Types of Data and Measurement Scales

**Data Science for Quality Management:
Module 1 - Data and Measurement**

with **Wendy Martin**

Learning objective:
Discern between qualitative and quantitative data, continuous and discrete data

Compare / contrast measurement and underlying characteristics

Using Data

"When you can measure what you are speaking about, and express it in numbers, you know something about it…

Using Data

…but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind."

Lord Kelvin (1883)

Why Do We Need Data?

In business, we have questions that need to be answered.

- We must ask ourselves, "What do I really want to know?" and
- "Do my data help answer these questions?"

Data Costs Money

We must make data both
- Efficient and
- Effective

Measurement and Data

How do we study, record and communicate an event? We assign numbers.

- **Measurement** is the *process*
- **Data** is the *output*

Two Basic Types of Data

Quantitative data are data measured along a numerical scale.

- Often referred to as **continuous**.

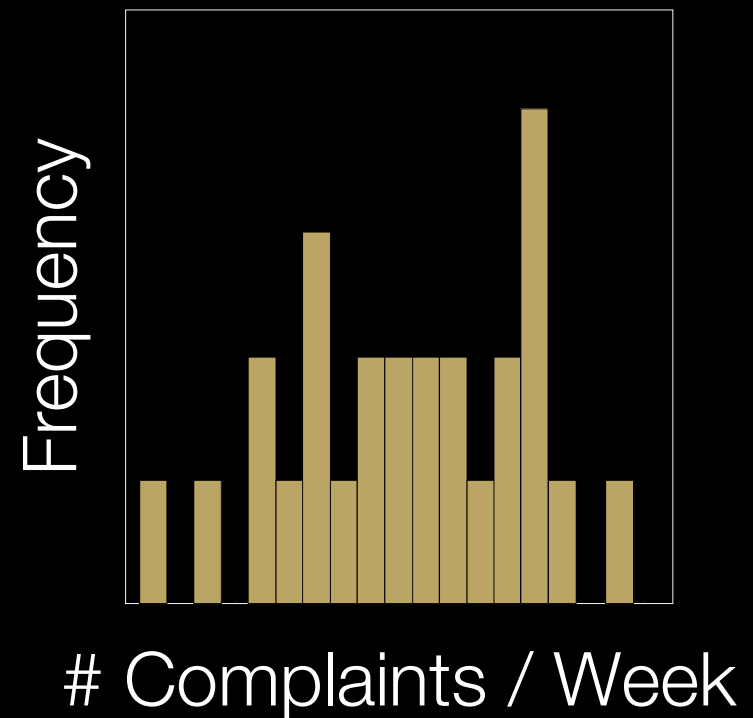# Two Basic Types of Data

Qualitative data are descriptions that fall into categories.

- Often referred to as **discrete**.
- Frequencies, proportions, or rates.

Discrete vs. Continuous Data

Discrete Data:
- Items/Units we count

Frequency
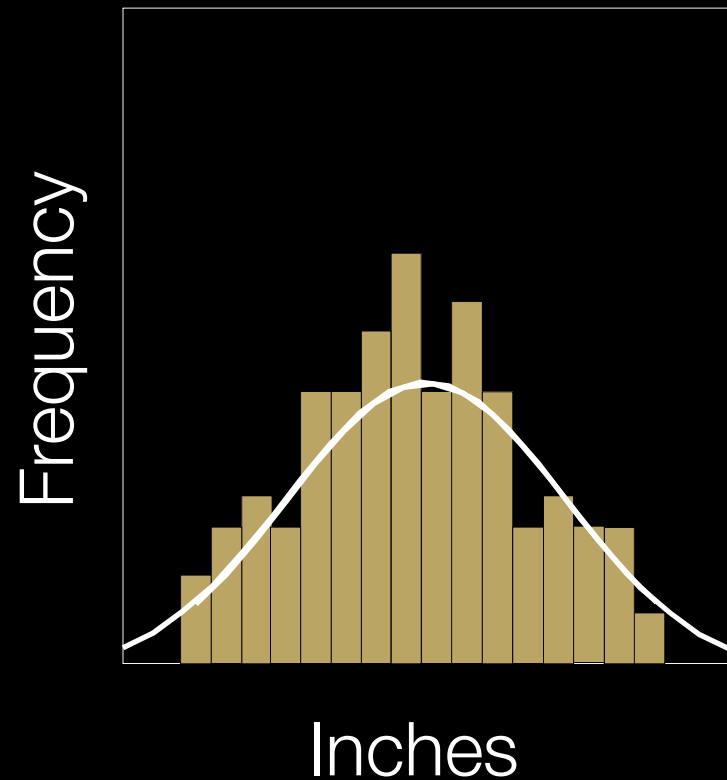
Complaints / Week

Discrete Data Examples

Examples of discrete data in business:
- Complaints per sales period
- Number of defects per unit
- Percent defective units
- Number of orders shipped on time

Discrete vs. Continuous Data

Continuous Data:
- Items/Units
  we measure

Continuous Data Examples

Examples of continuous data in business:
- Dimensions (height, length, width)
- Temperature
- Speed
- Volume of sales

Measurement & Measurement Scales

Measurement is the assignment of numbers or other symbols to an underlying attribute, characteristic or property.

Measurement & Measurement Scales

The numbers, or symbols, are assigned such that the relationships amongst the numbers or symbols reflect relationships in the attribute studied.

Measurement & Measurement Scale Example

Measurement & Measurement Scale Example

5 Dents Per Can

3 Dents Per Can

0 Dents Per Can

Measurement & Measurement Scales

Measurements are not the same as the attribute studied

To draw conclusions, we must consider how the measurement maps to the attribute

Data

Underlying Property → Criterion Measure

↓   ↓

Operational Definition   Data

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Measurement Scales

**Data Science for Quality Management:
Data and Measurement**
with **Wendy Martin**

Learning objectives:
Distinguish between nominal, ordinal, interval, ratio and absolute scales

Identify measurement level of data given background information

Measurement & Measurement Scales

Measurements are not the same as the attribute studied

To draw conclusions, we must consider how the measurement maps to the attribute

Measurement & Measurement Scales

The relationship between the assigned numbers or symbols and the corresponding underlying property yields a level of measurement.

Measurement Scales

Nominal

Ordinal

Interval

Ratio

Absolute

Nominal Scale

Numbers are assigned to categorize, identify or name attributes

- Zip codes
- Area codes
- Numbers assigned to types of nonconformity in products

Nominal Scale

- Numbers assigned to presence or absence of an attribute (e.g., 0, 1)
- Numbers assigned to sales territories

Nominal Scale values can only be used to indicate equal or not equal

Nominal Scale Example

```
> round.object(Nominal.Data.Example, digits = 4)
  Code        Label Response Relative.Frequency Cumulative.Frequency
1    1      Married      998             0.4250               0.4250
2    2      Widowed      200             0.0852               0.5102
3    3     Divorced      403             0.1716               0.6819
4    4    Separated       75             0.0319               0.7138
5    5 Never married      670             0.2853               0.9991
6    9    No answer        2             0.0009               1.0000
```

Be sure to compare the values in the Frequency column of the table and the bars of the Histogram.

Marital Status Data 2018
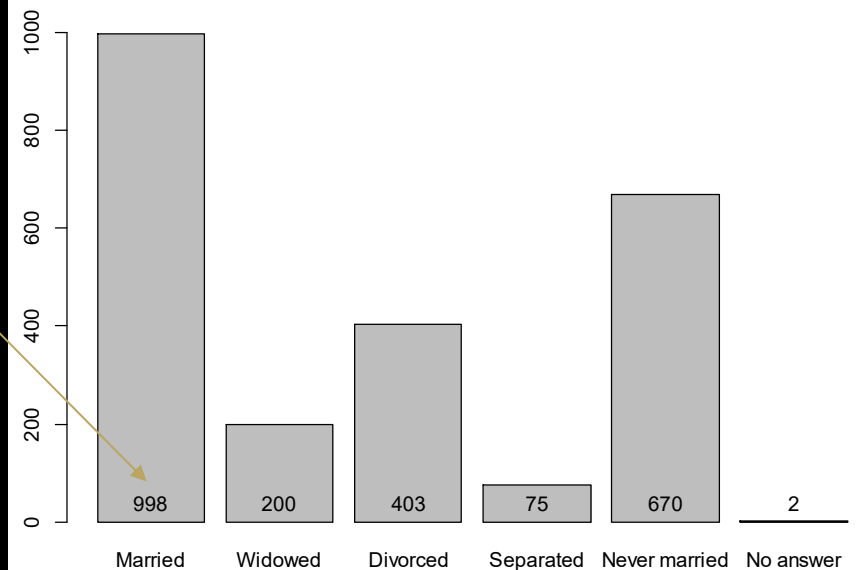
Ordinal Scale

Numbers are assigned to observations, such that the order of the numbers corresponds to the order of the underlying property studied

- 5-point scale to measure customer satisfaction
- Letter Grades for Academic performance

Ordinal Scale

- Sound intensity measured in decibels
- Socio-economic status
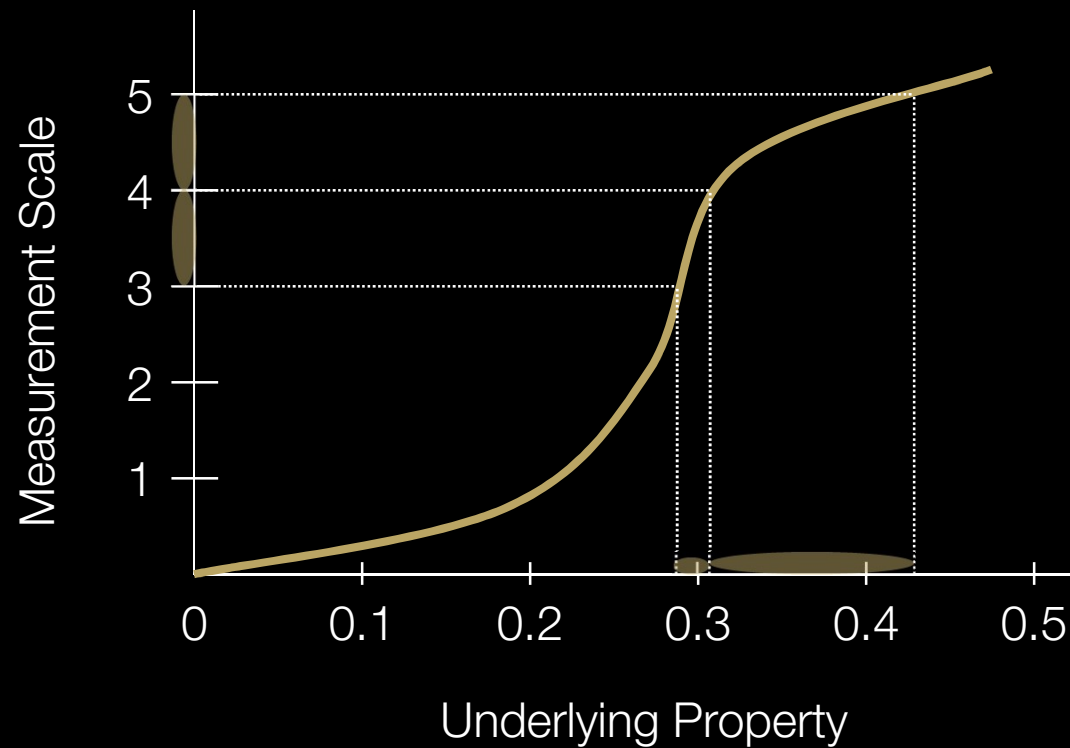- Project priority numbers

Ordinal Scale

Ratings and rankings are ordinal
- Ratings: assign a subjective score
- Rankings: result from sorting items

Ordinal Scale values can be used to determine = or ≠ and > or < as well, but NOT magnitude

Ordinal Scale Example

Ordinal Scale

Any order-preserving transformation may be used.

The median and mode may be used for "center" and range or interquartile range for dispersion.

Ordinal Scale

Numerous nonparametric hypothesis tests are available, which often involve ranking the data prior to conducting the inferential statistical analysis of interest.

Interval Scale

Numbers are assigned to observations such that differences between any two numbers correspond to proportional differences in the underlying property studied, and there are equal intervals along the scale.

Interval Scale

Examples:
- Temperature measured in degrees Fahrenheit
- Directional distance from a reference point
- Calendar date
- Height above sea level

Interval Scale

Permissible transformations include any linear transformation.

Statistics such as the mean and standard deviation may be used.

Interval Scale

Most of the parametric statistical tests may be employed as long as their underlying assumptions are met.

Interval Scale values can be used to determine = or ≠, > or <, and you can use sums (+) and differences (–) with meaning
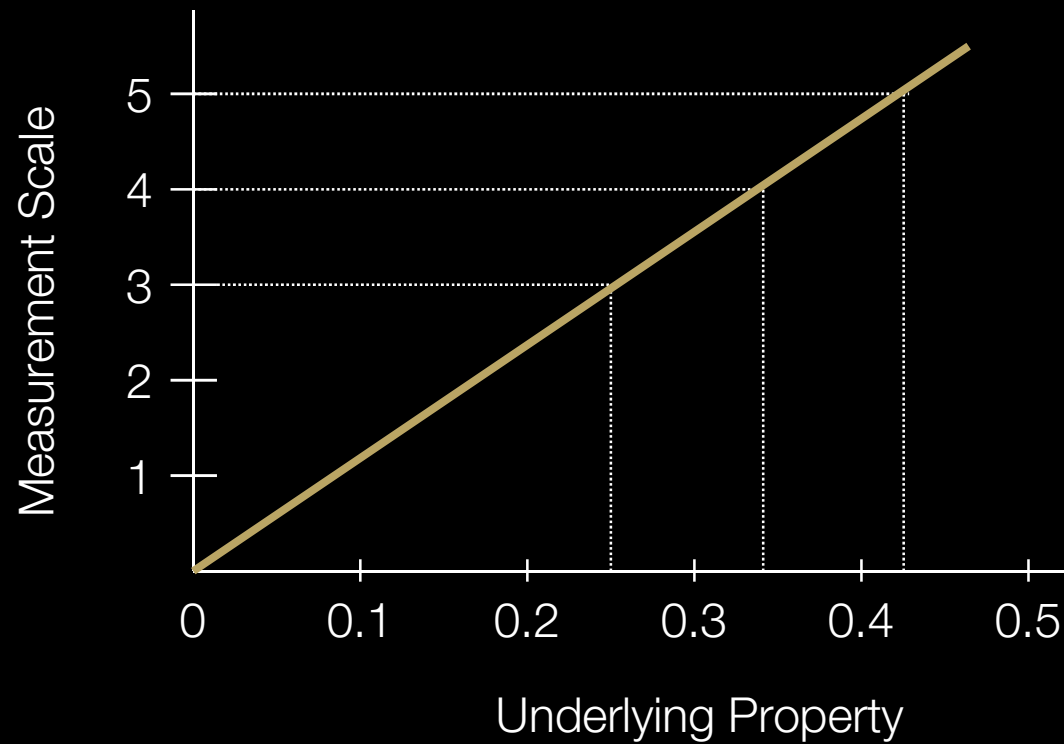
Ratio Scale

Numbers are assigned to observations such that an interval scale has been reached, and there is a zero point which corresponds to a zero, null state, or absence of the underlying property.

Ratio Scale

Examples:

- Length, height, width, distance
- Volume, weight
- Cycle-time and time-to-repair

Ratio Scale Example

Ratio Scale

Permissible transformations of ratio data involve simply multiplying by a constant.

A few more statistics are available:
- The geometric mean
- The coefficient of variation

Ratio Scale

Any parametric statistical test may be employed with ratio data as long as their underlying assumptions are met (or not significantly violated in the case of robust tests) .

Ratio Scale vs Interval Scale

Consider a bathroom scale:

Two people are weighed on the same bathroom scale, one weighs 200 pounds and the other weighs 100 pounds, is the first twice as heavy as the second?

Ratio Scale vs Interval Scale

What if I told you the scale was NOT set at zero to begin with, but at 50 pounds. Then what?

What are the corresponding Scales of Measurement for each situation?
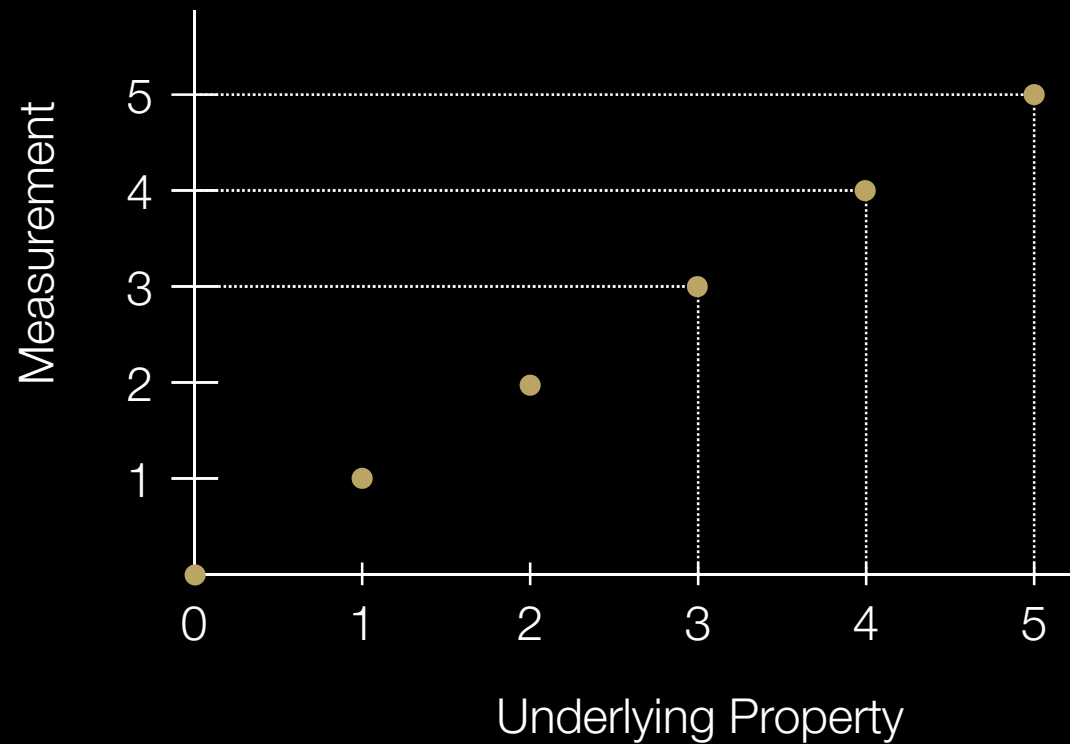
Absolute Scale (Count Data)

Numbers are assigned to observations such that the numbers directly correspond to the underlying property being studied.

Absolute Scale (Count Data)

Examples:

- The number of defects
- The number of scratches observed
- The number of parts made
- The number of safety accidents
- The number of customer complaints

Absolute Scale (Count Data)

Some transformations are permissible.

Data on an absolute scale have some of the properties of ratio data.

The mean, median and mode may be used.

Absolute Scale (Count Data)

Various statistical procedures may be used depending on resolution.

In many cases, standard parametric methods may be used, in other cases nonparametric methods must be employed.

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Measurement Scales

**Data Science for Quality Management:
Data and Measurement**
with **Wendy Martin**

Learning objectives:
Distinguish between nominal, ordinal, interval, ratio and absolute scales

Identify measurement level of data given background information

Measurement & Measurement Scales

Measurements are not the same as the attribute studied

To draw conclusions, we must consider how the measurement maps to the attribute

Measurement & Measurement Scales

The relationship between the assigned numbers or symbols and the corresponding underlying property yields a level of measurement.

Measurement Scales

Nominal

Ordinal

Interval

Ratio

Absolute

Interval Scale

Numbers are assigned to observations such that differences between any two numbers correspond to proportional differences in the underlying property studied, and there are equal intervals along the scale.

Interval Scale

Examples:

- Temperature measured in degrees Fahrenheit
- Directional distance from a reference point
- Calendar date
- Height above sea level

Interval Scale Example

Interval Scale

Permissible transformations include any linear transformation.

Statistics such as the mean and standard deviation may be used.

Interval Scale

Most of the parametric statistical tests may be employed as long as their underlying assumptions are met.

Interval Scale values can be used to determine = or ≠, > or <, and you can use sums (+) and differences (−) with meaning

Ratio Scale

Numbers are assigned to observations such that an interval scale has been reached, and there is a zero point which corresponds to a zero, null state, or absence of the underlying property.

Ratio Scale

Examples:
- Length, height, width, distance
- Volume, weight
- Cycle-time and time-to-repair

Ratio Scale Example

Ratio Scale

Permissible transformations of ratio data involve simply multiplying by a constant.

A few more statistics are available:
- The geometric mean
- The coefficient of variation

Ratio Scale

Any parametric statistical test may be employed with ratio data as long as their underlying assumptions are met (or not significantly violated in the case of robust tests) .

Ratio Scale vs Interval Scale

Consider a bathroom scale:

Two people are weighed on the same bathroom scale, one weighs 200 pounds and the other weighs 100 pounds, is the first twice as heavy as the second?

Ratio Scale vs Interval Scale

What if I told you the scale was NOT set at zero to begin with, but at 50 pounds. Then what?

What are the corresponding Scales of Measurement for each situation?

Absolute Scale (Count Data)

Numbers are assigned to observations such that the numbers directly correspond to the underlying property being studied.

Absolute Scale (Count Data)

Examples:

- The number of defects
- The number of scratches observed
- The number of parts made
- The number of safety accidents
- The number of customer complaints

Absolute Scale Example

Absolute Scale (Count Data)

Some transformations are permissible.

Data on an absolute scale have some of the properties of ratio data.

The mean, median and mode may be used.

Absolute Scale (Count Data)

Various statistical procedures may be used depending on resolution.

In many cases, standard parametric methods may be used, in other cases nonparametric methods must be employed.

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Measurement as a Process, The Big 5 Aspects of Data

**Data Science for Quality Management:
Data and Measurement**
with **Wendy Martin**

Learning objectives:

Describe the process of measurement
Demonstrate and recall the big five (5)
aspects of data

Data Costs Money

We must make data both
- Efficient and
- Effective

How Do We Make Data Efficient and Effective?

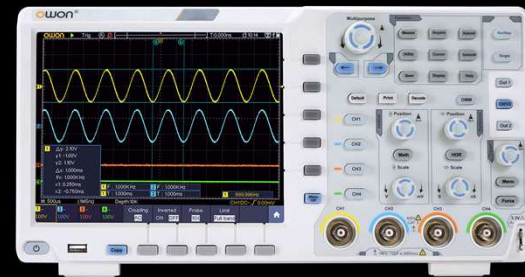Make certain that the measurement process itself is effective:

- Capable
- Acceptable

Measurement and Data

How do we study, record and communicate an event? We assign numbers.

- **Measurement** is the *process*
- **Data** is the *output*

1.5	1.3	1.6	1.5
1.6	1.4	1.2	1.3
1.2	1.7	1.3	1.2
1.4	1.4	1.1	1.7
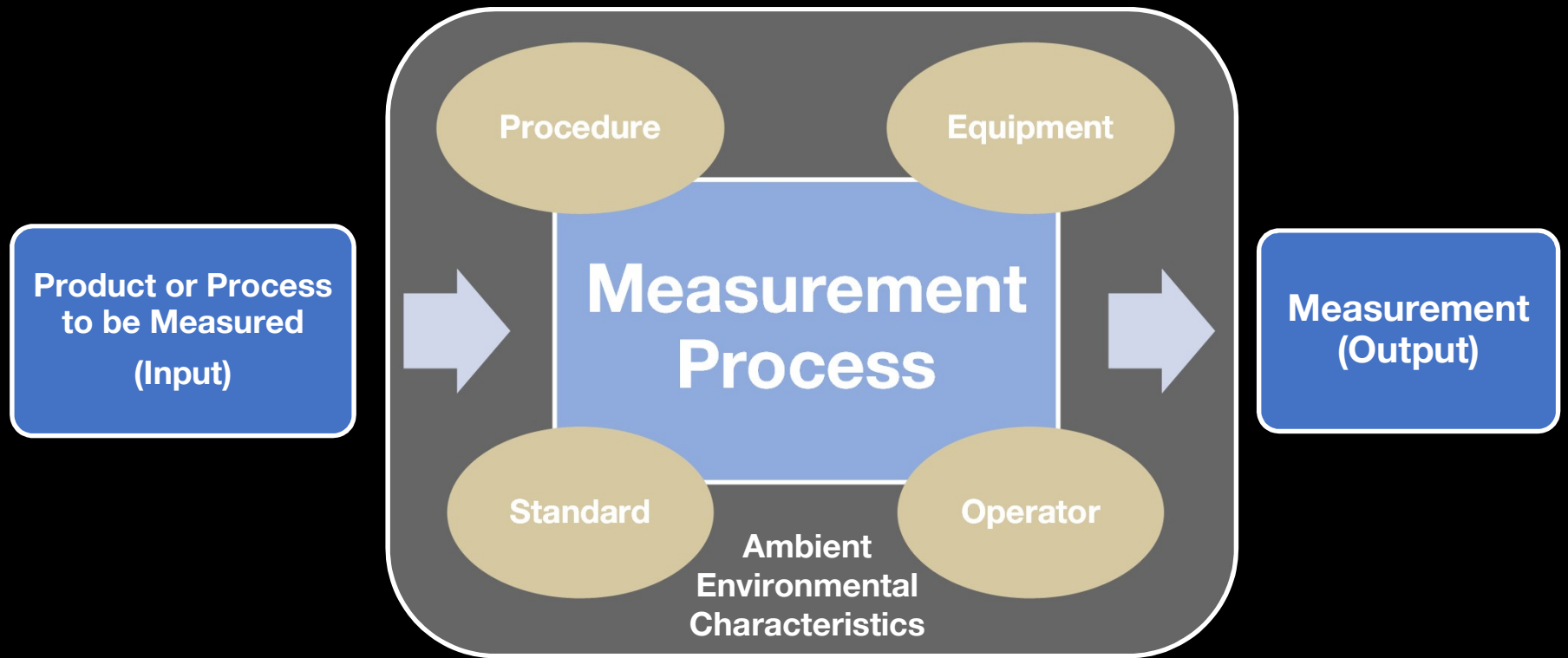
Common Myths of Measurement

The measurement system MUST be effective (capable) if it:

- Is expensive
- Has a digital readout
- Uses radiation
- Is made in Switzerland

- Is only used by a Supervisor
- Has a recent calibration sticker

Measurement as a Process

As in any process, measurement systems must demonstrate:

- Stability through time, or control
- The ability to generate reliable, or repeatable and reproducible measures
- The ability to generate valid measures

Measurement as a Process

Reliability in measurement is a measure of the Precision of the device / method

Validity in metrology is a measure of the Accuracy of the device / methods
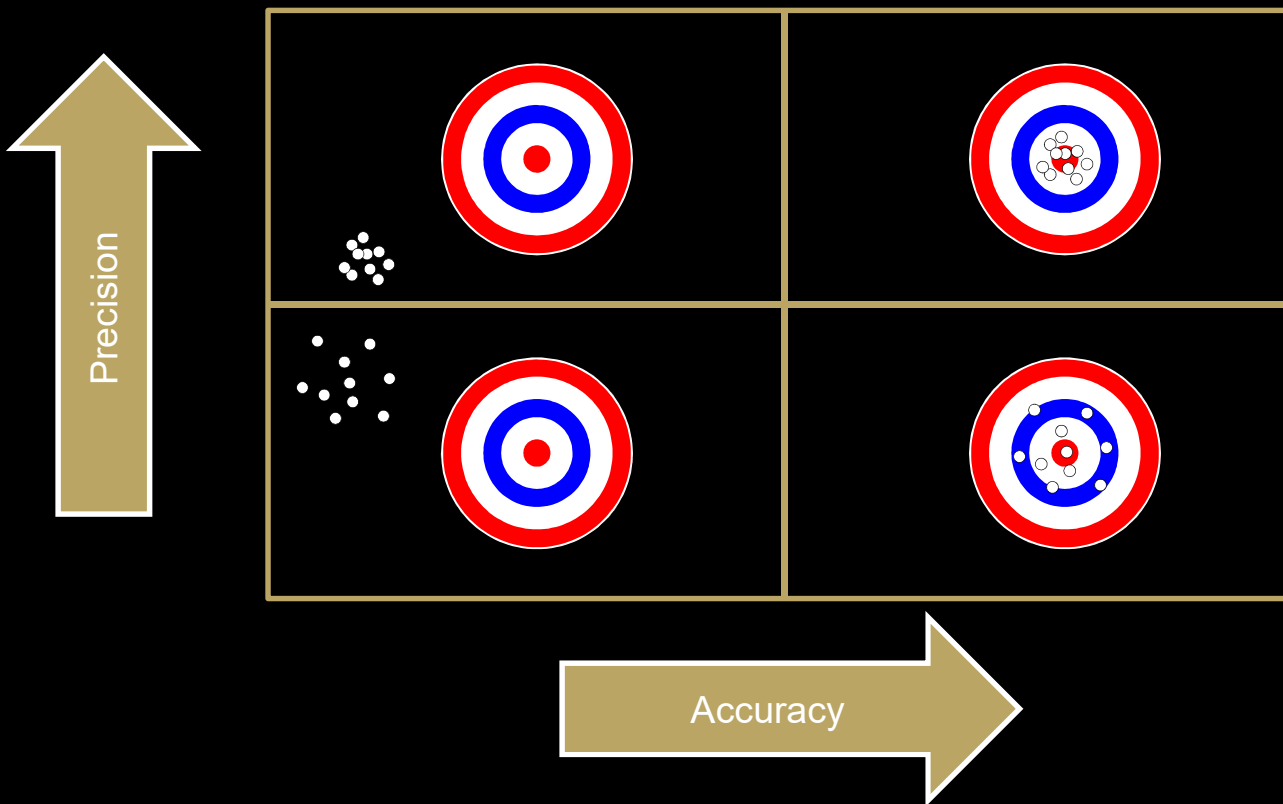
Tools for Understanding Data

Once the measurement process is found to be capable and acceptable (more on this later), we can begin to analyze the generated data

Tools for Understanding Data

There are many tools to help us understand what data are telling us, including

- Probability and statistics; to quantify and summarize data
- Control charts; to determine if a process is stable

Tools for Understanding Data

- Experimental design; to allow us to identify root causes of a problem so we can eliminate it; or to identify and properly manipulate Special and Common Causes of Variability to improve optimize the process output

Big 5 Aspects of Data

Location

Spread

Shape

Time Sequence

Relationship

5 Aspects of Data

Location

Spread

Shape

Time Sequence

Relationship

Measures of Location

Measures of location, sometimes called measures of central tendency, describe a middle or central point or tendency of a distribution.

- Mean, Median, Mode

Measures of Location

Mode

Median

Mean
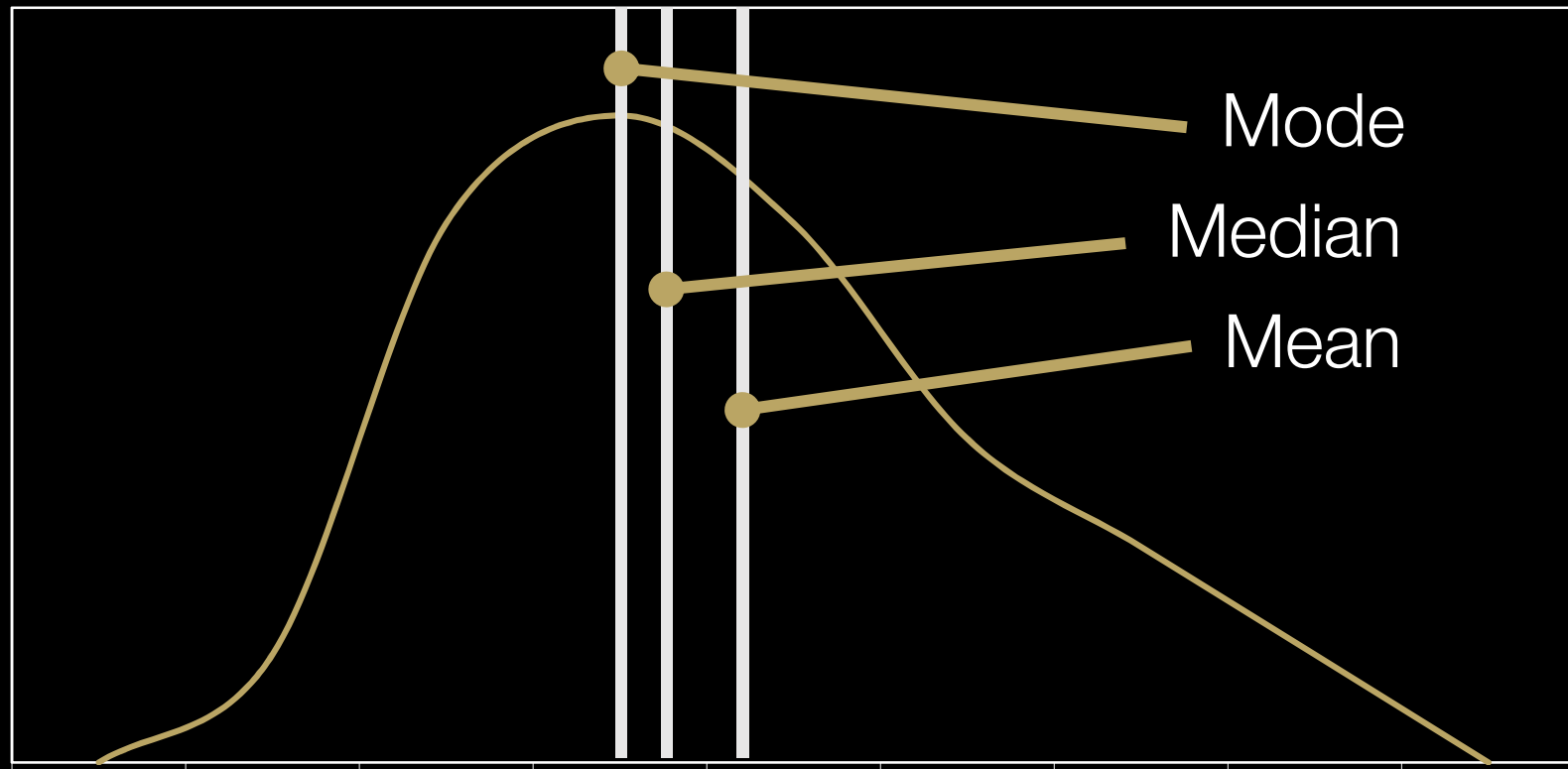
Measures of Spread

Measures of dispersion reflect the variation or spread in a data set or distribution. Some of the common measures of dispersion are:
- Range, Standard Deviation, Variance

Measures of Spread

Standard
Deviation

Range
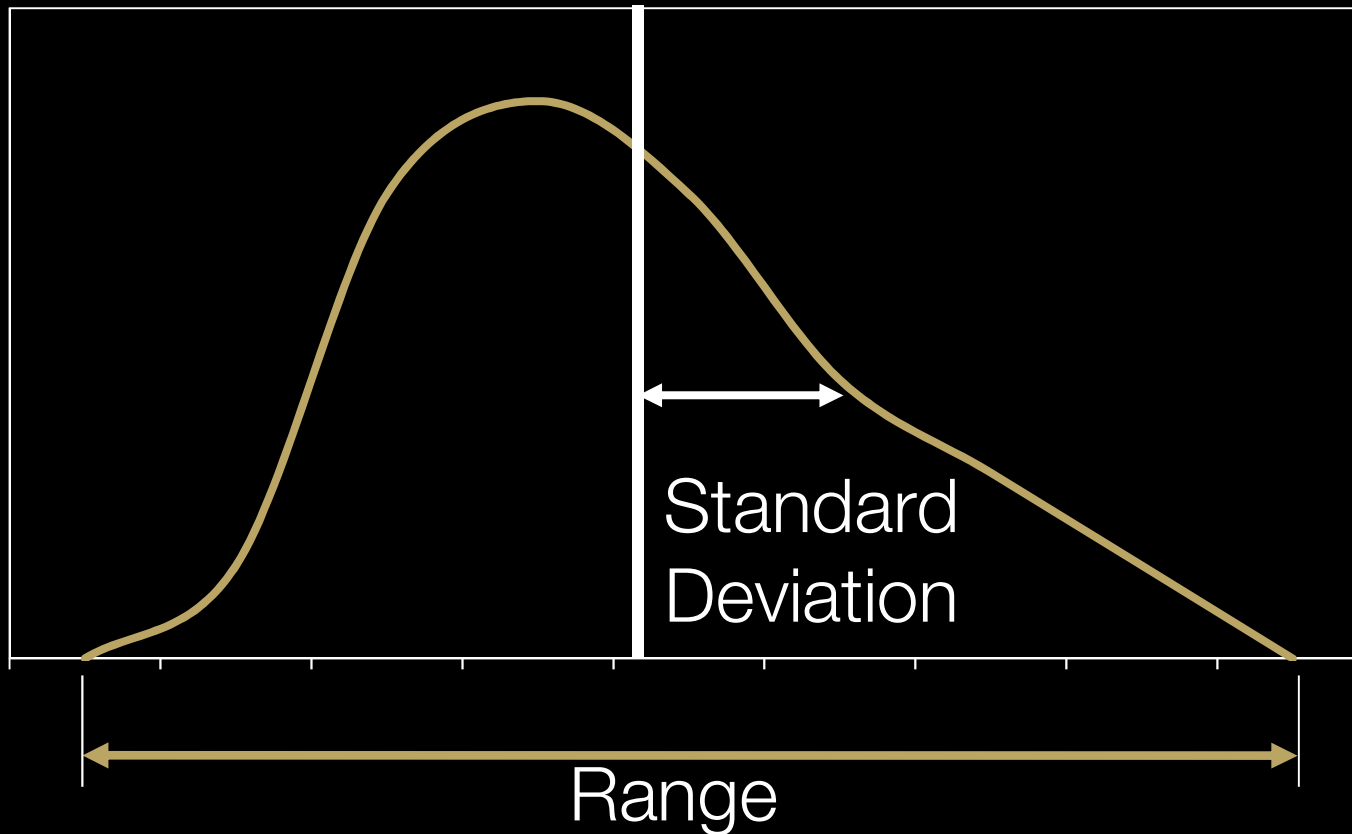
Measures of Shape

Measures of shape reflect the type of distribution sampled. There are two measures:
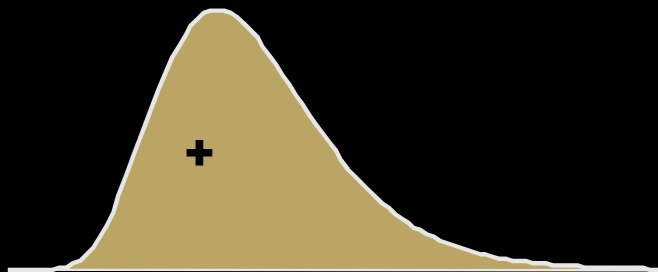
- Skewness
- Kurtosis

Skewness

Skewness is concerned with the symmetrical nature of the distribution, and is the degree of departure from symmetry of a distribution.
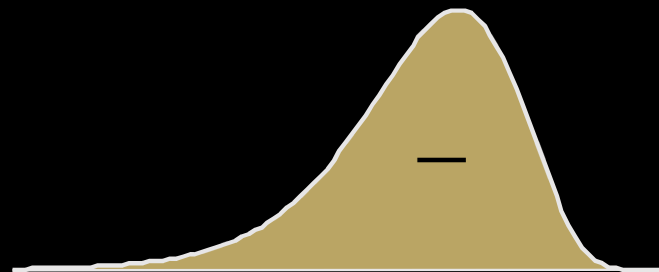
Skewness

Basically measures "lopsidedness."
Symmetric distributions have zero skewness.
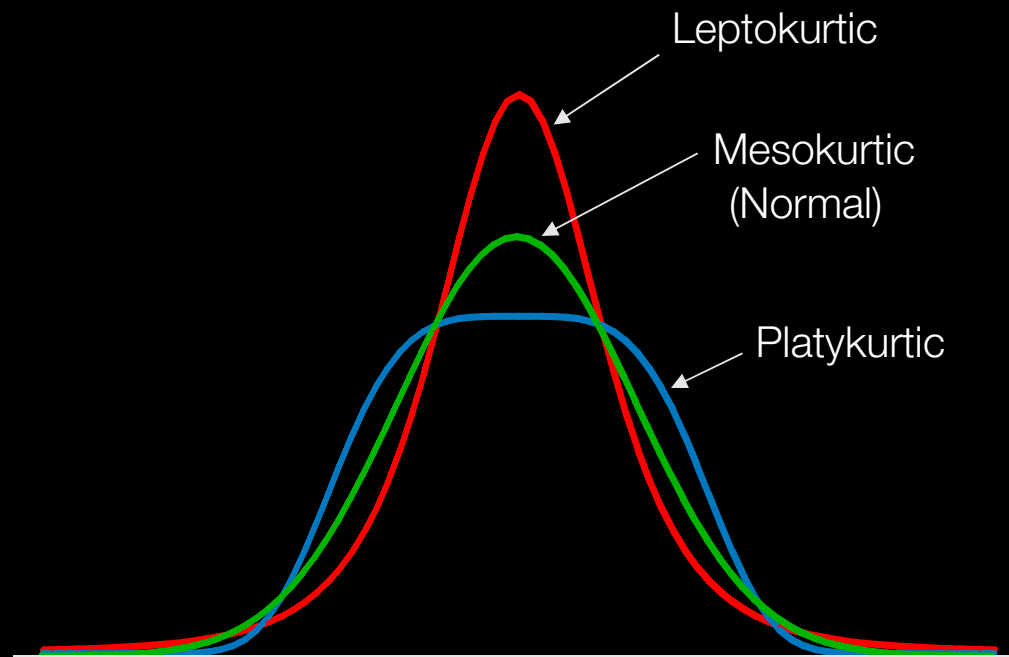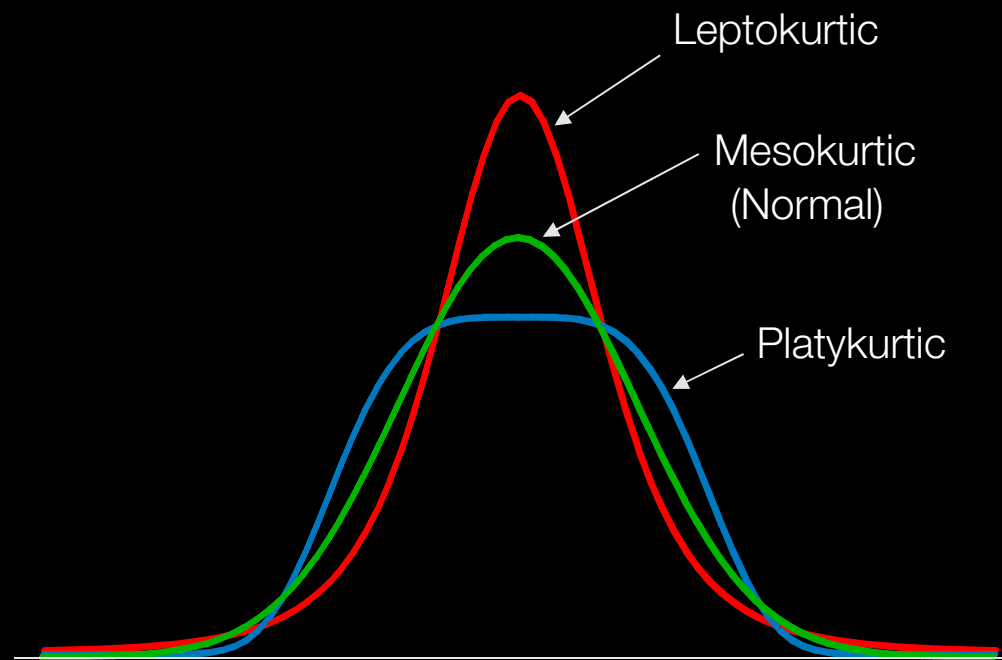

Positively skewed   Negatively skewed

Kurtosis

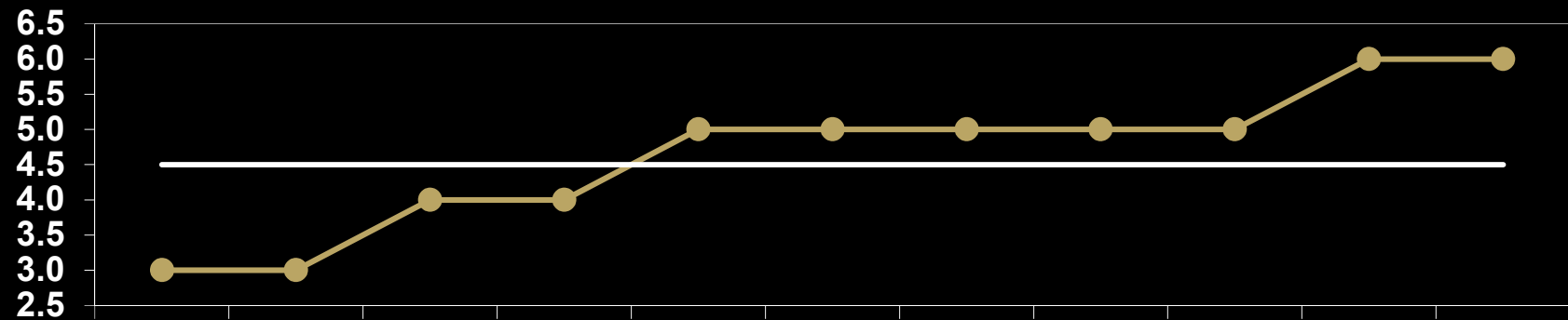Kurtosis is concerned with the peakedness of the distribution.
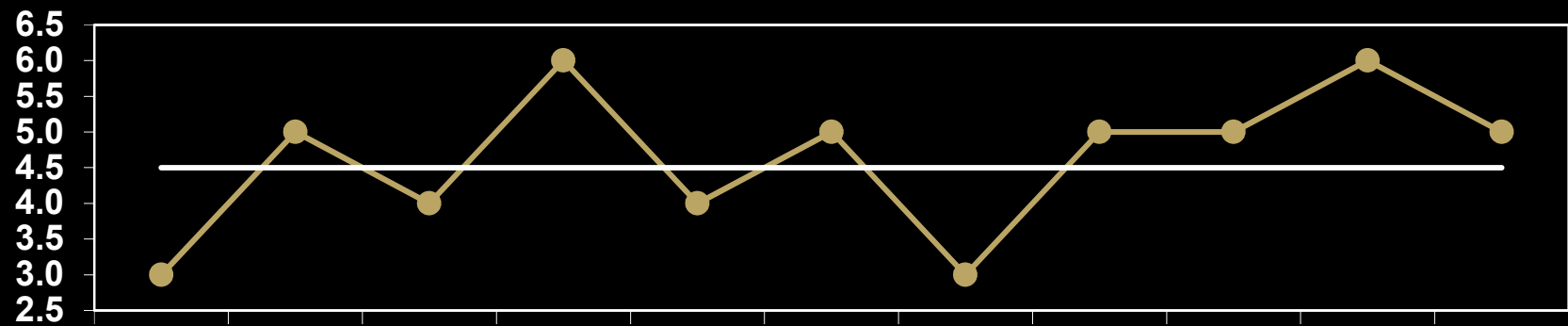
Kurtosis

Time Sequence

Time ordered data indicates the stability of the process through time.
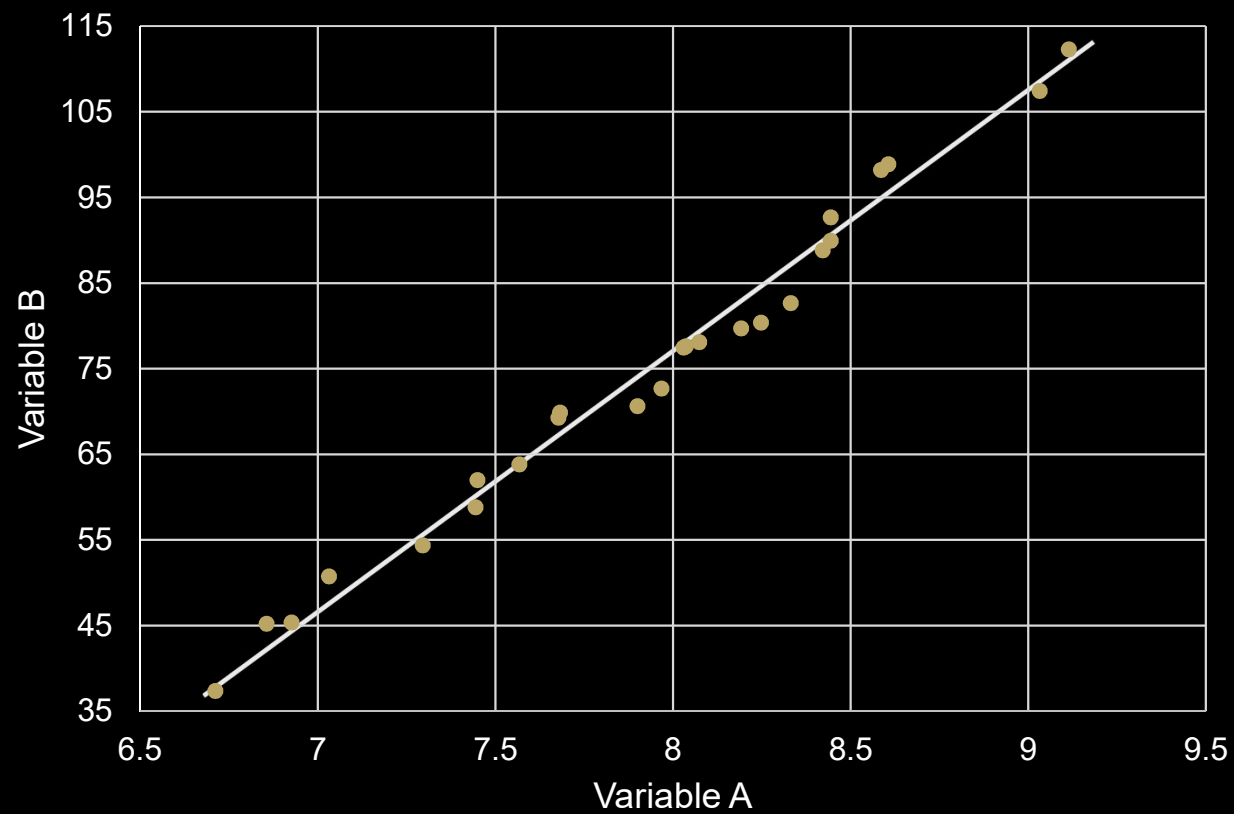
Time Sequence

Measures of Relationship

Measures of relationship quantify the "strength" of the relationship between two variables.

- Correlation
- Association

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Sampling Concepts

**Data Science for Quality Management:
Data and Measurement**
with **Wendy Martin**

Learning objectives:

Recall the approaches to sampling
Discern between population and sample
statistics

Populations and Samples

Population (Target Population)

- The entire group of objects, all with one characteristic of interest in common, and about which we want to make decisions
- Infinite, or finite but relatively huge

Populations and Samples

Research Population

- That portion of the Target Population available for sampling

Populations and Samples

Sample

- A subgroup of the population of interest, usually selected randomly.

- Random sampling is a prerequisite to using any type of inferential statistics!

General Approaches to Sampling

- Nonrandom or judgment sampling
- Random or probability sampling
    - Simple random Sampling
    - Systematic random sampling
    - Stratified random sampling
    - Cluster sampling

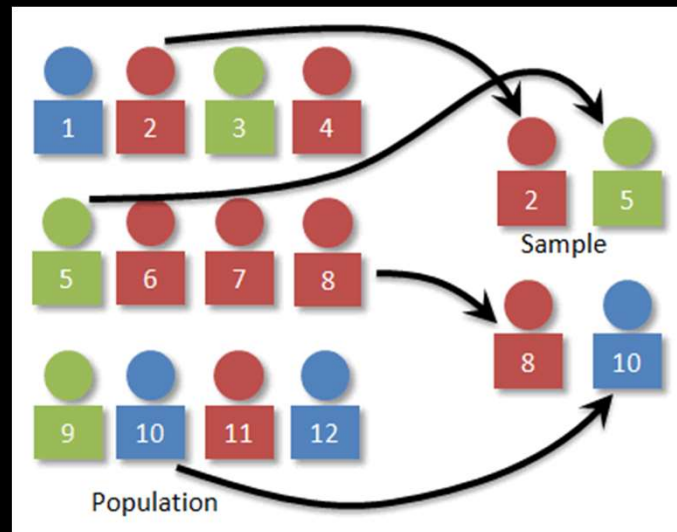# Nonrandom or Judgment Sampling

- Specimens or items are selected using personal judgment, reasoning, opinion, or convenience

Random or probability sampling

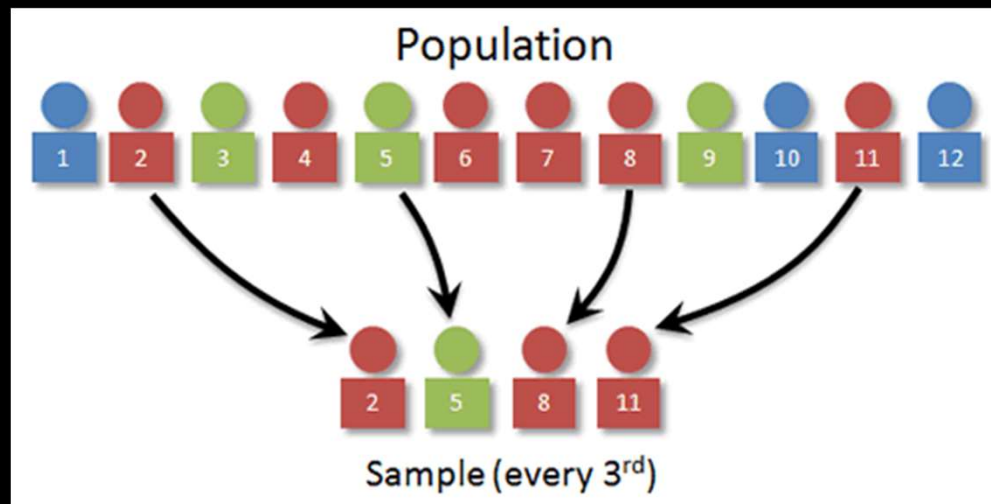- All specimens or items have a probability of being included in the sample

Simple Random Sampling

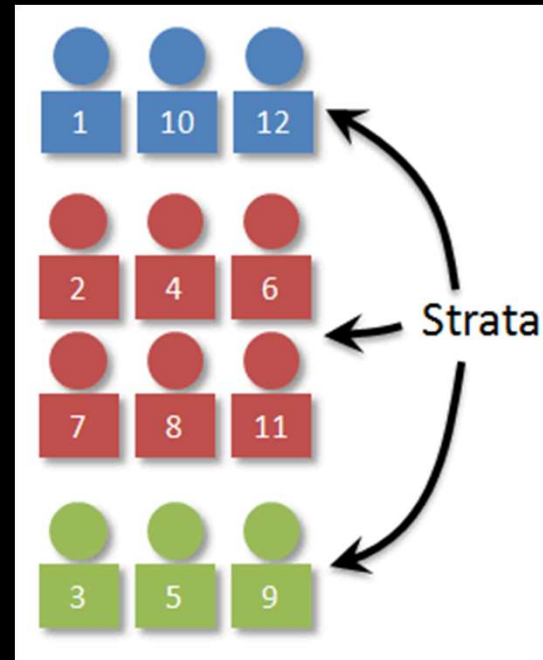- Every possible sample of size n has an equal chance of being selected

Systematic Random Sampling

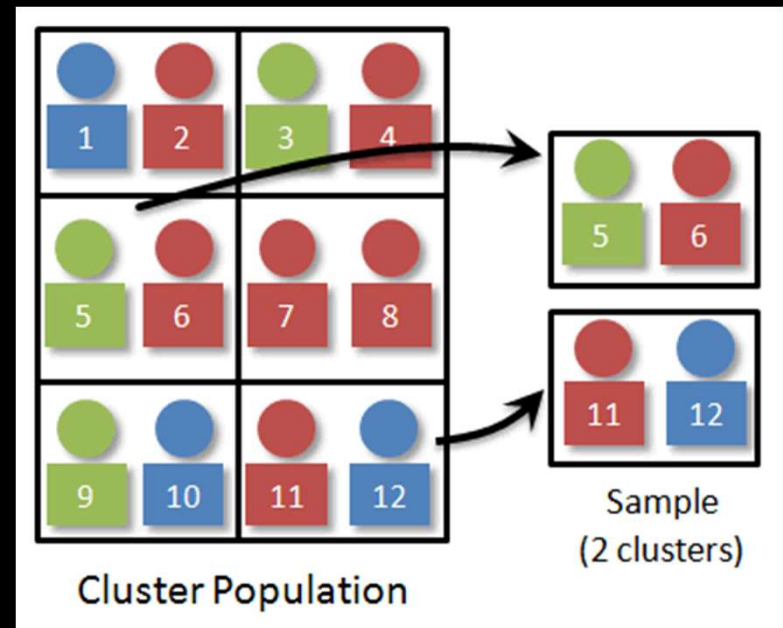- Specimens or items are selected at an interval

Stratified Random Sampling

- Specimens or items are divided into homogenous subsets, or strata

Cluster Sampling

- Specimens or items are divided into groups that are homogenous between each other, but heterogeneous within

Cluster Population

Sample
(2 clusters)

Statistics and Variation

A statistic is a measure calculated from sample data that may be used to make inferences about a population

- The average is a "statistic"
- The range is another "statistic"
- There are many more...

Statistics and Parameters

Descriptive Statistics

- Describe a characteristic of a sample
- Frequently used to make inferences about population parameters
- Represented by letters in English

Statistics and Parameters

Population parameters

- Describes a characteristic of the population
- Represented by Greek letters (with few exceptions)

Statistics and Parameters

Sample Statistics	Population Parameters	Description
$\bar{X}$	μ	Mean
$\tilde{X}$	M	Median
s	σ	Standard Deviation
s^2	σ^2	Variance
R	NT'	Range / Natural Tolerance
p	π	Count Per Unit
g_3	γ_3	Skewness
g_4	γ_4	Kurtosis

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Working in RStudio

**Data Science for Quality Management:
Data and Measurement**
with **Wendy Martin**

Learning objectives:

Create and import a data file into RStudio
Access R packages
Work with data frames

Introduction to R and RStudio

- R is the programming language
- RStudio is the user interface

Step 1: Install R

- Download R at http://www.r-project.org
- Once downloaded, double click to install R

Step 2: Install RStudio

- Download RStudio at http://www.rstudio.com

- Once downloaded, double click to install RStudio

Step 3: Open RStudio

Step 4: Install lolcat

```
install.packages("devtools")
require(devtools)
install_github("burrm/lolcat")
```

Packages

Once you have installed a package, it needs to be loaded any time you open RStudio

- require( )
- library( )

R Syntax

- Case sensitive
- Assignment operator: <-

Import Data into RStudio

Import Data into RStudio

The datasets we will be using in our course will be imported as a dataframe, with different variable names.

To select a variable: Dataframe$variable