \Phi\left(\frac{a-\mu}{\sigma}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the **standard normal distribution** (mean = 0, standard deviation = 1). A table of the cumulative standard normal distribution is given in Appendix Table II. The transformation (3.23) is usually called **standardization,** because it converts a $N(\mu, \sigma^2)$ random variable into an $N(0, 1)$ random variable.

EXAMPLE 3.7 Tensile Strength of Paper

The time to resolve customer complaints is a critical quality characteristic for many organizations. Suppose that this time in a financial organization, say, x—is normally distributed with mean $\mu = 40$ hours and standard deviation $\sigma = 2$ hours denoted $x \sim N(40, 2^2)$. What is the probability that a customer complaint will be resolved in less than 35 hours?

SOLUTION

The desired probability is

$$P\{x \leq 35\}$$

To evaluate this probability from the standard normal tables, we standardize the point 35 and find

$$P\{x \leq 35\} = P\left\{z \leq \frac{35-40}{2}\right\} =$$

$$P\{z \leq -2.5\} = \Phi(-2.5) = 0.0062$$

Consequently, the desired probability is

$$p\{x \geq 35\} = 0.0062$$

Figure 3.18 shows the tabulated probability for both the $N(40, 2^2)$ distribution and the standard normal distribution. Note that the shaded area to the left of 35 hr in Figure 3.18 represents the fraction of customer complaints resolved in less than or equal to 35 hours.

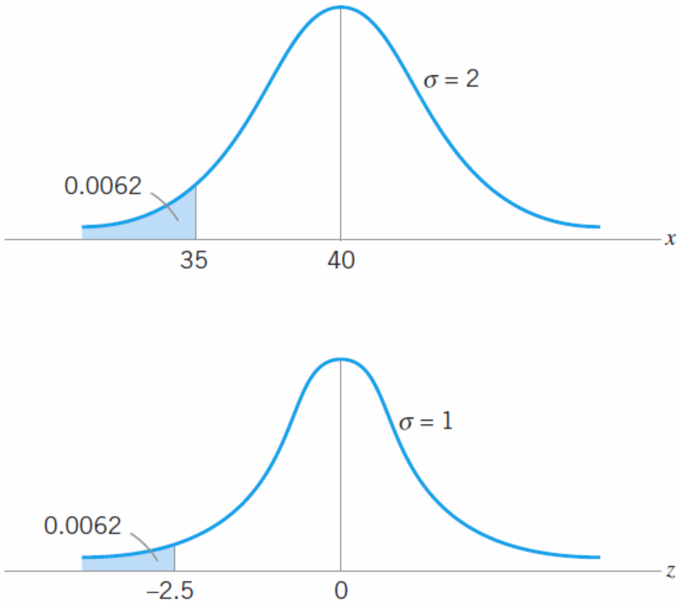


FIGURE 3.18 Calculation of $P\{x \leq 35\}$ in Example 3.7.

EXAMPLE 3.9 Another Use of the Standard Normal Table

Sometimes instead of finding the probability associated with a particular value of a normal random variable, we find it necessary to do the opposite—find a particular value of a normal random variable that results in a given probability. For example, suppose that $x \sim N(10, 9)$. Find the value of x—say, a—such that $P\{x > a\} = 0.05$.

SOLUTION

From the problem statement, we have

$$P\{x > a\} = P\left\{z > \frac{a-10}{3}\right\} = 0.05$$

or

$$P\left\{z \leq \frac{a-10}{3}\right\} = 0.95$$

From Appendix Table II, we have $P\{z \leq 1.645\} = 0.95$, so

$$\frac{a-10}{3} = 1.645$$

or

$$a = 10 + 3(1.645) = 14.935$$

The normal distribution has many useful properties. One of these is relative to **linear combinations** of normally and independently distributed random variables. If $x_1, x_2 \ldots, x_n$ are normally and independently distributed random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively, then the distribution of

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

is normal with mean

$$\mu_y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \tag{3.27}$$

and variance

$$\sigma_y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \tag{3.28}$$

where $a_1, a_2, \ldots, a_n$ are constants.

THE CENTRAL LIMIT THEOREM

Definition

The Central Limit Theorem If $x_1, x_2, \ldots, x_n$ are independent random variables with mean μ_i and variance σ_i^2, and if $y = x_1 + x_2 + \cdots + x_n$, then the distribution of

$$\frac{y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

approaches the $N(0, 1)$ distribution as n approaches infinity.

Practical interpretation – the sum of independent random variables is approximately normally distributed regardless of the distribution of each individual random variable in the sum

Introduction to Statistical Quality Control

3rd Hour

STATISTICAL QUALITY CONTROL

The Magnificent Seven

What are the Basic Seven Tools of Quality?

- Fishbone Diagrams
- Check Sheets
- Flowcharts
- Histograms
- Pareto Analysis
- Scatter Plots
- Run Charts

Where did the Basic Seven come from?

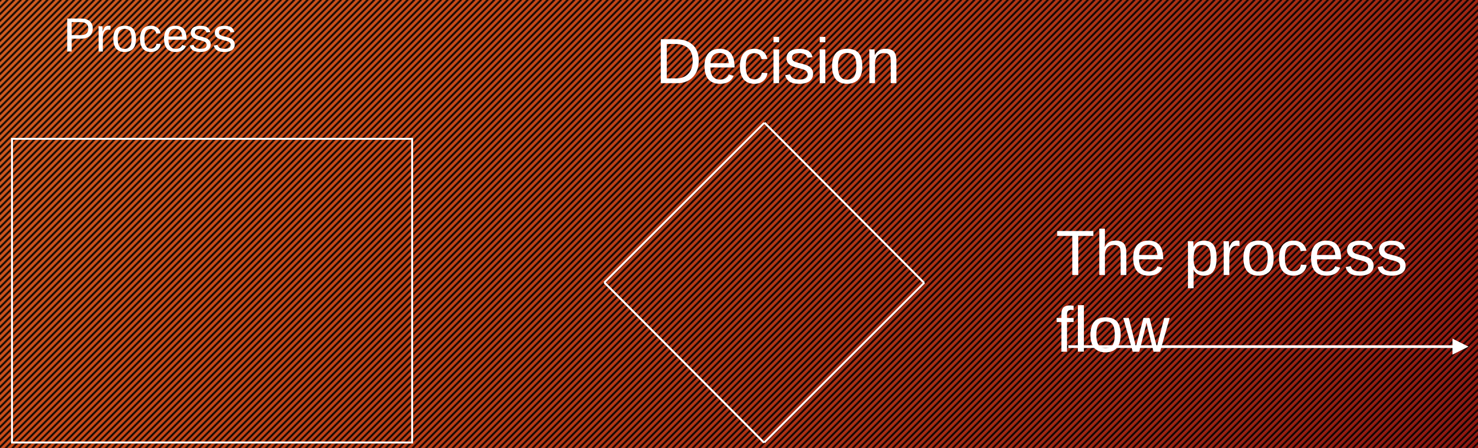
Kaoru Ishikawa

- Known for “Democratizing Statistics”
- The Basic Seven Tools made statistical analysis less complicated for the average person
- Good Visual Aids make statistical and quality control more comprehendible.

Flowcharts

Flowcharts

- No statistics involved
- A graphical picture of a PROCESS

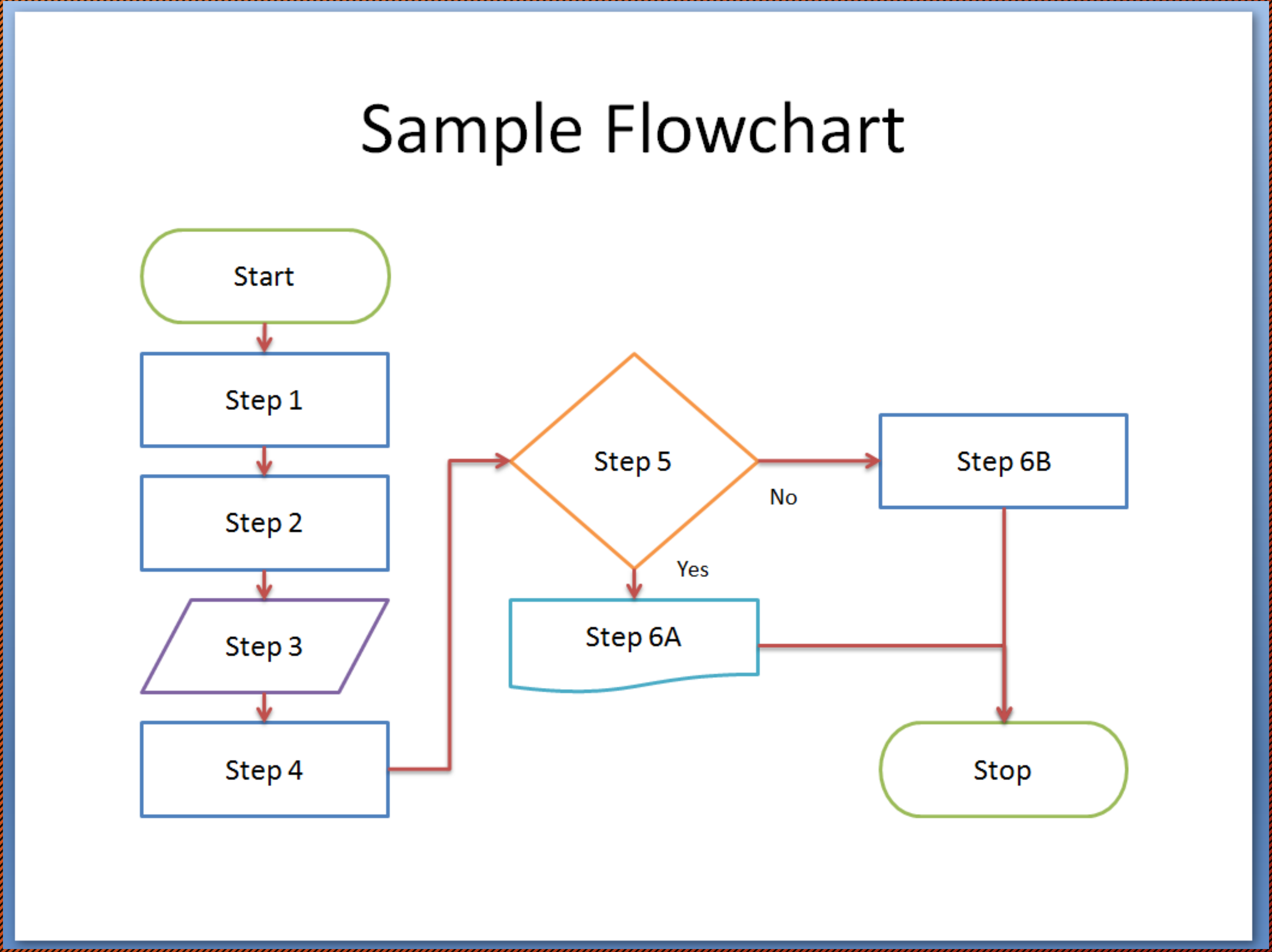


Flowcharts

Don't Forget to:

- Define symbols before beginning
- Stay consistent
- Check that process is accurate

Flowcharts


Sample Flowchart
Start
Step 1
Step 2
Step 3
Step 4
Step 5
No
Yes
Step 6A
Step 6B
Stop

Check Sheets

The check sheet is a form (document) used to collect data in real time at the location where the data is generated. The data it captures can be quantitative or qualitative. When the information is quantitative, the check sheet is sometimes called a tally sheet.

Check Sheets

How to Develop a type of Check Sheet?

Decide When Data will be Collected

Detect Types/Event Occurrences	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Totals
Crooked Needle		II	IIII	I	II	III		12
Crimped Needle Hub	IIIII	II	IIIII	II		II		16
Whole in needle Sheath		IIIII	IIIII	IIIII	IIIII	IIIII		25
No-fill volume	II	II	II			II		8
Low fill volume				I	I			2
Totals	7	11	21	9	8	12		

Fishbone Diagrams

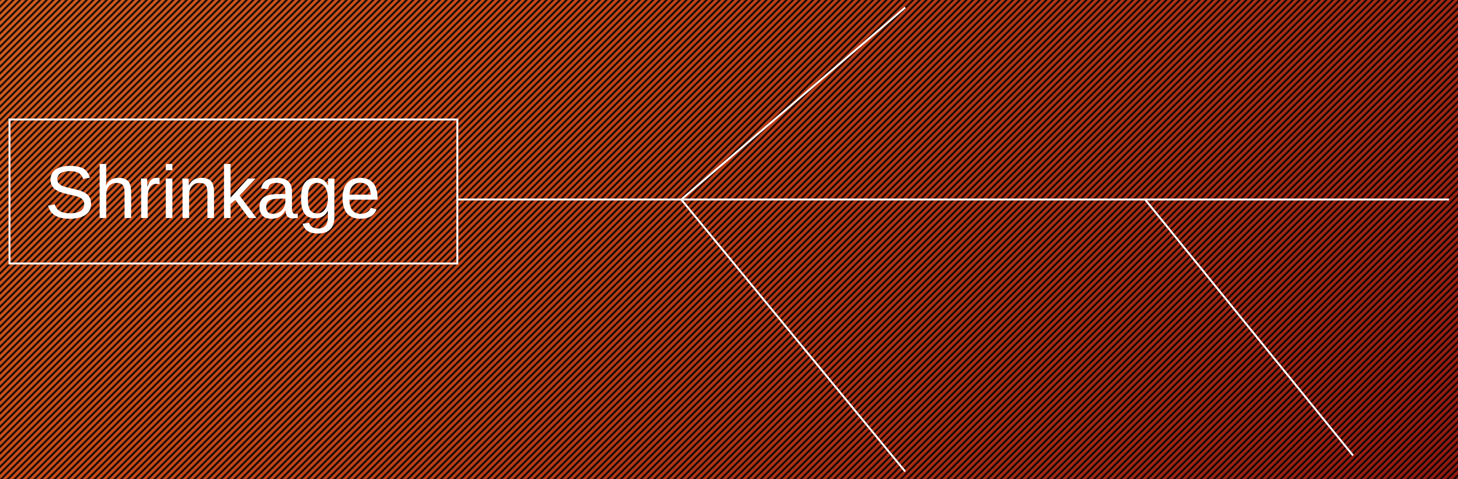
<u>Fishbone Diagrams</u>

- No statistics involved
- Maps out a process/problem
- Makes improvement easier
- Looks like a “Fish Skeleton”

Constructing a Fishbone Diagram

- Step 1 - Identify the Problem
- Step 2 - Draw “spine” and “bones”

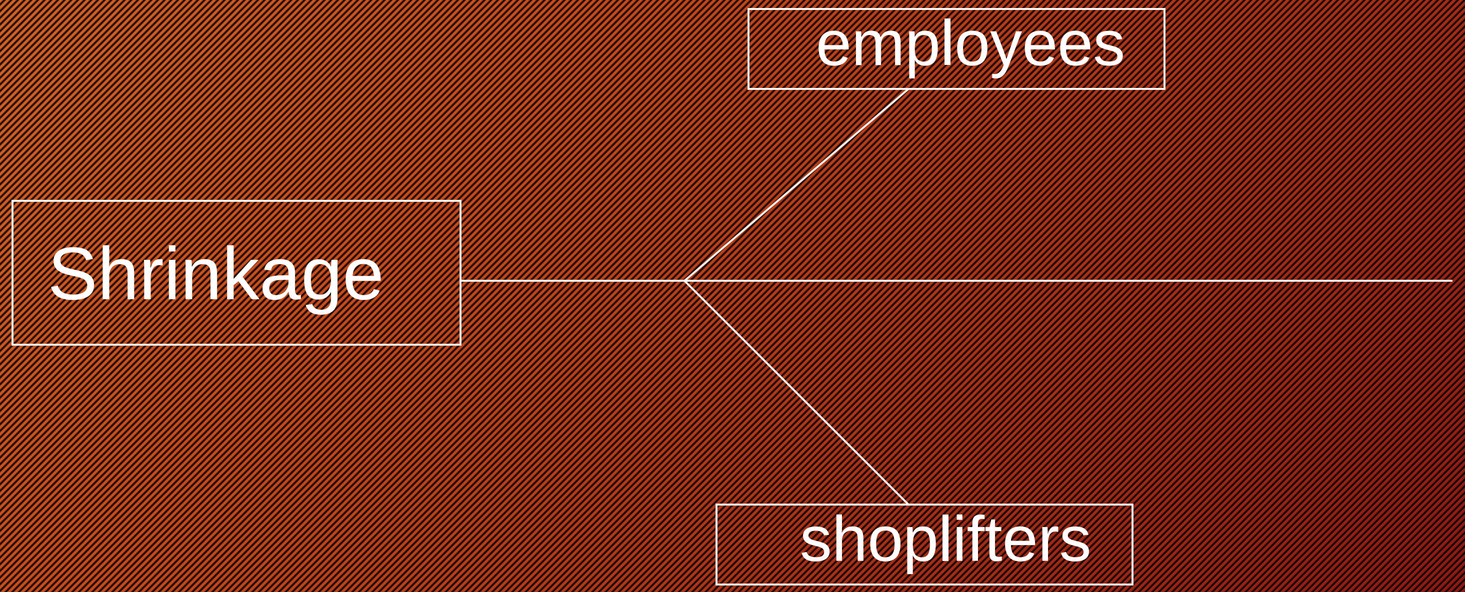
Example: High Inventory Shrinkage at local Drug Store



Constructing a Fishbone Diagram

- Step 3 - Identify different areas where problems may arise from

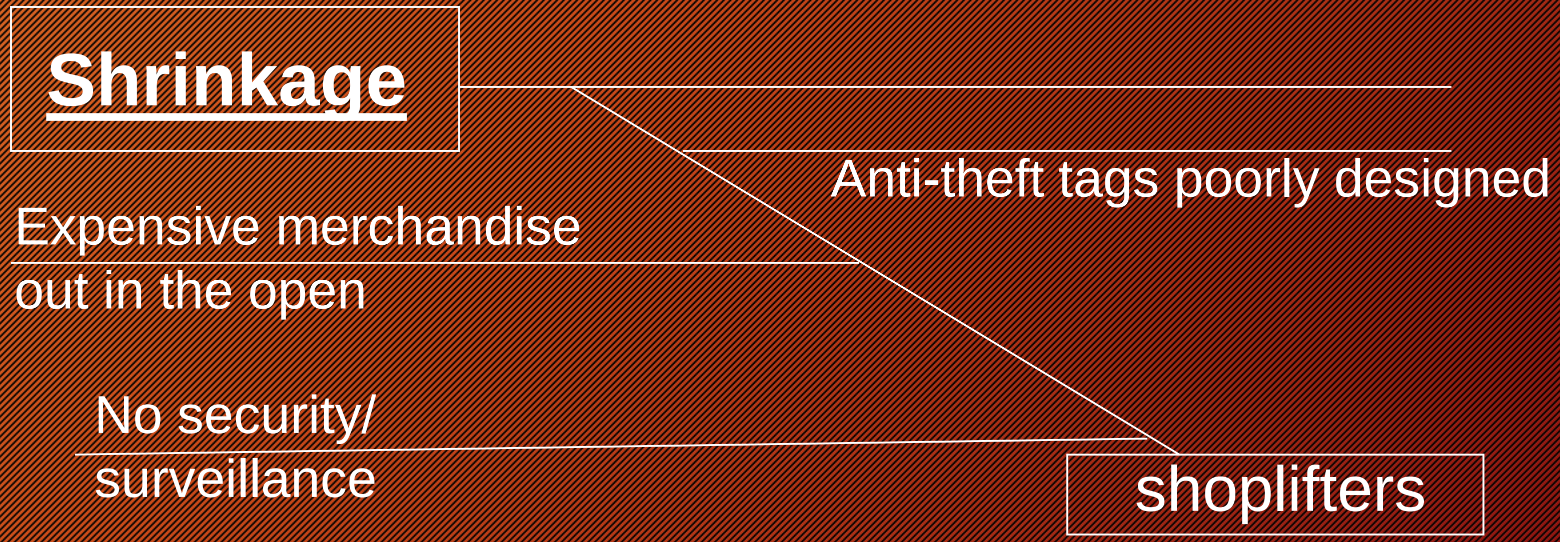
Ex. : High Inventory Shrinkage at local Drug Store



Constructing a Fishbone Diagram

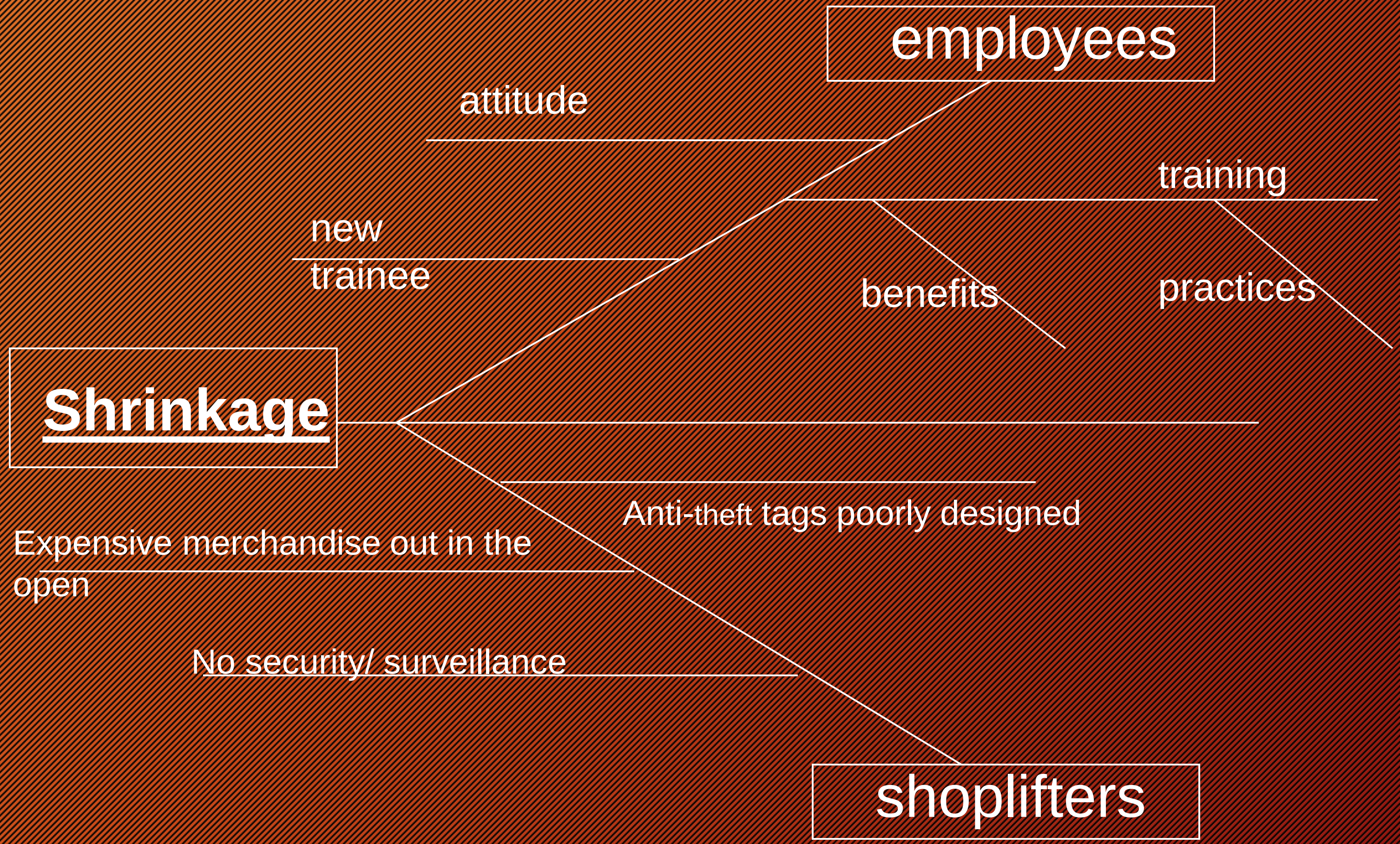
- Step 4 - Identify what these specific causes could be

Ex. : High Inventory Shrinkage at local Drug Store



Constructing a Fishbone Diagram

- Ex. : High Inventory Shrinkage at local Drug Store



Constructing a Fishbone Diagram

- Step 5 – Use the finished diagram to brainstorm solutions to the main problems.

The Basic Seven Tools of Quality

Histograms

- Bar chart
- Used to graphically represent groups of data

Why use a Histogram

To summarize data from a process that has been collected over a period of time, and graphically present its frequency distribution in bar form.

What Does a Histogram Do?

- Displays large amounts of data that are difficult to interpret in tabular form
- Shows the relative frequency of occurrence of the various data values
- Reveals the centering, variation, and shape of the data
- Illustrates quickly the underlying distribution of the data
- Provides useful information for predicting future performance of the process
- Helps to indicate if there has been a change in the process
- Helps answer the question “Is the process capable of meeting my customer requirements?”

How do I do it?

1. Decide on the process measure
   - **The data should be variable data, i.e., measured on a continuous scale. For example: temperature, time, dimensions, weight, speed.**

2. Gather data
   - **Collect at least 50 to 100 data points if you plan on looking for patterns and calculating the distribution's centering (mean), spread (variation), and shape. You might also consider collecting data for a specified period of time: hour, shift, day, week, etc.**
   - **Use historical data to find patterns or to use as a baseline measure of past performance.**

How do I do it? (cont'd)

3. Prepare a frequency table from the data

   a. **Count the number of data points, n, in the sample**

9.9	9.3	10.2	9.4	10.1	9.6	9.9	10.1	9.8
9.8	9.8	10.1	9.9	9.7	9.8	9.9	10	9.6
9.7	9.4	9.6	10	9.8	9.9	10.1	10.4	10
10.2	10.1	9.8	10.1	10.3	10	10.2	9.8	10.7
9.9	10.7	9.3	10.3	9.9	9.8	10.3	9.5	9.9
9.3	10.2	9.2	9.9	9.7	9.9	9.8	9.5	9.4
9	9.5	9.7	9.7	9.8	9.8	9.3	9.6	9.7
10	9.7	9.4	9.8	9.4	9.6	10	10.3	9.8
9.5	9.7	10.6	9.5	10.1	10	9.8	10.1	9.6
9.6	9.4	10.1	9.5	10.1	10.2	9.8	9.5	9.3
10.3	9.6	9.7	9.7	10.1	9.8	9.7	10	10
9.5	9.5	9.8	9.9	9.2	10	10	9.7	9.7
9.9	10.4	9.3	9.6	10.2	9.7	9.7	9.7	10.7
9.9	10.2	9.8	9.3	9.6	9.5	9.6	10.7	

In this example, there are 125 data points, n = 125. For our example, 125 data points would be divided into 7-12 class intervals.

b. Determine the range, R, for the entire sample. The range is the smallest value in the set of data subtracted from the largest value. For our example:

$$R = x_{max} - x_{min} = 10.7 - 9.0 = 1.7$$

c. Determine the number of class intervals, k, needed.

Use the table below to provide a guideline for dividing your sample into reasonable number of classes.

Number of Data Points	**Number of Classes (k)**
Under 50	**5-7**
50-100	**6-10**
100-250	**7-12**
Over 250	**10-20**

How do I do it? (cont'd)

Tip: The number of intervals can influence the pattern of the sample. Too few intervals will produce a tight, high pattern. Too many intervals will produce a spread out, flat pattern.

d. Determine the class width, H.

- **The formula for this is:**

$$H = \frac{R}{k} = \frac{1.7}{10} = 0.17$$

- **Round your number to the nearest value with the same decimal numbers as the original sample. In our example, we would round up to 0.20. It is useful to have intervals defined to one more decimal place than the data collected.**

e. Determine the class boundaries, or end points.

- **Use the smallest individual measurement in the sample, or round to the next appropriate lowest round number. This will be the lower end point for the *first* class interval. In our example this would be 9.0.**

How do I do it? (cont'd)

- **Add the class width, H, to the lower end point. This will be the lower end point for the next class interval. For our example:**

  9.0 + H = 9.0 + 0.20 = 9.20

  Thus, the first class interval would be 9.00 and everything up to, *but not including* 9.20, that is, 9.00 through 9.19. The second class interval would begin at 9.20 and everything up to, but not including 9.40.

Tip: Each class interval would be mutually exclusive, that is, every data point will fit into *one, and only one* class interval.

- **Consecutively add the class width to the lowest class boundary until the K class intervals and/or the range of all the numbers are obtained.**

How do I do it? (cont'd)

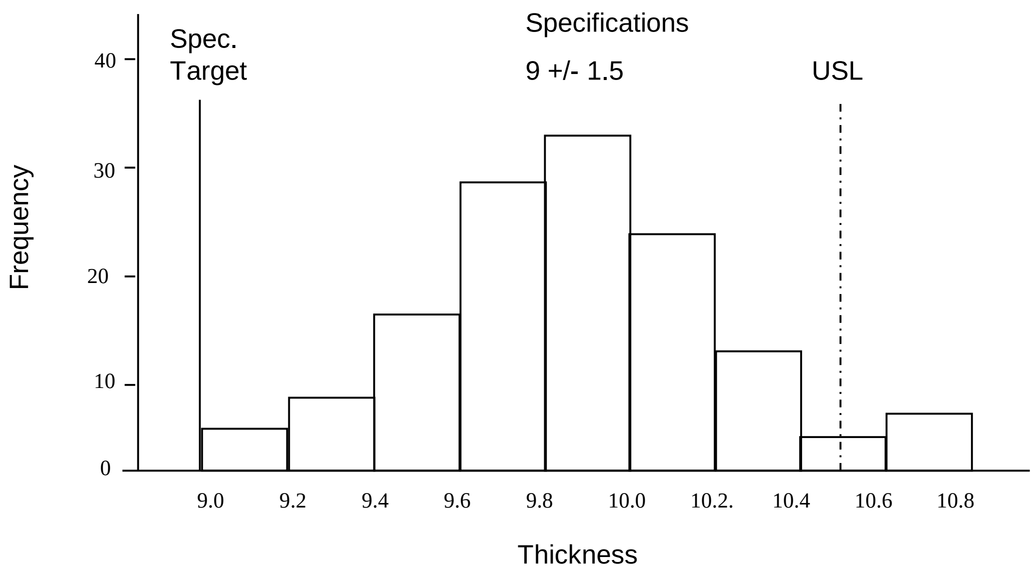
f. Construct the frequency table based on the values you computed in item "e".

A frequency table based on the data from our example is show below.

Class #	Class Boundaries	Mid-Point	Frequency	Total
1	9.00-9.19	9.1	|	1
2	9.20-9.39	9.3	𝍸 ||||	9
3	9.40-9.59	9.5	𝍸 𝍸 𝍸 |	16
4	9.60-9.79	9.7	𝍸 𝍸 𝍸 𝍸 𝍸 ||	27
5	9.80-9.99	9.9	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |	31
6	10.00-10.19	10.1	𝍸 𝍸 𝍸 𝍸 ||	22
7	10.20-10.39	10.3	𝍸 𝍸 ||	12
8	10.40-10.59	10.5	||	2
9	10.60-10.79	10.7	𝍸	5
10	10.80-10.99	10.9		0

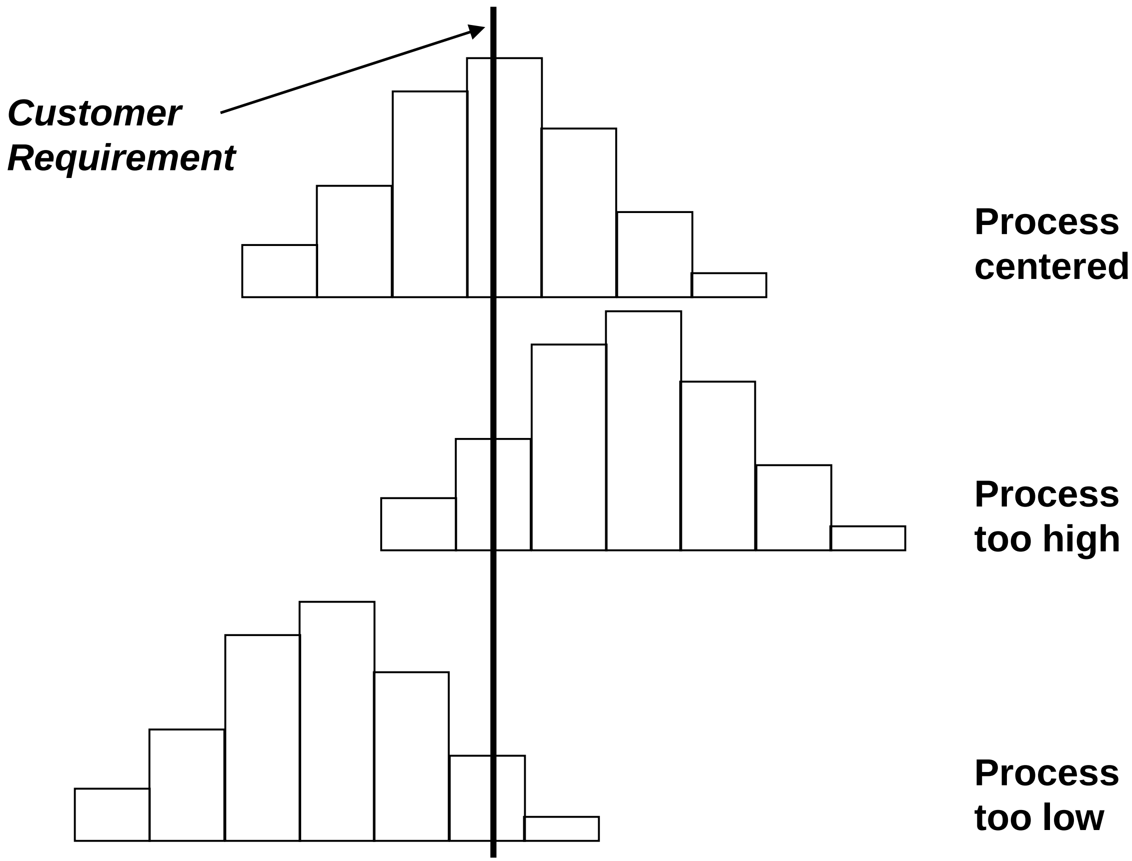
How do I do it? (cont'd)

4. Draw a Histogram from the frequency table
   - **On the vertical line, (y axis), draw the frequency (count) scale to cover class interval with the highest frequency count.**
   - **On the horizontal line, (x axis), draw the scale related to the variable you are measuring.**
   - **For each class interval, draw a bar with the height equal to the frequency tally of that class.**



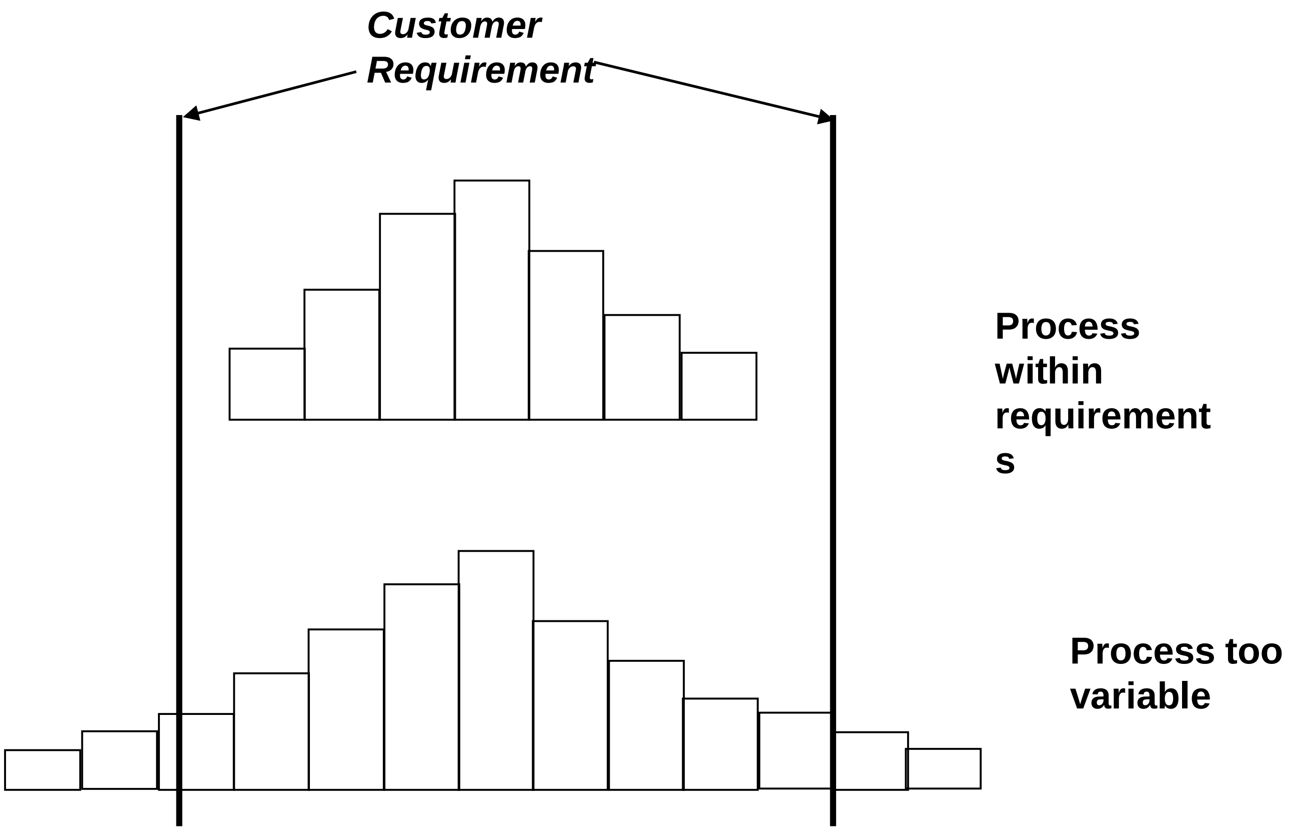
How do I do it? (cont'd)

5. Interpret the Histogram
   ***a. Centering*. Where is the distribution centered?**
   b. Is the process running too high? Too low?



How do I do it? (cont'd)

b. *Variation*. What is the variation or spread of the data? Is it too variable?

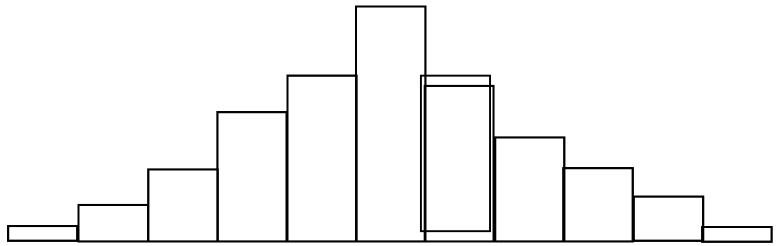


How do I do it? (cont'd)

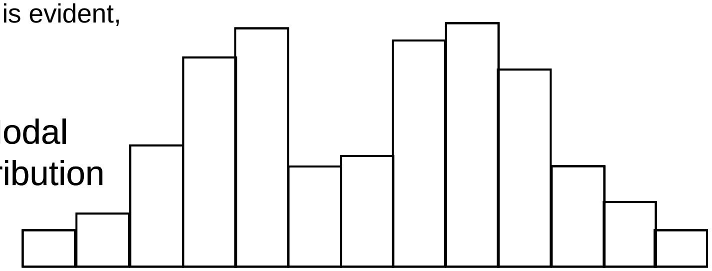
c. Shape. What is the shape? Does it look like a normal, bell-shaped distribution? Is it positively or negatively skewed, that is, more data values to the left or to the right? Are there twin (bi-modal) or multiple peaks?

Tip: Some processes are naturally skewed; don't expect every distribution to follow a bell-shaped curve.

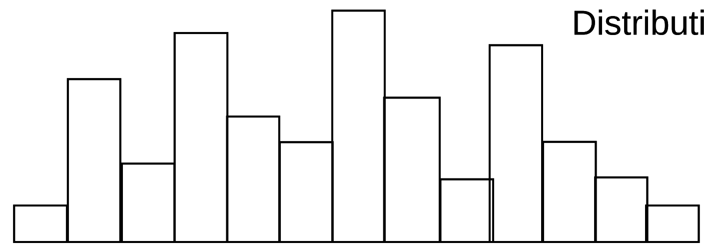
Tip: Always look for twin or multiple peaks indicating that the data is coming from two or more different sources, e.g., shifts, machines, people, suppliers. If this is evident, stratify the data.



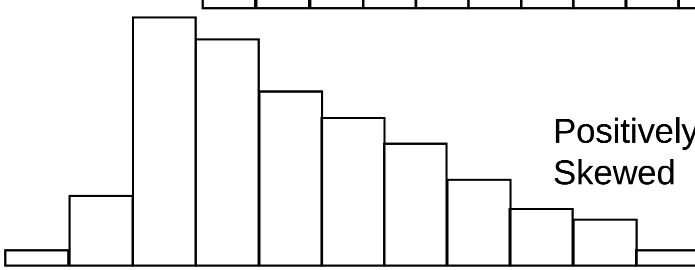
Normal Distribution



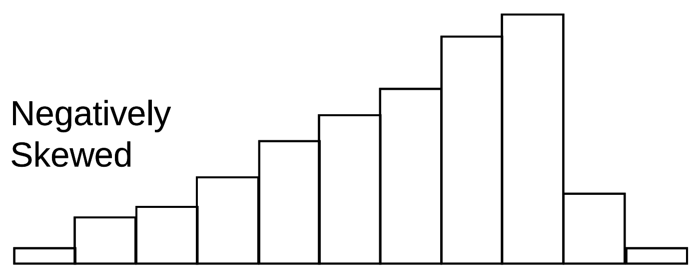
Bi-Modal Distribution



Mulit-Modal Distribution



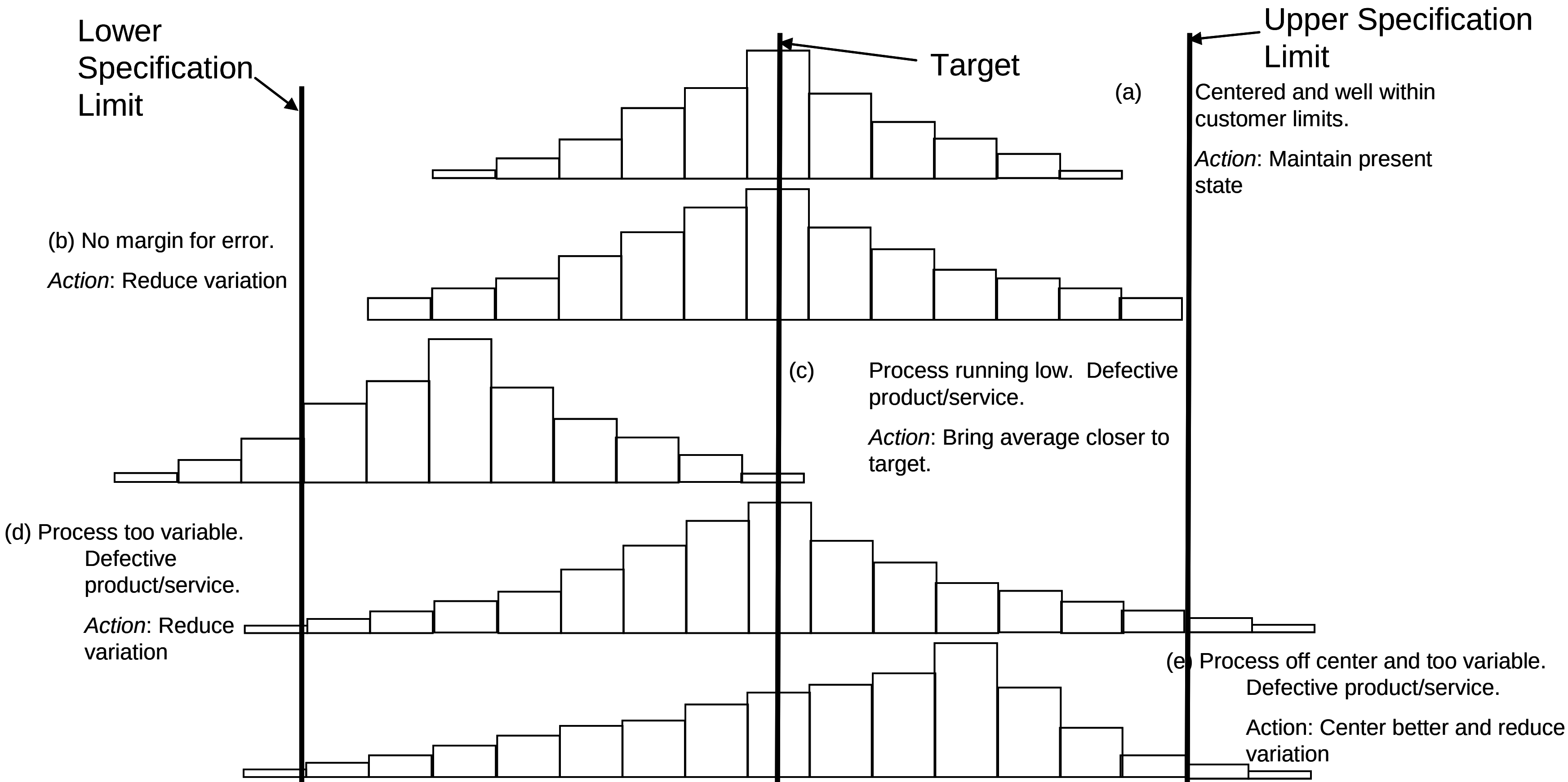
Positively Skewed



Negatively Skewed

How do I do it? (cont'd)

d. Process Capability. Compare the results of your Histogram to your customer requirements or specifications. Is your process capable of meeting the requirements, i.e., is the Histogram centered on the target and within the specification limits?



How do I do it? (cont'd)

Tip: Get suspicious of the accuracy of the data if the Histogram suddenly stops at one point (such as a specification limit) without some previous decline in the data. It could indicate that defective product is being sorted out and is not included in the sample.

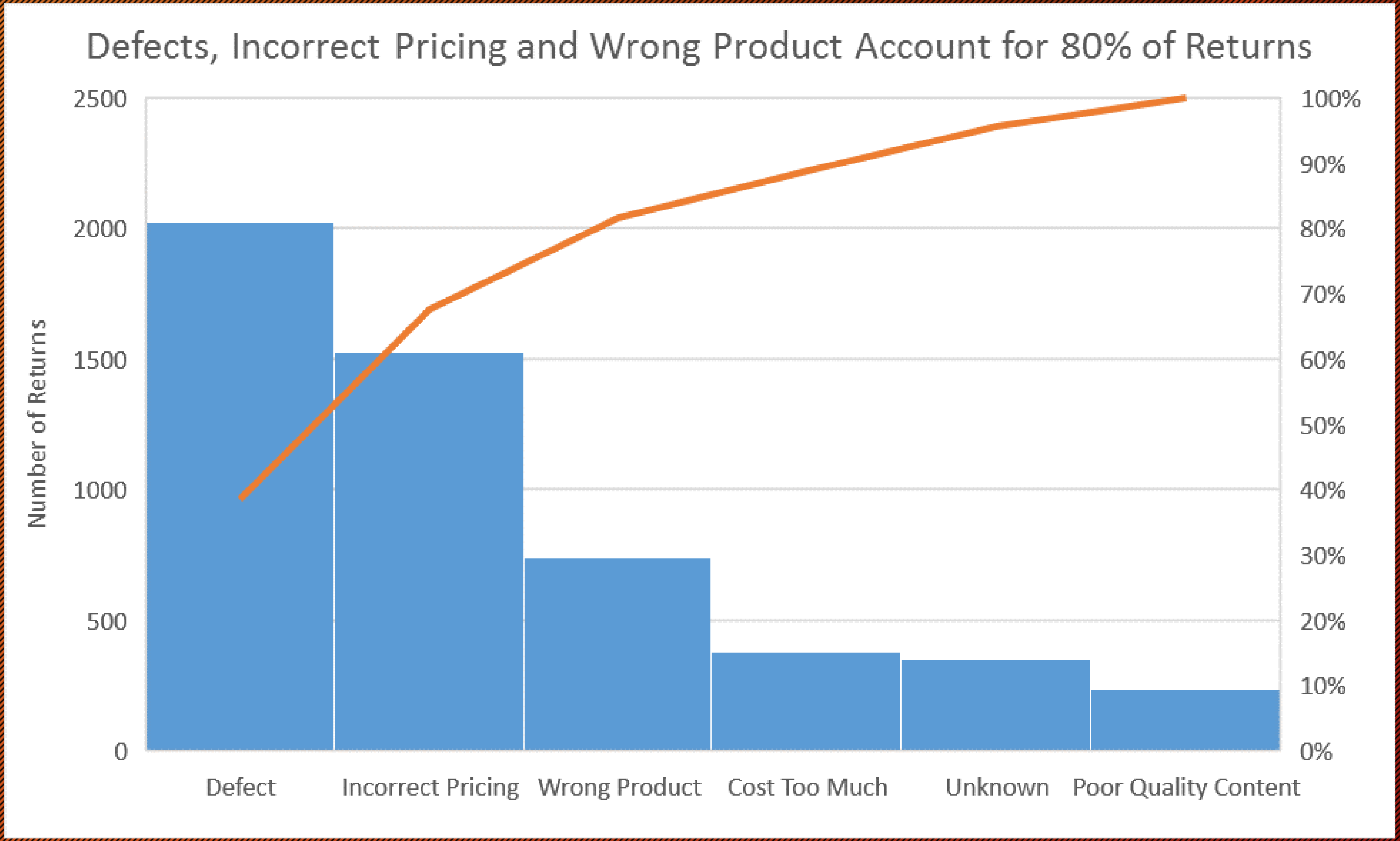
Tip: The Histogram is related to the Control Chart. Like a Control Chart, a normally distributed Histogram will have almost all its values within +/-3 standard deviations of the mean. See Process Capability for an illustration of this.

The Basic Seven Tools of Quality

Pareto Analysis

- Very similar to Histograms
- Use of the 80/20 rule
- Use of percentages to show importance

Pareto Chart



The Basic Seven Tools of Quality

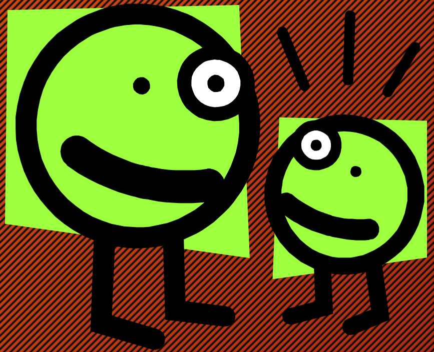
Scatter Plots

- 2 Dimensional X/Y plots
- Used to show relationship between independent(x) and dependent(y) variables

Scatter Plot

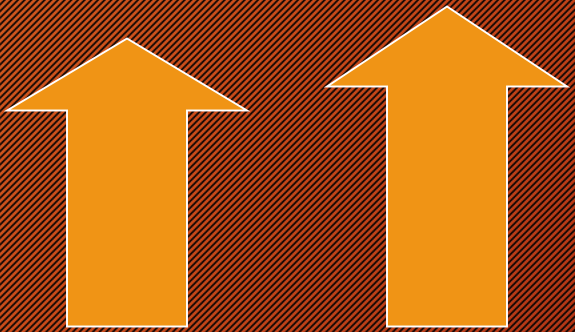
- A scatter plot is a graph of a collection of ordered pairs (x,y).

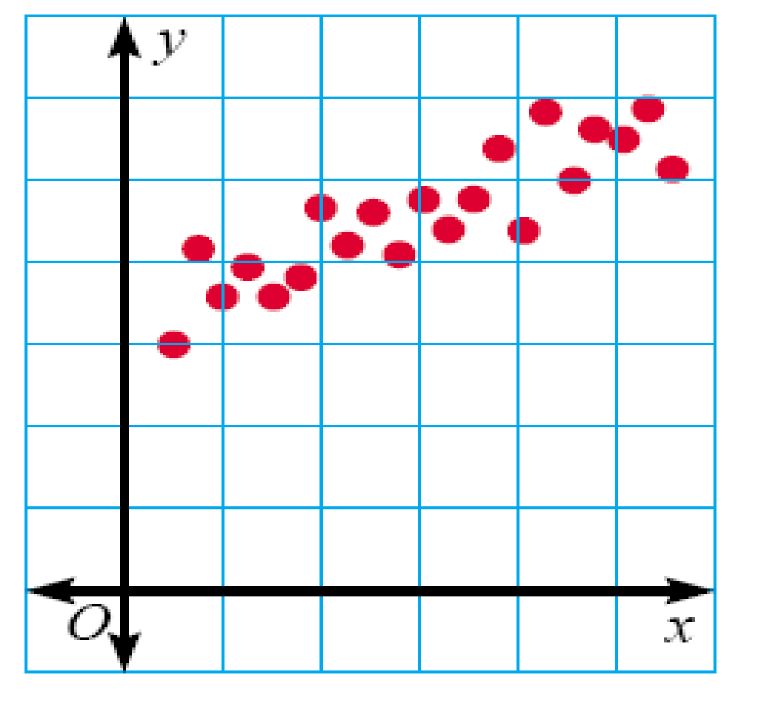
- The graph looks like a bunch of dots, but some of the graphs are a general shape or move in a general direction.



Positive Correlation

- If the x-coordinates and the y-coordinates both increase, then it is POSITIVE CORRELATION.
- This means that both are going up, and they are related.





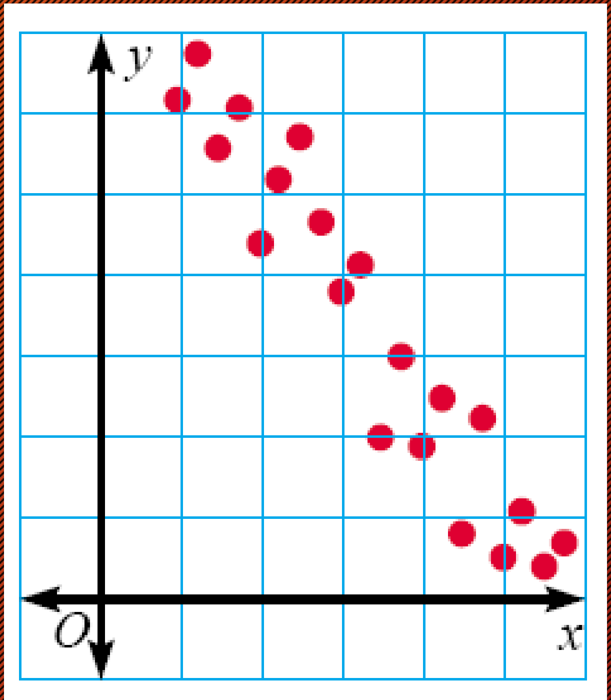
Positive Correlation

- If you look at the age of a child and the child's height, you will find that as the child gets older, the child gets taller. Because both are going up, it is positive correlation.

Age	1	2	3	4	5	6	7	8
Height "	25	31	34	36	40	41	47	55

Negative Correlation

- If the x-coordinates and the y-coordinates have one increasing and one decreasing, then it is NEGATIVE CORRELATION.
- This means that 1 is going up and 1 is going down, making a downhill graph. This means the two are related as opposites.



Negative Correlation

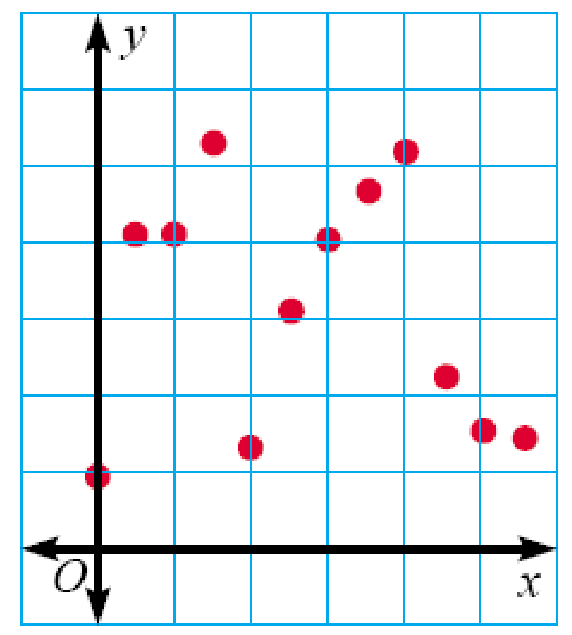
- If you look at the age of your family's car and its value, you will find as the car gets older, the car is worth less. This is negative correlation.

Age of car	1	2	3	4	5
Value	$30,000	$27,000	$23,500	$18,700	$15,350

No Correlation

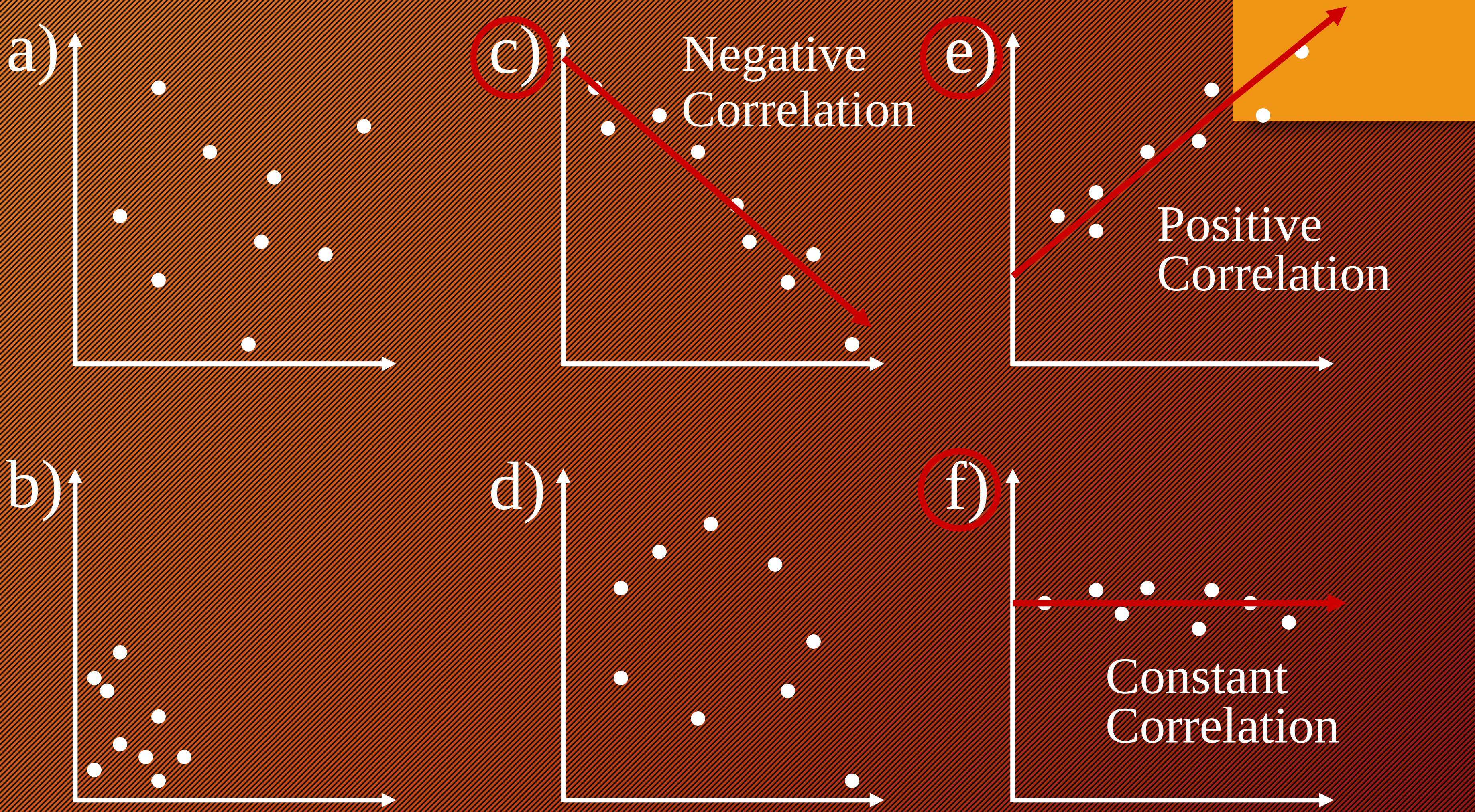
- If there seems to be no pattern, and the points looked scattered, then it is no correlation.

- This means the two are not related.



Scatterplots

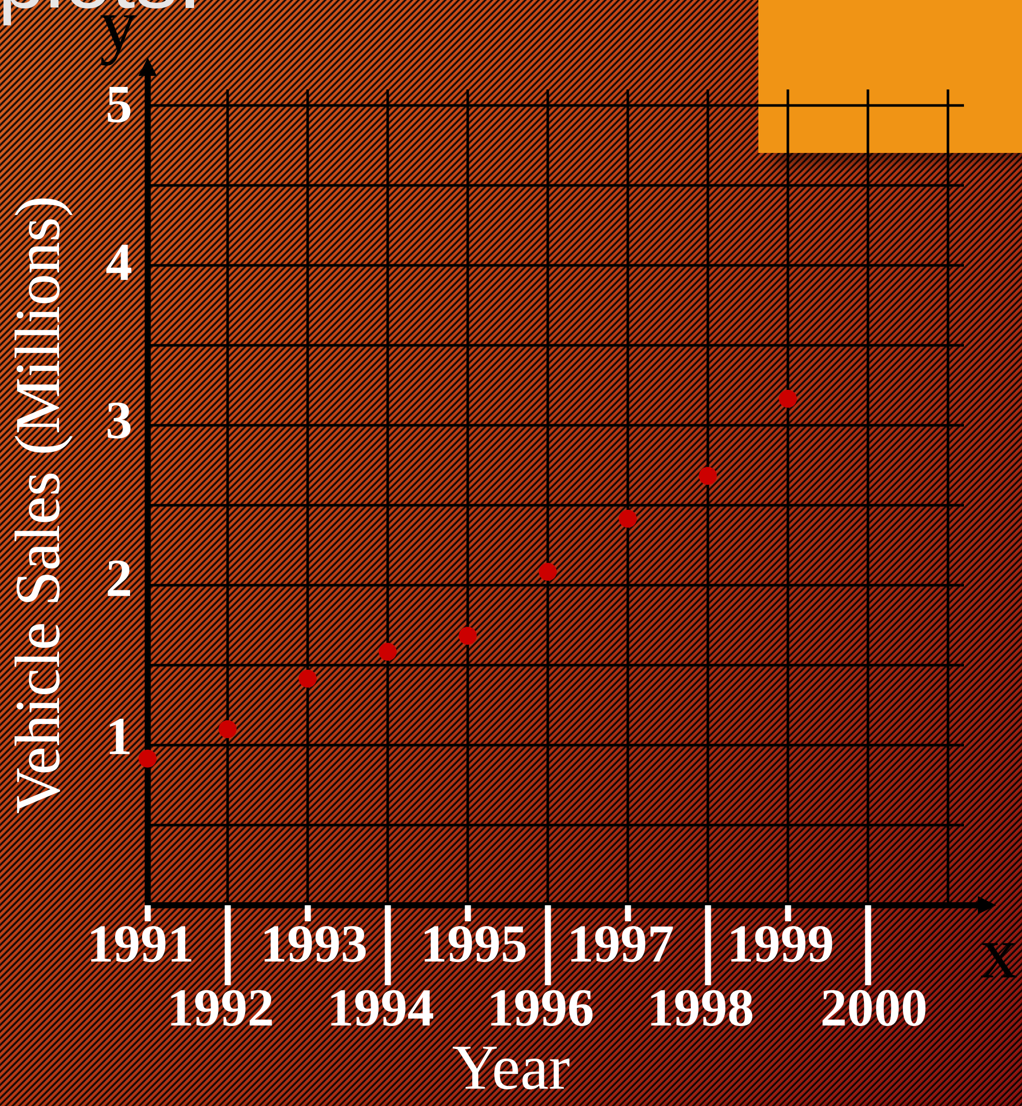
Which scatterplots below show a linear trend?



Objective - To plot data points in the coordinate plane and interpret scatter plots.

Sport Utility Vehicles (SUVs) Sales in U.S.

Year	Sales (in Millions)
1991	0.9
1992	1.1
1993	1.4
1994	1.6
1995	1.7
1996	2.1
1997	2.4
1998	2.7
1999	3.2



Scatterplot - a coordinate graph of data points.

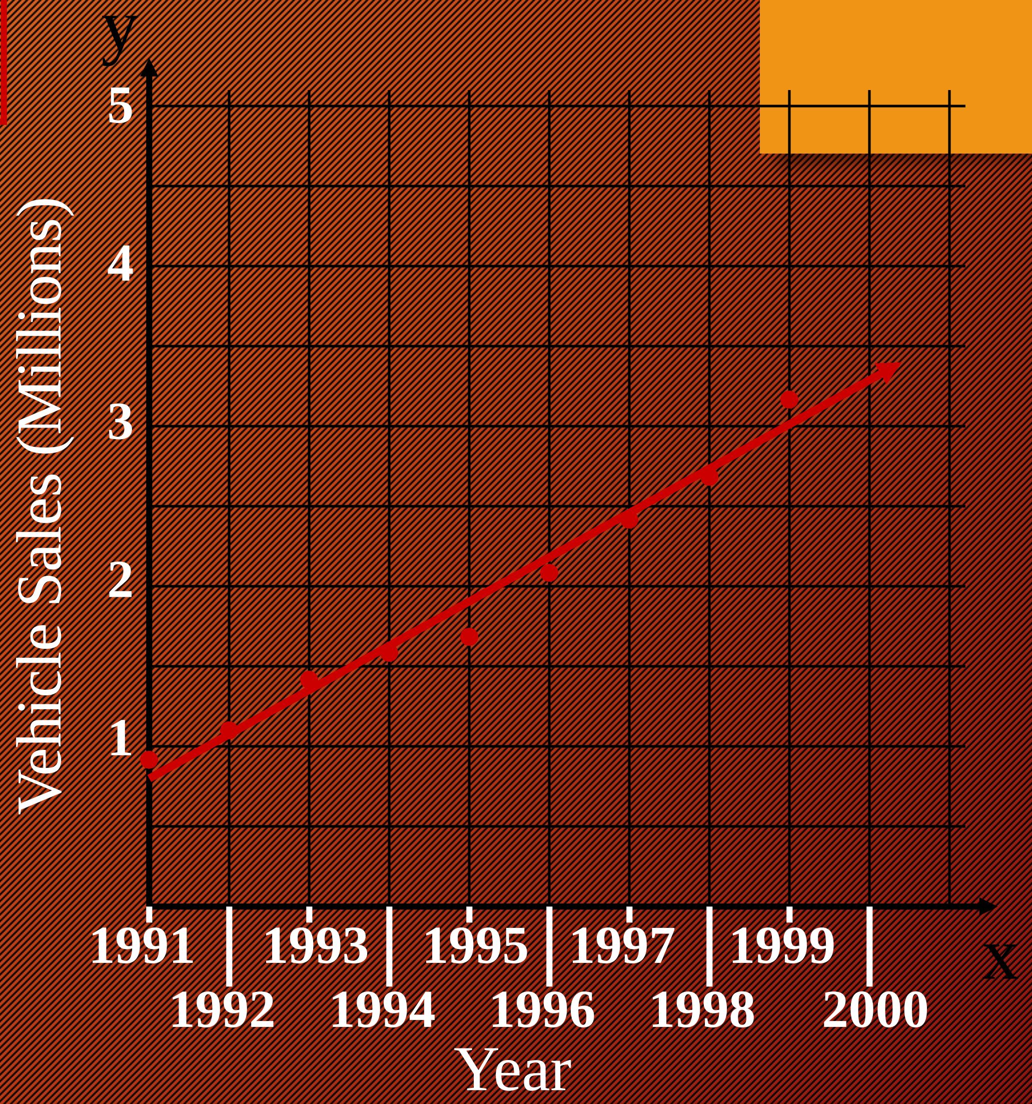
Trend appears linear.

Trend is increasing.

Year ↑
SUV Sales ↑

Positive correlation.

Predict the sales in 2001.

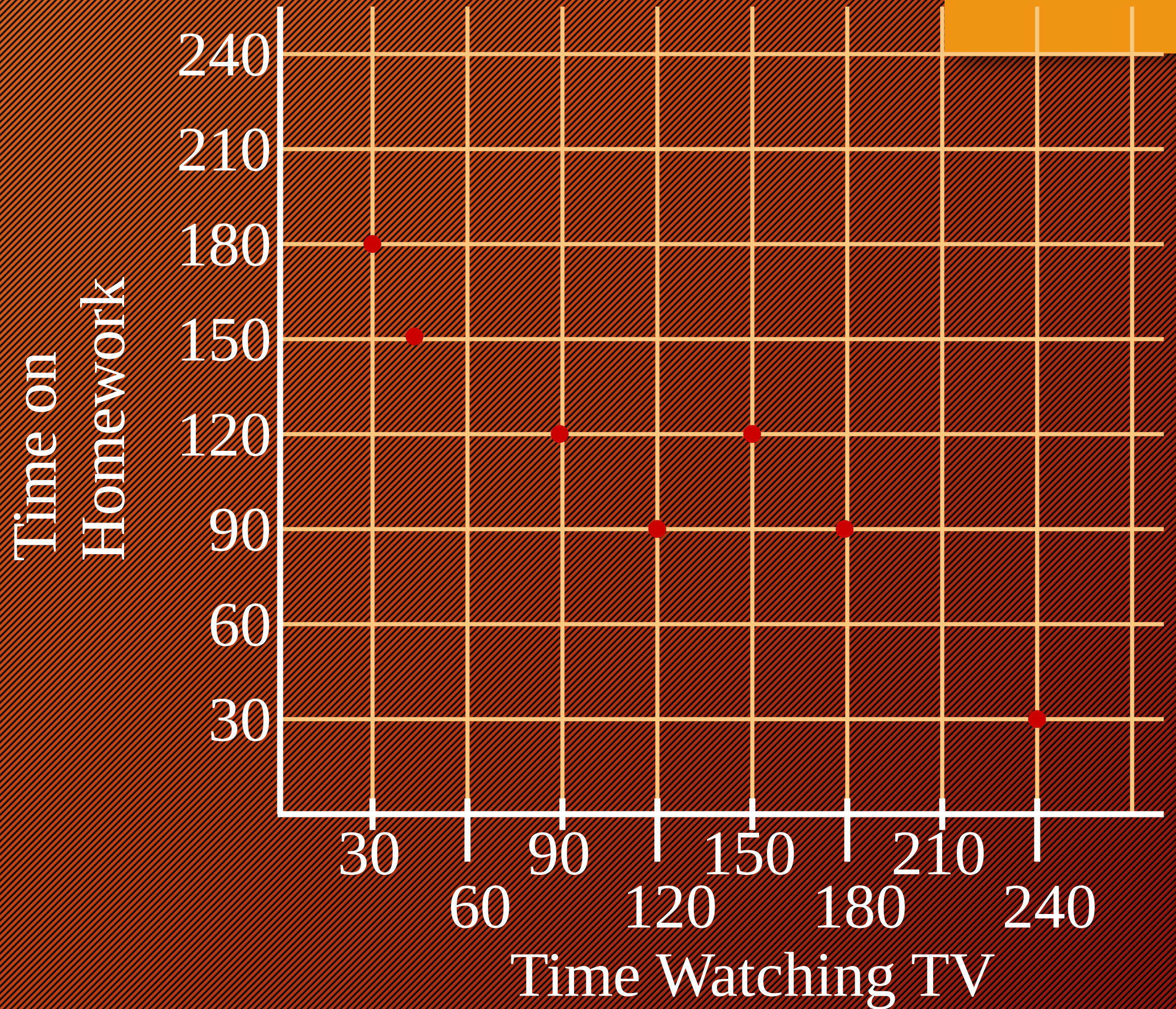


Plot the data on the graph such that homework time is on the y-axis and TV time is on the x-axis..

Student	Time Spent Watching TV	Time Spent on Homework
Sam	30 min.	180 min.
Jon	45 min.	150 min.
Lara	120 min.	90 min.
Darren	240 min.	30 min.
Megan	90 min.	90 min.
Pia	150 min.	90 min.
Crystal	180 min.	90 min.

Plot the data on the graph such that homework time is on the y-axis and TV time is on the x-axis.

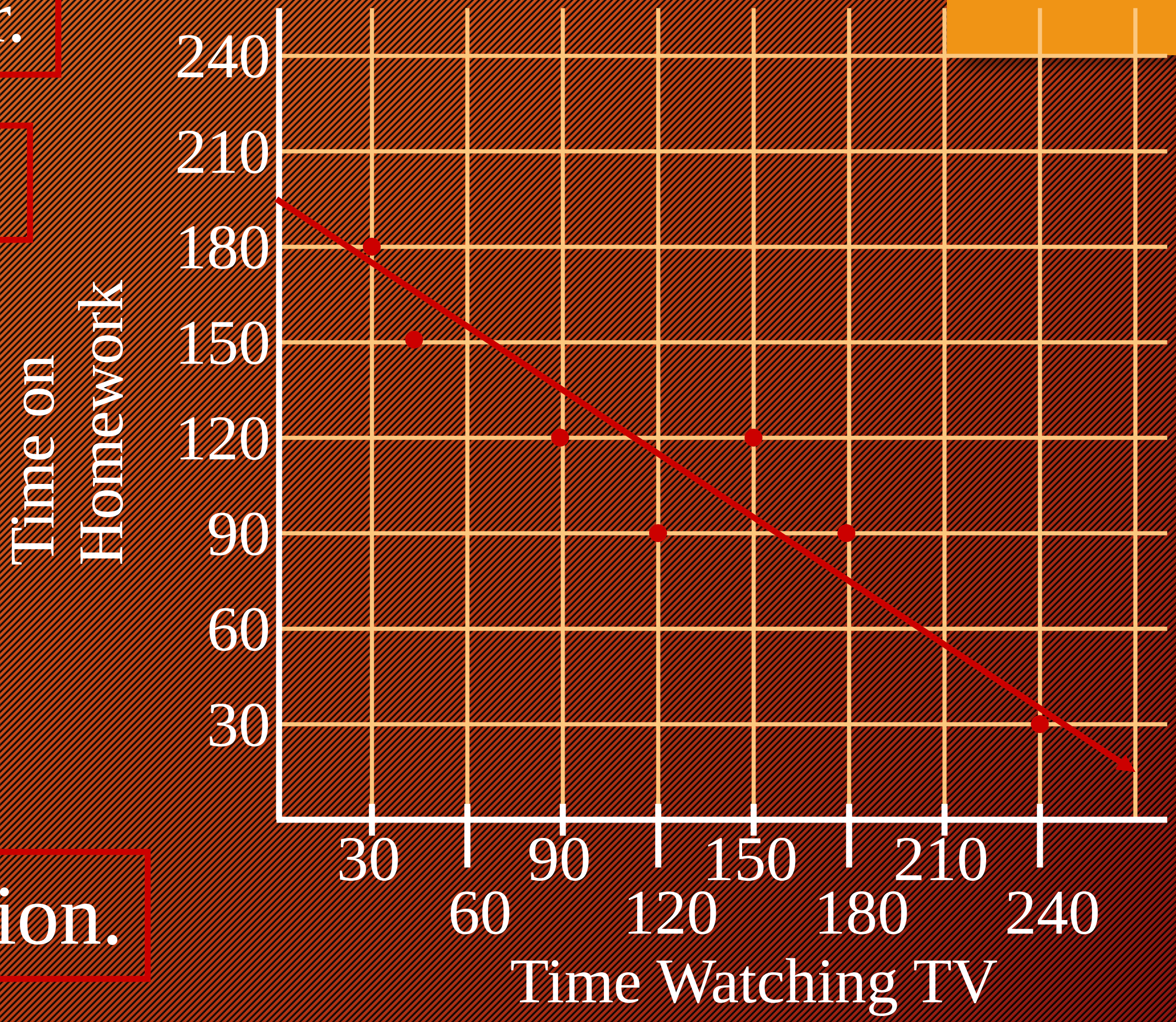
TV	Homework
30 min.	180 min.
45 min.	150 min.
120 min.	90 min.
240 min.	30 min.
90 min.	120 min.
150 min.	120 min.
180 min.	90 min.



Describe the relationship between time spent on homework and time spent watching TV.

Trend appears linear.

Trend is decreasing.



Time on TV ↑

Time on HW ↓

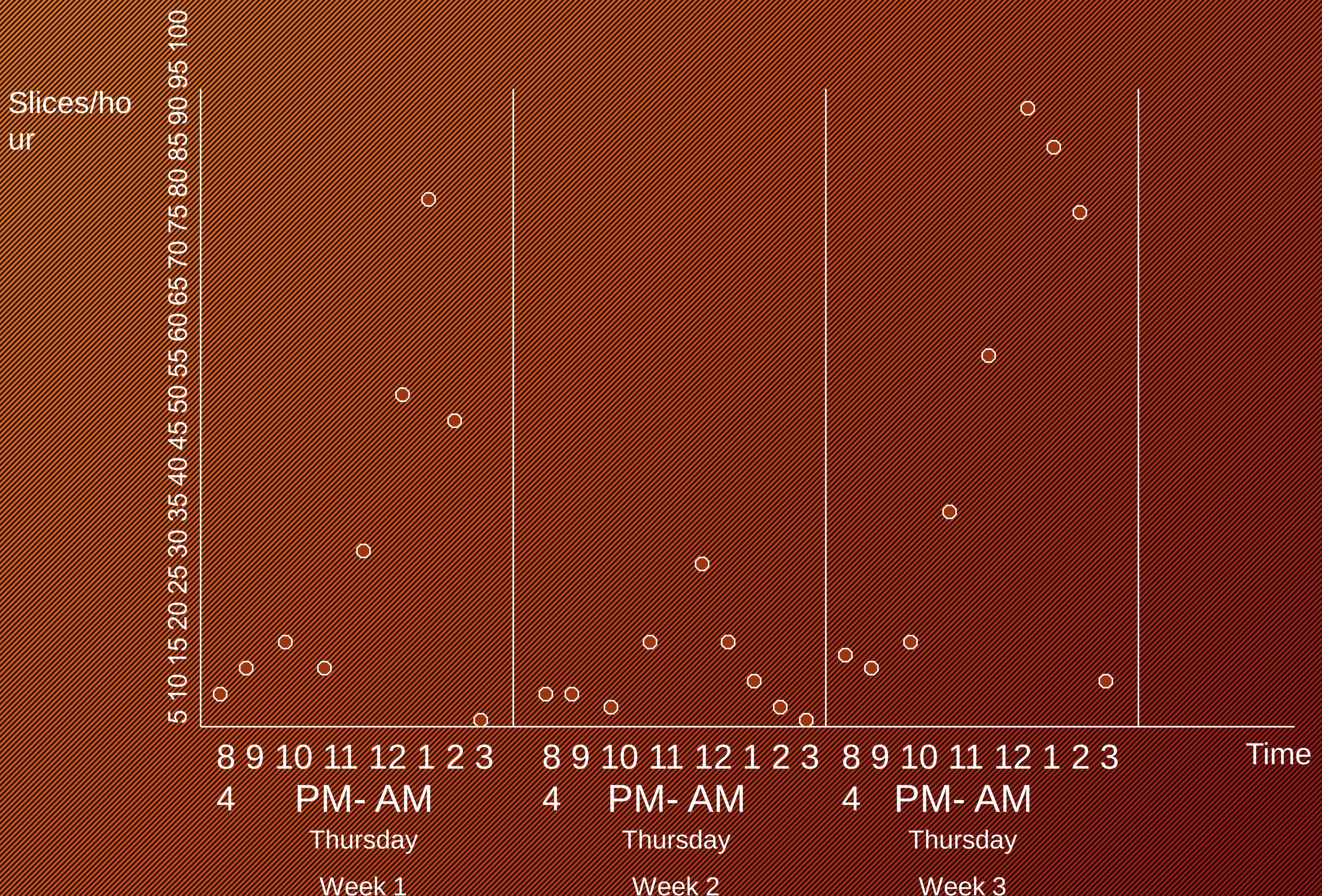
Negative correlation.

The Basic Seven Tools of Quality

Run charts

- Time-based (x-axis)
- Cyclical
- Look for patterns

Run Charts

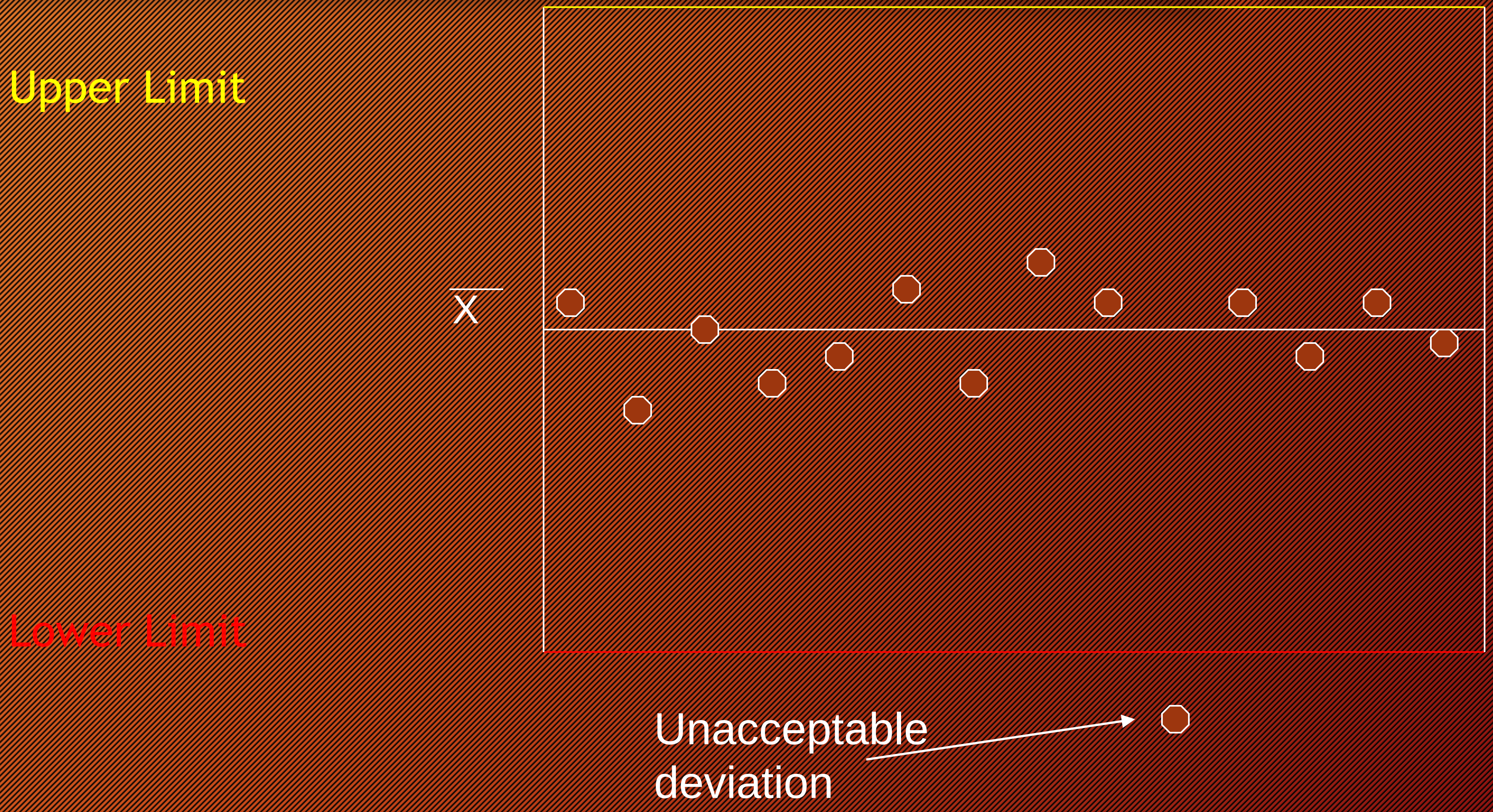

Slices/hour
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
8 9 10 11 12 1 2 3
4 PM- AM
Thursday
Week 1
8 9 10 11 12 1 2 3
4 PM- AM
Thursday
Week 2
8 9 10 11 12 1 2 3
4 PM- AM
Thursday
Week 3
Time

The Basic Seven Tools of Quality

Control Charts

- Deviation from Mean
- Upper and Lower Spec's
- Range

Control Charts



STATISTICAL QUALITY CONTROL

Understanding Variability

Statistical Process Control (SPC)

- ☑ Variability is inherent in every process
  - ☑ Natural or common causes
  - ☑ Special or assignable causes
- ☑ Provides a statistical signal when assignable causes are present
- ☑ Detect and eliminate assignable causes of variation

Natural Variations

- ☑ Also called common causes
- ☑ Affect virtually all production processes
- ☑ Expected amount of variation
- ☑ Output measures follow a probability distribution
- ☑ For any distribution there is a measure of central tendency and dispersion
- ☑ If the distribution of outputs falls within acceptable limits, the process is said to be "in control"

Assignable Variations

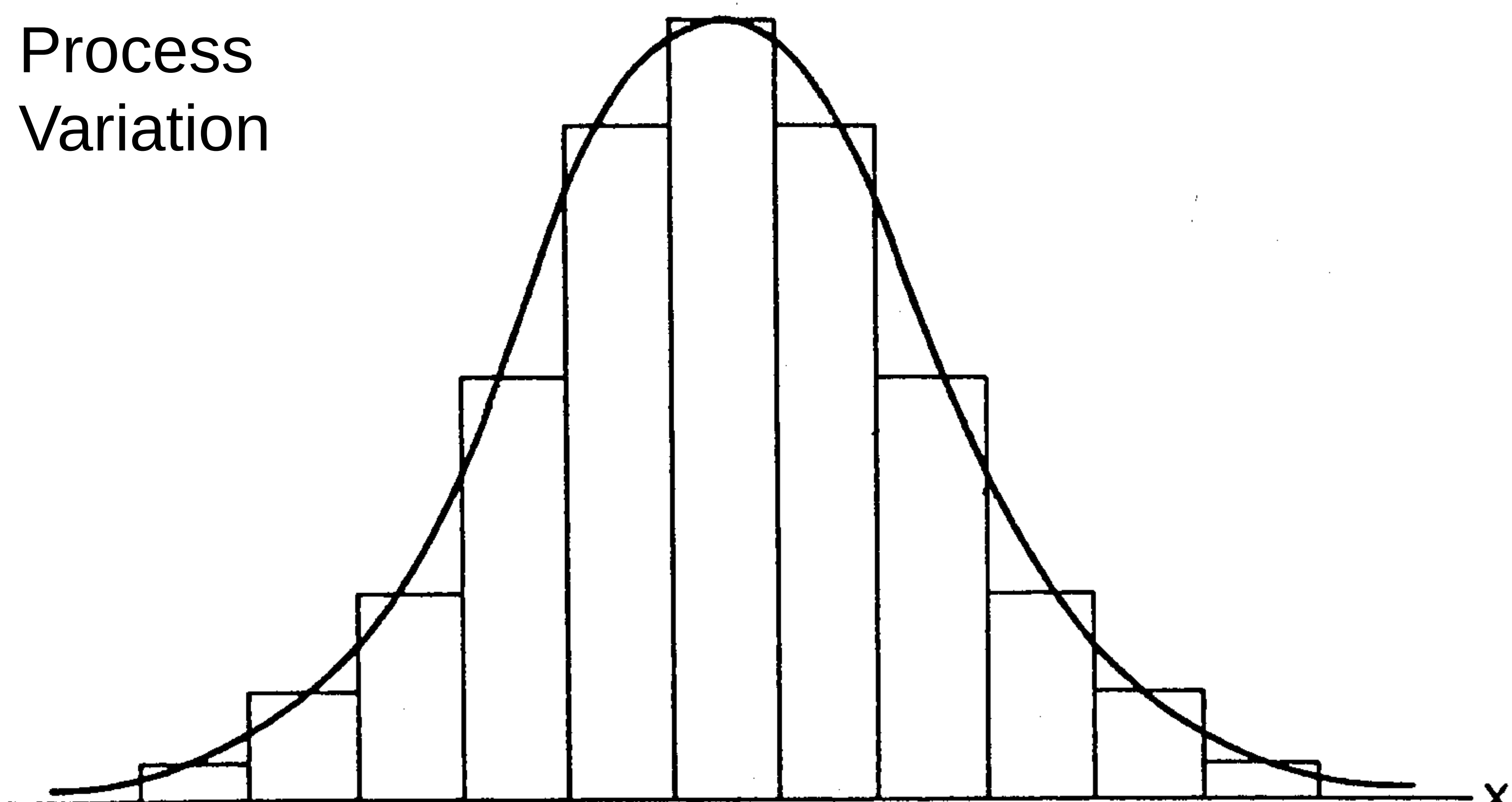
- ☑ Also called special causes of variation
  - ☑ Generally this is some change in the process
- ☑ Variations that can be traced to a specific reason
- ☑ The objective is to discover when assignable causes are present
  - ☑ Eliminate the bad causes
  - ☑ Incorporate the good causes

Types of Variations

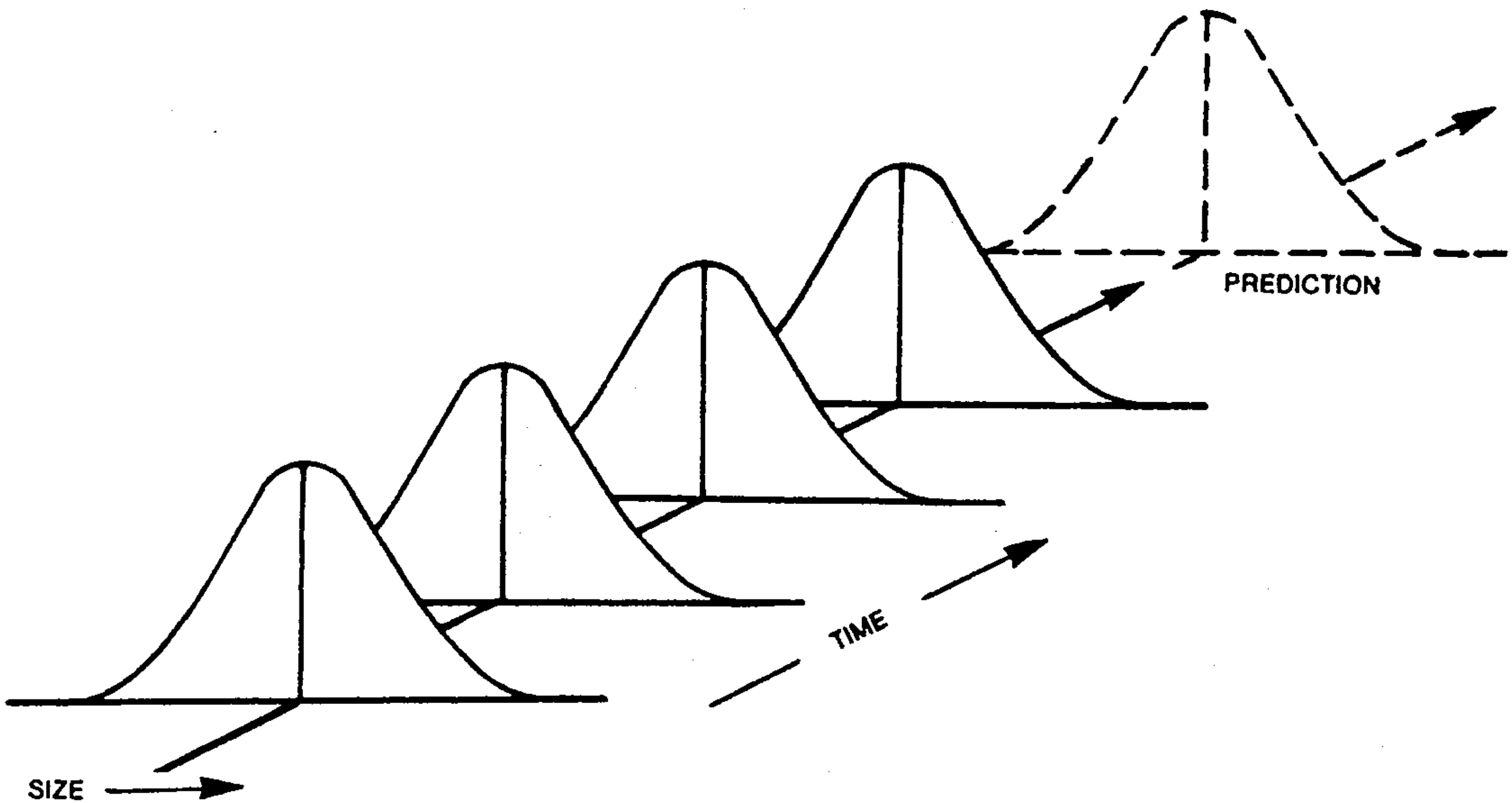
- <u>Common Cause</u>
- Random
- Chronic
- Small
- System problems
- Mgt controllable
- Process improvement
- Process capability

- <u>Special Cause</u>
- Situational
- Sporadic
- Large
- Local problems
- Locally controllable
- Process control
- Process stability

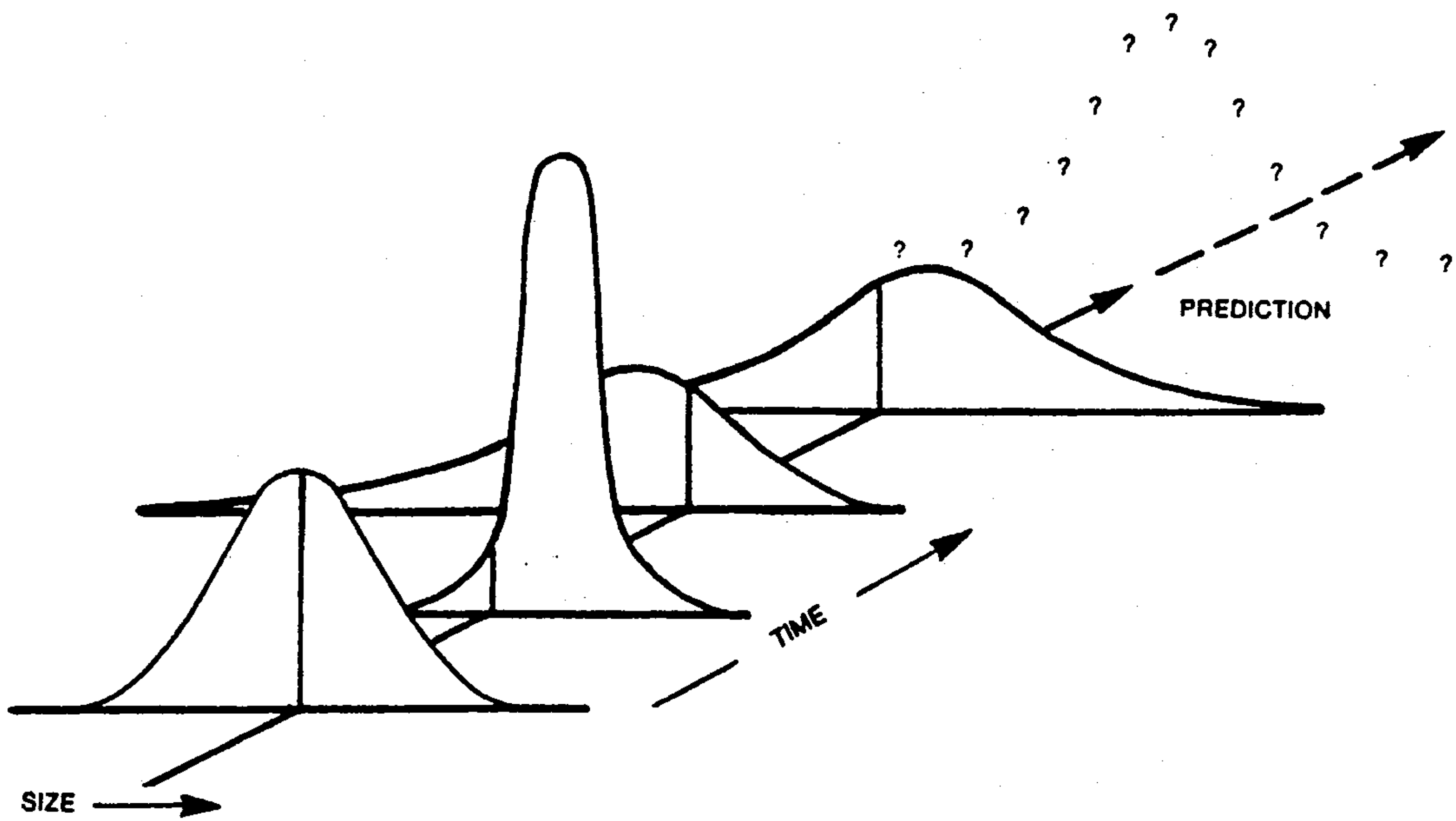
Process
Variation



Variation from Common Causes



Variation from Special Causes



Samples

To measure the process, we take samples and analyze the sample statistics following these steps

(a) Samples of the product, say five boxes of cereal taken off the filling machine line, vary from each other in weight

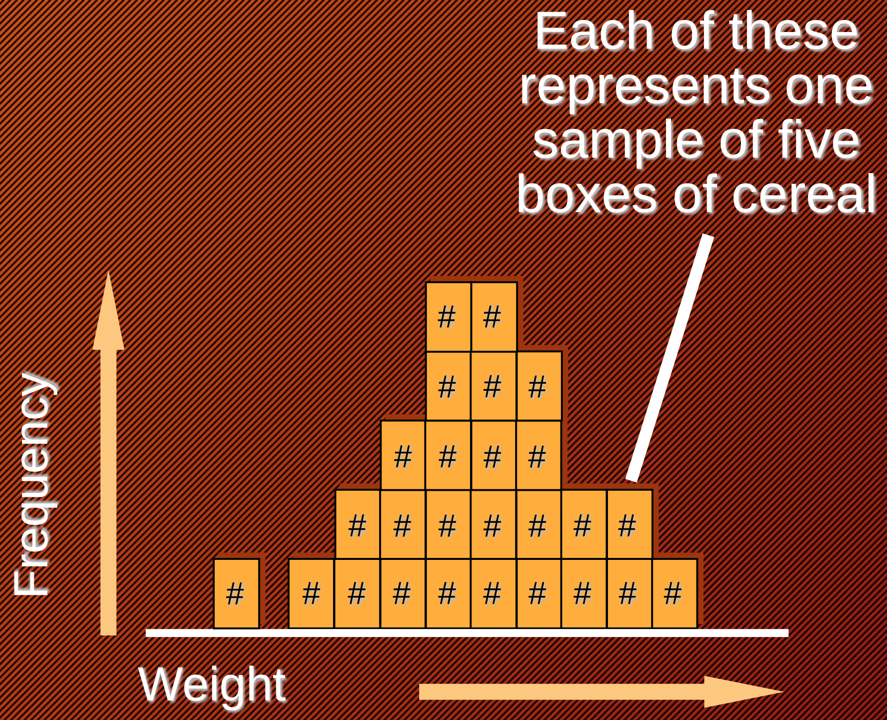


Figure S6.1

Samples

To measure the process, we take samples and analyze the sample statistics following these steps

(b) After enough samples are taken from a stable process, they form a pattern called a distribution

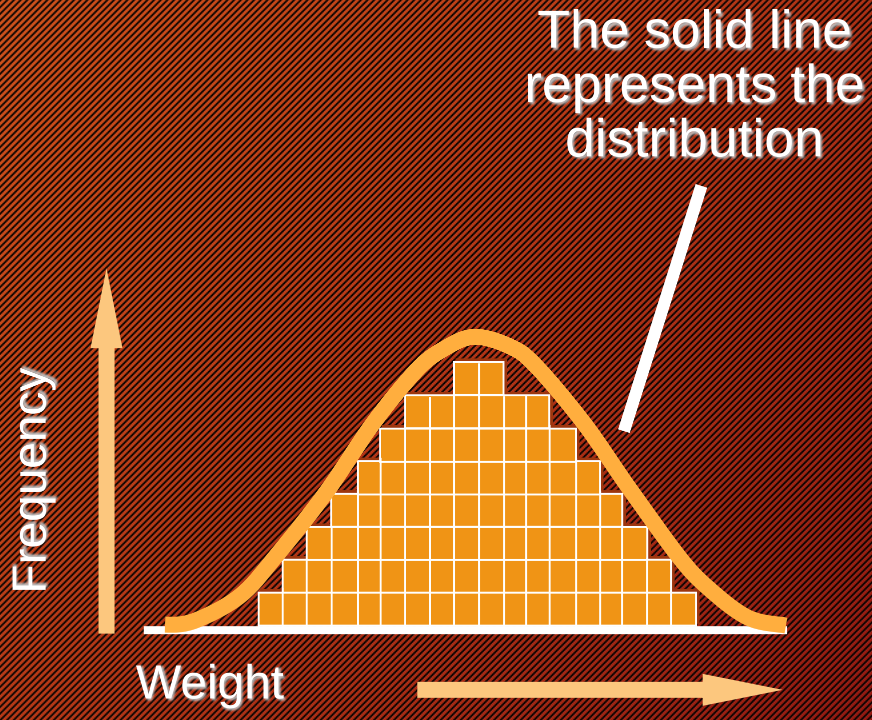


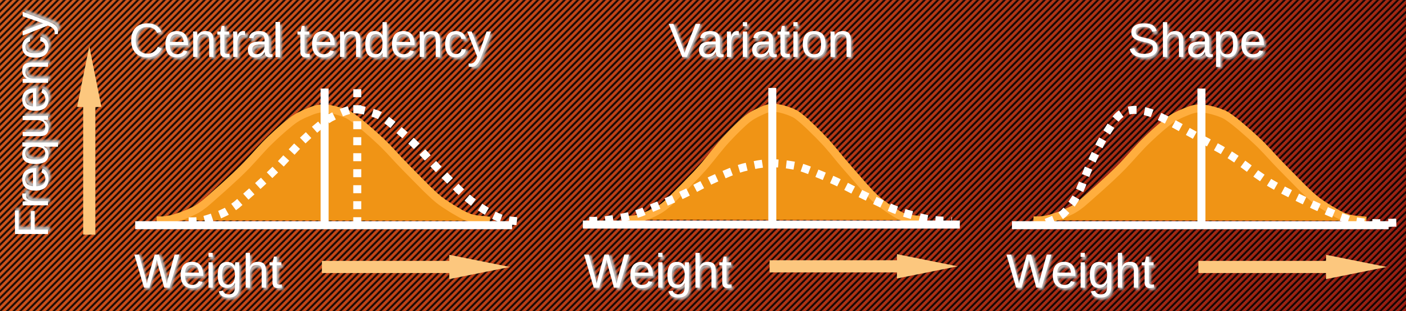
Figure S6.1

Samples

To measure the process, we take samples and analyze the sample statistics following these steps

(c) There are many types of distributions, including the normal (bell-shaped) distribution, but distributions do differ in terms of central tendency (mean), standard deviation or variance, and shape

Figure S6.1



Samples

To measure the process, we take samples and analyze the sample statistics following these steps

(d) If only natural causes of variation are present, the output of a process forms a distribution that is stable over time and is predictable

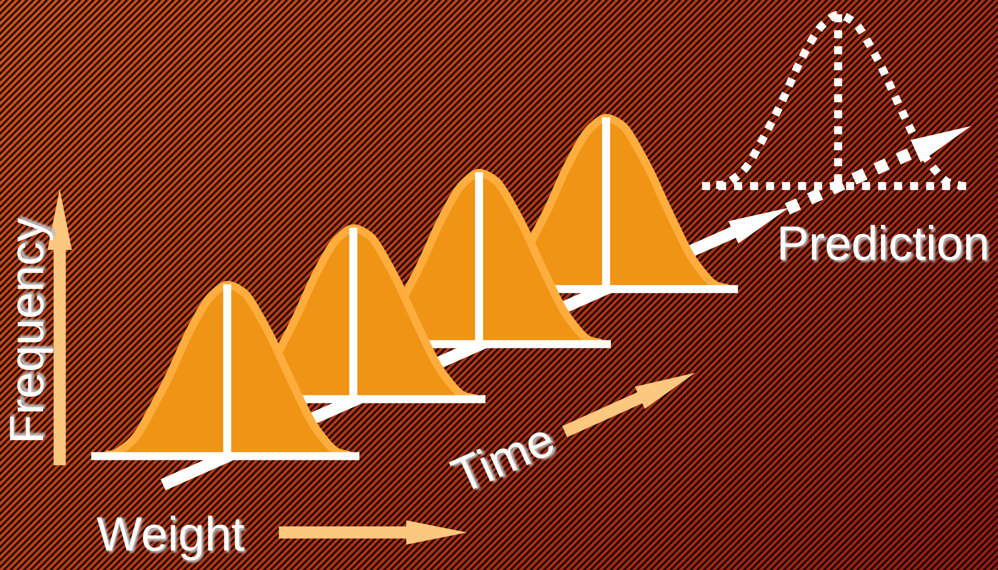


Figure S6.1

Samples

To measure the process, we take samples and analyze the sample statistics following these steps

(e) If assignable causes are present, the process output is not stable over time and is not predicable

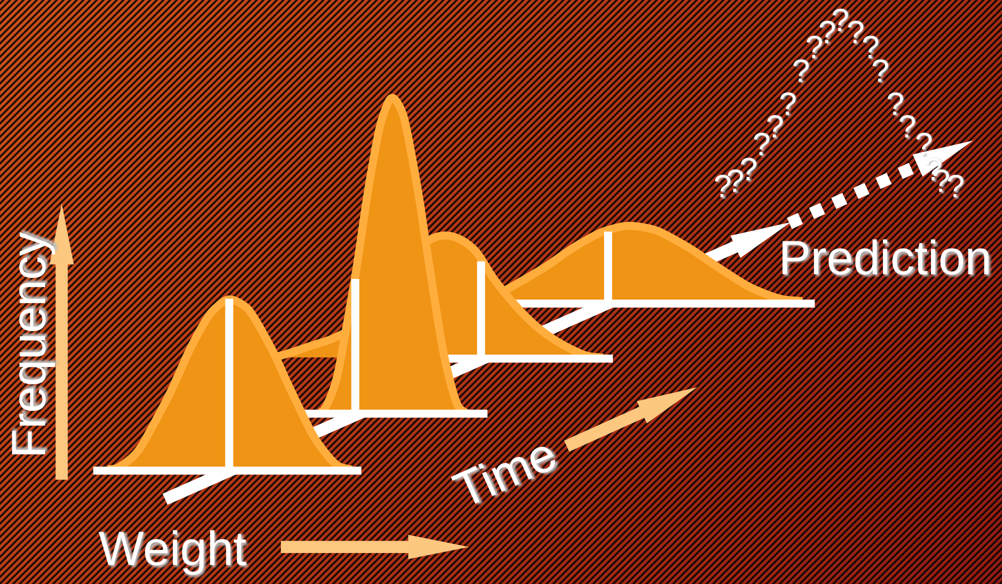
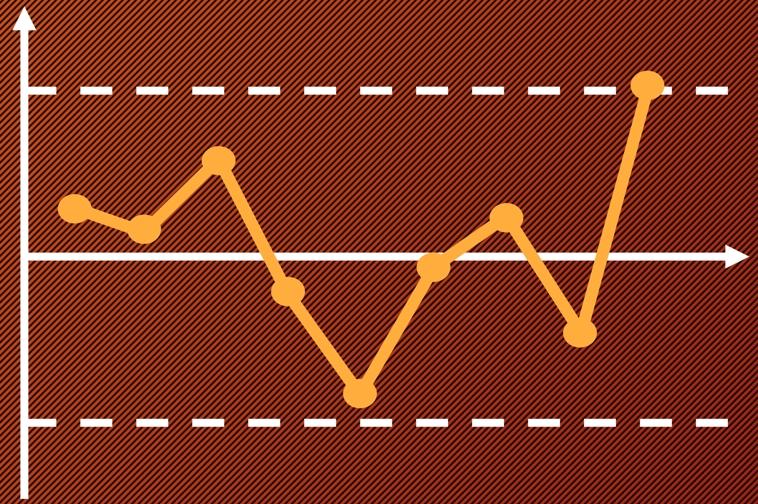


Figure S6.1

Control Charts

Constructed from historical data, the purpose of control charts is to help distinguish between natural variations and variations due to assignable causes



Statistical Process Control

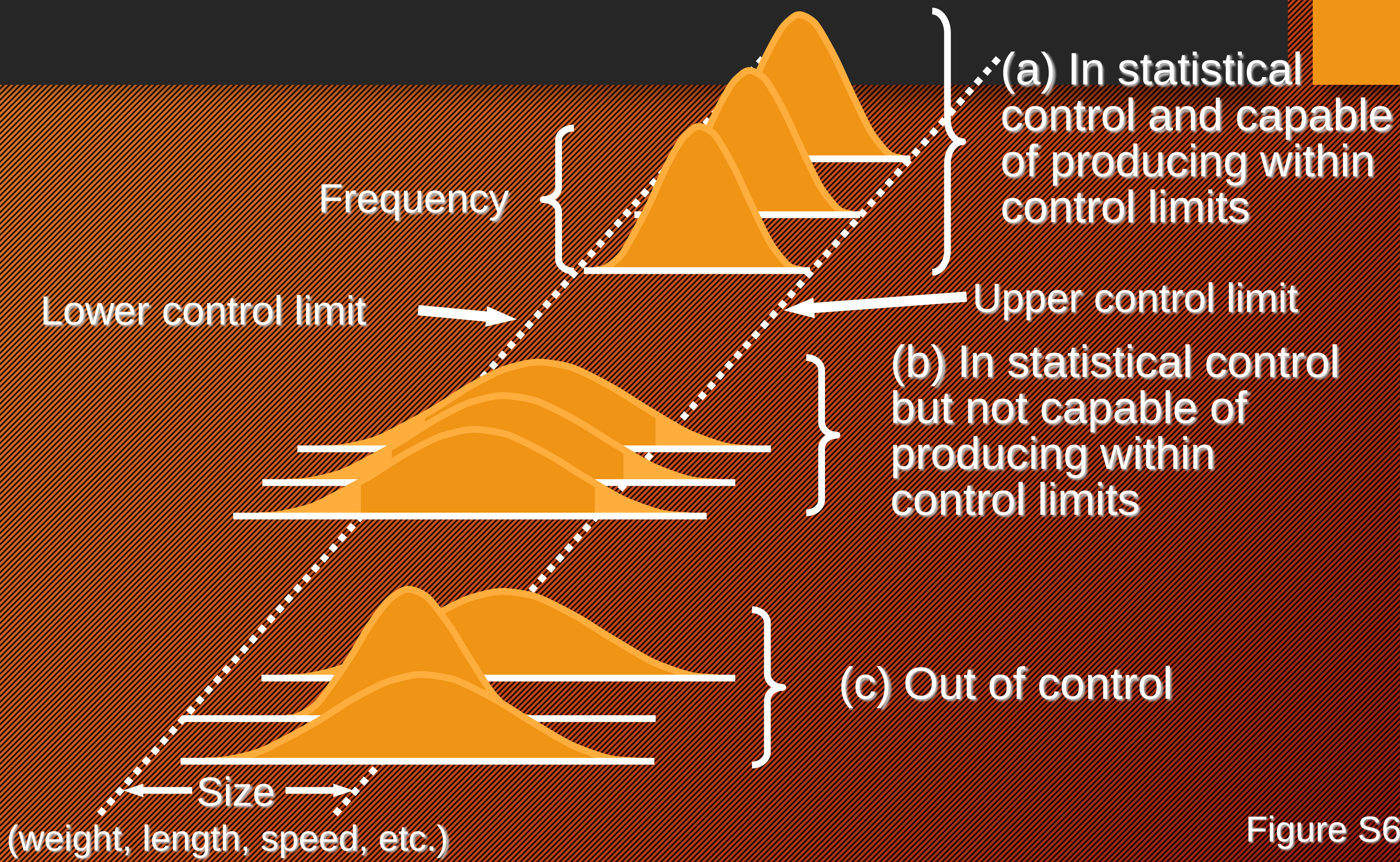


Figure S6.2

Central Limit Theorem

Regardless of the distribution of the population, the distribution of sample means drawn from the population will tend to follow a normal curve

1. The mean of the sampling distribution ($\bar{\bar{x}}$) will be the same as the population mean μ

$$\bar{\bar{x}} = \mu$$

2. The standard deviation of the sampling distribution ($\sigma_{\bar{x}}$) will equal the population standard deviation (σ) divided by the square root of the sample size, n

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Population and Sampling Distributions

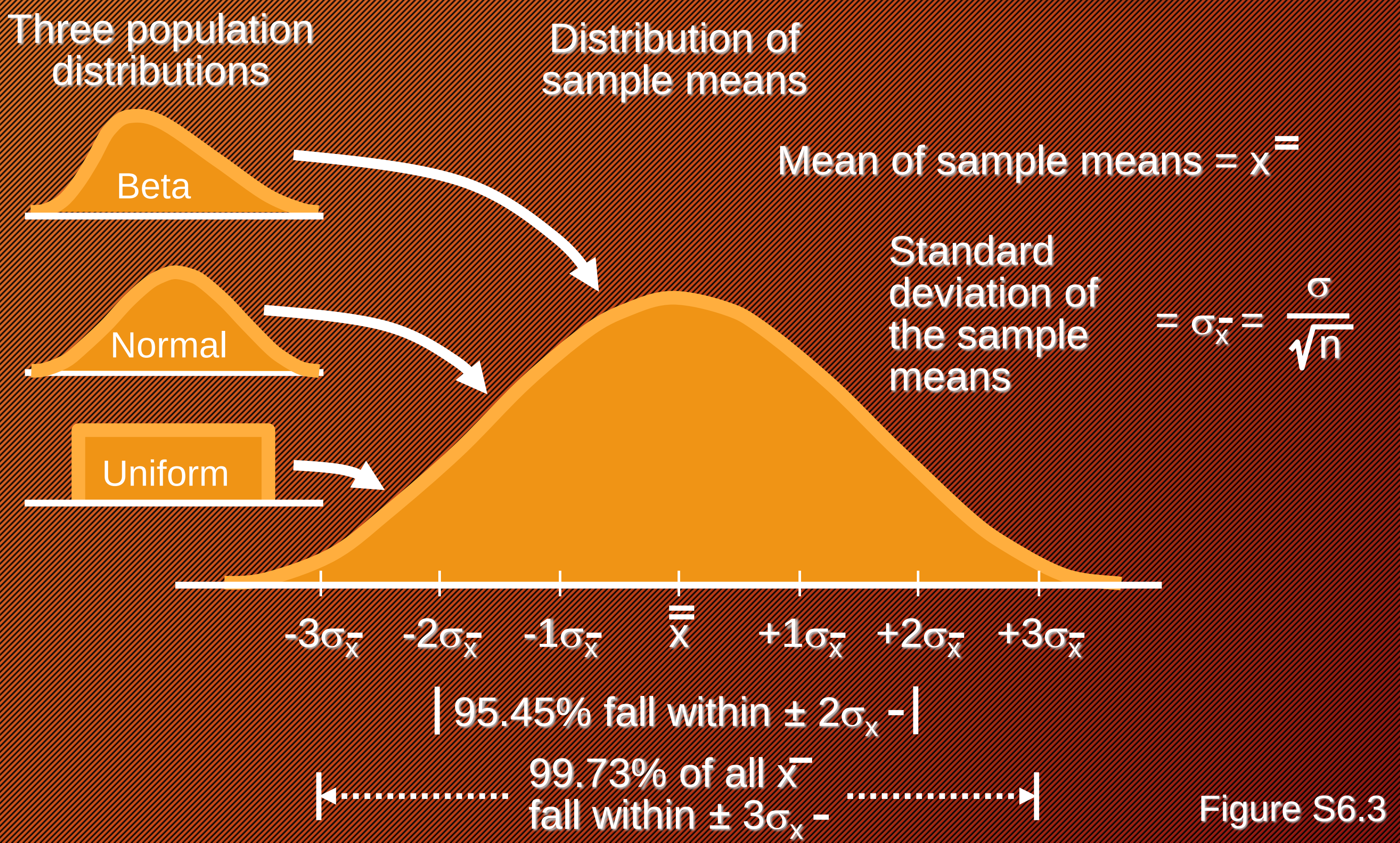


Figure S6.3

Sampling Distribution

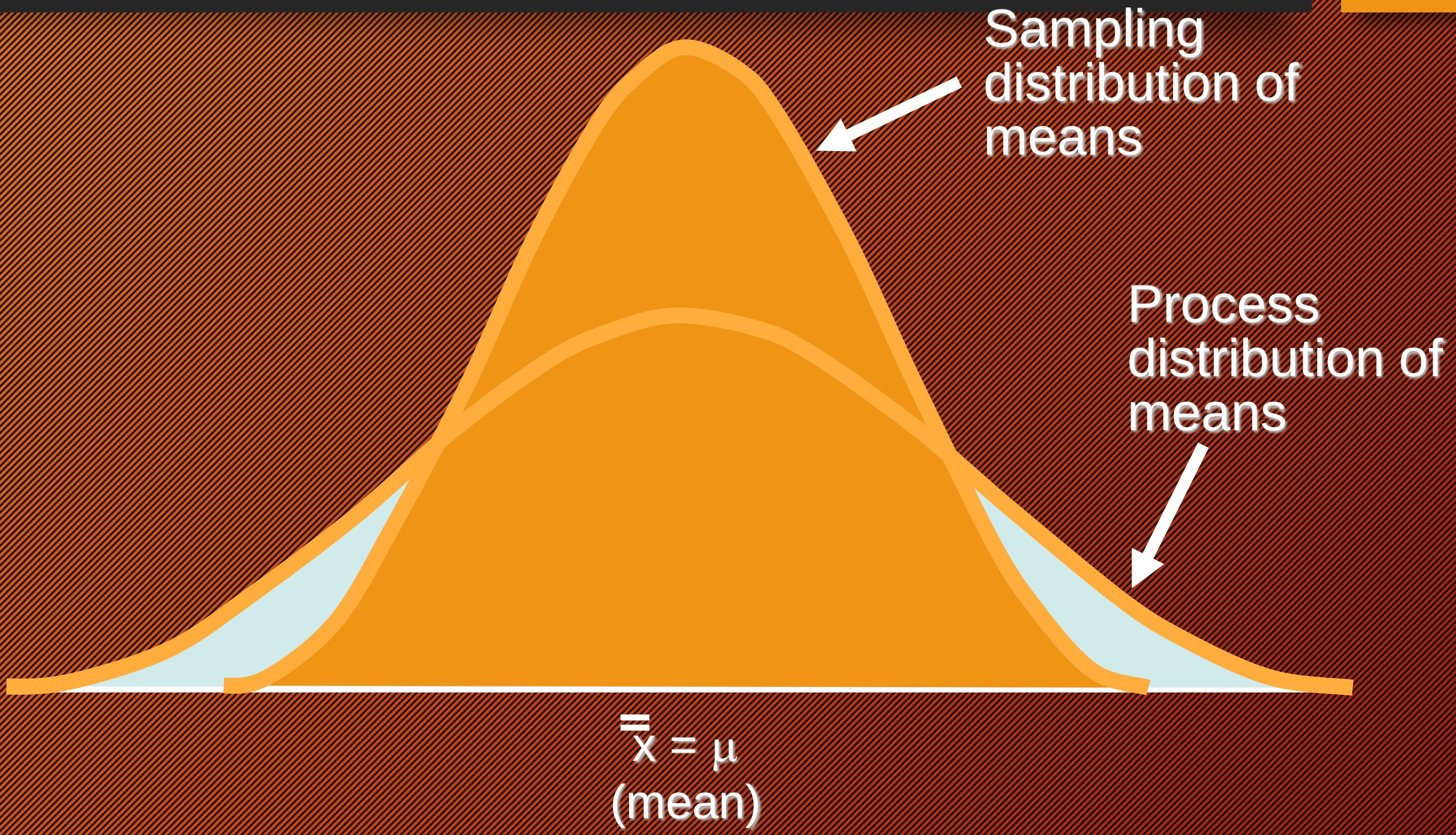


Figure S6.4

Types of Data

Variables

- ☑ ***Characteristics that can take any real value***
- ☑ ***May be in whole or in fractional numbers***
- ☑ ***Continuous random variables***

Attributes

- ☑ Defect-related characteristics
- ☑ Classify products as either good or bad or count defects
- ☑ Categorical or discrete random variables

Introduction to Statistical Quality Control

4th Hour

STATISTICAL QUALITY CONTROL

Process Cabability

Process Capability

Process capability refers to the **uniformity** of the process. Obviously, the variability of critical-to-quality characteristics in the process is a measure of the uniformity of output. There are two ways to think of this variability:

1. The natural or inherent variability in a critical-to-quality characteristic at a specified time; that is, "instantaneous" variability
2. The variability in a critical-to-quality characteristic over time

Natural tolerance limits are defined as follows:

$$\text{UNTL} = \mu + 3\sigma$$
$$\text{LNTL} = \mu - 3\sigma$$

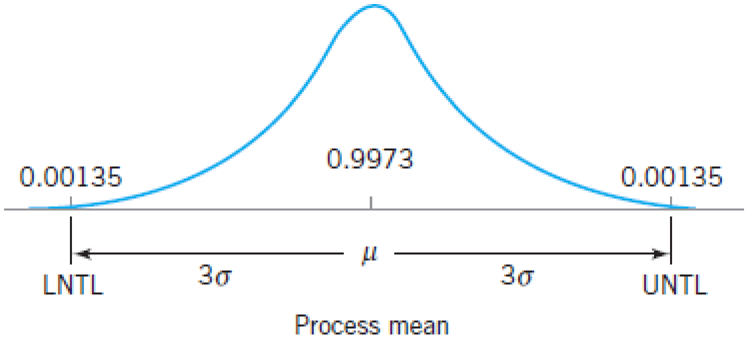


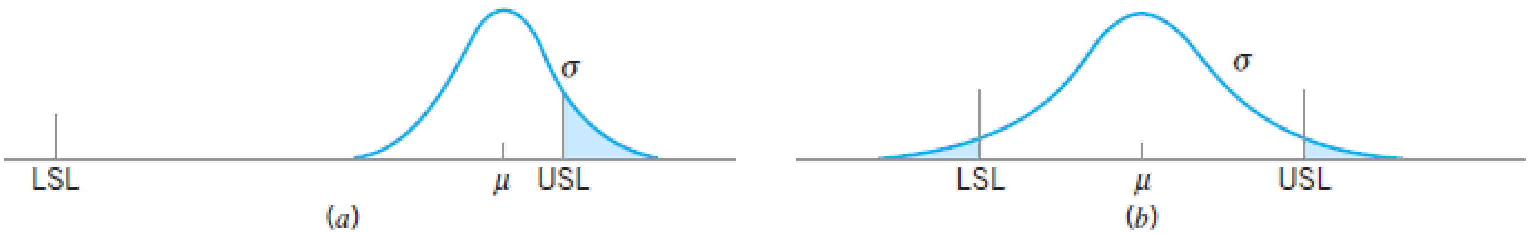
FIGURE 8.1 Upper and lower natural tolerance limits in the normal distribution.

We define **process capability analysis** as an engineering study to estimate process capability. The estimate of process capability may be in the form of a probability distribution having a specified shape, center (mean), and spread (standard deviation). For example, we may determine that the process output is normally distributed with mean $\mu = 1.0$ cm and standard deviation $\sigma = 0.001$ cm. In this sense, a process capability analysis may be performed **without regard to specifications on the quality characteristic.** Alternatively, we may express process capability as a percentage outside of specifications. However, specifications are not *necessary* to process capability analysis.

Uses of process capability data:

1. Predicting how well the process will hold the tolerances
2. Assisting product developers/designers in selecting or modifying a process
3. Assisting in establishing an interval between sampling for process monitoring
4. Specifying performance requirements for new equipment
5. Selecting between competing suppliers and other aspects of supply chain management
6. Planning the sequence of production processes when there is an interactive effect of processes on tolerances
7. Reducing the variability in a manufacturing process

Reasons for Poor Process Capability



■ **FIGURE 8.3** Some reasons for poor process capability. (*a*) Poor process centering. (*b*) Excess process variability.

Process may have good potential capability

8.2 Process Capability Analysis Using a Histogram or a Probability Plot

8.2.1 Using the Histogram

The histogram can be helpful in estimating process capability. Alternatively, a stem-and-leaf plot may be substituted for the histogram. At least 100 or more observations should be available for the histogram (or the stem-and-leaf plot) to be moderately stable so that a reasonably reliable estimate of process capability may be obtained. If the quality engineer has access to the process and can control the data-collection effort, the following steps should be followed prior to data collection:

1. Choose the machine or machines to be used. If the results based on one (or a few) machines are to be extended to a larger population of machines, the machine selected should be representative of those in the population. Furthermore, if the machine has multiple workstations or heads, it may be important to collect the data so that head-to-head variability can be isolated. This may imply that designed experiments should be used.
2. Select the process operating conditions. Carefully define conditions, such as cutting speeds, feed rates, and temperatures, for future reference. It may be important to study the effects of varying these factors on process capability.
3. Select a representative operator. In some studies, it may be important to estimate *operator* variability. In these cases, the operators should be selected at random from the population of operators.
4. Carefully monitor the data-collection process, and record the time order in which each unit is produced.

The histogram, along with the sample average $\bar{x}$ and sample standard deviation s, provides information about process capability. You may wish to review the guidelines for constructing histograms in Chapter 3.

EXAMPLE 8.1 Estimating Process Capability with a Histogram

Figure 8.2 presents a histogram of the bursting strength of 100 glass containers. The data are shown in Table 8.1. What is the capability of the process?

SOLUTION

Analysis of the 100 observations gives

$$\bar{x} = 264.06 \qquad s = 32.02$$

Consequently, the process capability would be estimated as

$$\bar{x} \pm 3s$$

or

$$264.06 \pm 3(32.02) \simeq 264 \pm 96 \text{ psi}$$

Furthermore, the shape of the histogram implies that the distribution of bursting strength is approximately normal. Thus, we can estimate that approximately 99.73% of the bottles manufactured by this process will burst between 168 and 360 psi. Note that we can estimate process capability *independently of the specifications on bursting strength.*

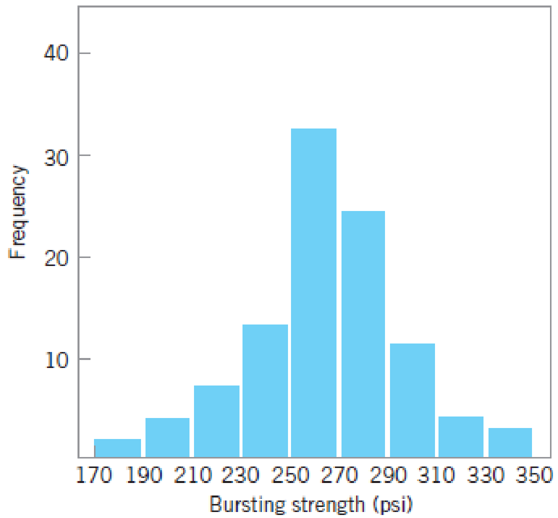


FIGURE 8.2 Histogram for the bursting-strength data.

TABLE 8.1
Bursting Strengths for 100 Glass Containers

265	197	346	280	265	200	221	265	261	278
205	286	317	242	254	235	176	262	248	250
263	274	242	260	281	246	248	271	260	265
307	243	258	321	294	328	263	245	274	270
220	231	276	228	223	296	231	301	337	298
268	267	300	250	260	276	334	280	250	257
260	281	208	299	308	264	280	274	278	210
234	265	187	258	235	269	265	253	254	280
299	214	264	267	283	235	272	287	274	269
215	318	271	293	277	290	283	258	275	251

8.3 Process Capability Ratios

8.3.1 Use and Interpretation of C_p

It is frequently convenient to have a simple, quantitative way to express process capability. One way to do so is through the **process capability ratio (PCR)** C_p first introduced in Chapter 6. Recall that

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma} \tag{8.4}$$

where USL and LSL are the upper and lower specification limits, respectively. C_p and other process capability ratios are used extensively in industry. They are also widely *misused*. We will point out some of the more common abuses of process capability ratios. An excellent recent book on process capability ratios that is highly recommended is Kotz and Lovelace (1998). There is also extensive technical literature on process capability analysis and process capability ratios. The review paper by Kotz and Johnson (2002) and the bibliography (papers) by Spiring, Leong, Cheng, and Yeung (2003) and Yum and Kim (2011) are excellent sources.

In a practical application, the process standard deviation σ is almost always unknown and must be replaced by an estimate σ. To estimate σ we typically use either the *sample standard deviation s* or $\overline{R}/d_2$ (when variables control charts are used in the capability study). This results in an estimate of C_p—say,

$$\hat{C}_p = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} \tag{8.5}$$

- C_p does not take process centering into account

- It is a measure of **potential** capability, not **actual** capability

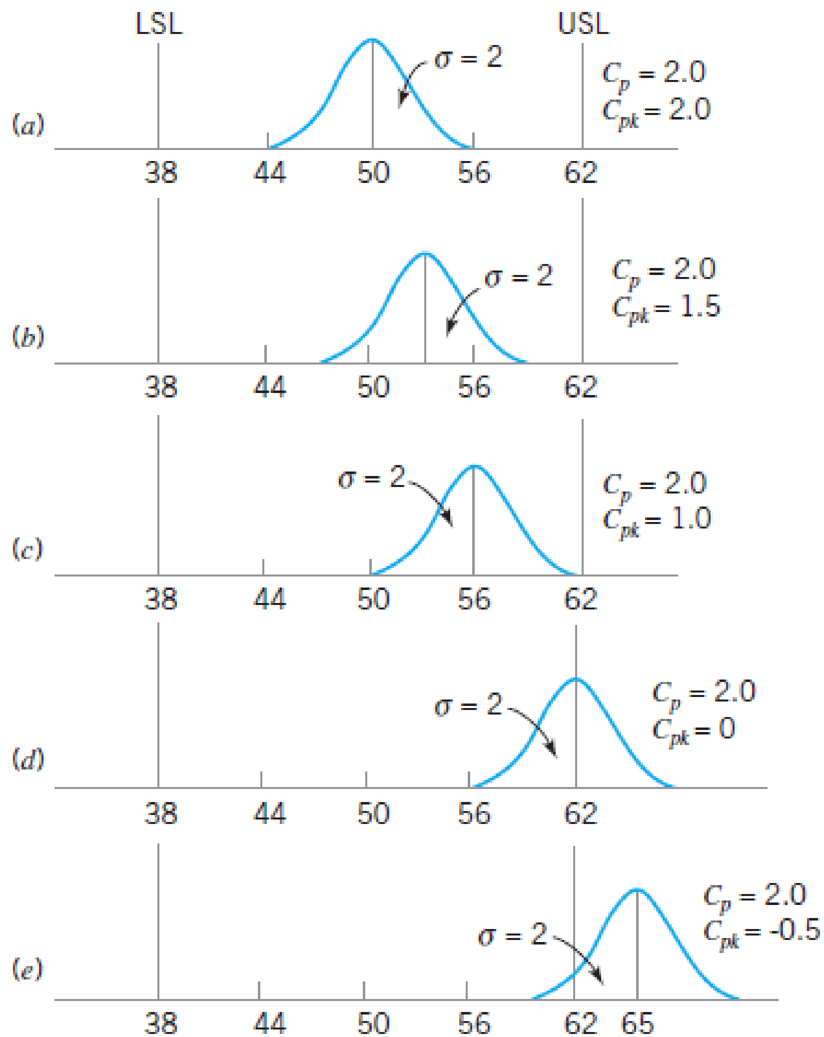


FIGURE 8.8 Relationship of C_p and C_{pk}.

A Measure of **Actual** Capability

$$C_{pk} = \min\left(C_{pu}, C_{pl}\right) \tag{8.9}$$

Note that C_{pk} is simply the one-sided PCR for the specification limit nearest to the process average. For the process shown in Figure 8.8*b*, we would have

$$\begin{aligned}
C_{pk} &= \min\left(C_{pu}, C_{pl}\right) \\
&= \min\left(C_{pu} = \frac{\text{USL} - \mu}{3\sigma}, C_{pl} = \frac{\mu - \text{LSL}}{3\sigma}\right) \\
&= \min\left(C_{pu} = \frac{62 - 53}{3(2)} = 1.5, C_{pl} = \frac{53 - 38}{3(2)} = 2.5\right) \\
&= 1.5
\end{aligned}$$

Generally, if $C_p = C_{pk}$, the process is centered at the midpoint of the specifications, and when $C_{pk} < C_p$ the process is off center.

Process Capability

Product Specifications

- Preset product or service dimensions, tolerances: bottle fill might be 16 oz. ±.2 oz. (15.8oz.-16.2oz.)
- Based on how product is to be used or what the customer expects

Process Capability – Cp and Cpk

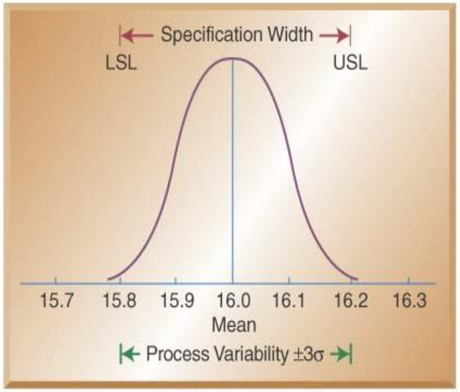
- Assessing capability involves evaluating process variability relative to preset product or service specifications
- **Cp** assumes that the process is centered in the specification range

$$\mathbf{Cp} = \frac{\textbf{specification width}}{\textbf{process width}} = \frac{\mathbf{USL - LSL}}{\mathbf{6\sigma}}$$

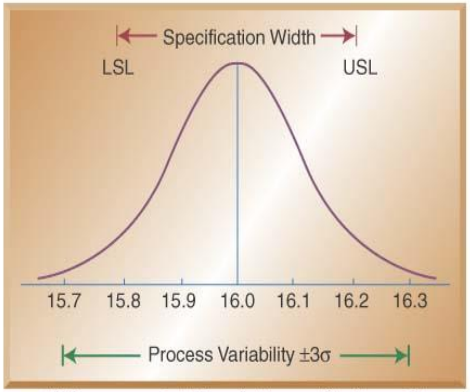
- **Cpk** helps to address a possible lack of centering of the process

$$\mathbf{Cpk} = \mathbf{min}\left(\frac{\mathbf{USL - \mu}}{\mathbf{3\sigma}}, \frac{\mathbf{\mu - LSL}}{\mathbf{3\sigma}}\right)$$

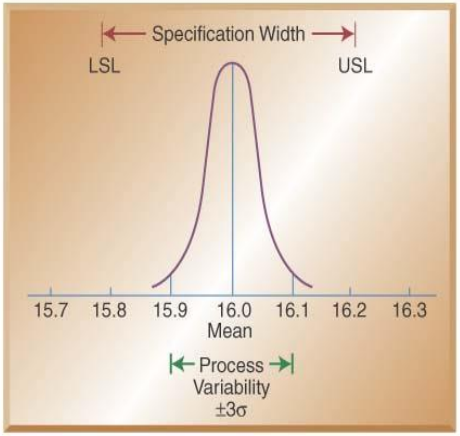
Relationship between Process Variability and Specification Width



(a) Process variability meets specification width



(b) Process variability outside specification width



(c) Process variability within specification width

- **Three possible ranges** for **Cp**
  - **Cp = 1, as in Fig. (a), process variability just meets specifications**
  - **Cp ≤ 1, as in Fig. (b), process not capable of producing within specifications**
  - **Cp ≥ 1, as in Fig. (c), process exceeds minimal specifications**
- **One shortcoming, Cp assumes that the process is centered on the specification range**
- **Cp=Cpk when process is centered**

Computing the Cp Value at Cocoa Fizz: **3 bottling machines** are being evaluated for possible use at the Fizz plant. The machines **must be capable of meeting the design specification of 15.8-16.2 oz**. with at least a **process capability index of 1.0 ($C_p \geq 1$)**

The table below shows the information gathered from production runs on each machine. **Are they all acceptable?**

Machine	σ	USL-LSL	6σ
A	.05	.4	.3
B	.1	.4	.6
C	.2	.4	1.2

Solution:

– **Machine A**

$$\mathbf{Cp}\ \frac{\mathbf{USL-LSL}}{\mathbf{6\sigma}} = \frac{\mathbf{.4}}{\mathbf{6(.05)}} = \mathbf{1.33}$$

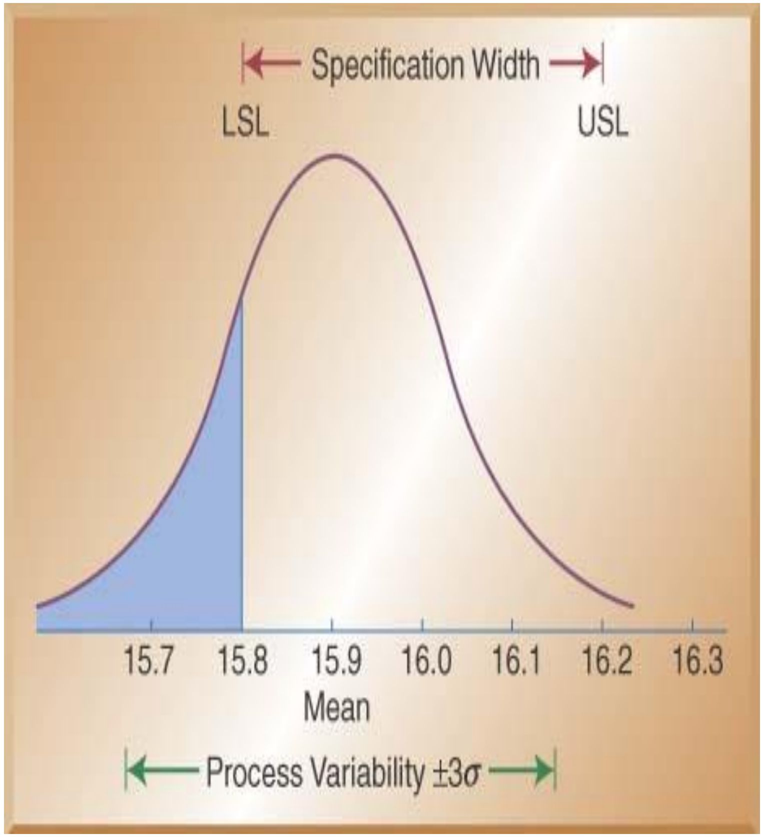
– **Machine B**

Cp=

– **Machine C**

Cp=

Computing the C_{pk} Value at Cocoa Fizz



- Design specifications call for a target value of **16.0 ±0.2 OZ.**

(USL = 16.2 & LSL = 15.8)

- Observed process output has now shifted and has a **μ of 15.9** and a **σ of 0.1 oz.**

- **Cpk is less than 1**, revealing that the process **is not capable**

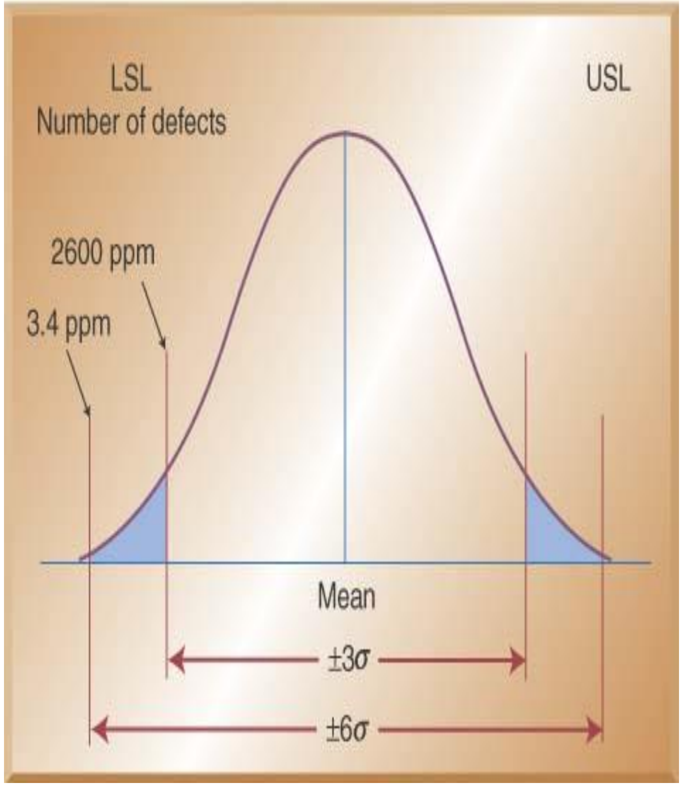
±6 Sigma versus ± 3 Sigma

- **In 1980's, Motorola coined "six-sigma" to describe their higher quality efforts**

Six-sigma quality standard is now a benchmark in many industries

  - Before design, marketing ensures customer product characteristics
  - Operations ensures that product design characteristics can be met by controlling materials and processes to 6σ levels
  - Other functions like finance and accounting use 6σ concepts to control all of their processes

- **PPM Defective for ±3σ versus ±6σ quality**



STATISTICAL QUALITY CONTROL

The End