Lean Six Sigma Green Belt Certification Course

simplilearn

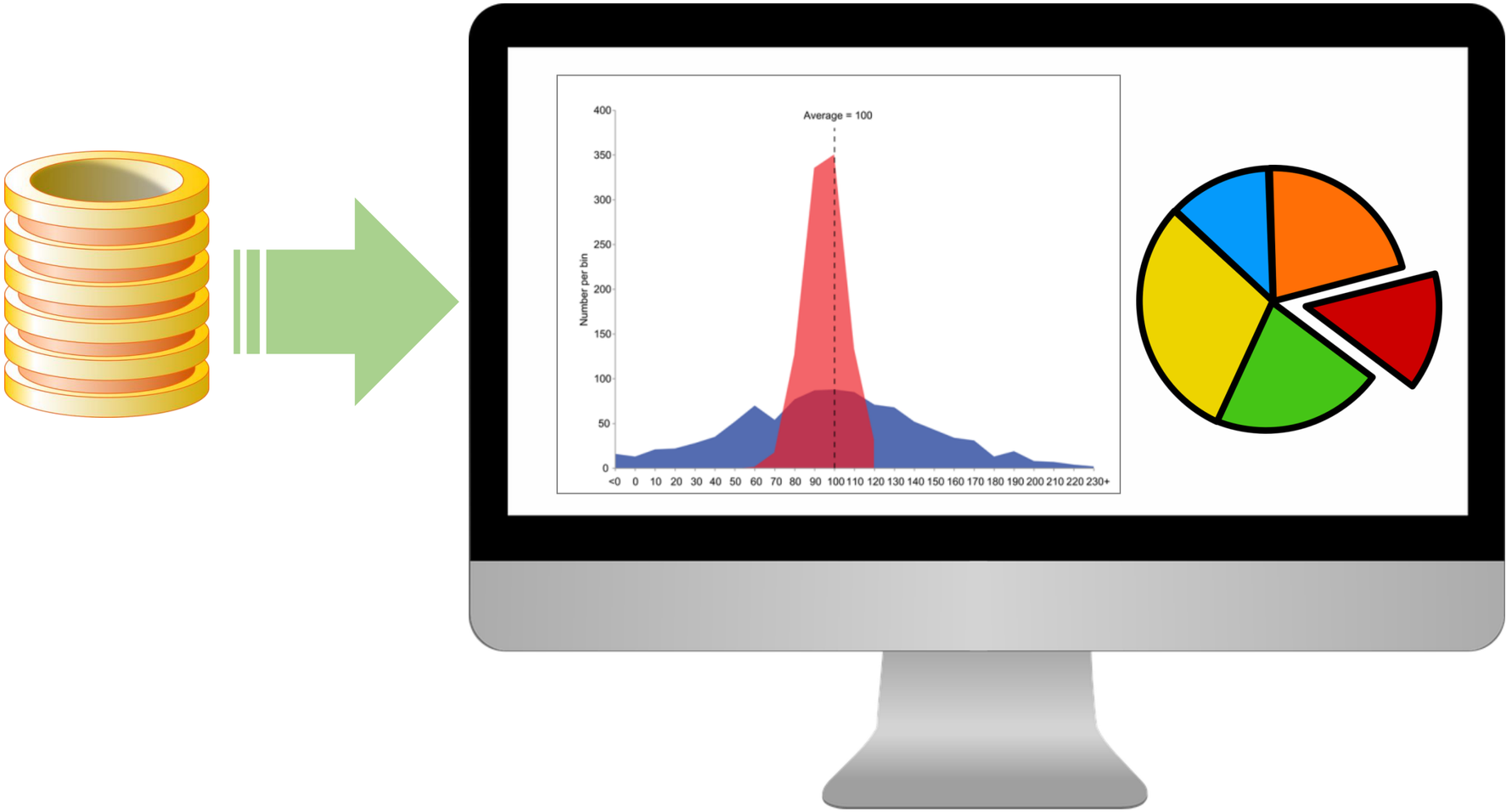
Collecting and Summarizing Data

Learning Objectives

By the end of this lesson, you will be able to:

- List the types of measuring scales
- Identify the types of data
- Identify the measures used to summarize data
- Describe the types of data collection methods
- Explain the concept of descriptive statistics
- Present the data visually using graphical methods

Introduction

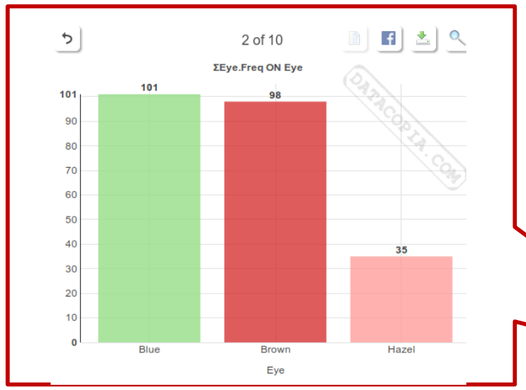


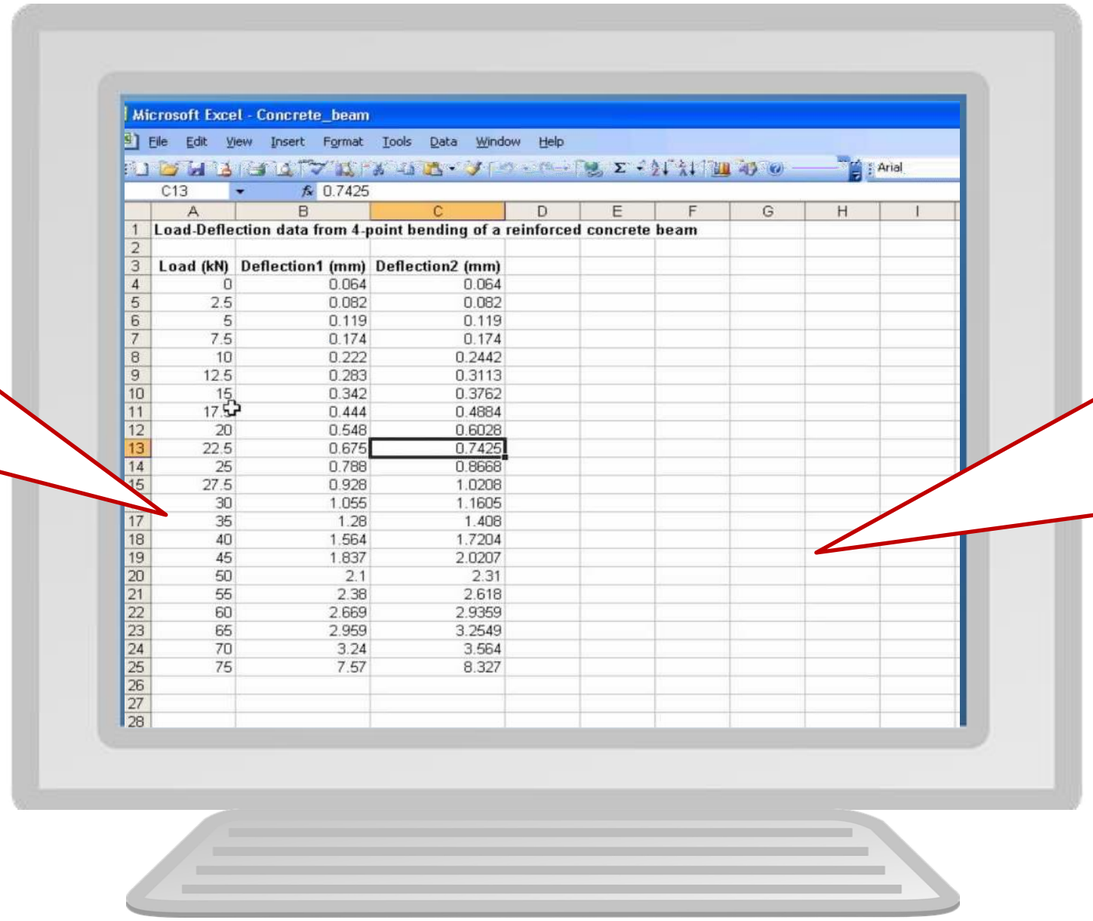
DIGITAL OPERATIONS

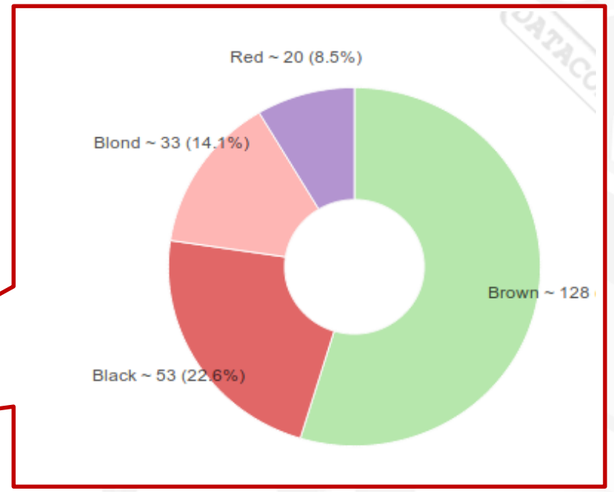
Collecting Data

Meaning of Data

Data is a collection of facts from which conclusions may be drawn.



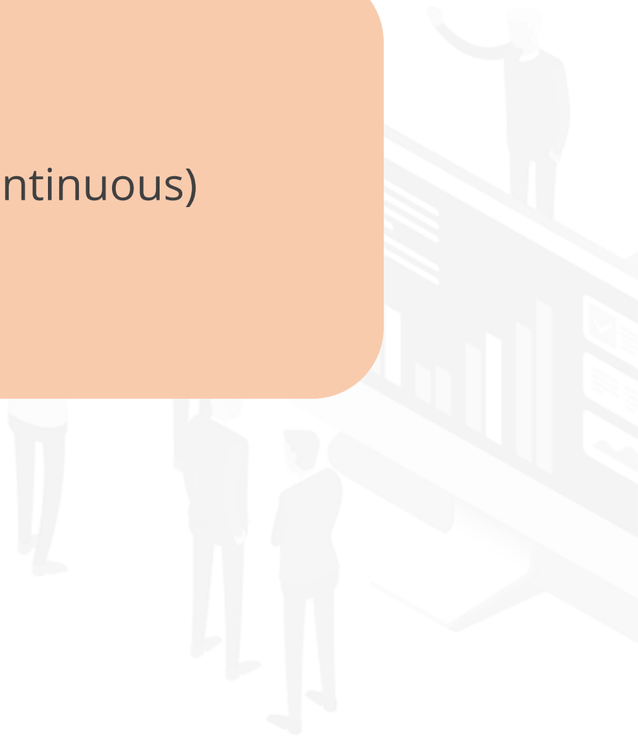




Types of Data

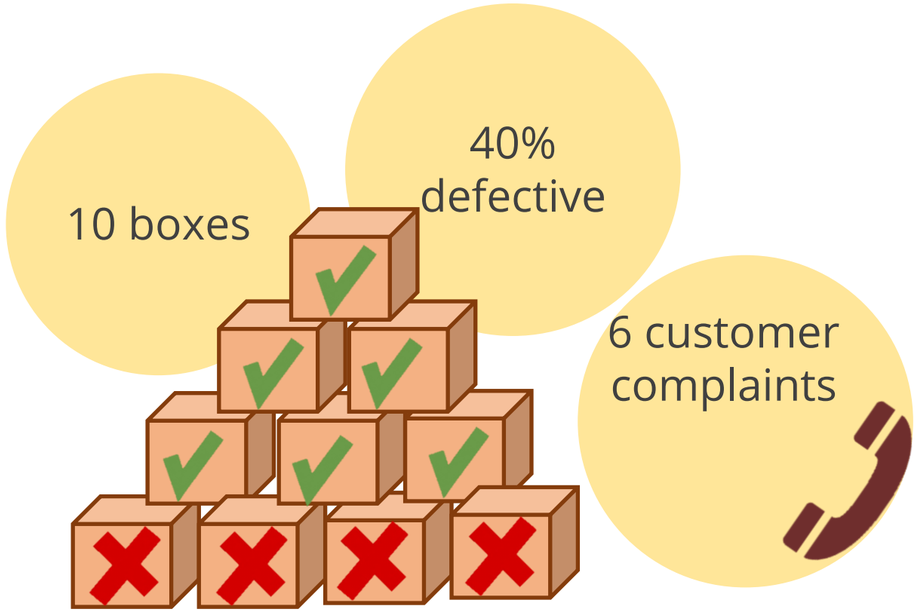
Attribute data (Discrete)

Variable data (Continuous)



Attribute Data (Discrete)

Attribute data is countable and only includes integers such as 5, 100, and 500.

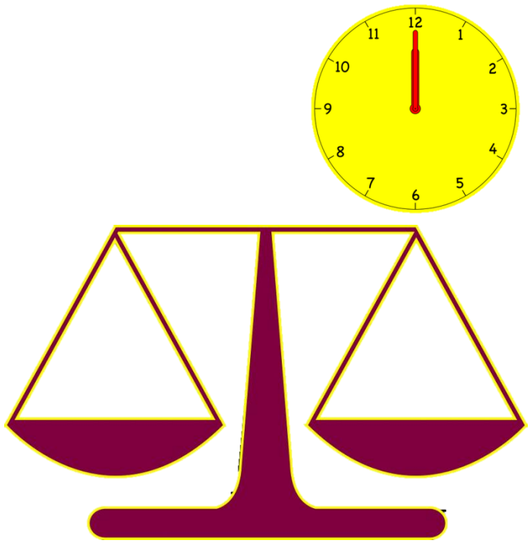


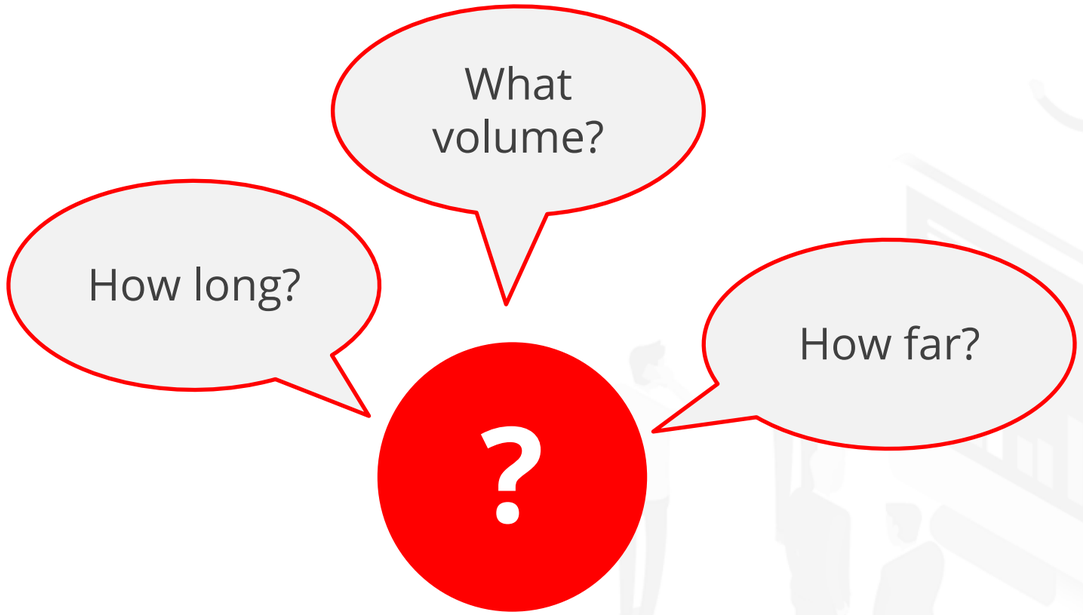
Examples

Number of defective products | Percentage of defective products | Frequency of customer complaints | Type of call received

Variable Data (Continuous)

Variable data can be measured on a continuous scale.



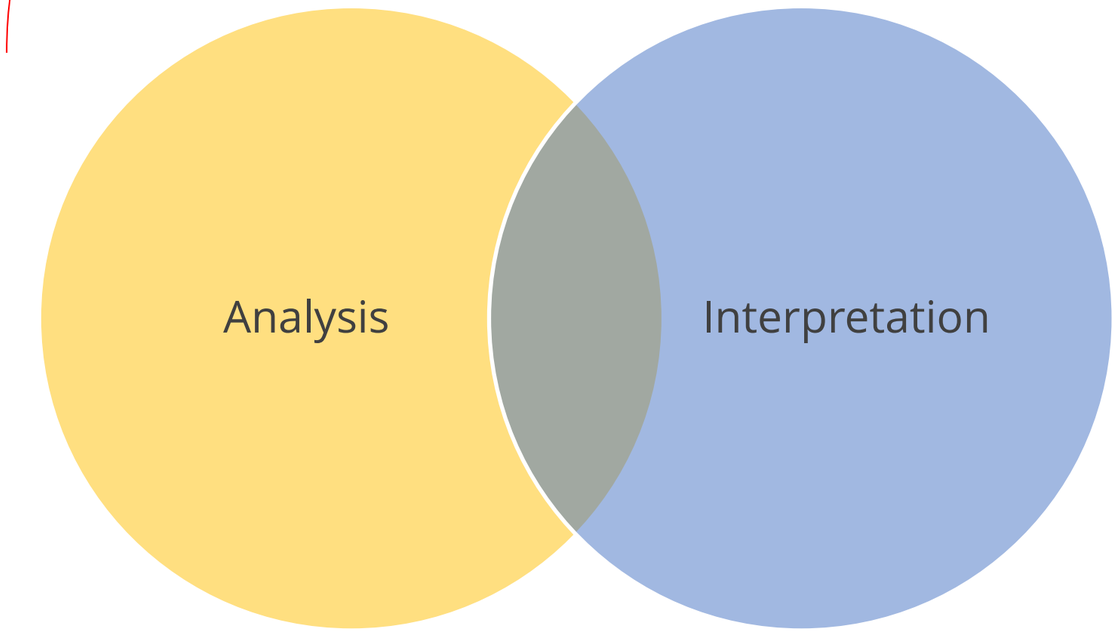


Examples

Height | Weight | Time taken to complete a task

Selecting Data Type

What type of data should be collected?



Selecting Data Type

- Critical to Quality (CTQs) parameters
- Key Process Output Variables (KPOVs)
- Key Process Input Variables (KPIVs)

What variables have been identified for the process?

What type of data is selected?

- Data that fits the metrics for the key variables

- To enable collecting, analyzing, and drawing inferences from the right set of data

Why should the data type be identified?

It may not be easy to convert attribute data to variable data in the absence of assumptions or additional information, which can include retesting all units.

Measurement Scales

Nominal

Ordinal
(ranking)

Interval

Ratio

Measurement Scales

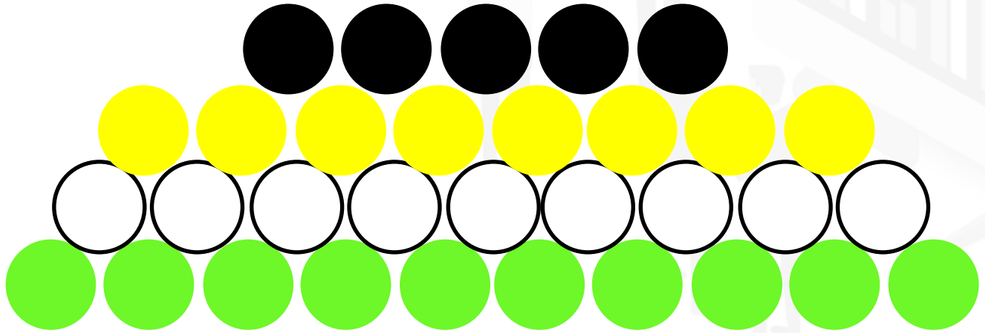
- Nominal
- Ordinal (ranking)
- Interval
- Ratio

Description

- Data consists of only names or categories and ordering is not possible.
- It is considered the least informative of all scales.
- The mode is used as the measure of central tendency.

Example

A bag of colored balls



Measurement Scales

Nominal

Ordinal (ranking)

Interval

Ratio

Description

- Data is arranged in order and values can be compared.
- Median or mode is the measure of central tendency.

Example

Restaurant ratings

Measurement Scales

Nominal

Ordinal (ranking)

Interval

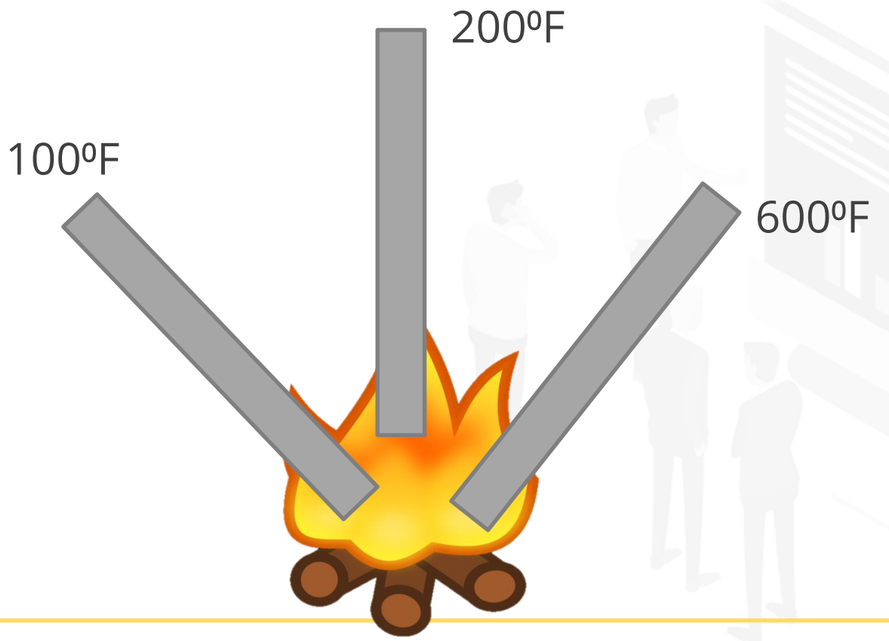
Ratio

Description

- This is used for ranking items in step order along a scale of equidistant points.
- Mean, median, or mode can be used for central tendency.

Example

Temperature of three metal rods



Measurement Scales

Nominal

Ordinal (ranking)

Interval

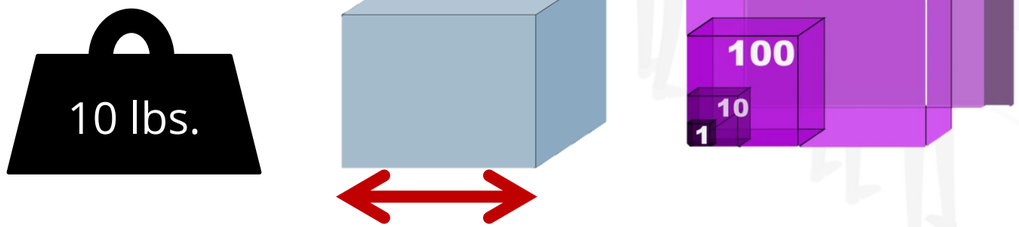
Ratio

Description

- This represents variable data and is measured against a known standard or increment. However, this scale also has an absolute zero (no numbers exist below zero).
- Median or mode, as well as arithmetic, geometric, or harmonic means can be used.

Examples

- Mass
- Length
- Volume



Data Collection Methods

Census

Sampling

Experiment

Observations

Data Collection Methods

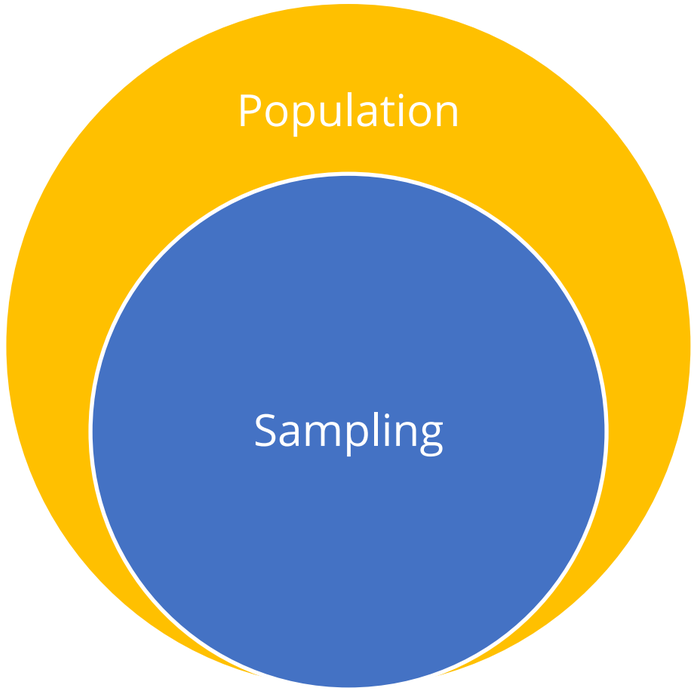
- Census
- Sampling
- Experiment
- Observations

A study that obtains data from every member of a population

Data Collection Methods

- Census
- **Sampling**
- Experiment
- Observations

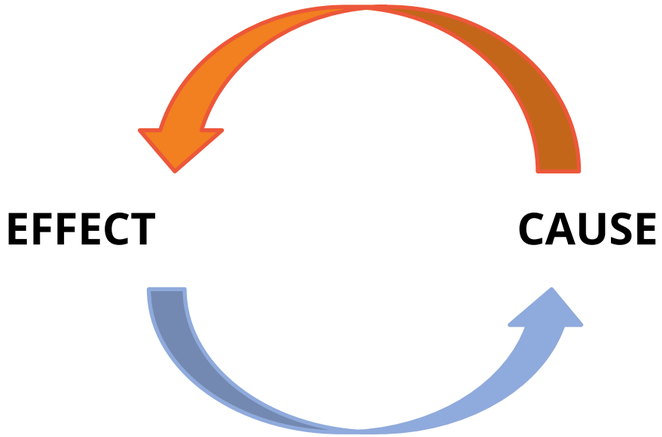
A study that obtains data from a subset of a population



Data Collection Methods

- Census
- Sampling
- Experiment
- Observations

A controlled study to understand cause and effect relationships



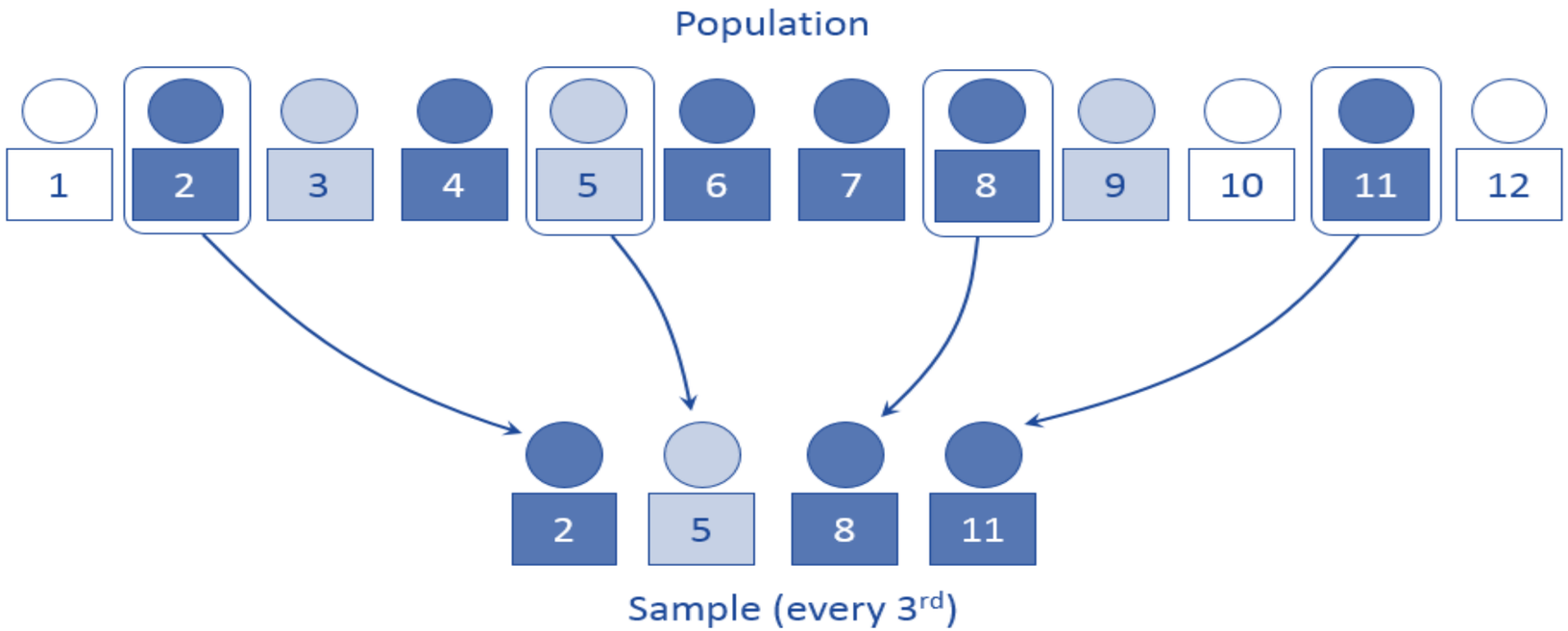
Data Collection Methods

- Census
- Sampling
- Experiment
- Observations

A study of cause and effect relationships without any controls

Meaning of Sampling

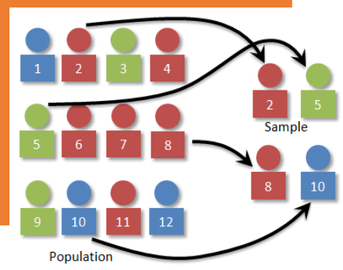
Sampling is the act, process, or technique of selecting an appropriate subset from a group or population.



Types Of Sampling

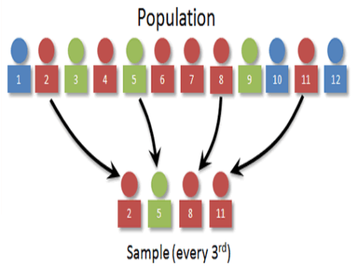
Sample group chosen from a population at random

Random sampling



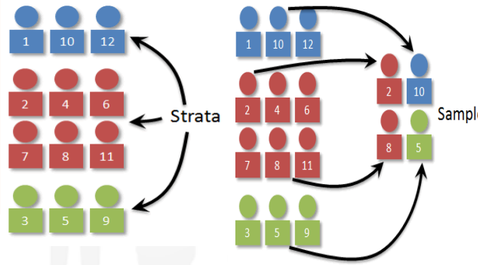
Non-probability technique where a single subject or group of subjects is selected for a period, analyzed, and followed by the next sample

Sequential sampling



Samples taken from each stratum or sub-group of a population

Stratified sampling



Simple Random Sampling vs. Stratified Sampling

Simple Random Sampling	vs.	Stratified Sampling
• Simple random sampling is easy to carry out.		• Stratified sampling is time consuming and requires more effort.
• Possibility of erroneous results is high.		• Possibility of errors is minimized.
• This type of sampling cannot indicate possible causes of variation.		• When done correctly, it is capable of showing causes of variation.

In practice, you will calculate means and standard deviations from the collected sample.

Data Collection Plan

Helps the team determine how process performance will be measured or quantified, which will help drive decision-making

Provides structure and consistency around the process of collecting data, which can be variable (continuous) or attribute (discrete)

Provides a common language for the team

Data Collection Plan: Example

Develop a data collection plan based on the process map and input priority matrix.

Measure	Type of Measure	Type of Data	Operational Definition	Specification	Target	Data Collection Form	Sampling	Baseline Sigma
Input, output, and process measure the team will collect	Select as an input (I), Process measure (P) or output (O)	Select if measure is continuous or discrete	Description to enable a common understanding and prevention of ambiguity	The least acceptable product or service performance to the customer	Customer's detail performance against some requirement	How to collect the data	The method to obtain information from the entire population	DPO or DPMO
Example: Speed	O, P	Continuous	From end of call to knock on door	30 minutes	As soon as possible	Histogram	Representative	Unit: Meal Defect: Late Oppty: 3
Process time	P	Continuous	Time to process report once database is populated	12 minutes	8 minutes	Customized checklist that contains: Date, region, number of reports, process time and approval time	100% daily	Unit: Report Defect: Process time > 12 minutes Oppty: 1
Approval time	P	Continuous	Time to review and approve report	5 minutes	3 minutes	Customized checklist that contains: Date, region, number of reports, process time and approval time	100% daily	Unit: Report Defect: Approval time > 5 minutes Oppty: 1

Source: *http://asq.org/service/body-of-knowledge/tools-data-collection-plan*

Data Collection Methods: Check Sheets

A check sheet is a structured, prepared form for collecting and analyzing data.

Day	Absences	Total
Monday	𝍸 𝍸 𝍸 𝍸 𝍸 ||	27
Tuesday	𝍸 𝍸 ||	12
Wednesday	𝍸 |||	8
Thursday	𝍸 𝍸 |||	13
Friday	𝍸 𝍸	10

Use of Check Sheets

- Data can be observed and collected repeatedly by the same individual or from the same location.
- Data is collected from a production process.

Paint Job Quality Control Checklist

Job: 629555
Inspector: Al Kyder

Problem	Frequency
Clip	
Bubble	
Run	
Scrape or scratch	
Inadequate coverage	
Other	

Data Coding

Data coding is the process of converting and condensing raw data into categories and sets so that they can be used for further analysis.

Male	☐ 1	Ill	☐ 1
Female	☐ 2	Not ill	☐ 0
Don't know	☐ 3	Don't know	☐ 9
Single	☐ 1	Separated	☐ 3
Married	☐ 2	Divorced	☐ 4
Widowed	☐ 5	Don't know	☐ 9

Data Coding: Benefits

Data coding

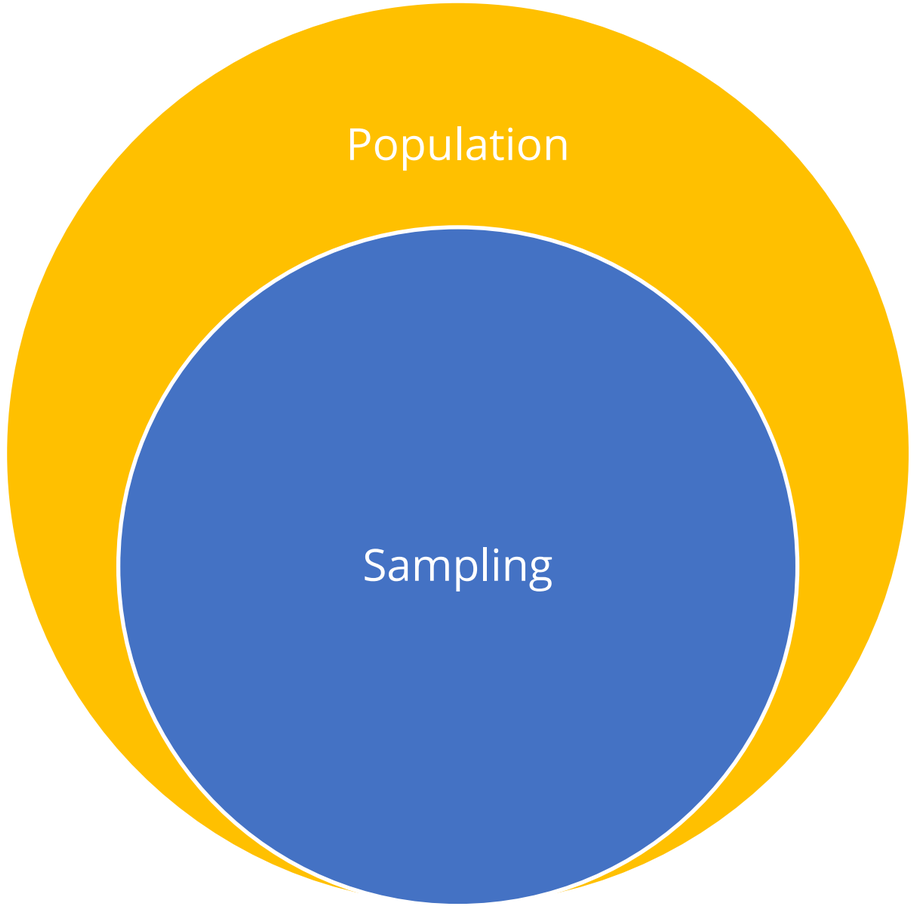
- Simplifies large quantities of data
- Enables easy entry of process data into computers
- Facilitates data analysis
- Enables organized data representation
- Eliminates data repetition

Population vs. Sample

Greek letters refer to values that represent a population's characteristic, also known as a parameter.

$\mu = population\ mean$

$\sigma = population\ standard\ deviation$



Roman letters represent values for a sample's characteristic, also known as a statistic

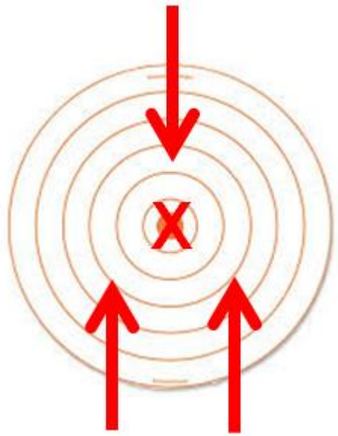
$\bar{X}$ = sample mean

s or $\hat{\sigma}$ = sample standard deviation

DIGITAL OPERATIONS

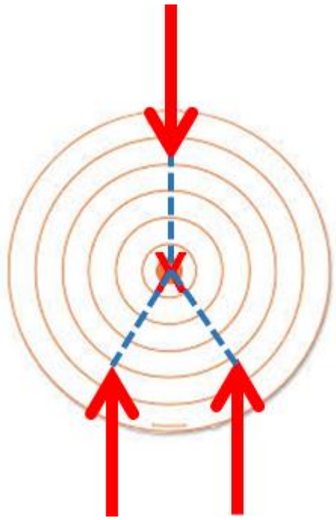
Summarizing Data

Descriptive Statistics: Central Tendency and Dispersion



A measure of central tendency is a single value that indicates the central point in a set of data.

Measure of central tendency



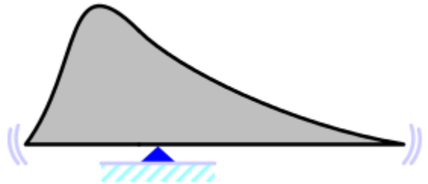
Measures of dispersion describe the spread of values.

Higher the variation of data points, the higher the spread of the data.

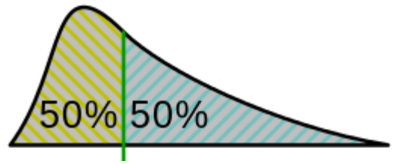
Measures of dispersion

Measures of Central Tendency

The three most common measures of central tendency:



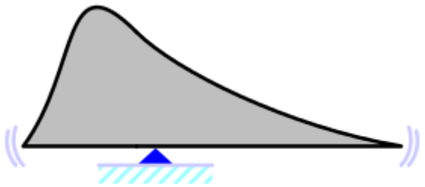
Mean



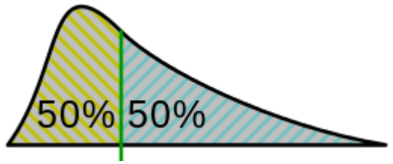
Median

Mode

Measures of Central Tendency



Mean



Median

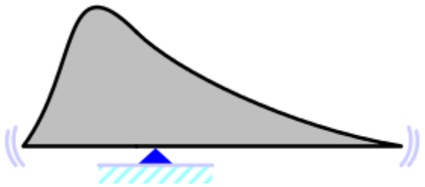
Mode

- Most common measure of central tendency
- Given by the sum of entries in a dataset and divided by the number of entries
- Also called the average or arithmetic mean

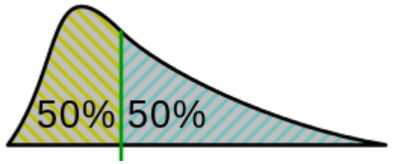
Formula:

$$\text{Mean} = \sum_{1}^{n} \frac{x_i}{n}$$

Measures of Central Tendency



Mean



Median

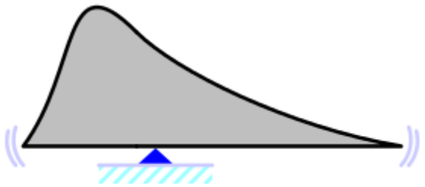
Mode

- Also known as positional mean
- Number in the middle of the dataset
- Mean of the middle two numbers in an even dataset
- Calculated using the following formula for odd dataset

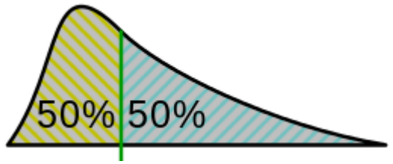
Formula:

$$\text{Median} = \frac{n+1}{2}$$

Measures of Central Tendency



Mean



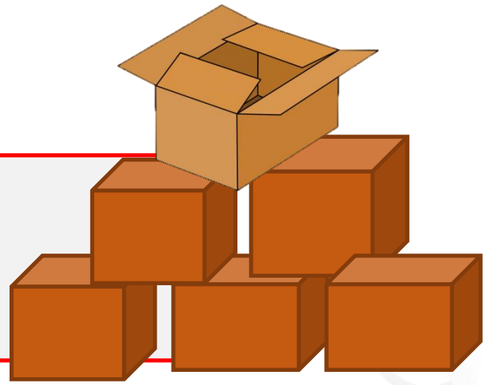
Median

Mode

- Also known as frequency mean
- Value that occurs most frequently in a dataset

Measures of Central Tendency: Example

The following is a sample dataset *for the thickness of boxes.*
Calculate the mean, median, and mode.



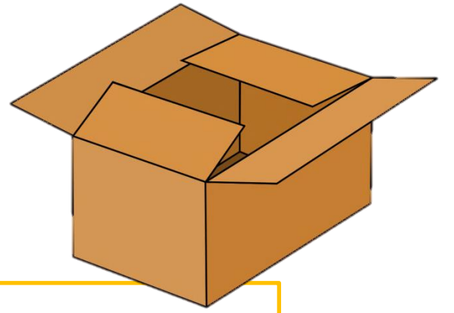
4, 8, 1, 6, 6, 2, 9, 3, 6, 7 → Sorted → 1, 2, 3, 4, 6, 6, 6, 7, 8, 9

Mean = $\frac{1+2+3+4+6+6+6+7+8+9}{10} = \frac{52}{10} = \mathbf{5.2}$ ✓

Median = **6** ✓

Mode = **6** ✓

Impact of Outliers On Mean



1, 2, 3, 4, 6, 6, 6, 7, 8, 9

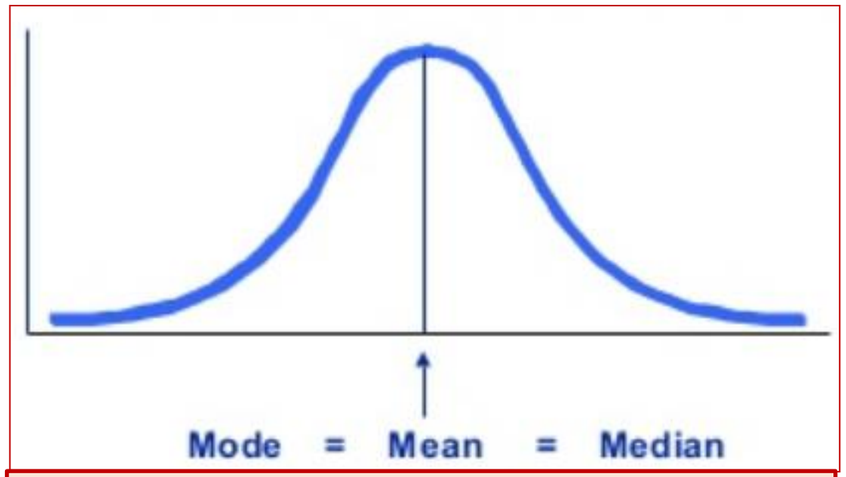
Changed

1, 2, 3, 4, 6, 6, 6, 7, 8, 100

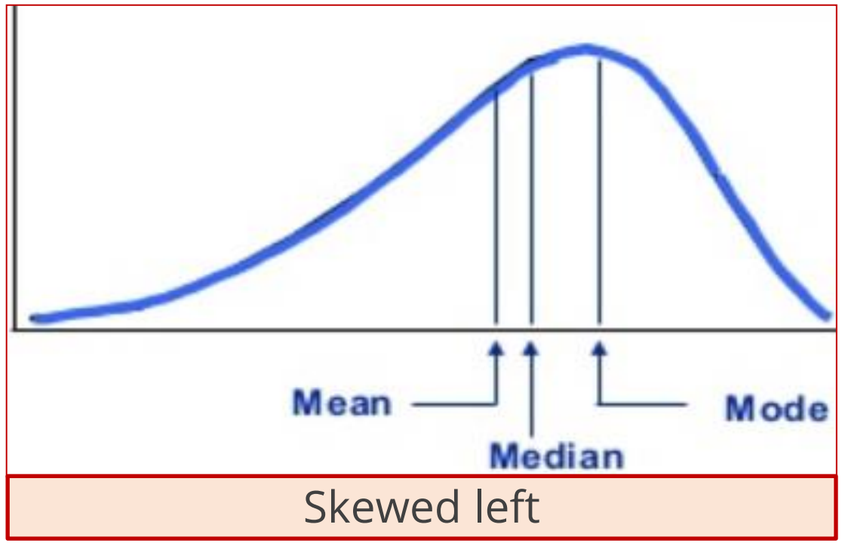
$$\text{Mean} = \frac{1 + 2 + 3 + 4 + 6 + 6 + 6 + 7 + 8 + 100}{10} = \mathbf{14.3}$$

When the dataset has outliers, median is preferred over mean as a measure of central tendency.

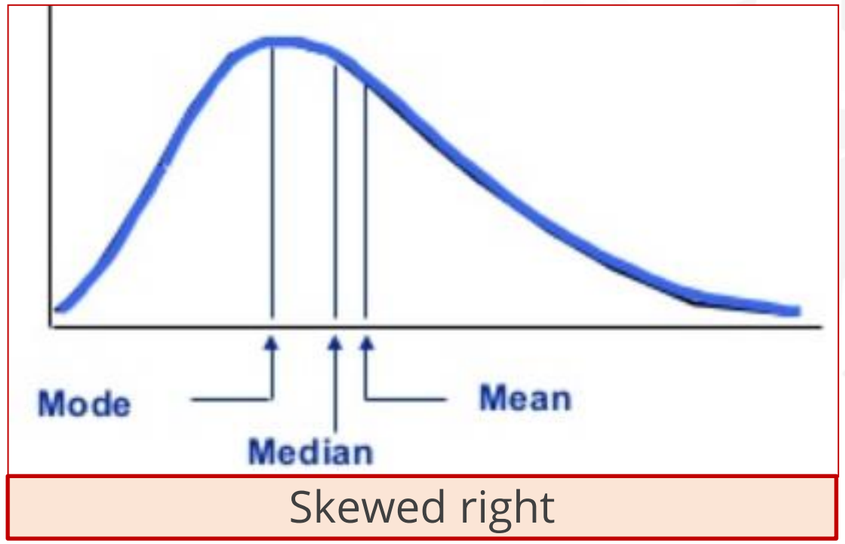
Illustration based on Distributions



Symmetric

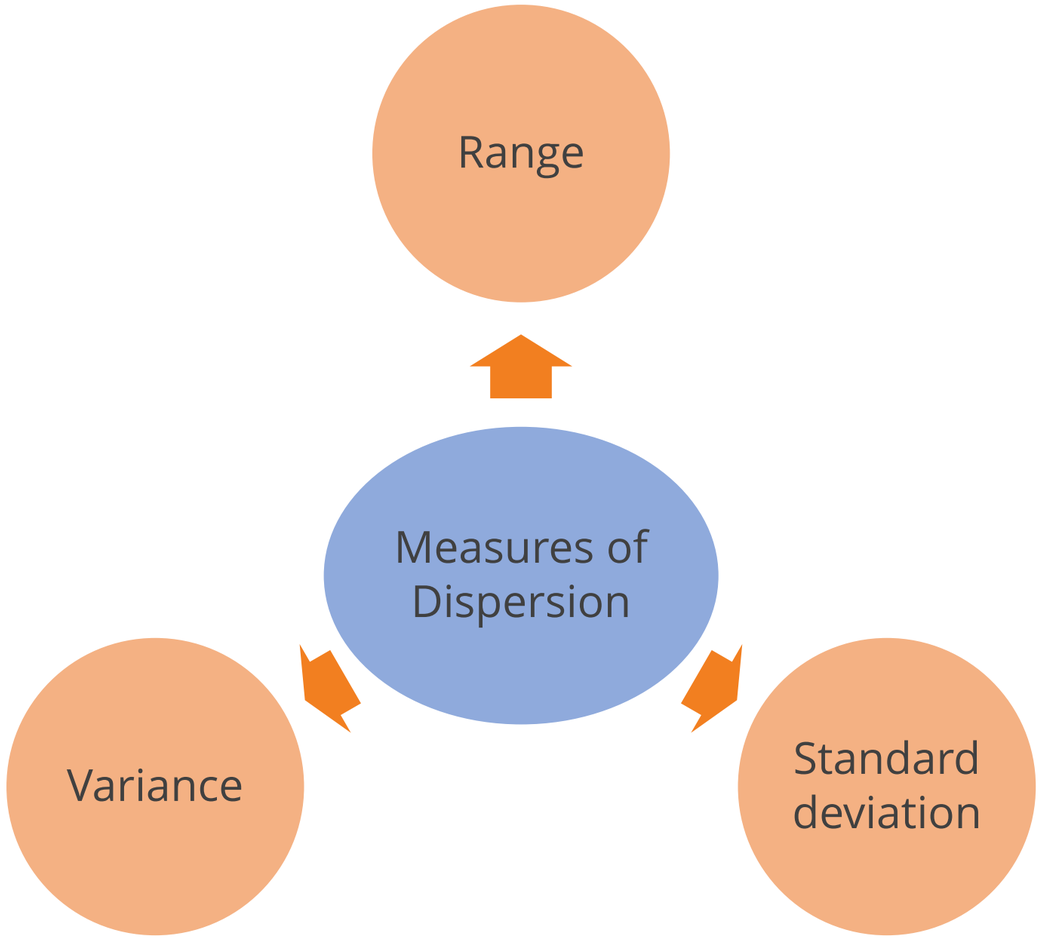


Skewed left



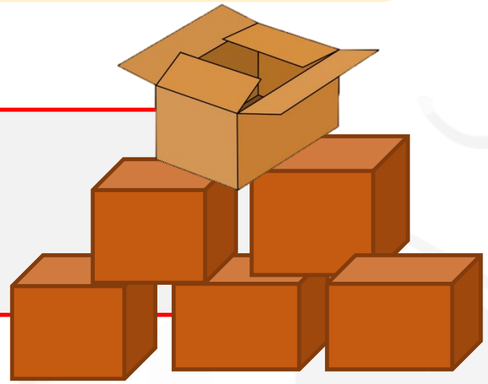
Skewed right

Measures of Dispersion



Range

The range of a particular set of data is the difference between the largest value and the smallest value in the data.



The following is a sample dataset *for the thickness of boxes*.
Calculate the range.

1, 2, 3, 4, 6, 6, 6, 7, 8, **9**

Range = Maximum – Minimum

= 9 – 1

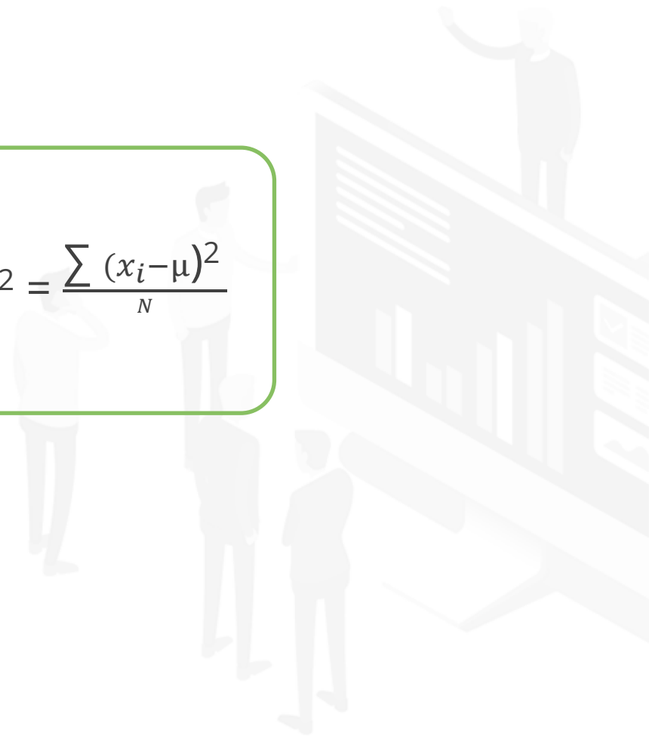
= 8 ✓

Variance

Variance is the average of squared mean differences and shows the variation in a dataset.

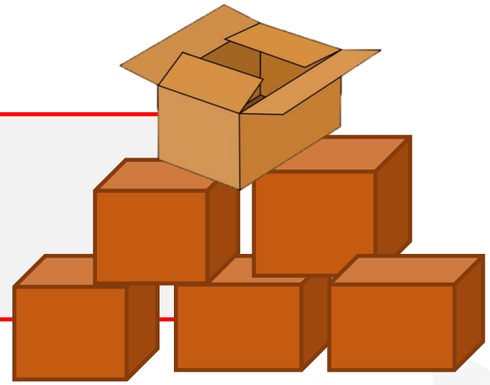
Sample Variance = $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

Population Variance = $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$



Variance: Example

The following is a sample dataset *for the thickness of boxes.*
Calculate the variance.



1, 2, 3, 4, 6, 6, 6, 7, 8, 9

$$\frac{(1-5.2)^2+(2-5.2)^2+(3-5.2)^2+(4-5.2)^2+(6-5.2)^2+(6-5.2)^2+(6-5.2)^2+(7-5.2)^2+(8-5.2)^2+(9-5.2)^2}{10-1}$$

61.6/(10-1)

= **6.84 cm^2**

Standard Deviation

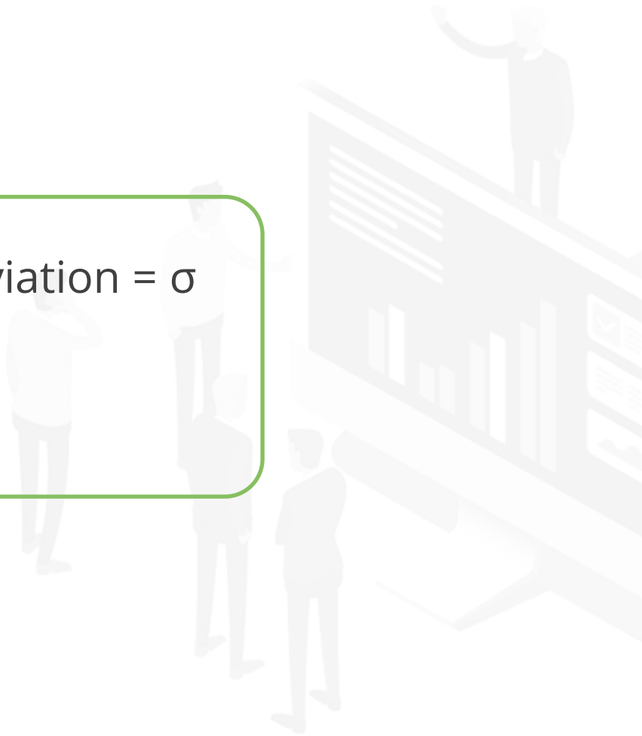
The range of a particular set of data is the difference between the largest and the smallest values in the data.

Sample Standard Deviation = s

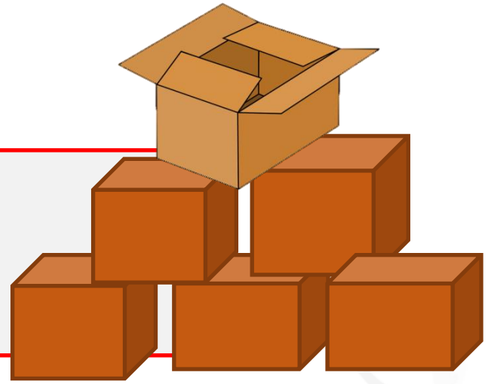
$$= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Populations Standard Deviation = σ

$$= \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$



Standard Deviation: Example



The following is a sample dataset *for the thickness of boxes*.
Calculate the standard deviation.

1, 2, 3, 4, 6, 6, 6, 7, 8, 9

Variance = 6.84

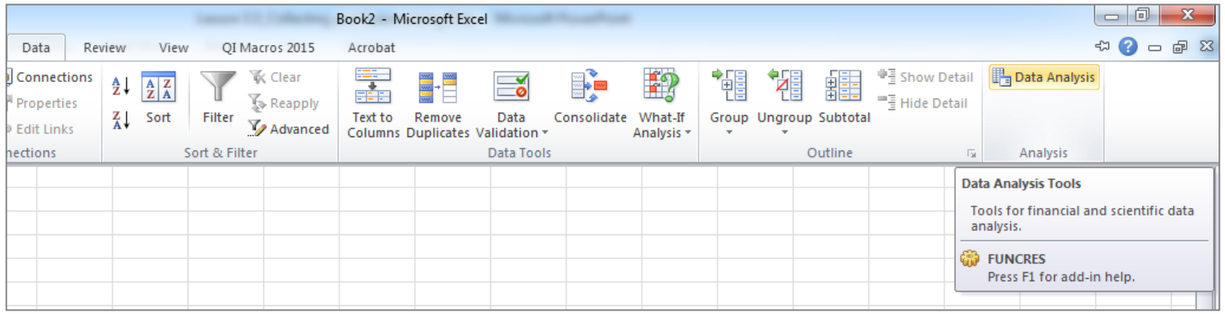
Standard deviation = $\sqrt{6.84} = \mathbf{2.62}$

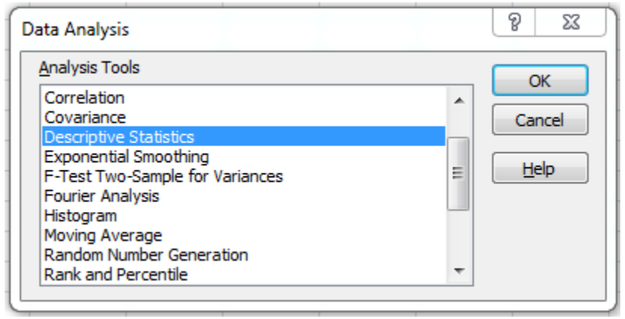
Microsoft Excel for Descriptive Statistics

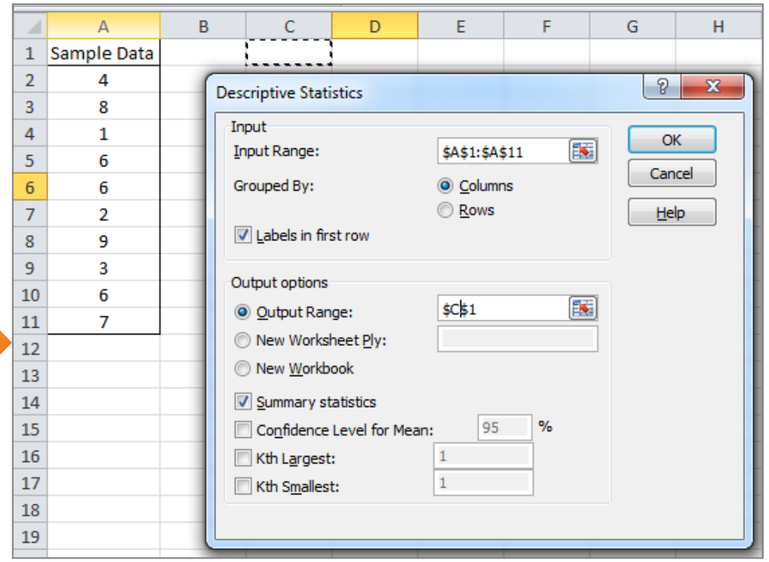
	A	B	C	D
1	Sample Data	Descriptive Stat	Excel Function	Results
2	4	Mean	=AVERAGE(A2:A11)	5.2
3	8	Median	=MEDIAN(A2:A11)	6
4	1	Mode	=MODE(A2:A11)	6
5	6	Range	=MAX(A2:A11)-MIN(A2:A11)	8
6	6	Sample Variance	=VAR.S(A2:A11)	6.84444444444445
7	2	Standard Deviation	=STDEV.S(A2:A11)	2.61618891604648
8	9			
9	3	Population Variance	=VAR.P(A2:A11)	6.16
10	6	Pop. Std Dev.	=STDEV.P(A2:A11)	2.48193472919817
11	7			

Microsoft Excel for Descriptive Statistics

Steps to use Data Analysis Tool in Microsoft Excel







C	D
Sample Data	
Mean	5.2
Standard Error	0.827312
Median	6
Mode	6
Standard Deviation	2.616189
Sample Variance	6.844444
Kurtosis	-0.9698
Skewness	-0.27178
Range	8
Minimum	1
Maximum	9
Sum	52
Count	10

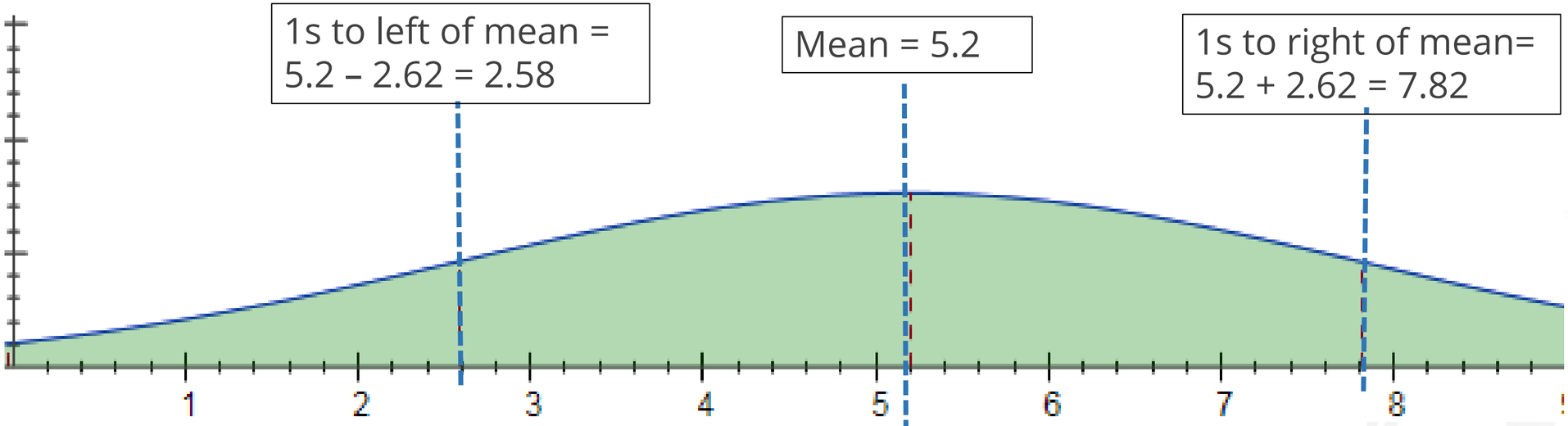
Normal Curve: Derivation from Mean and Standard Deviation

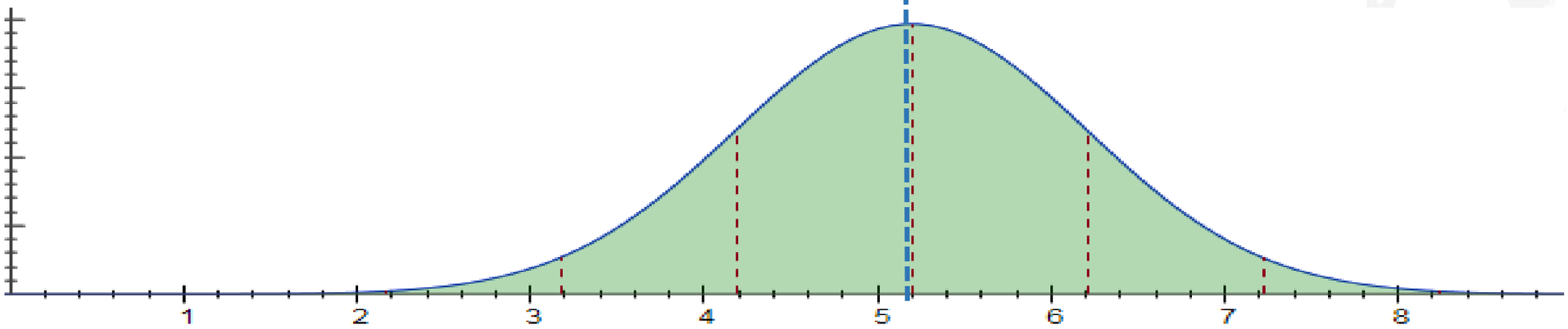
1, 2, 3, 4, 6, 6, 6, 7, 8, 9

1s to left of mean =
5.2 – 2.62 = 2.58

Mean = 5.2

1s to right of mean=
5.2 + 2.62 = 7.82





Frequency Distribution

Frequency distribution is the grouping of data into mutually exclusive categories, showing the number of observations in each class.

Number	Tally	Frequency
0	IIII	4
1	𝍸 I	6
2	𝍸	5
3	III	3
4	II	2

Divide the results into **intervals** and count the number of results in each interval.

Make a **table with separate columns** for the interval numbers, the tallied results, and the frequency of results in each interval.

Record the number of observations in each interval with a tally mark.

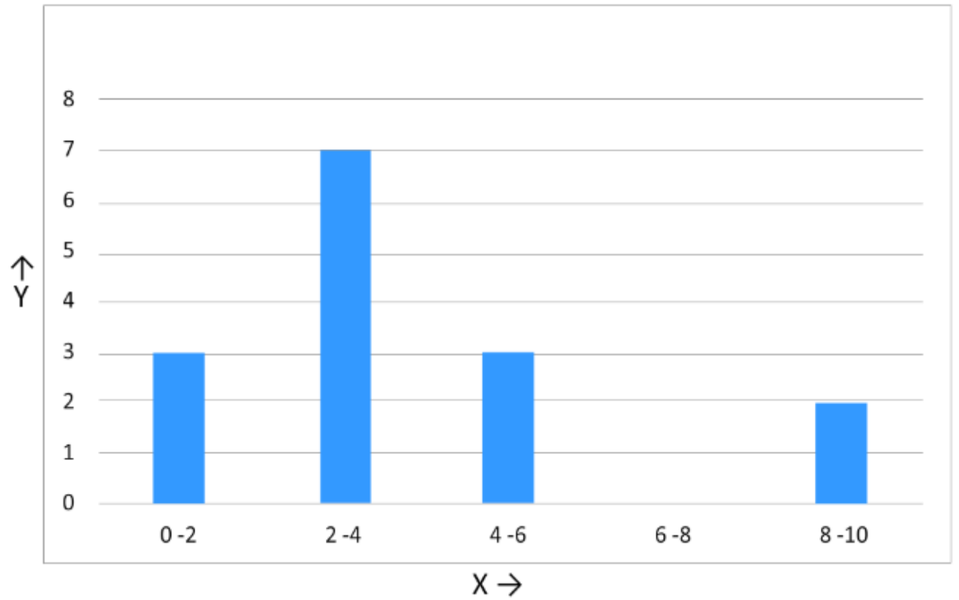
Add the number of tally marks in each interval and record them in the Frequency column.

Histogram

A histogram is similar to a bar graph, except that the data in a histogram is grouped into intervals.

1.5, 1.5, 2, 3, 3, 3, 5, 4, 4, 4, 4.5, 5, 6, 9.5, 10

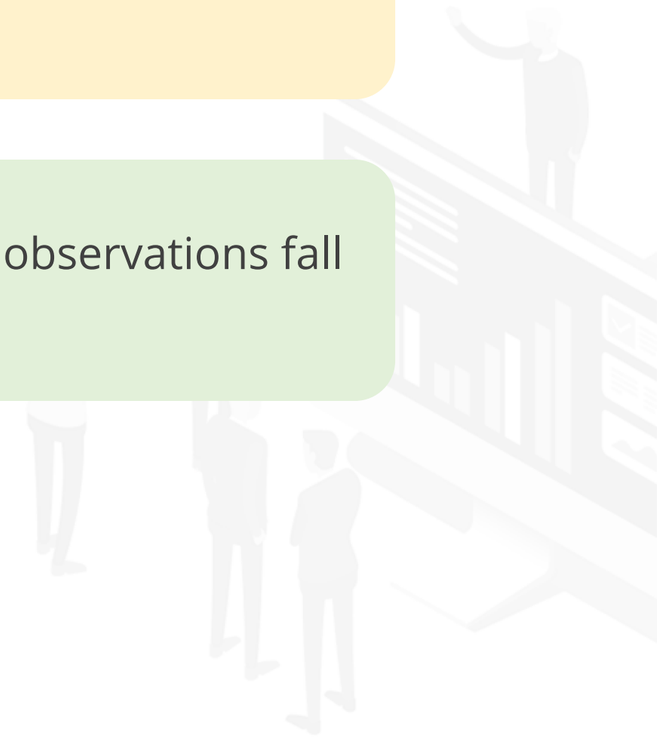
Minutes spent (X)	Number of orders (Frequency) (Y)
0 - 2	3
2 - 4	7
4 - 6	3
6 - 8	0
8 - 10	2



Cumulative Frequency Distribution

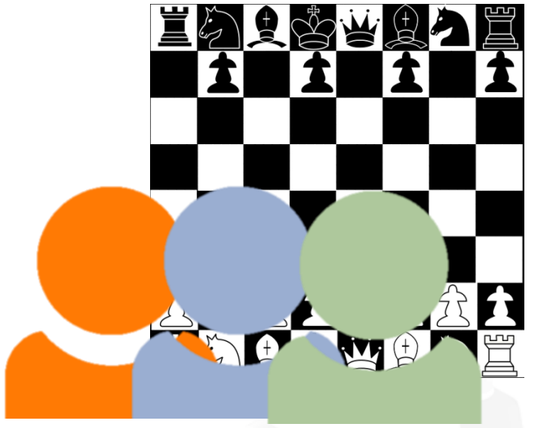
A cumulative frequency distribution is the sum of the category and all categories below it in a frequency distribution.

The main purpose is to study a population or sample to find out how many observations fall lower or higher than a particular category of interest.



Cumulative Frequency Distribution

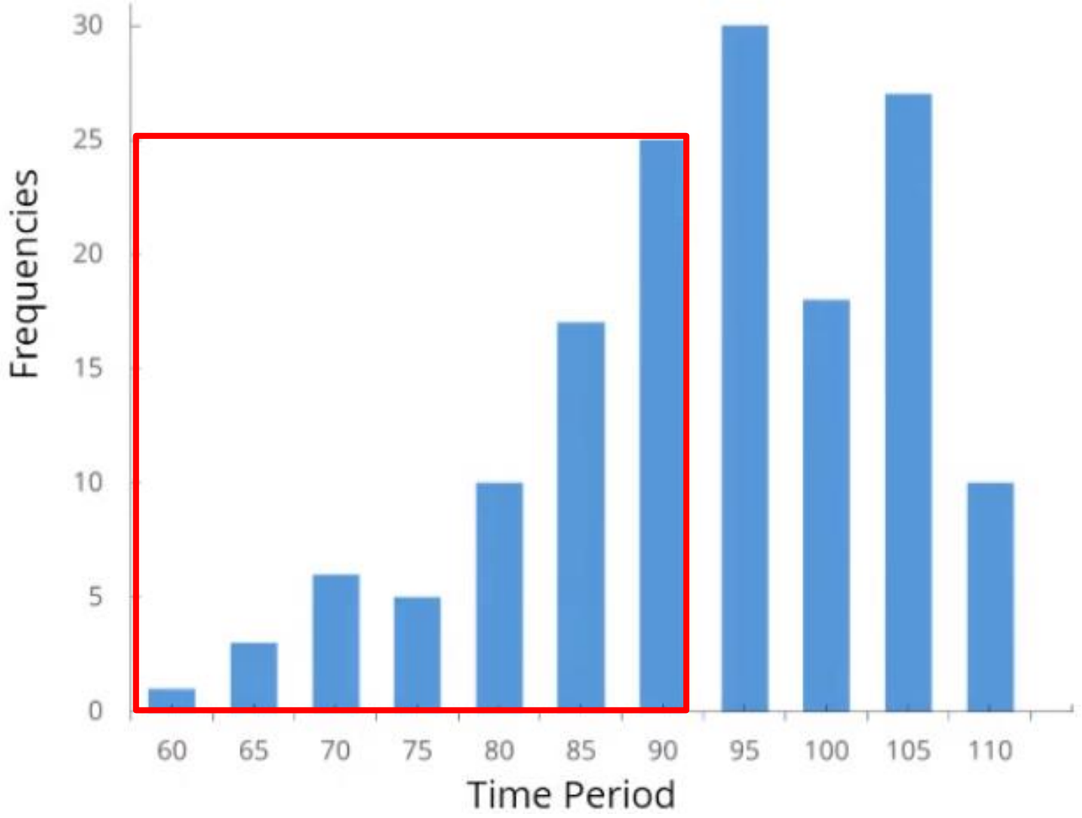
37, 49, 54, 91, 60, 62, 65, 77, 67, 81



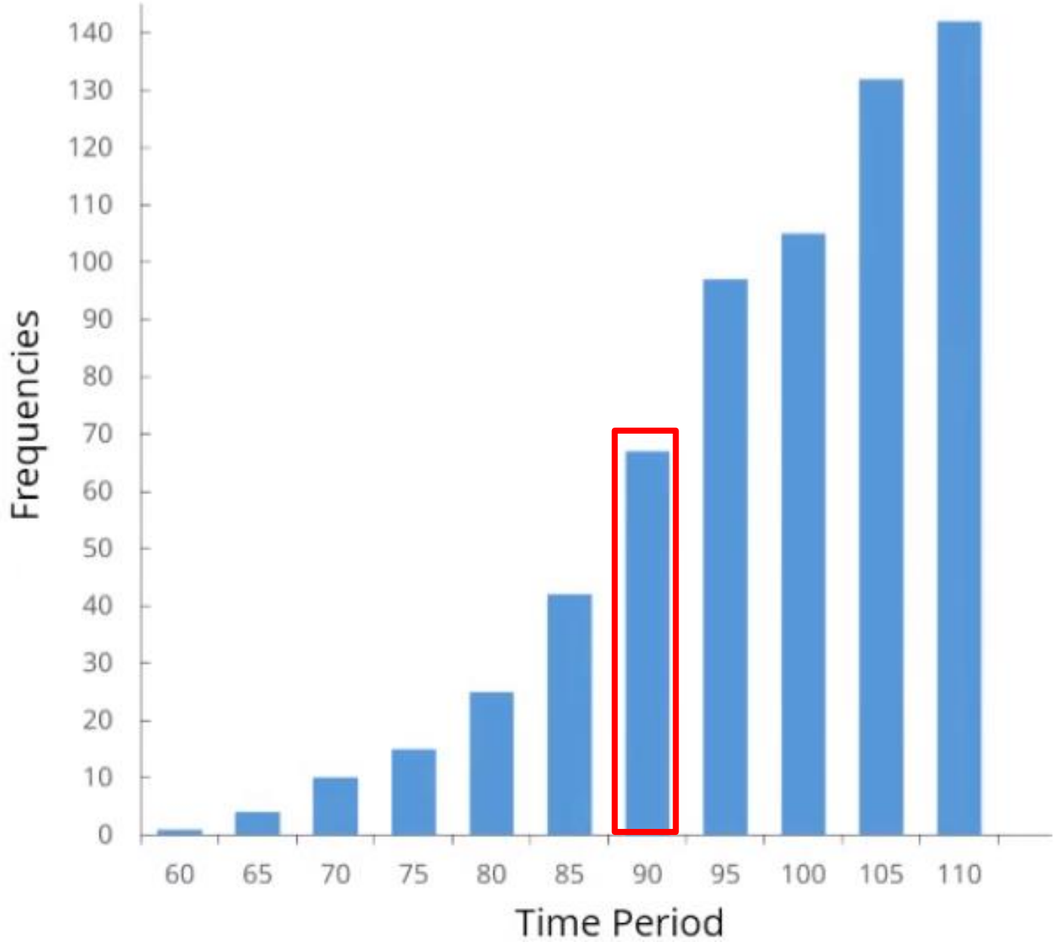
Lower Value	Upper Value	Frequency	Cumulative Frequency	Percentage	Cumulative Percentage
35	44	1	1	10	10
45	54	2	3	20	30
55	64	2	5	20	50
65	74	2	7	20	70
75	84	2	9	20	90
85	94	1	10	10	100

Frequency vs. Cumulative Frequency Distributions

Frequency Distribution Graph



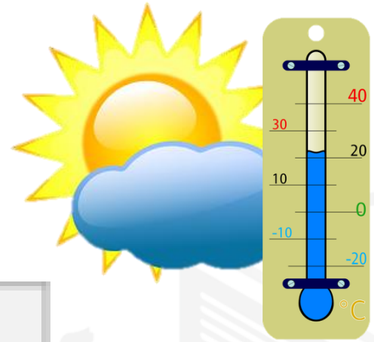
Cumulative Frequency Distribution Graph



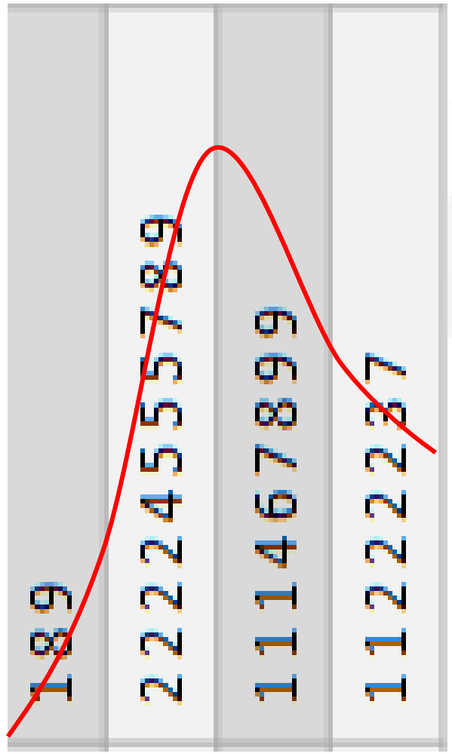
Stem and Leaf Plots

A stem and leaf plot is used to present data in a graphical format to enable visualizing the shape of a distribution.

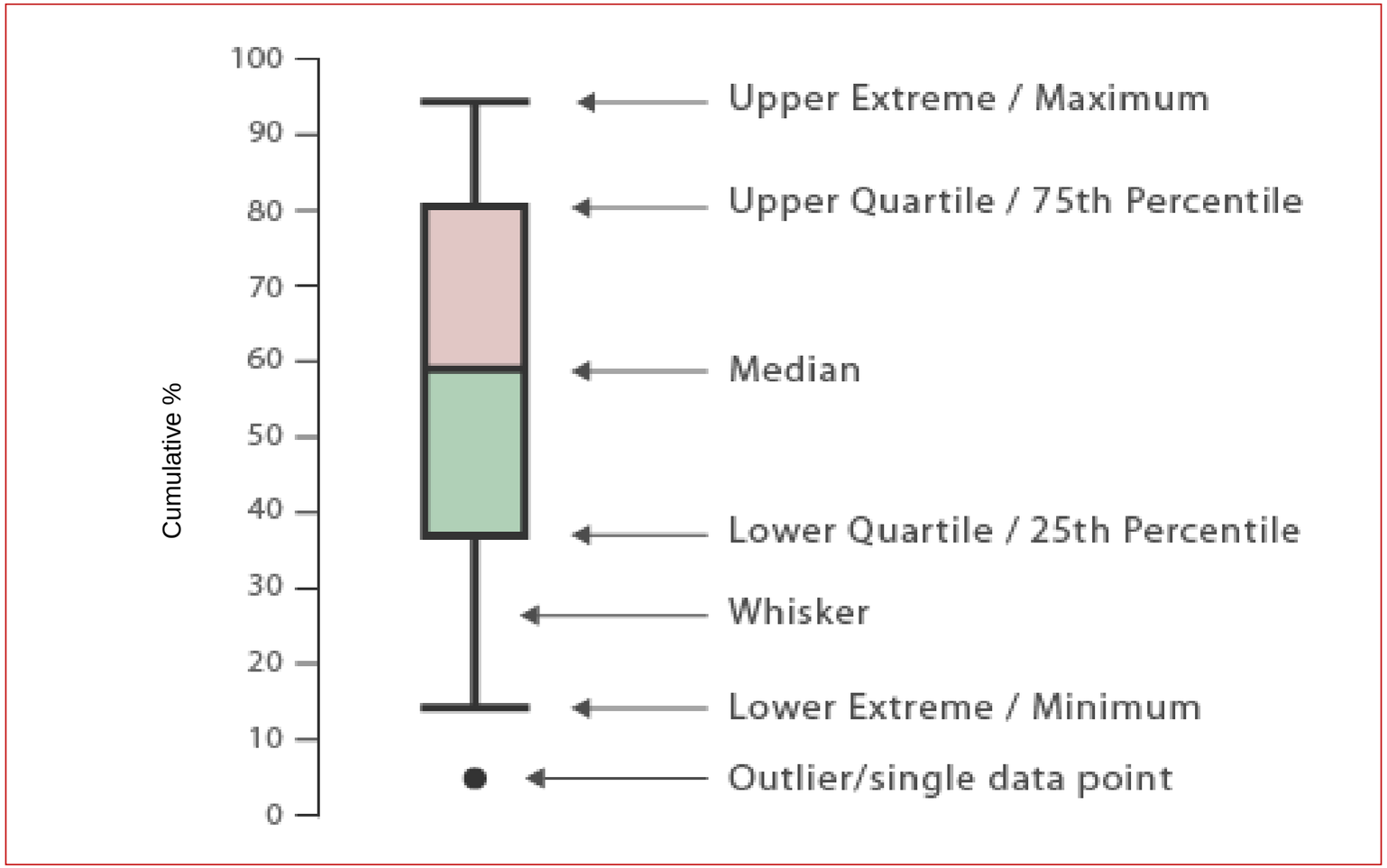
78, 81, 82, 68, 65, 59, 62, 58, 51, 62, 62, 71, 69, 64, 67, 71, 62, 65, 65, 74, 76, 87, 82, 82, 83, 79, 79, 71, 82, 77, 81



Stem	Leaf
5	1 8 9
6	2 2 2 2 4 5 5 5 7 8 9
7	1 1 1 4 6 7 8 9 9
8	1 1 2 2 2 2 3 7

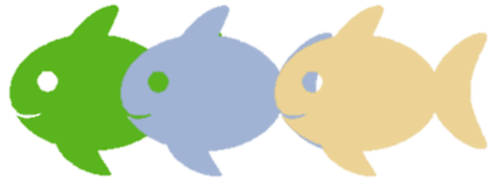


Box and Whisker Plots



IQR = Upper Quartile Value – Lower Quartile Value
Outlier is any value 1.5 *IQR units or more away from the upper or lower quartile

Box and Whisker Plots



Rewrite the data in increasing order.

12, 13, 5, 8, 9, 20, 16, 14, 14, 6, 9, 12, 12

5, 6, 8, 9, 9, 12, 12, 12, 13, 14, 14, 16, 20

Find the median for the dataset.

5, 6, 8, 9, 9, 12, 12, 12, 13, 14, 14, 16, 20

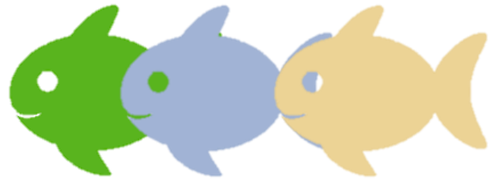
Find the lower and upper quartiles.

5, 6, 8, 9, 9, 12, 12, 12, 13, 14, 14, 16, 20

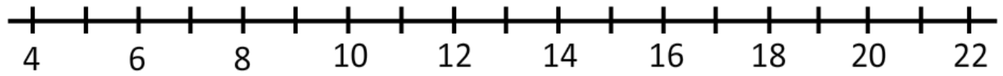
Lower quartile = 8.5

Upper quartile = 14

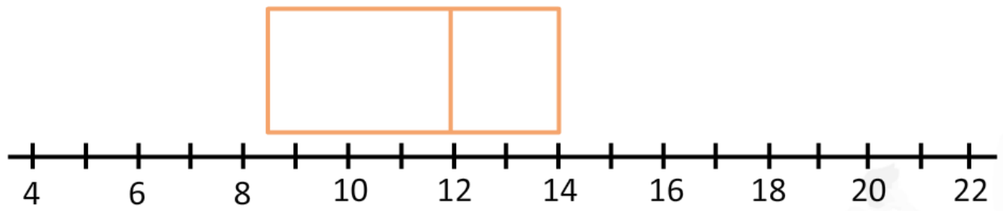
Box and Whisker Plots



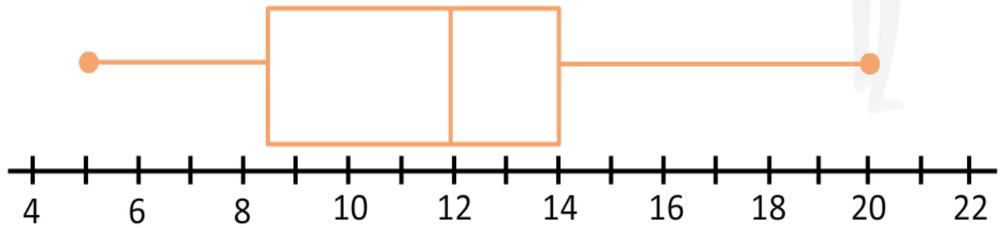
Draw a number line to include all the data points.



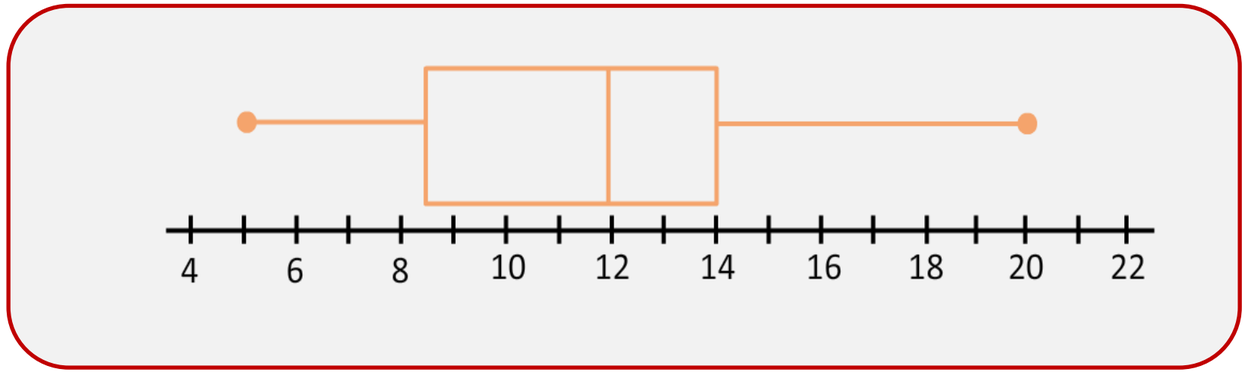
Locate the main median, 12, using a vertical line. Locate the lower and upper quartiles (8.5 and 14) and join them with the median by drawing boxes.



Extend whiskers from either ends of the boxes to the smallest and largest numbers (5 and 20) in the dataset.



Box and Whisker Plots



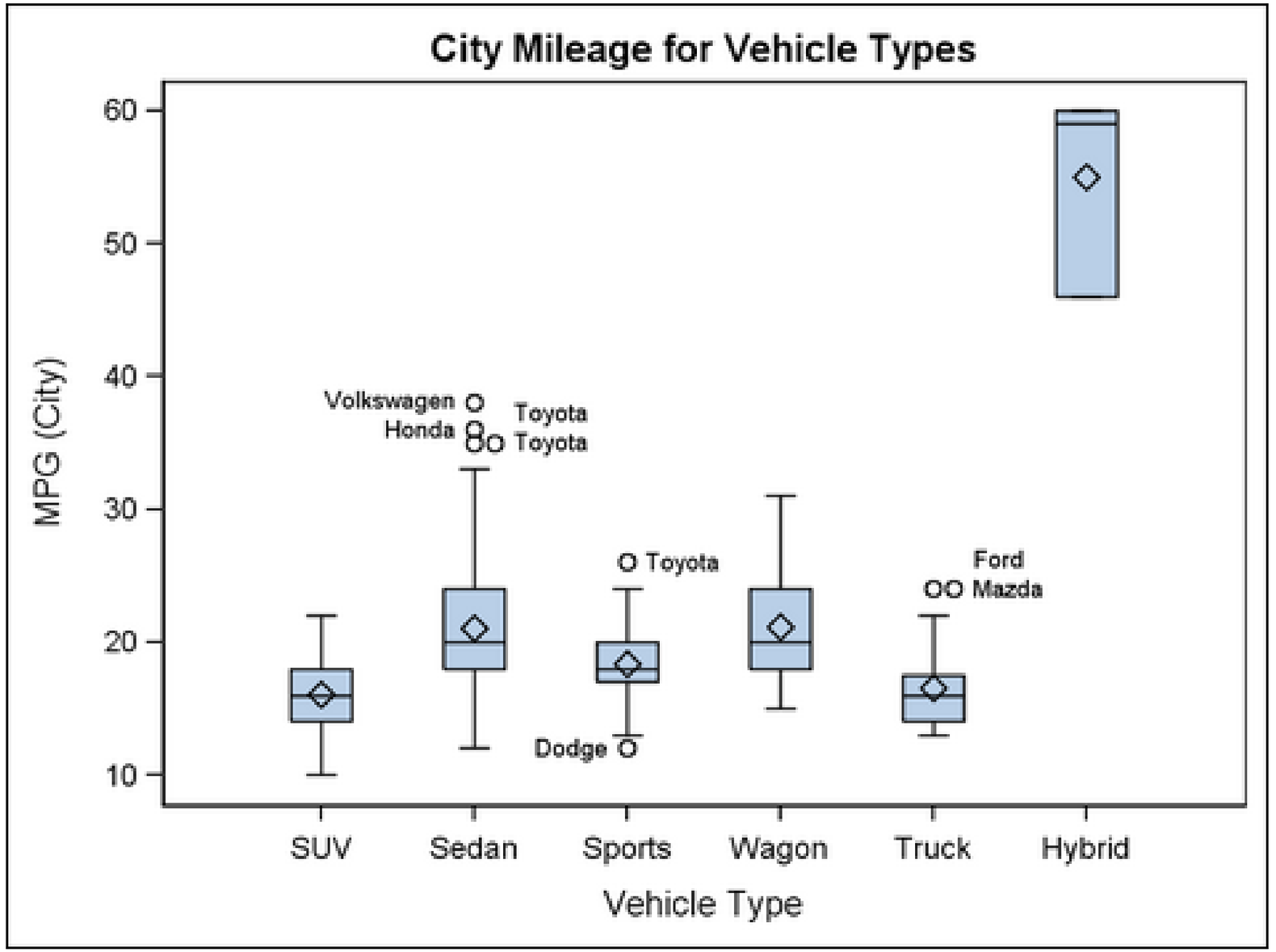
Inference

Range = 20 – 5 = 15

The quartiles split the data into four equal parts:

- Numbers lesser than 8.5
- Numbers between 8.5 and 12
- Numbers between 12 and 14
- Numbers greater than 14

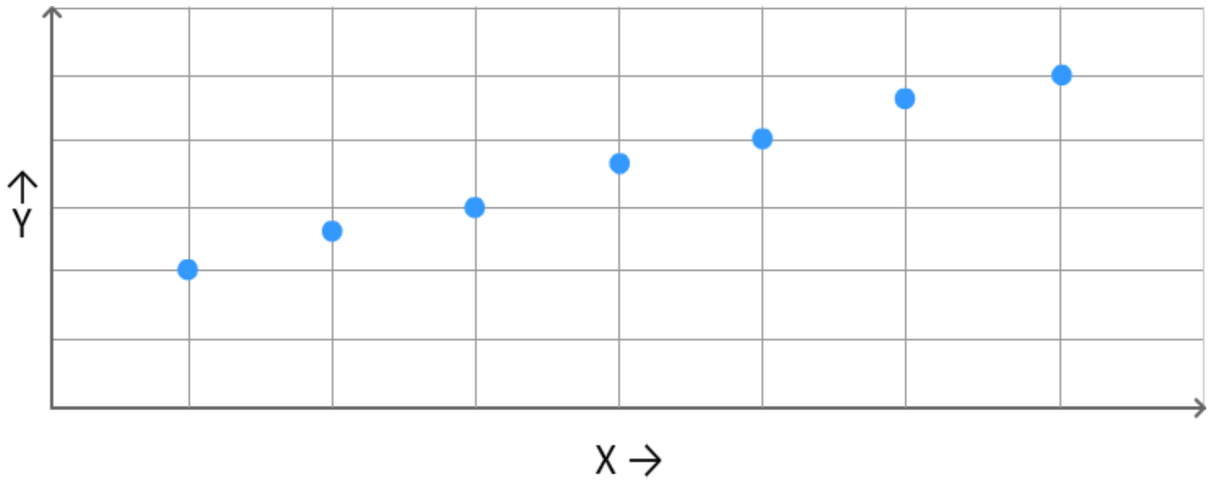
Comparison of Box and Whisker Plots



Scatter Diagrams: Purpose

A scatter diagram can be used to:

Understand the correlation between two variables

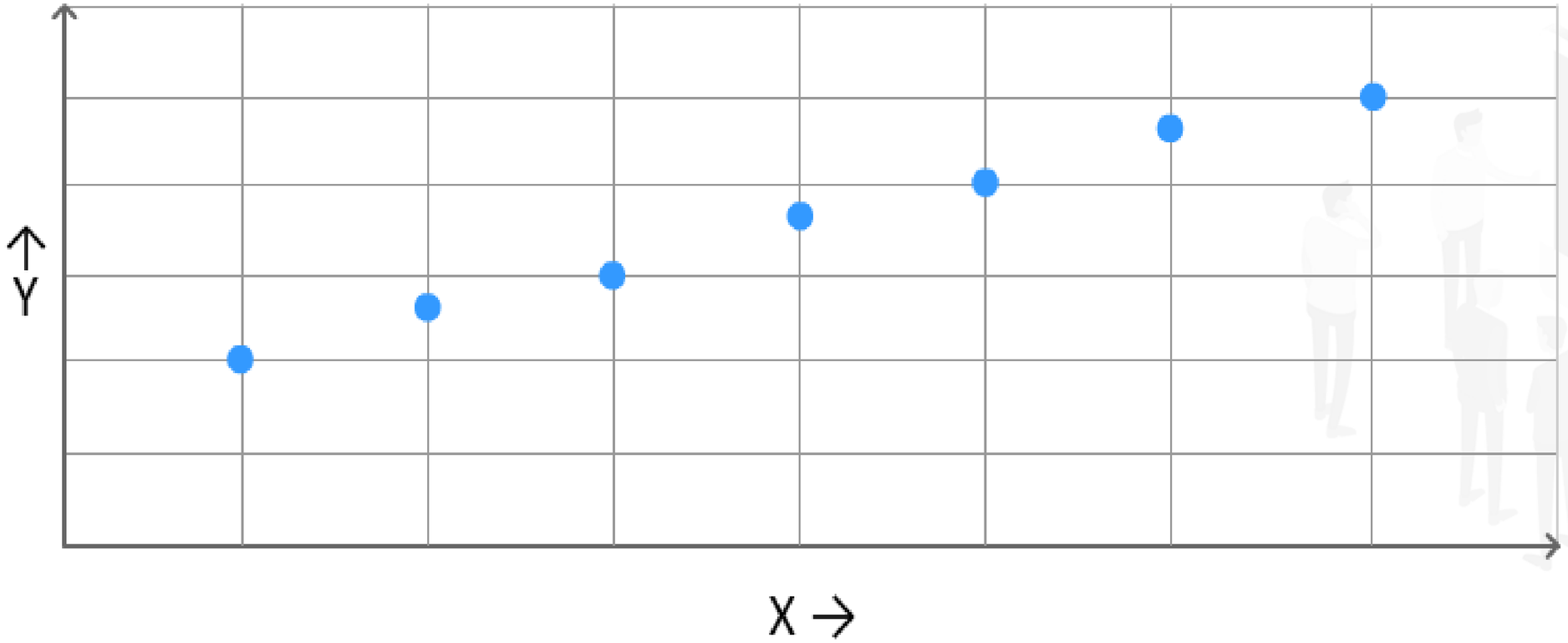


Identify the root cause

Examine cause-and-effect relationships

Scatter Diagrams: Types of Correlation

Perfect positive correlation	Moderate positive correlation	No relation	Moderate negative correlation	Perfect negative correlation

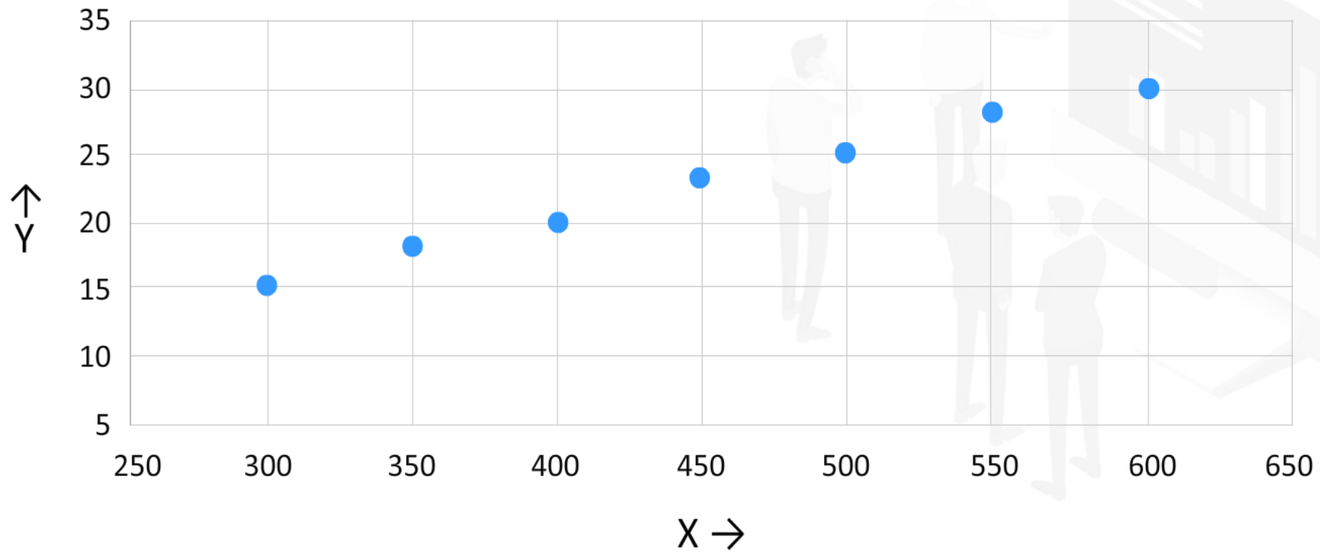


Scatter Diagrams: Types of Correlation

Perfect positive correlation	Moderate positive correlation	No relation	Moderate negative correlation	Perfect negative correlation

In perfect positive correlation, as the value of X increases, the value of Y also increases proportionally.
Example: Correlation between consumption of coffee and consumption of milk

Coffee Consumption in ml (X)	Milk Consumption in L (Y)
300	15
350	17.5
400	20
450	22.5
500	25
550	27.5
600	30

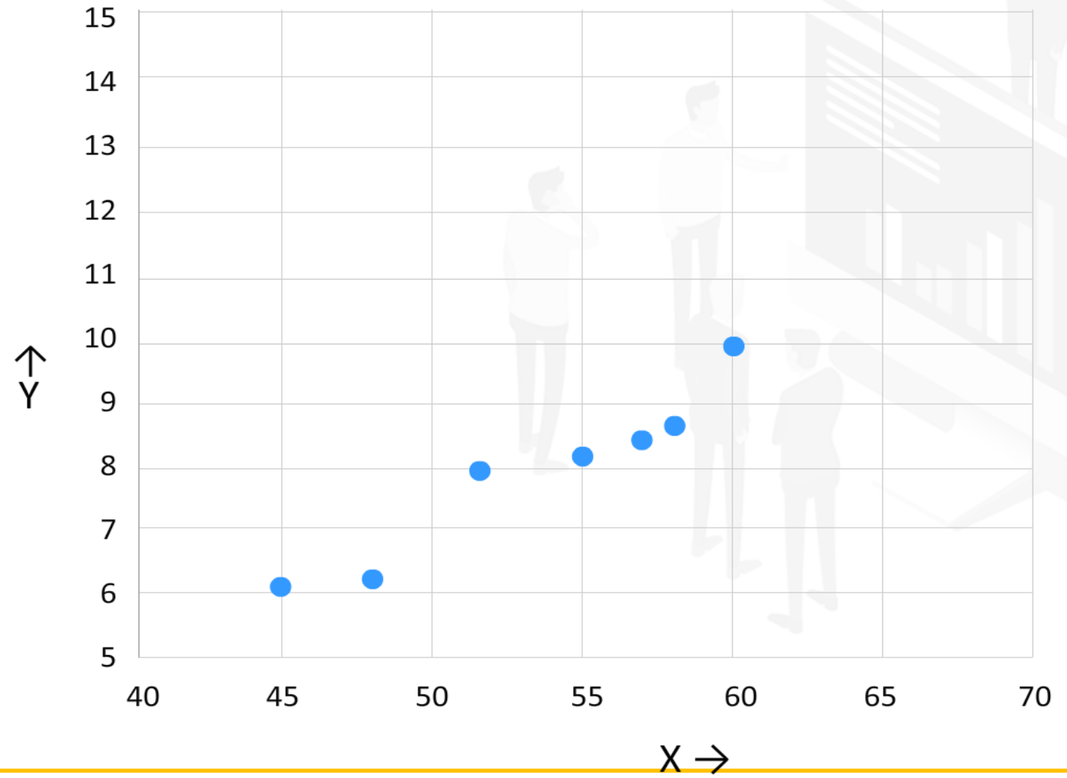


Scatter Diagrams: Types of Correlation

Perfect positive correlation	Moderate positive correlation	No relation	Moderate negative correlation	Perfect negative correlation

In moderate positive correlation, as the value of X increases, the value of Y also increases but not in the same proportion. **Example**: Correlation between monthly salary and monthly savings

Salary (in thousands) (X)	Savings (in thousands) (Y)
45	6
48	6.2
52	8
55	8.2
57	8.5
58	8.6
60	10
65	12

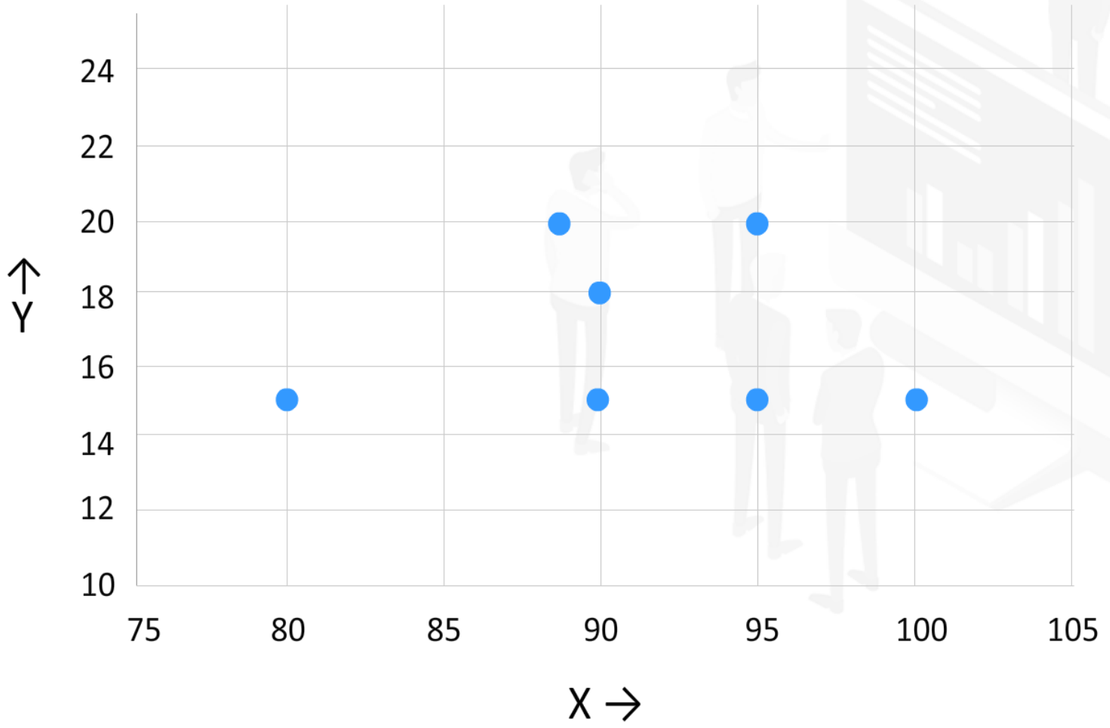


Scatter Diagrams: Types of Correlation

Perfect positive correlation	Moderate positive correlation	**No relation**	Moderate negative correlation	Perfect negative correlation

When a change in one variable has no impact on the other, there is no correlation between them.
Example: Relation between number of recent graduates and open job positions in a city

Recent Graduates (in thousands) (X)	Open Job Positions (in thousands) (Y)
80	15
100	15
90	18
95	20
89	20
90	15
95	15

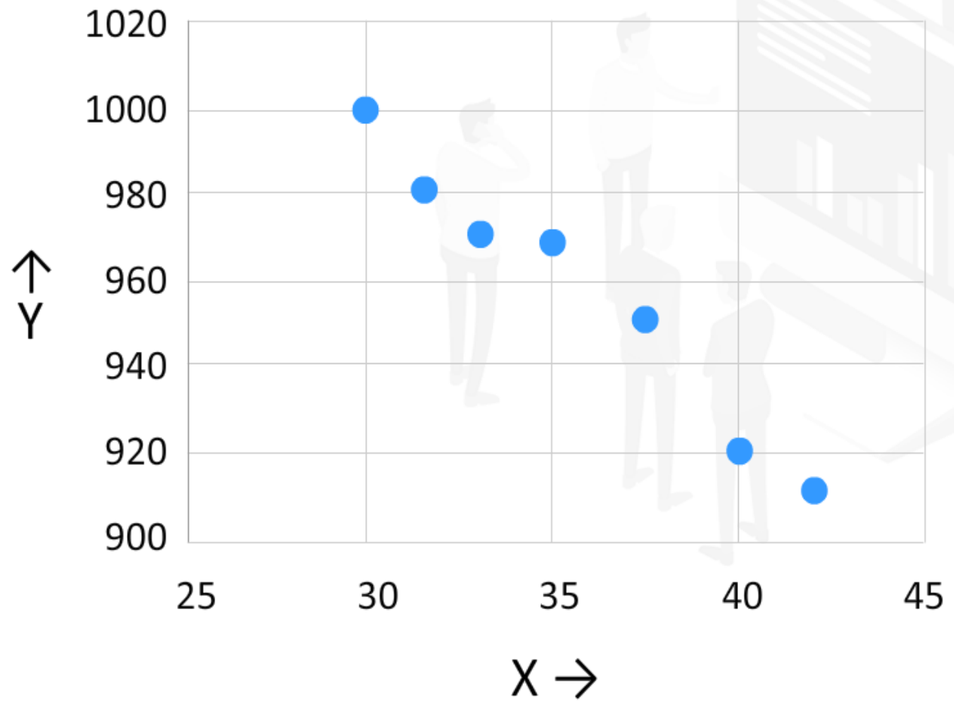


Scatter Diagrams: Types of Correlation

Perfect positive correlation	Moderate positive correlation	No relation	Moderate negative correlation	Perfect negative correlation

In moderate negative correlation, as the value of X increases, the value of Y decrease but not in the same proportion. **Example**: Correlation between the price of a product and the number of units sold

Unit Price of Product (in thousands) (X)	Units Sold (Y)
30	1000
32	980
33	970
35	965
38	950
40	920
42	910

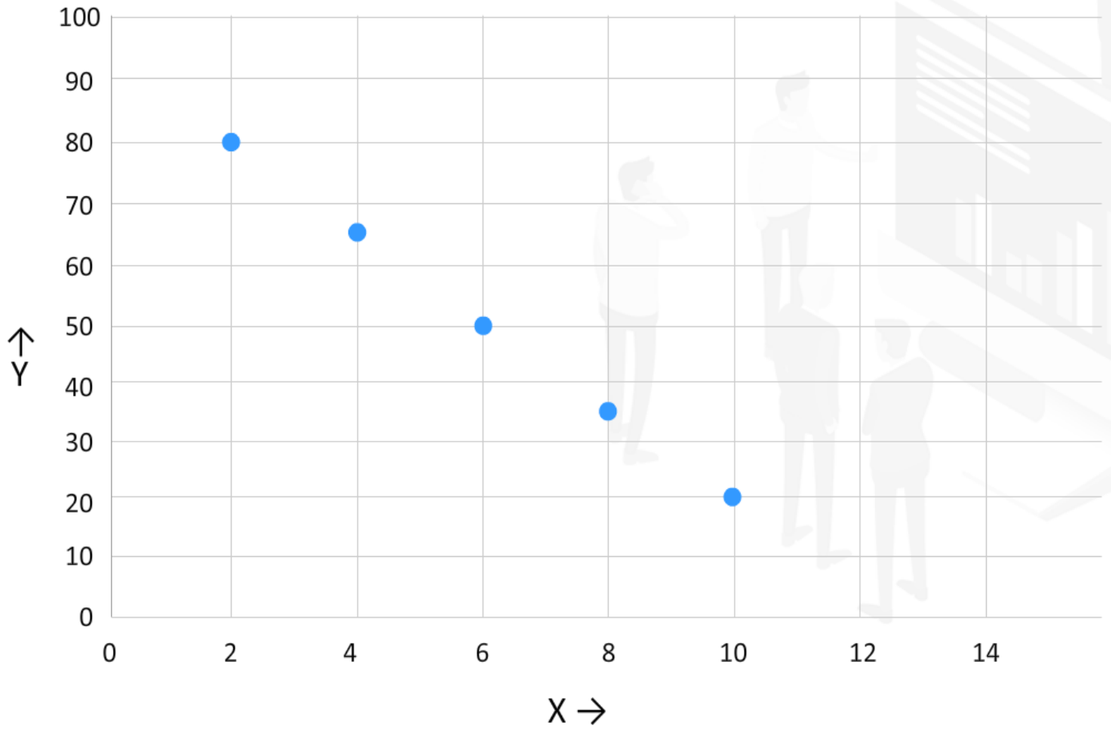


Scatter Diagrams: Types of Correlation

Perfect positive correlation	Moderate positive correlation	No relation	Moderate negative correlation	Perfect negative correlation

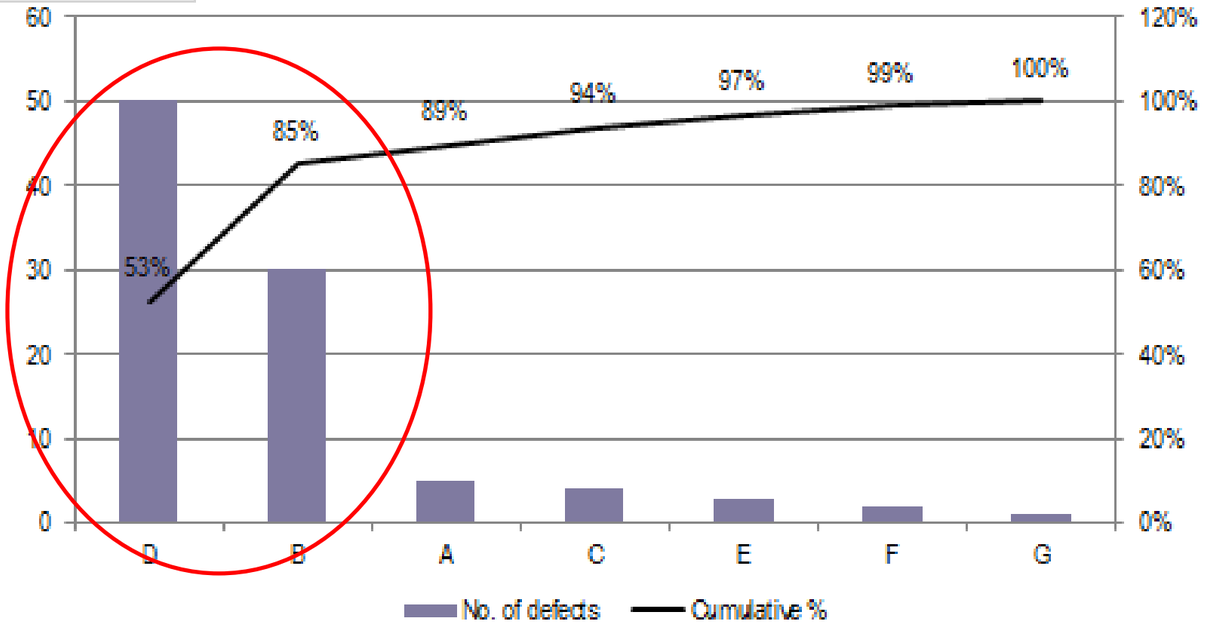
In perfect negative correlation, as X increases, Y decreases proportionally.
Example: Correlation between project time extension and project success

Time Extension (in days) (X)	Project Success Probability (in percentage) (Y)
2	80
5	60
7	40
10	20
13	00

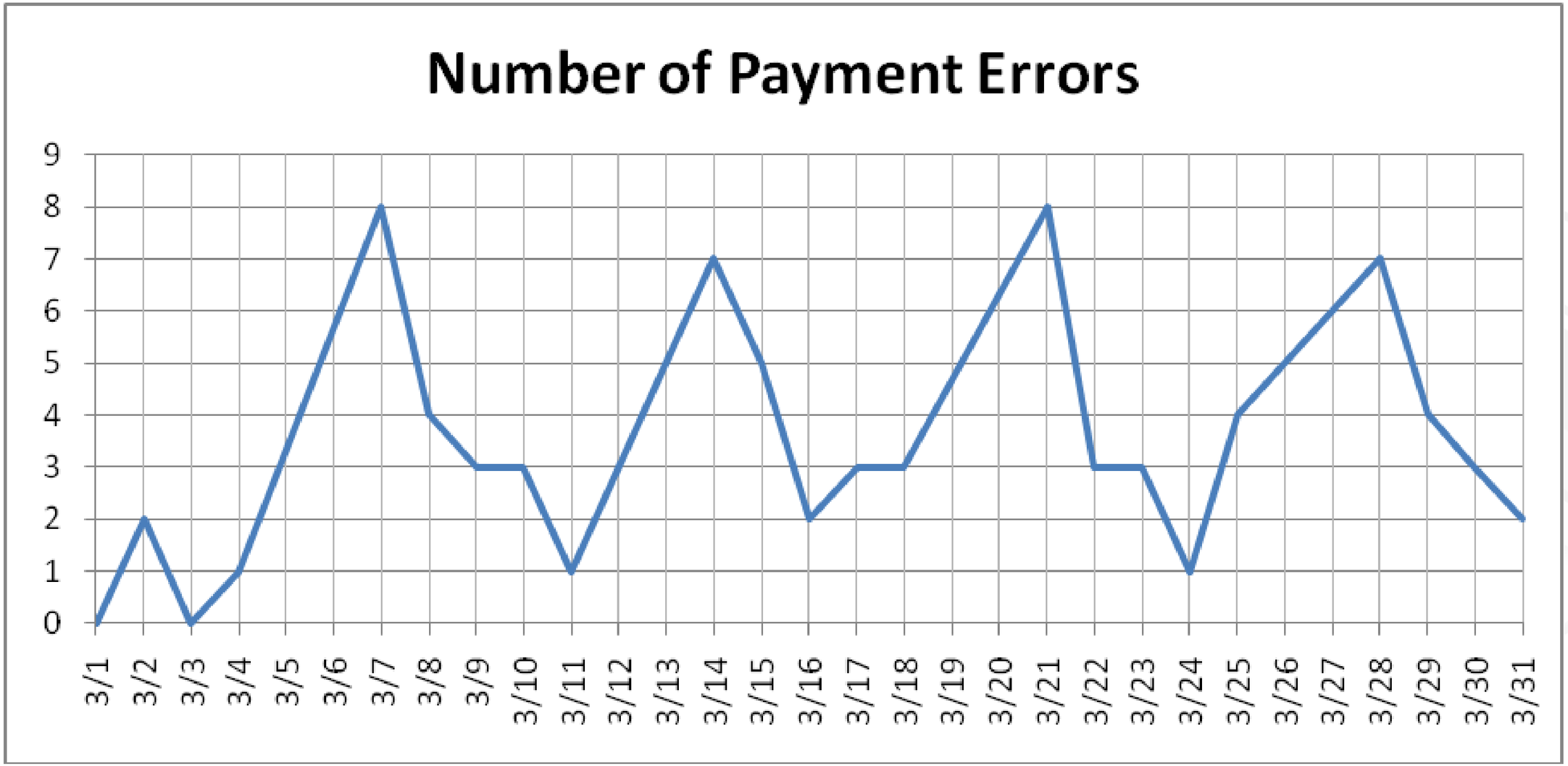


Other Visualization Tools: Pareto Charts

Module	No. of defects	% contribution	Cumulative %
D	50	52.6%	52.6%
B	30	31.6%	84.2%
A	5	5.3%	89.5%
C	4	4.2%	93.7%
E	3	3.2%	96.8%
F	2	2.1%	98.9%
G	1	1.1%	100.0%
Total	95		



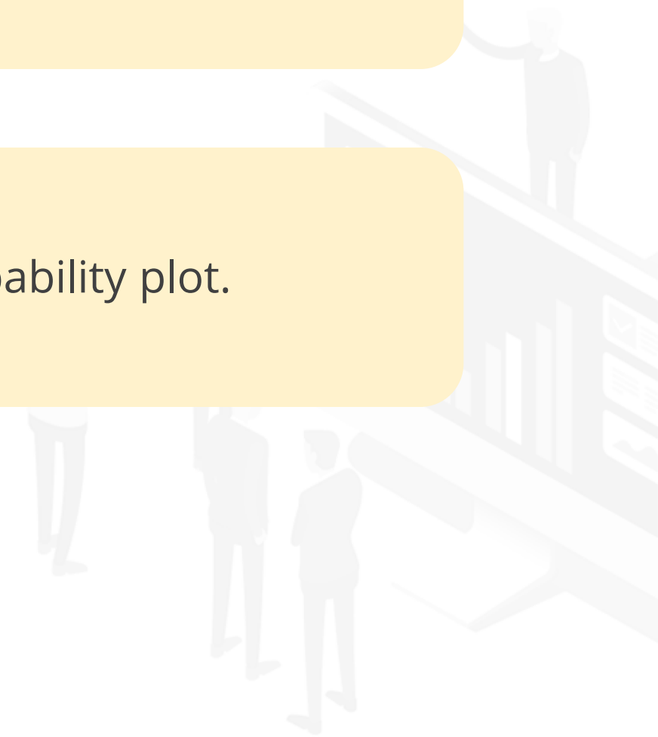
Other Visualization Tools: Run Charts



Normal Probability Plots

Normal probability plots are used to identify if a dataset is normally distributed.

A normally distributed dataset forms a straight line in a normal probability plot.



Normal Probability Plots

.127, .125, .123, .123, .120, .124, .126, .122, .123, .125, .121, .123, .122, .125, .124, .122, .123, .123, .126, .121, .124, .121, .124, .122, .126, .125, .123

Construct a cumulative frequency distribution table.

Calculate the mean rank probability estimate.

$$\text{Mean rank probability estimate} = \frac{\text{Cumulative frequency}}{(n + 1)} * 100$$

Where n = sample size

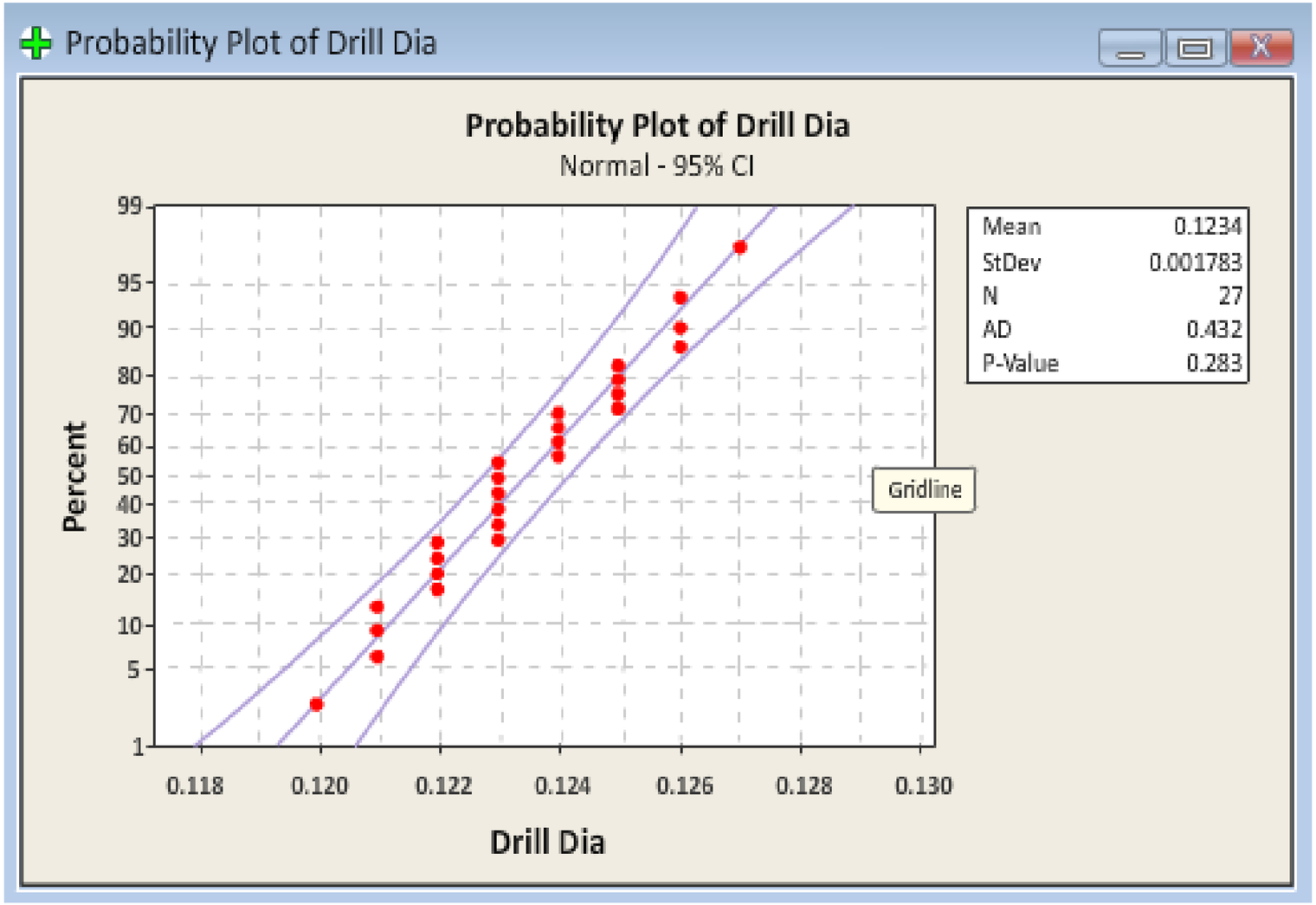
Normal Probability Plots

Construct a cumulative frequency distribution table and calculate the mean rank.

X	Frequency	Cumulative Frequency	(Cumulative Frequency)/(n+1)	Mean Rank (%)
0.120	1	1	1/28	4
0.121	3	4	4/28	14
0.122	4	8	8/28	29
0.123	7	15	15/28	54
0.124	4	19	19/28	68
0.125	4	23	23/28	82
0.126	3	26	26/28	93
0.127	1	27	27/28	96
	n = 27			

Normal Probability Plots

Plot the graph on log paper or using Minitab, a statistical software used in Six Sigma.



Good to Know

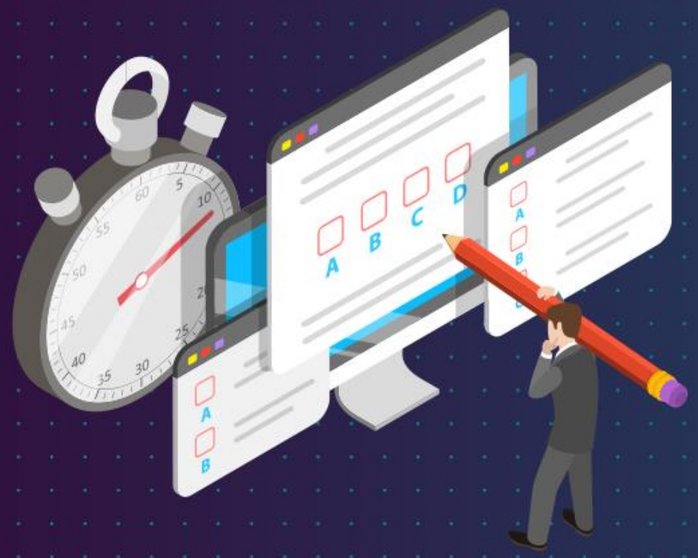
A set of raw numbers is not easily understood by most people.

The key is to easily and visually depict centrality, dispersion, and the data distribution shape.

Remember the point of summarizing data is to make a set of numbers more understandable.

Key Takeaways

- The two types of data are attribute (discrete) and variable (continuous).
- The scales of measurement are nominal, ordinal, interval, and ratio.
- Data may be collected by census, sampling, experiment, and observations.
- The types of sampling are random, sequential, and stratified.
- Data can be summarized using measures: measures of central tendency and measures of dispersion.
- Mean, median, and mode are the measures of central tendency.
- Range, variance, and standard deviation are the measures of dispersion.



Knowledge Check

Knowledge Check

1

A team counts how many times a defect occurs in a product and every part is measured. What type of data collection is used?

A. Sampling

B. Experiment

C. Observation

D. Census

Knowledge Check

1

A team counts how many times a defect occurs in a product and every part is measured. What type of data collection is used?

A. Sampling

B. Experiment

C. Observation

D. Census

The correct answer is **D**

When every item of a population is considered it is considered a census.

Knowledge Check 2	A coding team has documented the number errors in the system modules over 9 days and wants to create a box blot. What is the median, lower quartile, and upper quartile values for the following? 14,10,12,10,13,13,15,16,13

A. 13,11,14

B. 11,13,14

C. 13,11,14.5

D. 13,14.5,11

Knowledge Check 2

A coding team has documented the number errors in the system modules over 9 days and wants to create a box blot. What is the median, lower quartile, and upper quartile values for the following? 14,10,12,10,13,13,15,16,13

A. 13,11,14

B. 11,13,14

C. 13,11,14.5

D. 13,14.5,11

The correct answer is **C**

This data sorted is [10,10,12,13,13,13,14,15,16] in which the median is at 13, the lower quartile value is the average of 10 and 12, and the upper quartile is the average of 14 and 15.

Knowledge Check

3

When Y increases in the same proportion as the decrease in X, what type of correlation is represented?

A. Negative

B. Positive

C. Perfect Positive

D. Perfect Negative

Knowledge Check 3	When Y increases in the same proportion as the decrease in X, what type of correlation is represented?

A. Negative

B. Positive

C. Perfect Positive

D. Perfect Negative

The correct answer is **D**

In a perfect correlation, there is a 1:1 relationship between the Y and X variables in which both change in the same magnitude. In a negative relationship, one variable increases, while the other decreases.

Knowledge Check 4	You want to know how satisfied customers who purchased your products are. To find out, you select names of customers who purchased products in the last three months. Each customer has an equal chance of being selected. What sampling method is used in this scenario?

A. Stratified

B. Random

C. Sequential

D. Mean

Knowledge Check 4	You want to know how satisfied customers who purchased your products are. To find out, you select names of customers who purchased products in the last three months. Each customer has an equal chance of being selected. What sampling method is used in this scenario?

A. Stratified

B. Random

C. Sequential

D. Mean

The correct answer is **B**

With random sampling, each unit has the same chance of being selected.

Knowledge Check 5	What is the standard deviation for the data subset below? 23, 45, 34, 46, 66, 51, 24, 33, 44, 26

A. 39.2

B. 13.01

C. 13.72

D. 39

Knowledge Check 5	What is the standard deviation for the data subset below? 23, 45, 34, 46, 66, 51, 24, 33, 44, 26

A. 39.2

B. 13.01

C. 13.72

D. 39

The correct answer is **C**

The sample standard deviation using Excel's =STDEV.S function is 13.72.

Knowledge Check

6

A team wants to compare the distributions of different processes. Which visual tool should it use?

A. Pareto Chart

B. Scatter Plot

C. Normality Plot

D. Box and Whiskers Plot

Knowledge Check

6

A team wants to compare the distributions of different processes. Which visual tool should it use?

A. Pareto Chart

B. Scatter Plot

C. Normality Plot

D. Box and Whiskers Plot

The correct answer is **D**

A box and whiskers plot is a simple tool to display the distributions of many groups.