Mathematics & Statistics Essentials

Descriptive Statistics

Learning Objectives

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

- Determine a statistical tool to compare and evaluate data security
- Define mathematical and positional averages
- Learn about mean, median, decile, percentile, mode, and quartiles
- Explain the concepts of outliers

Learning Objectives

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

- Explain the measures of dispersion, such as range, interquartile range, and outliers
- Describe mean absolute deviation (MAD), standard deviation, and variance
- Describe the Z-score
- Elaborate the measures of shape and how to summarize data

Business Scenario

ABC is an organization that stores a large amount of data. The organization wants to analyze the data to determine meaningful insights from it.

However, the organization is supposed to perform multiple calculations to determine the central tendency, mathematical and positional averages, and much more.

To do this, the organization will have to go through a few concepts of descriptive statistics that will help them determine the calculations and visualize data to get meaningful insights.

Measures of Central Tendency and Mathematical and Positional Averages

Data for Businesses

Businesses collect a lot of data.

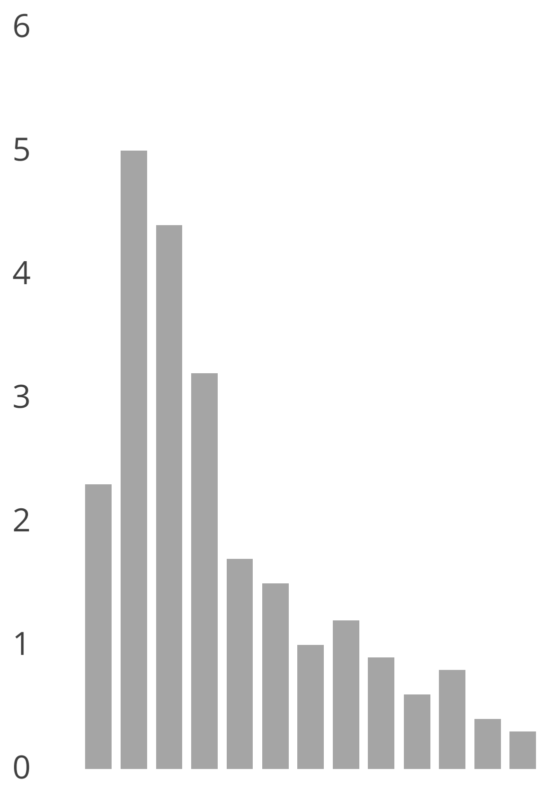
Data is used for the planning and execution of operations.

Data collected is analyzed to gain insights and take proper action.

Data is depicted in many executive reports through visualizations along with summary measures, such as measures of central tendency.

Central Tendency

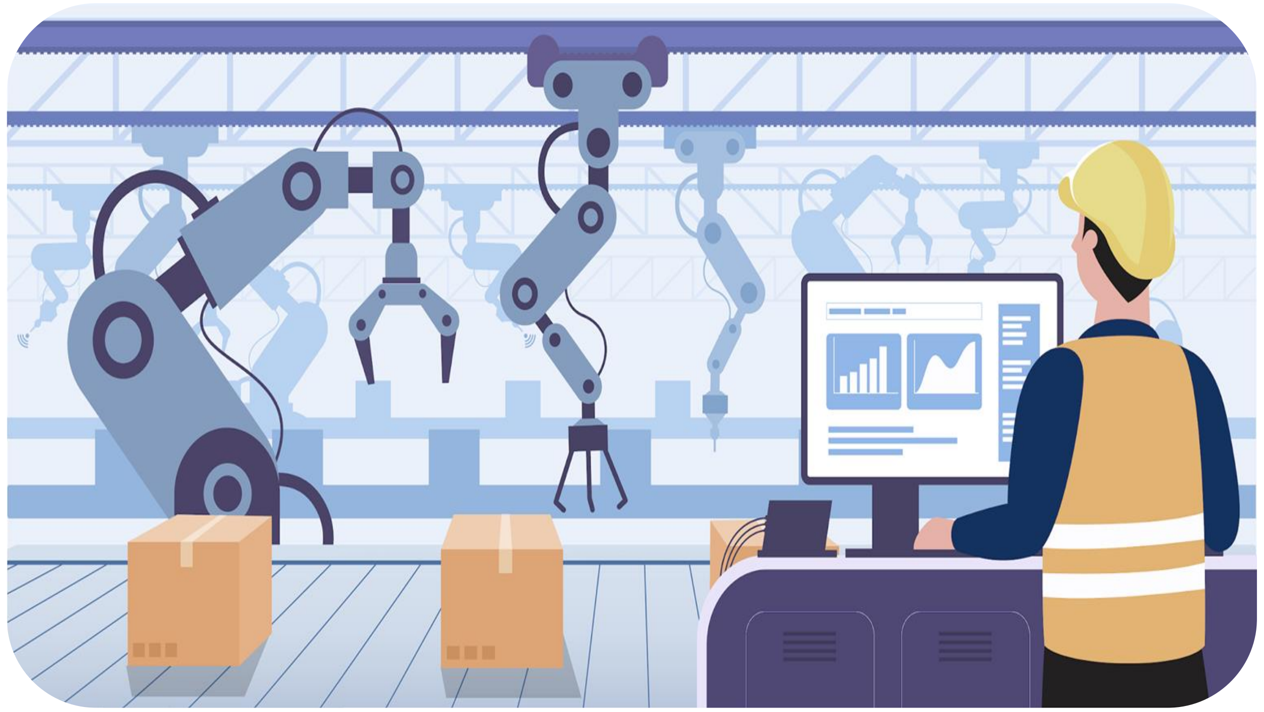
Central tendency is a descriptive summary of a dataset through a single value that reflects the center of the data distribution.



Measures of central tendency indicate where most values in a distribution fall and are also referred to as the central location of a distribution.

Example for Central Tendency

Example 1: A factory's production capacity can be estimated using data on the output collected over time.



Example for Central Tendency

Example 2: Real estate agents calculate the average price of homes in each area so that they can inform their clients what they can expect if they want to buy a house.

Significance of Central Tendency

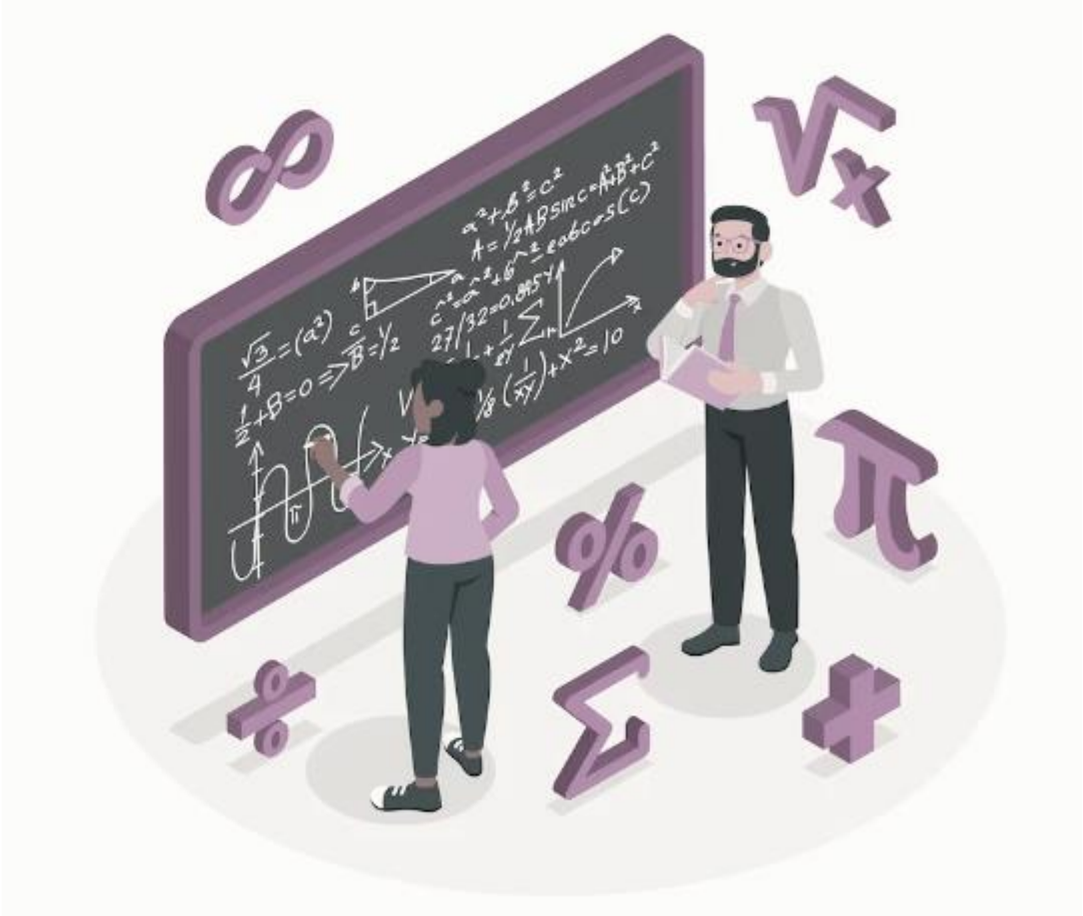
The two ways to look at the measure of the central tendency of a dataset are:

Mathematical average

Positional average

Mathematical Average

The mathematical average is also known as the mean value.



It can be calculated by:

$$\frac{\text{Sum of all the values in a dataset}}{\text{Number of values in a dataset}}$$

Positional Average

It is derived by arranging data points in an ascending or a descending order and identifying the value in the middle.

It gives the median value.

Mathematical and Positional Averages

Example: The heights of five people are shown below:0

Person	Height in cms
Person A	180
Person B	200
Person C	150
Person D	175
Person E	190

Calculating Mathematical Average

The average height of these five people is calculated as given below:

Person	Height in cm
Person A	180
Person B	200
Person C	150
Person D	175
Person E	190

$$\frac{\text{Individual heights}}{\text{Number of people}}$$

$$\frac{180+200+150+175+190}{5}$$

Mathematical average = 179 cm

Calculating Positional Average

To find the positional average, arrange the set of values either in an ascending or a descending order.

Person	Height in cm
Person A	180
Person B	200
Person C	150
Person D	175
Person E	190

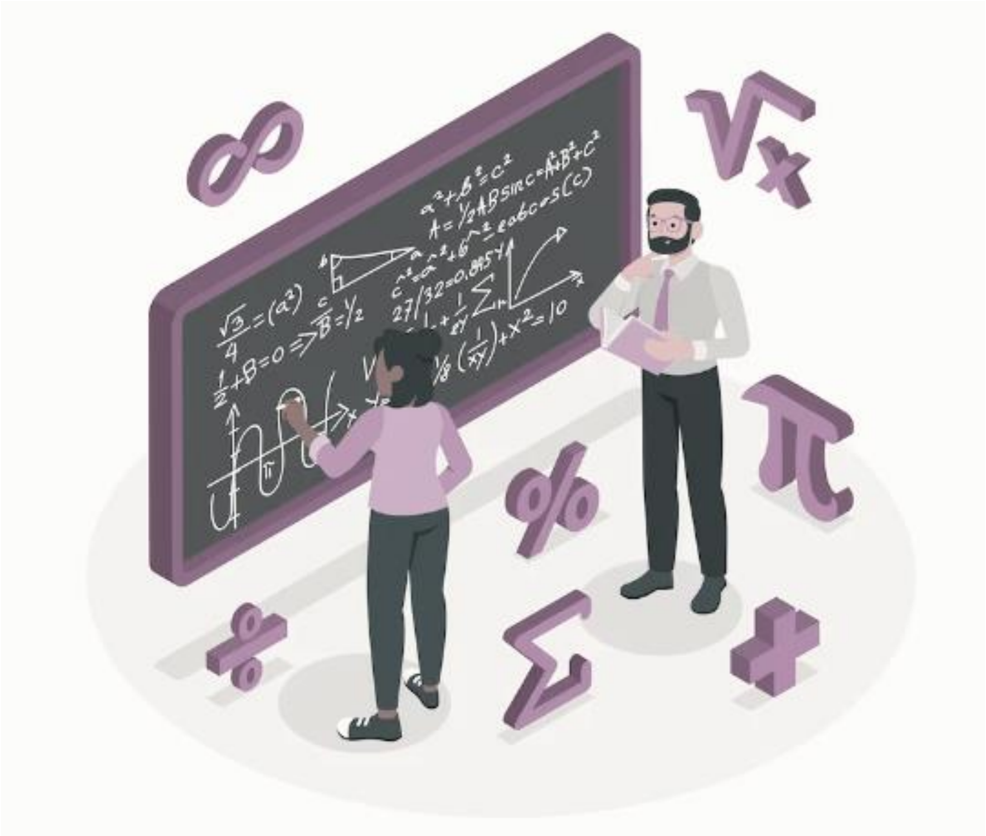
Person	Height in cm (ascending)
Person C	150
Person D	175
Person A	180
Person E	190
Person B	200

Person	Height in cm (descending)
Person B	200
Person E	190
Person A	180
Person D	175
Person C	150

Positional average = 180cm

Mathematical and Positional Averages

In mathematical averages, the calculated value might not be in the series of the respective dataset.

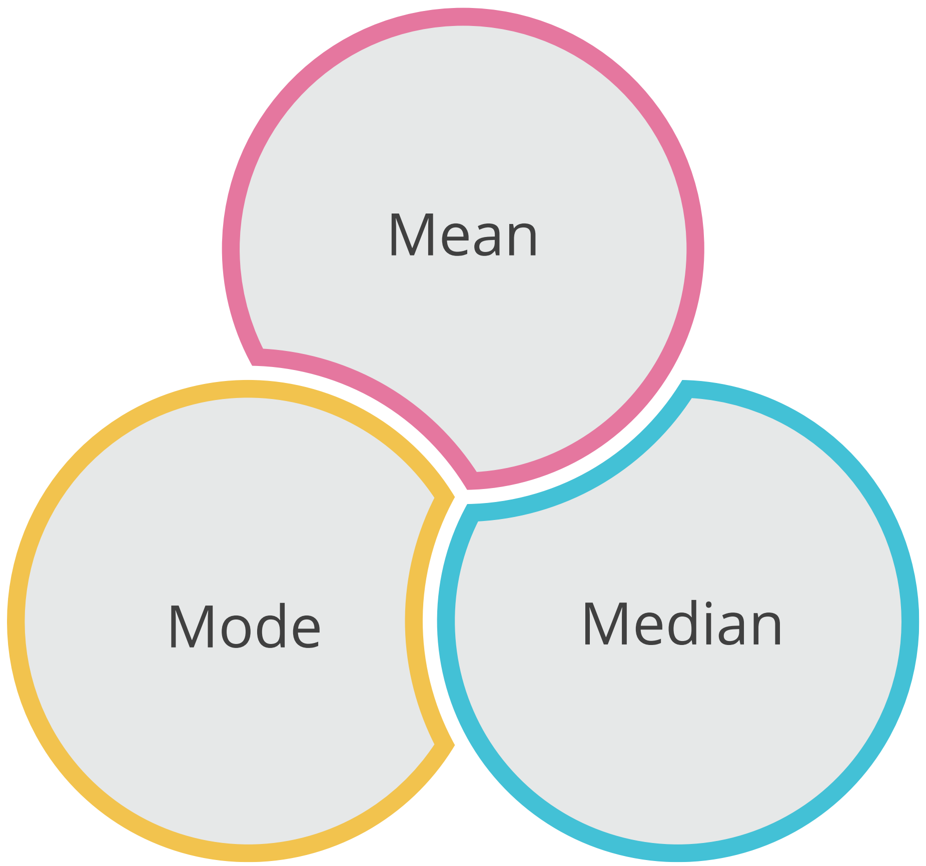


In positional averages, the calculated average value must be a value that lies within the set of observed data.

Measures of Central Tendency: Part 1

3 Ms

The following are the three Ms:



Measures of Central Tendency

There are three types of means. They are defined below:

Arithmetic mean	It is calculated by summing up a set of values and dividing the sum by the number of values in the set.
Geometric mean	It is the mean or average that indicates the central tendency of a finite set of real numbers by using the product of their values.
Harmonic mean	It is the reciprocal of the arithmetic mean of the reciprocals of a set of values.

Distortion of Mean

Mean is easily distorted by extreme values.

Such values may be present when data from unusual or nonrepresentative situations is included.

Distortion of Mean

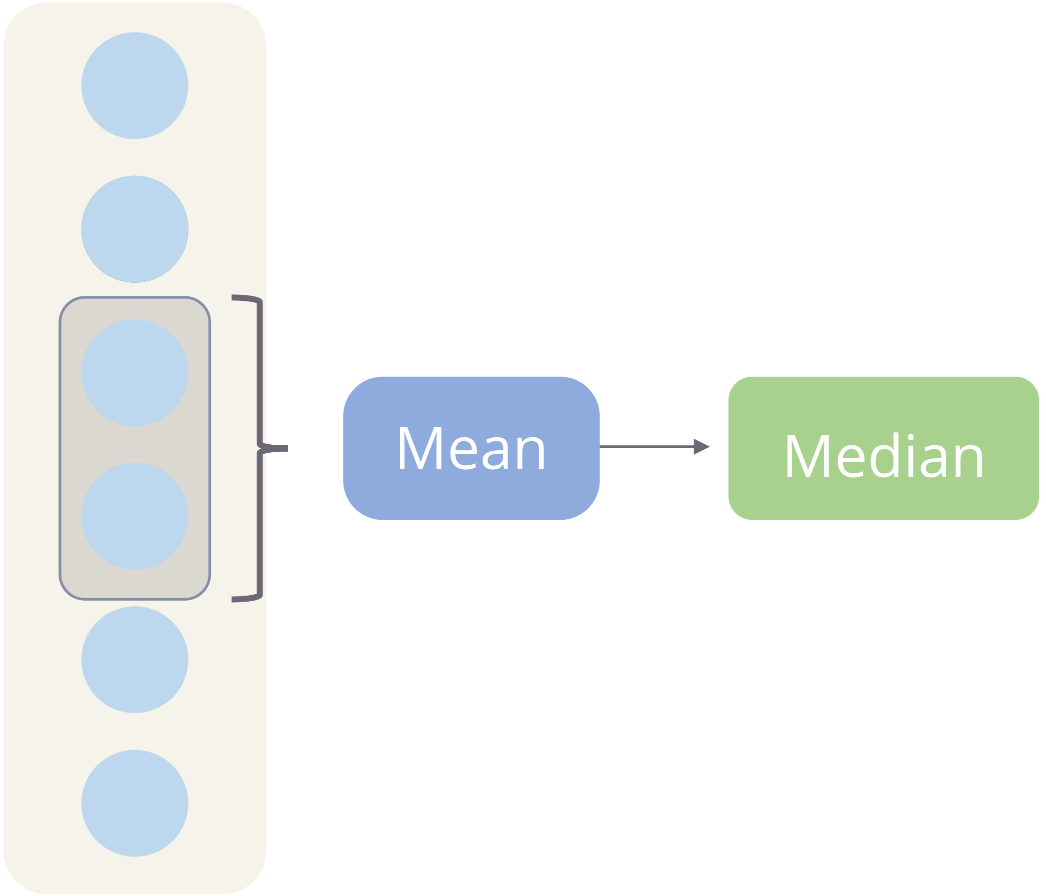
Example: In a dataset, an outlier or an extreme value can distort the mean and give a misleading representation of the typical value of the dataset.

- Consider the dataset, Age = (12, 15, 14, 13, 11, 45)
- The arithmetic mean of these values is:
  (12 + 15 + 14 + 13 + 11 + 45) / 6 = 110 / 6 = 18.33
- However, in this dataset, the value 45 is significantly higher than the other values.
- It is an outlier that does not represent the typical age in the group.

As a result, the mean of 18.33 is distorted upwards, giving a false impression of the central tendency of the ages.

Median

A median is the middle value or observation of a given set of data.



If the number of observations is even, the mean of the two middlemost observations is taken as the median.

Median and Mean

When compared to the mean, the median:

Is less amenable to further mathematical treatment

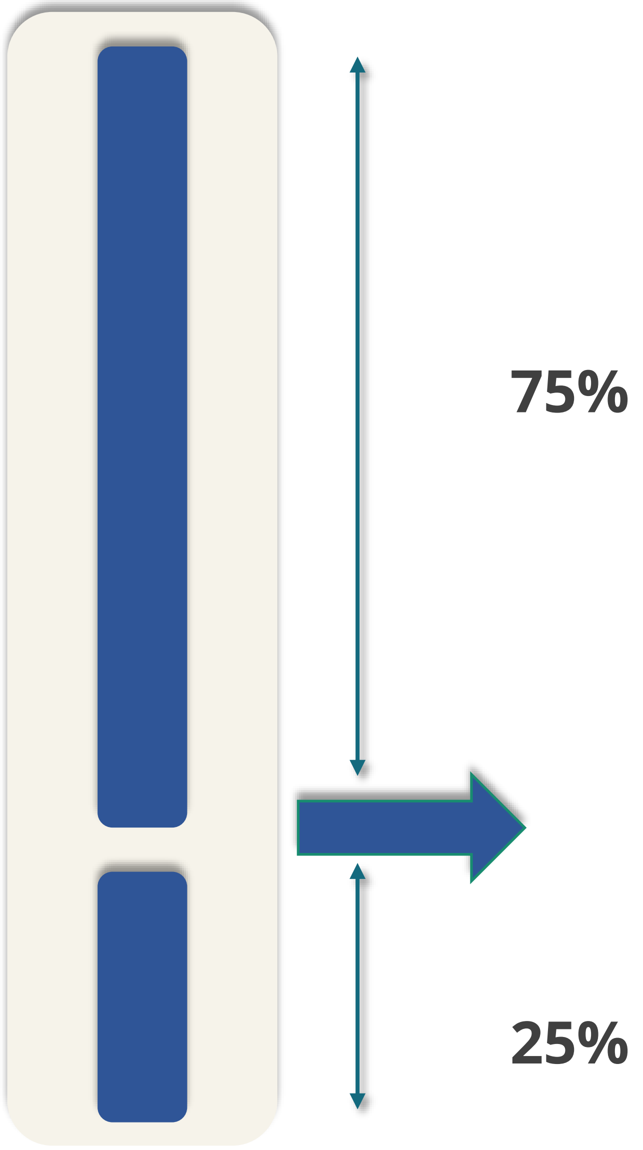
Is not impacted by extreme values

Quartile

A quartile divides the number of data points into four parts or quarters. In the first quartile:

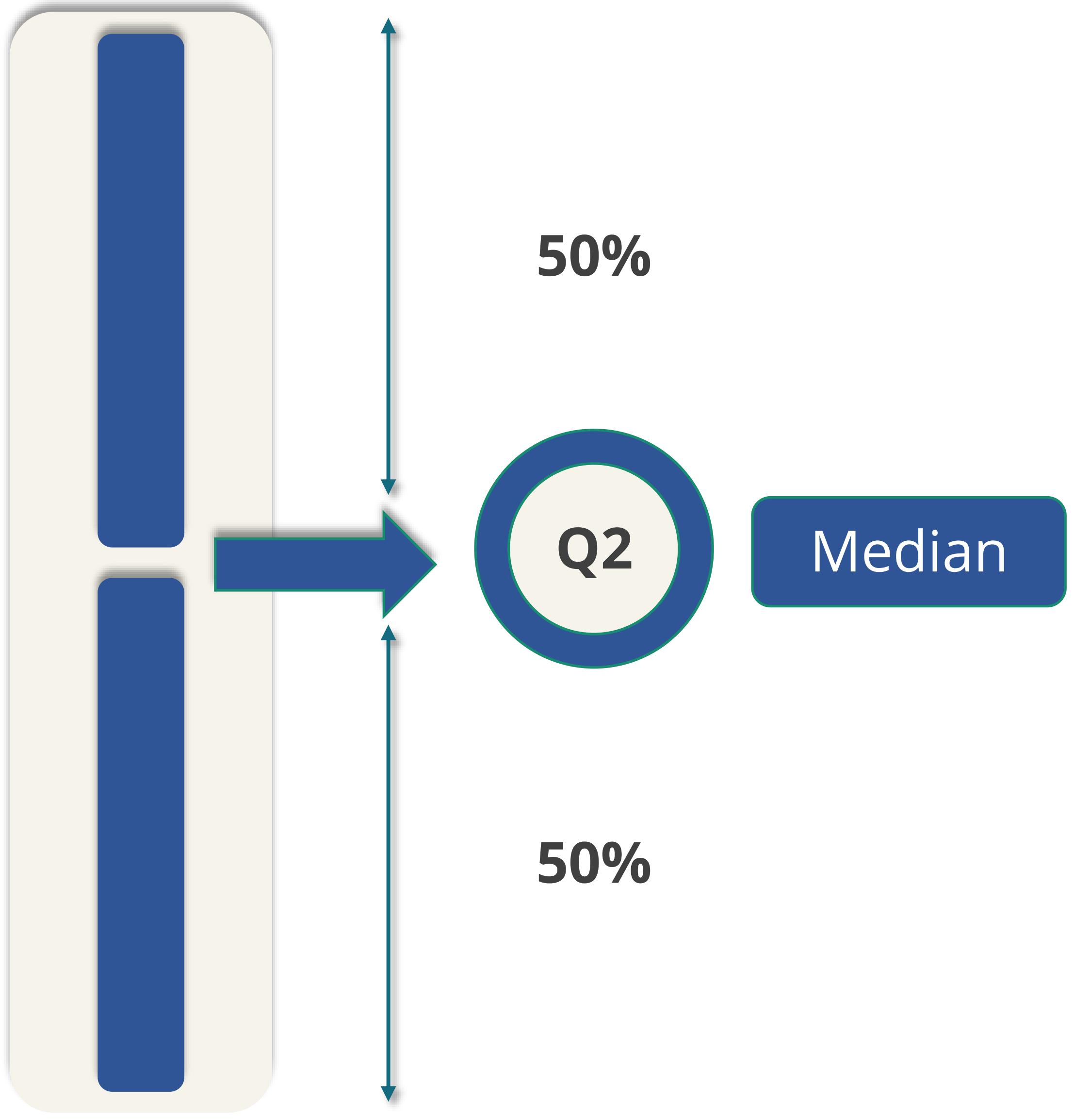
Q1

- 25% of observations are below that value.
- 75% of observations are above that value.



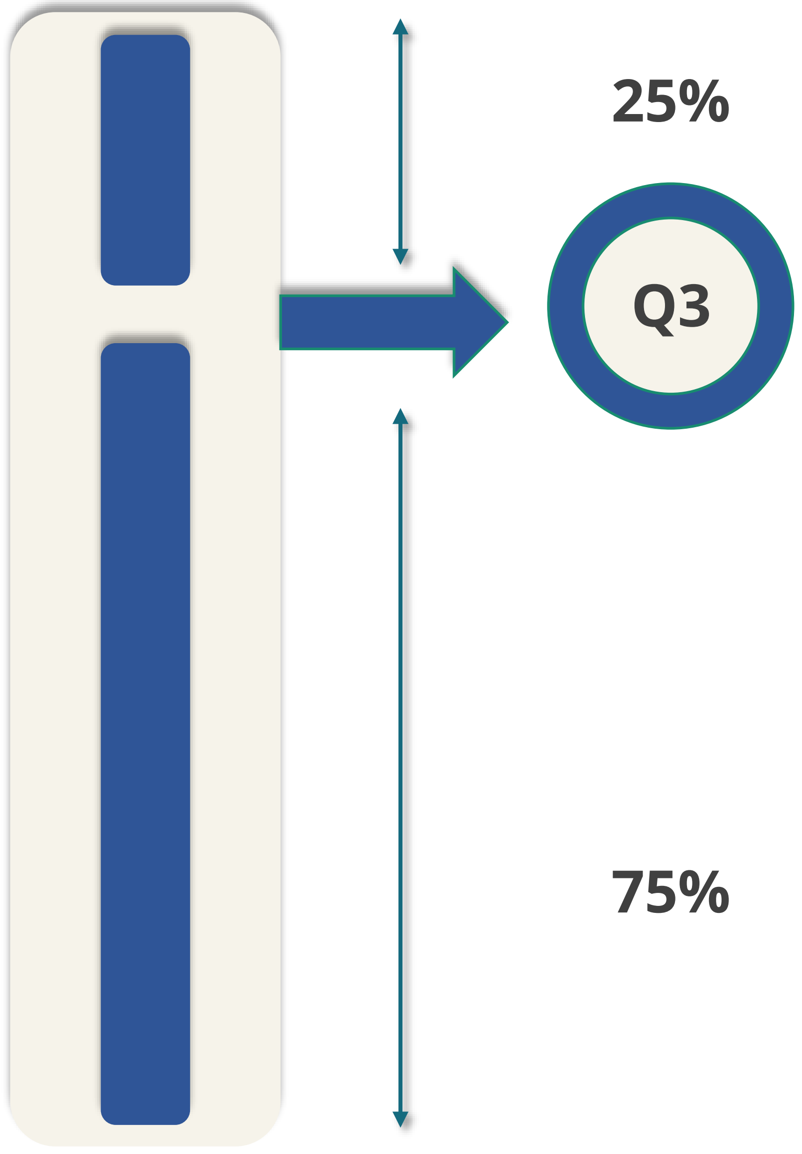
Quartile

The second quartile is the median.



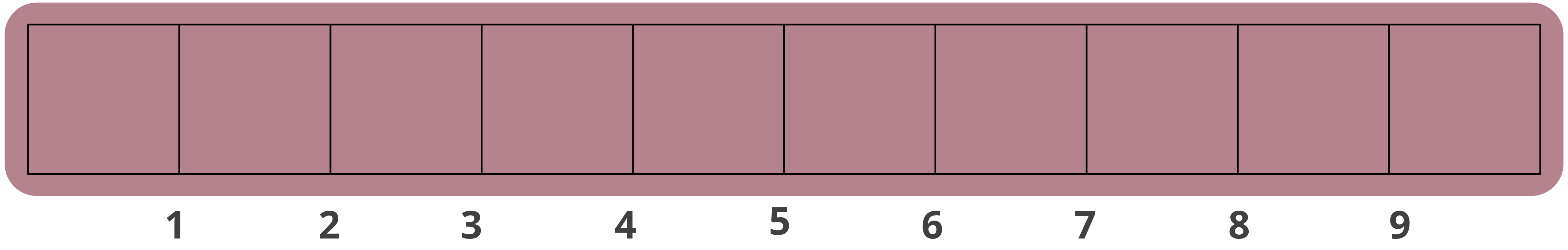
Quartile

In the third quartile, 75% of observations are below the value, and 25% are above it.



Decile

A decile is any of the nine values that divide a dataset into ten equal parts.



Each part represents one-tenth of the sample.

Percentile

A percentile is a number that represents the percentage of data below a given value.

1   2   3   99

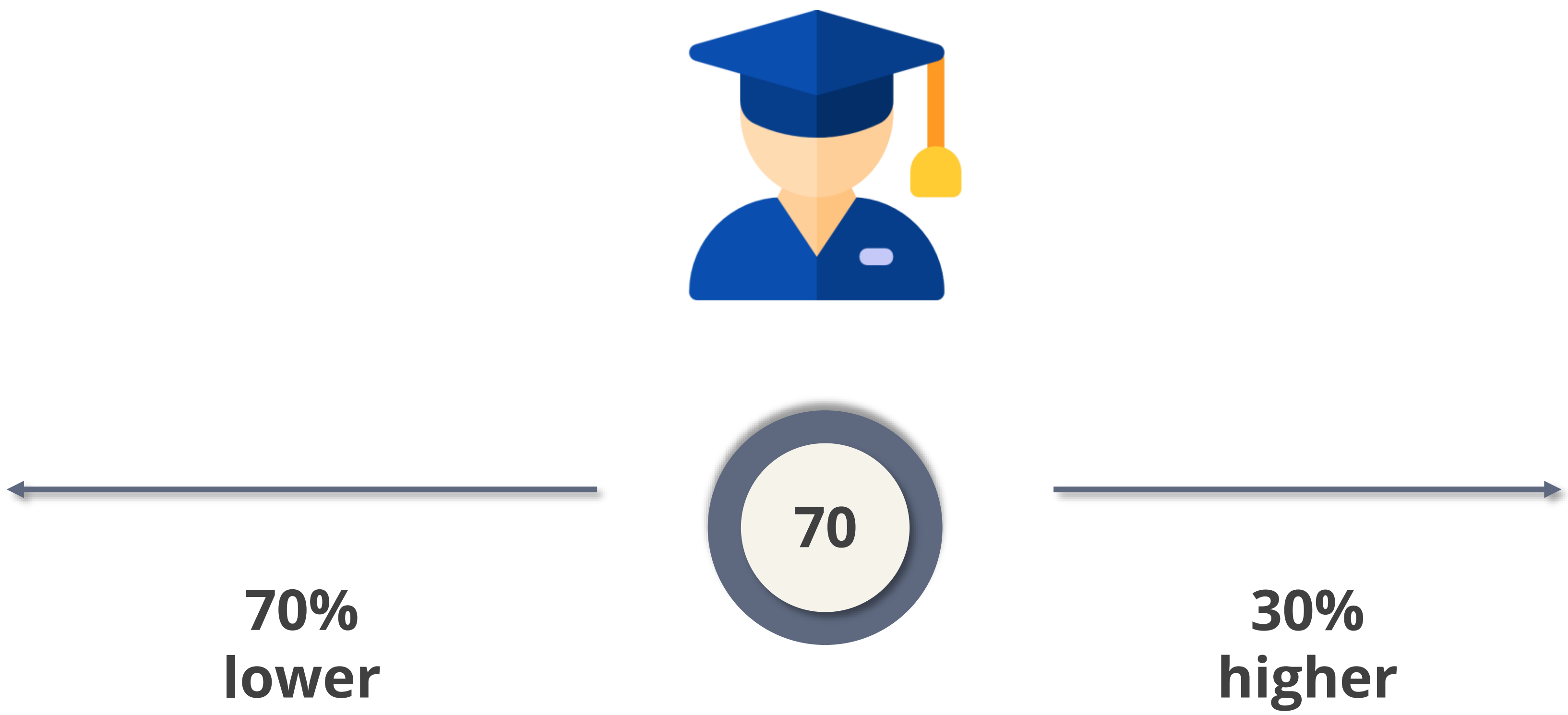
p^{th} percentile value

p percent of the data lies below it

In some selection tests, the candidates' raw scores are given as a percentile to show their relative position.

Percentile: Example

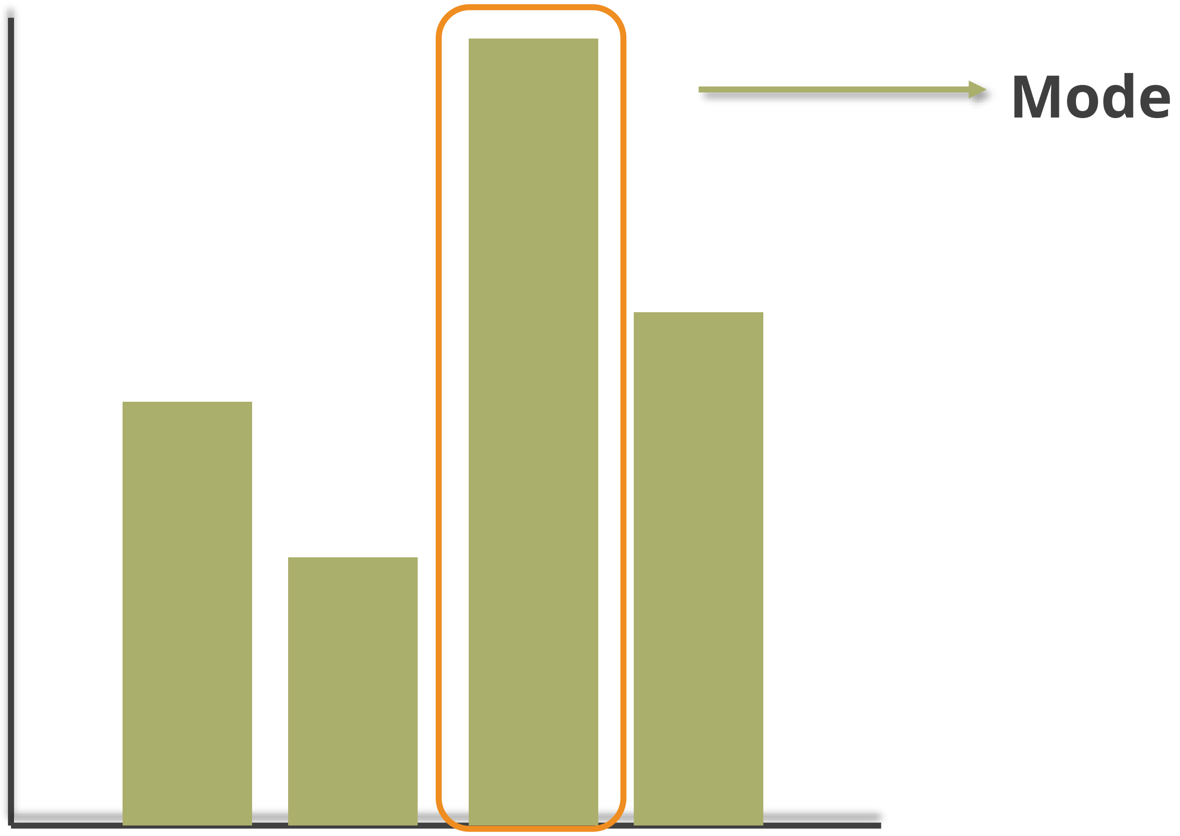
If a candidate’s percentile is 70, then 70% of other candidates scored lower than that student, while 30% of the candidates scored higher.



The percentile score is used in shortlisting and screening of candidates.

Mode

The mode is a measure of central tendency that identifies the most frequently occurring value in a dataset.



Consider a dataset: {1, 2, 2, 3, 4, 4, 4, 5, 6}

In this set, the number 4 appears three times, which is more frequently than any other number.

Therefore, 4 is the mode of this dataset.

Measures of Dispersion

Measures of Dispersion: Application

Consider these two datasets here

5, 15, 15, 25

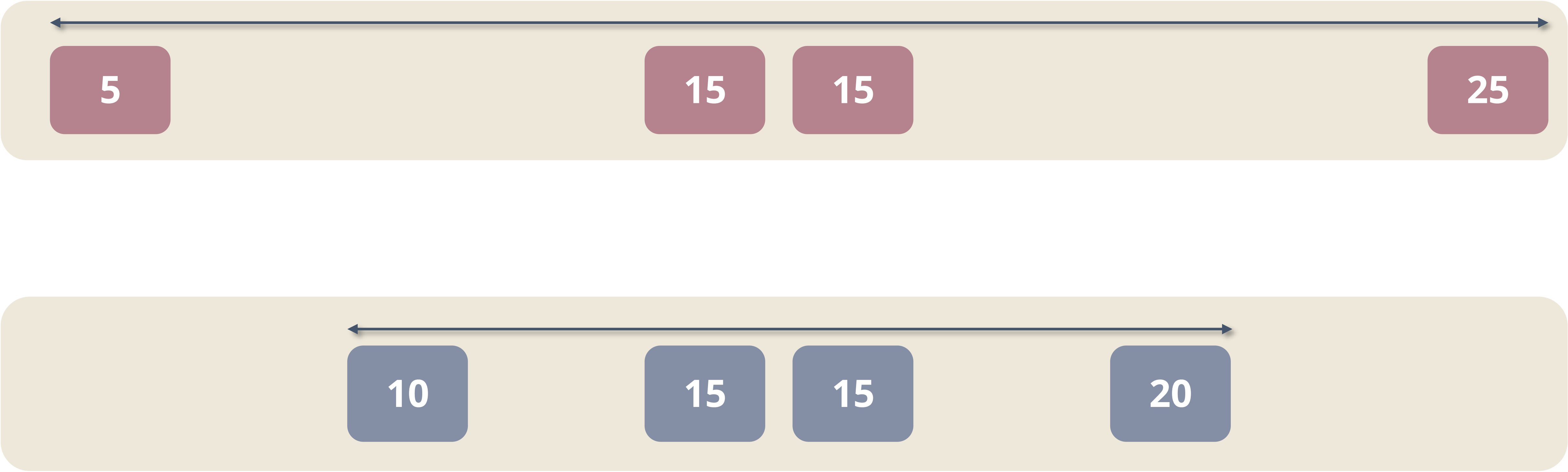
10, 15, 15, 20

The mean, median, and mode of the two datasets are:

15

Measures of Dispersion: Application

The first dataset is spread over a broader range than the second dataset.



The measures of dispersion are used to depict this difference statistically.

Example: Dispersion

Example: Scores of two players on four occasions:

Player 1	Player 2
5, 15, 15, 25	10, 15, 15, 20

The second player is more consistent as the dispersion of their score is less spread.

Range, Outliers, and Quartile Deviation

Range

The range is the simplest measure of dispersion.

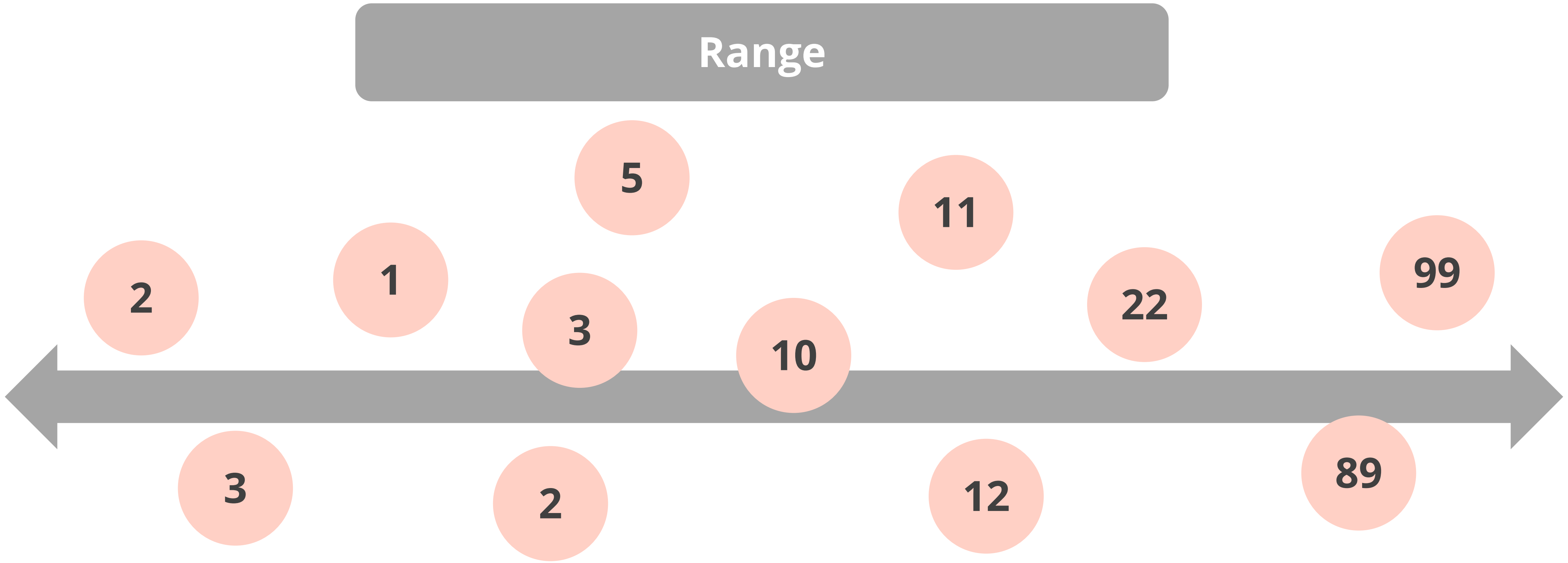
The range of data is defined as:

↓

The difference between the highest and lowest values in the dataset

Range

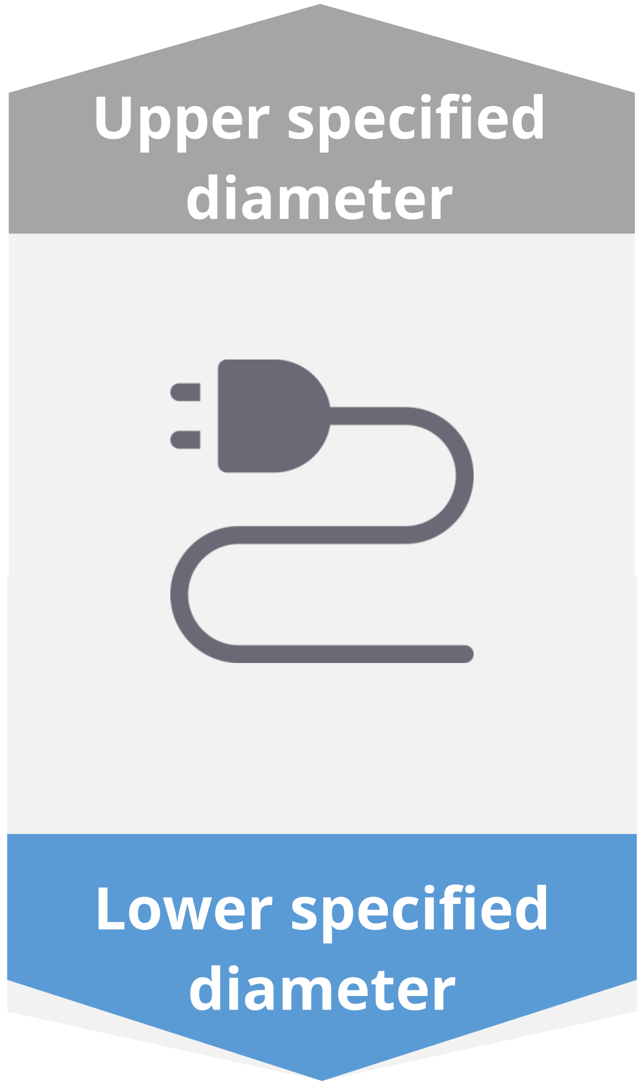
Range helps to detect and eliminate outliers.



This makes it a valuable tool in quality control studies.

Application of Range

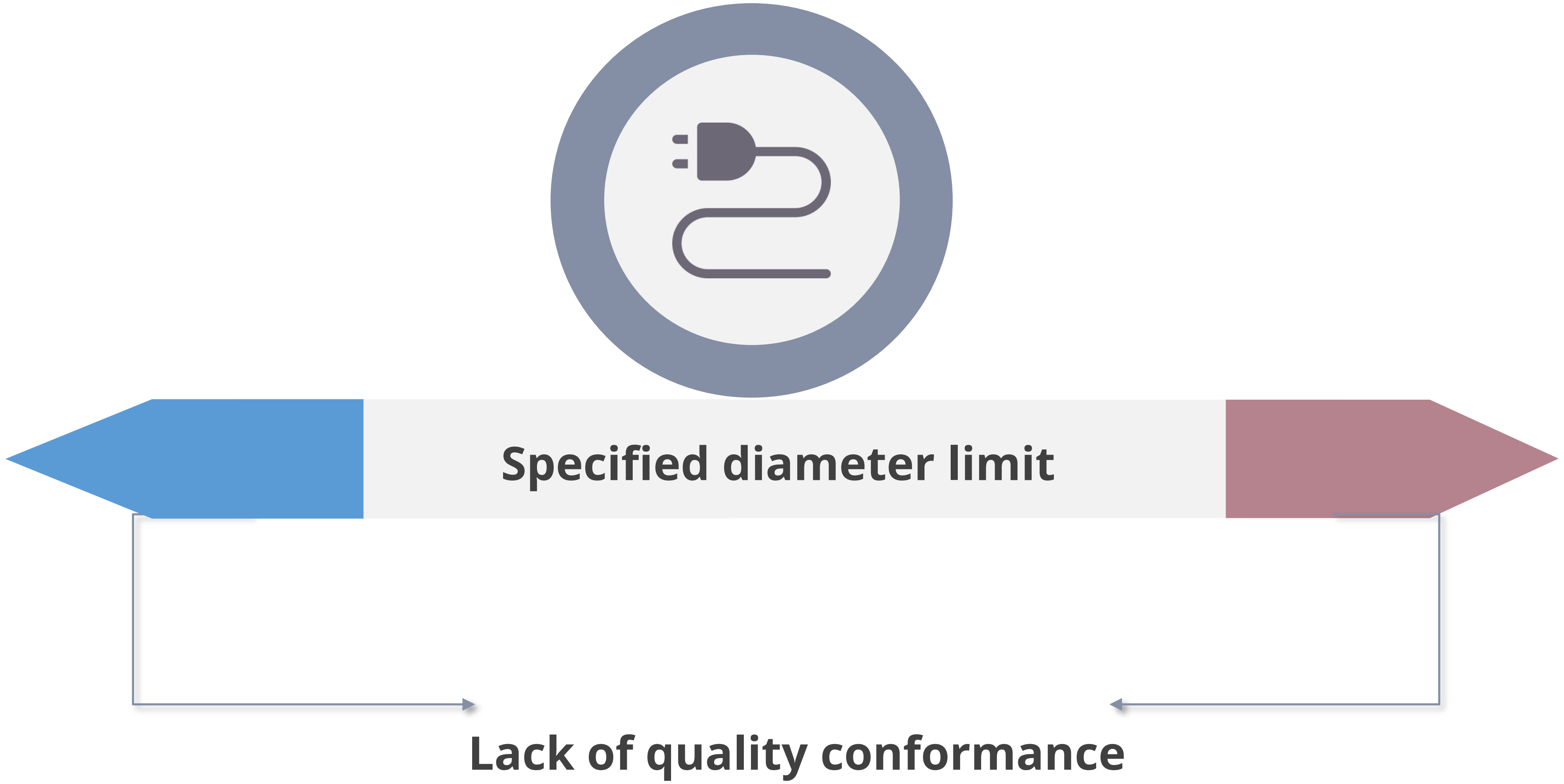
Example: A factory produces wires with lower and upper specification limits of specified diameters.



Periodically, samples of fixed size are drawn, and the diameters are noted.

Application of Range

A lack of quality conformance is detected when the range of diameters exceeds a predetermined limit.



This implies that not all diameters comply with specified limits.

Inter-Quartile and Quartile Deviations

The inter-quartile deviation is the difference between the first and third quartiles.

Inter-quartile deviation = $(Q_3 - Q_1)$

100% of the data

Q1 Q3

Quartile deviation = $\left(\frac{1}{2}\right)(Q_3 - Q_1)$

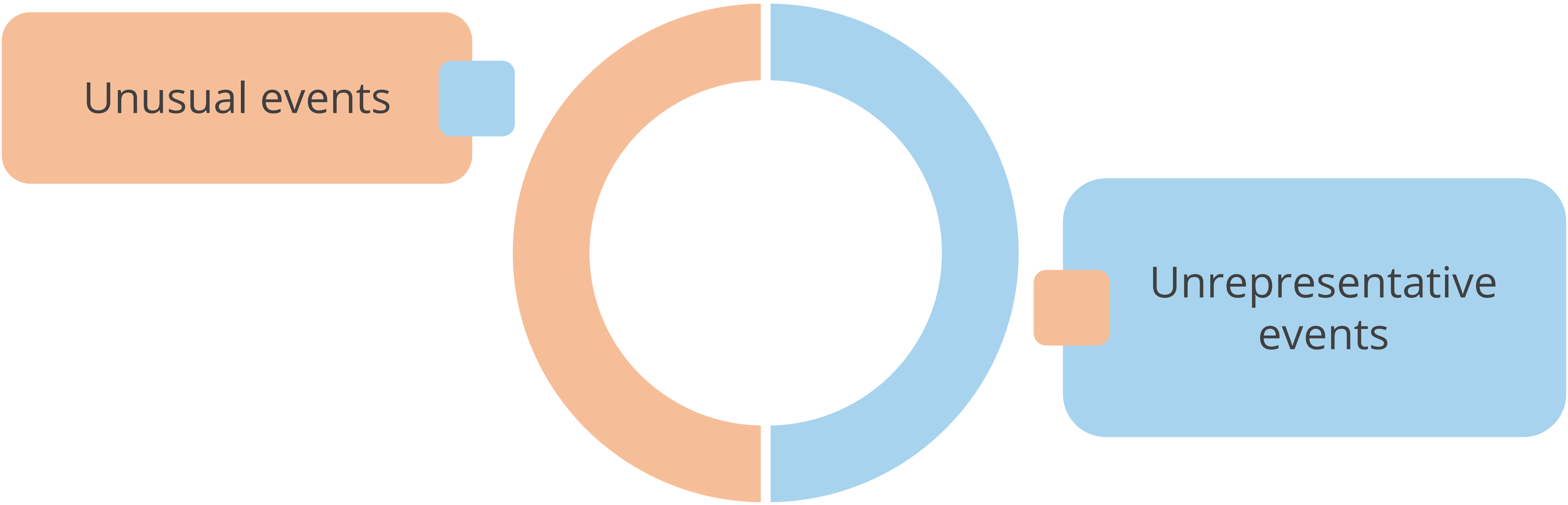
50% of the data

Q1 Q3

Quartile deviation, like the range, is not based on all observations.

Outliers

Outliers are extreme values in a dataset that do not represent the overall pattern.



Application of Outliers

Example: A machine broke down without being fixed for an unusually long time, as the external service mechanic's arrival was delayed.

The utilization of the machine is significantly lower during this period.

Application of Outliers

The observed value of capacity utilization is considered unrepresentative of reality and should be excluded.

Unrepresentative of reality

Leads to a distorted picture

Mean Absolute Deviation, Standard Deviation, and Variance

Measures of Dispersion

The two sets of data are:

5, 15, 15, 25

10, 15, 15, 20

Arithmetic mean

15

Measures of Dispersion

The values of deviations in the two datasets from the mean are:

-10, 0, 0, 10

-5, 0, 0, 5

Total deviations

0

Deviation Values

Sum of deviations from mean = 0

↓

Average of deviations from mean = 0

↓

Average deviations cannot be used to measure dispersion

Mean Deviation

The average of absolute deviations is called the mean deviation.

Mean Deviation = $\frac{1}{n}\Sigma|x_i - \bar{x}|$

$\sum$ = Sigma is a summation of all points in the sample or set

x = Observations

$\bar{x}$ =Mean

n = The number of observations

Standard Deviation

Standard deviation is a measure of how dispersed the data is in relation to the mean.

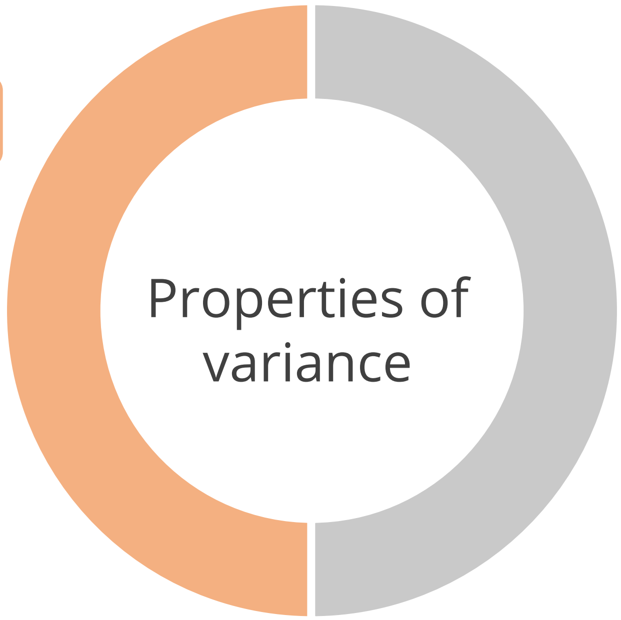
$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

The square of the standard deviation is called variance.

Variance

Variance is a measure of the scatter of the squared distances measured from the mean. It is indicated by the symbol σ^2.

The result is always non-negative.



Variance always has a squared unit.

Variance Formula: Population

The formula for the variance of the population dataset is:

$$\sigma^2 = \left(\frac{1}{n}\right) \Sigma (x_i - \mu)^2$$

σ^2 = Population variance
$\sum$ = The total sum
n = The number of observations
x_i = i^{th} observation in the population
μ = Population mean

Variance Formula: Sample

The formula for the variance of a sample dataset is:

$$s^2 = \left(\frac{1}{n-1}\right)\Sigma(x_i - \bar{x})^2$$

$\sum$ = The total sum

x_i = Observations

$\bar{x}$ = Sample Mean

n = Number of observations

Variance and Standard Deviation

The variances and standard deviations of the two datasets are:

5, 15, 15, 25

Variances = 50

SD = 7.07

10, 15, 15, 20

Variances =12.5

SD = 3.53

Mean Deviation vs. Standard Deviation

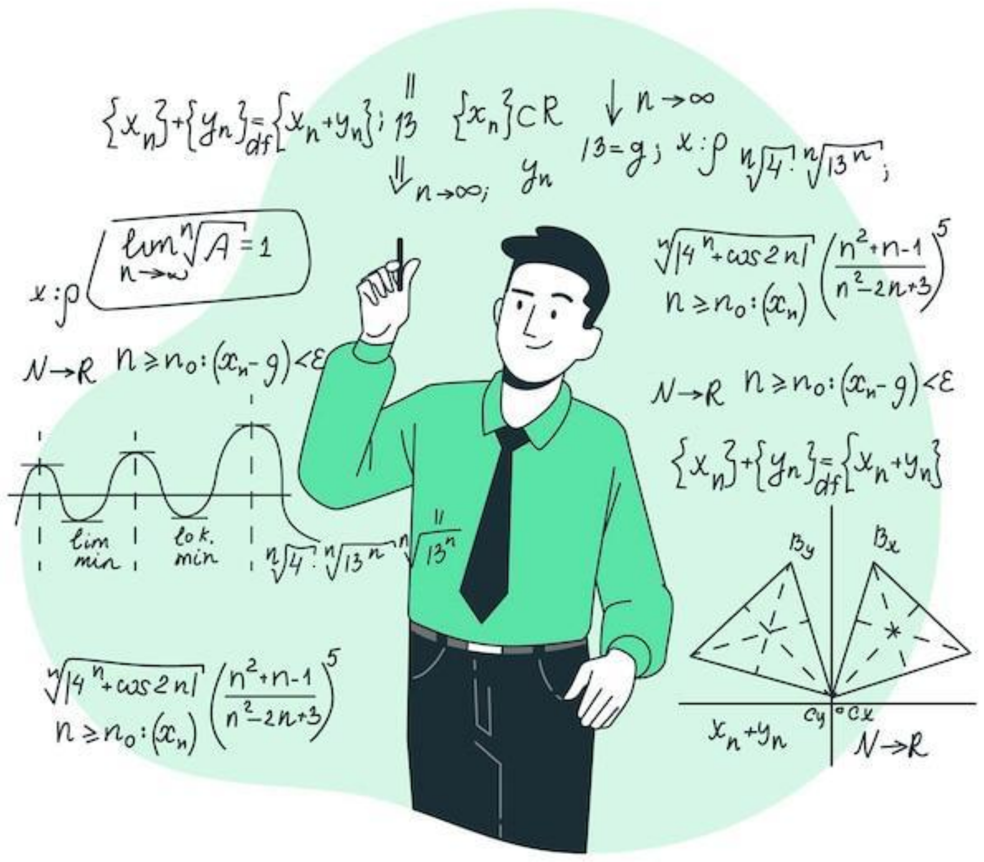
When compared with mean deviations, the standard deviation magnifies larger deviations.

Mean deviations	Standard deviations
5	7.07
2.5	3.53

Therefore, SD is a better measure to capture the magnitude of spread or dispersion.

Mean Deviation vs. Standard Deviation

Like the arithmetic mean, the standard deviation is also amenable to further mathematical treatment.

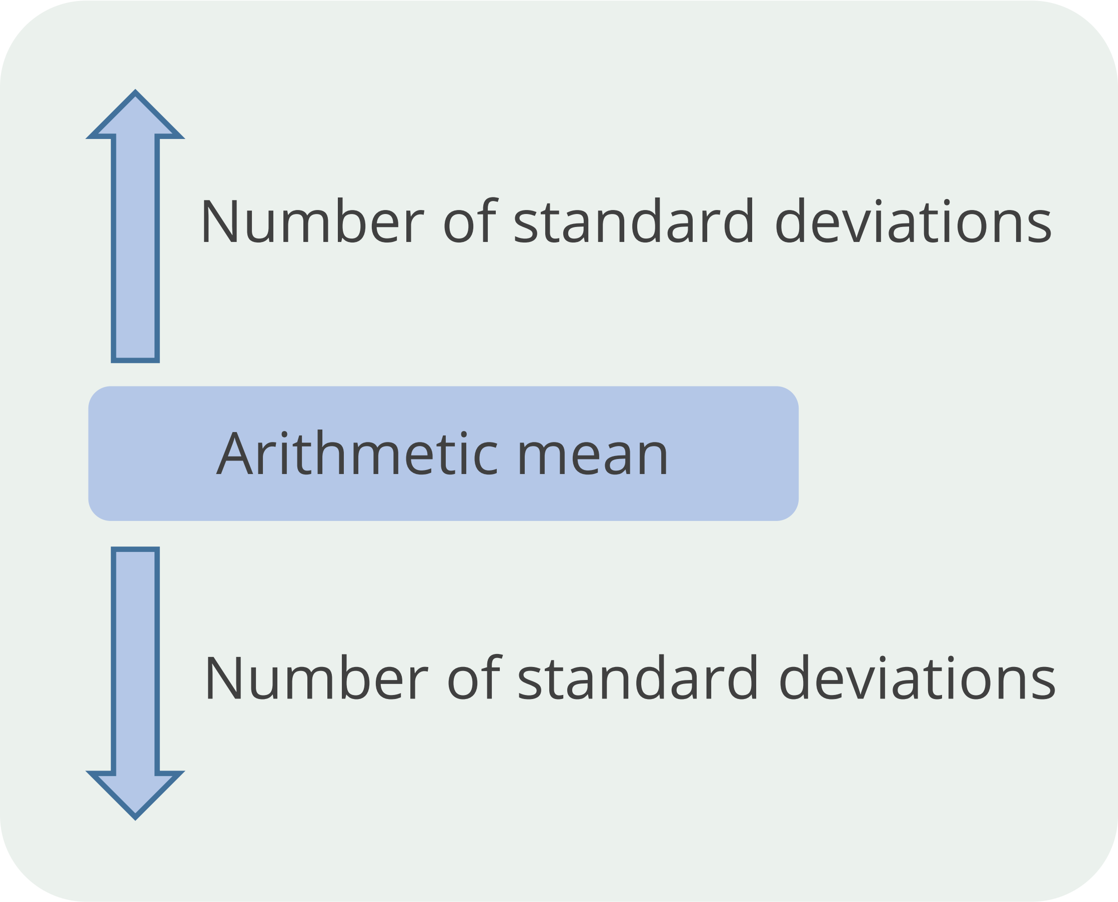


It is the most used measure of dispersion.

Z-Score or Standard Score and Empirical Rule

Z-Score

The Z-score is a statistical measure that describes the relationship of a value to the mean of a set of values.



Z-score is also known as a standard score.

Z-Score Formula

Suppose a dataset consists of n values:

$$x_1 + x_2 + x_{3+\ldots} \ldots + x_n$$

Mean ($\bar{x}$)

Standard deviation (s)

The standard score of the jth value is obtained as:

$$z_j = \frac{(x_j - \bar{x})}{s}$$

Standard Score

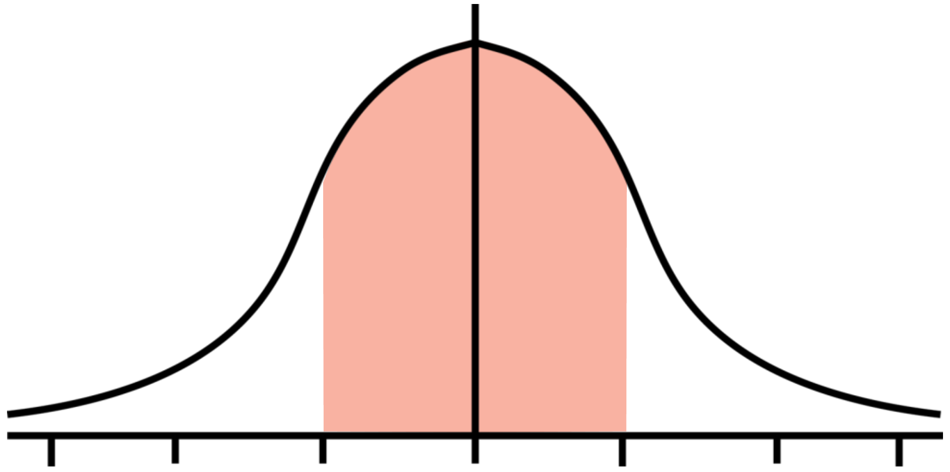
A standard score of two indicates that the value in the dataset exceeds the arithmetic mean by two times the standard deviation.

Values in the dataset above the arithmetic mean have positive standard scores.

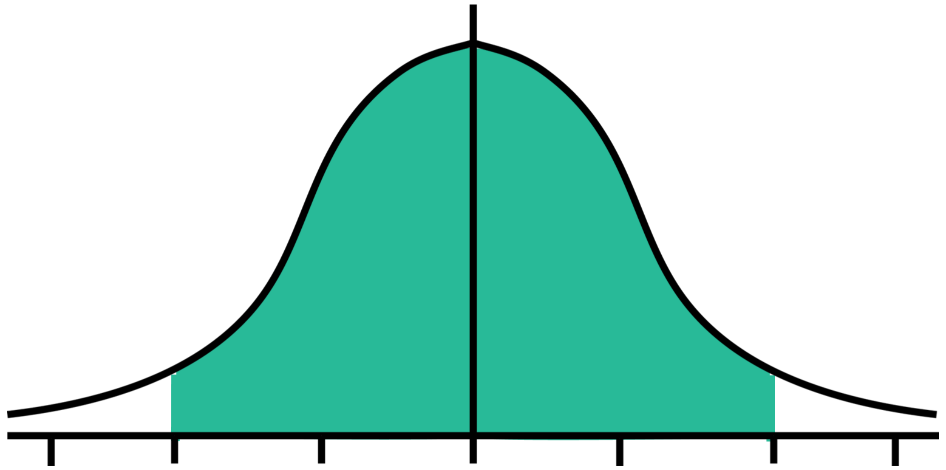
Values below the arithmetic mean have negative standard scores.

Empirical Rule

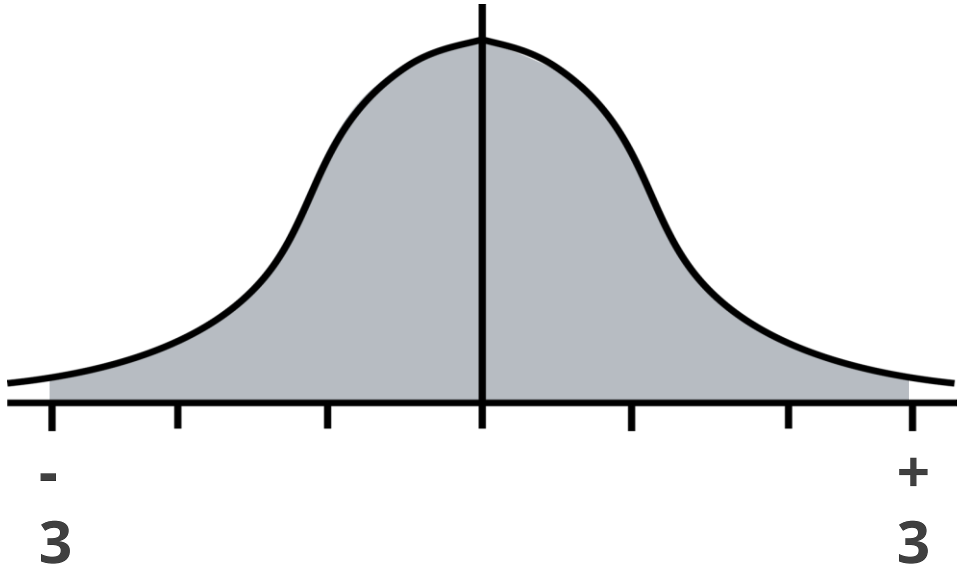
The Empirical Rule tells us how much of the data lies within the one, two, and three standard deviations.



About 65% of the data have Z-scores between -1 and +1



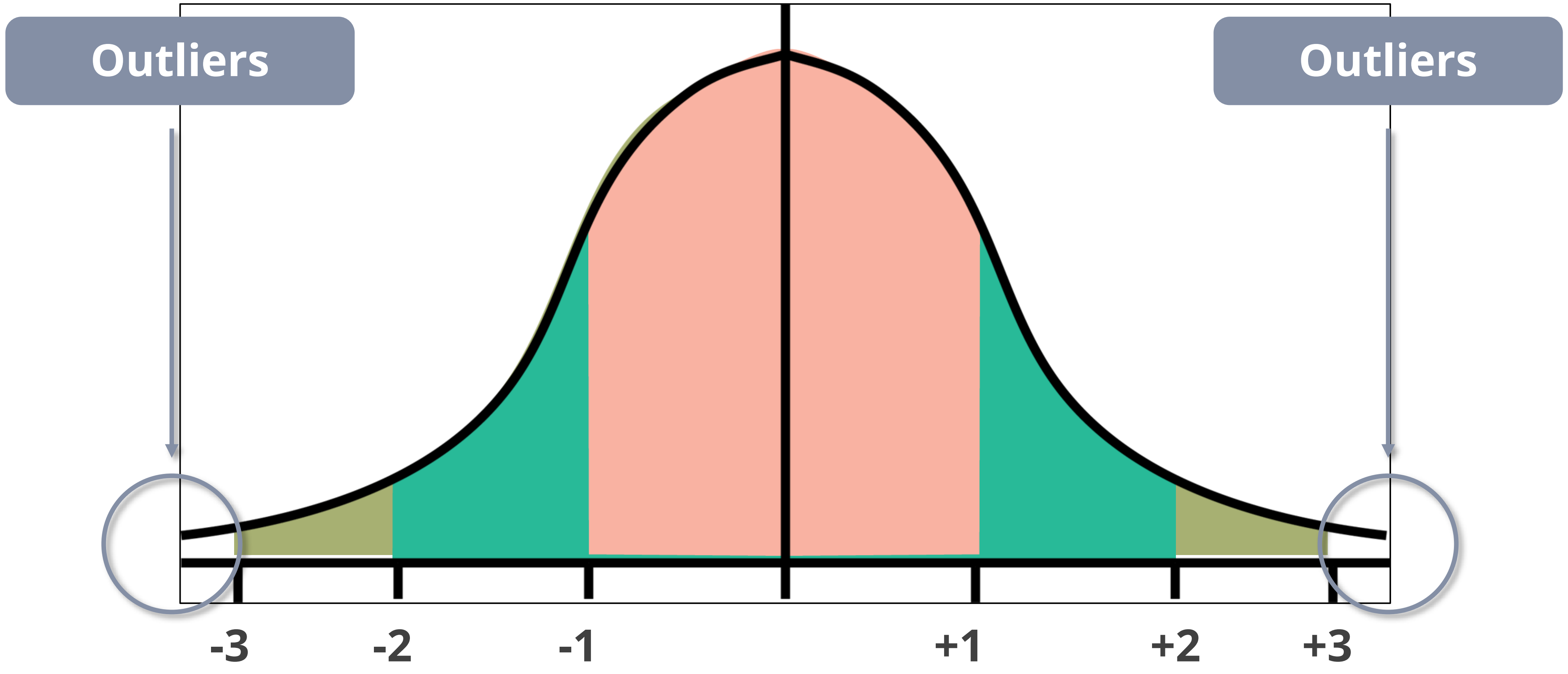
About 95% of the data have Z-scores between -2 and +2



About 99.7% of the data have Z-scores between -3 and +3

Outliers

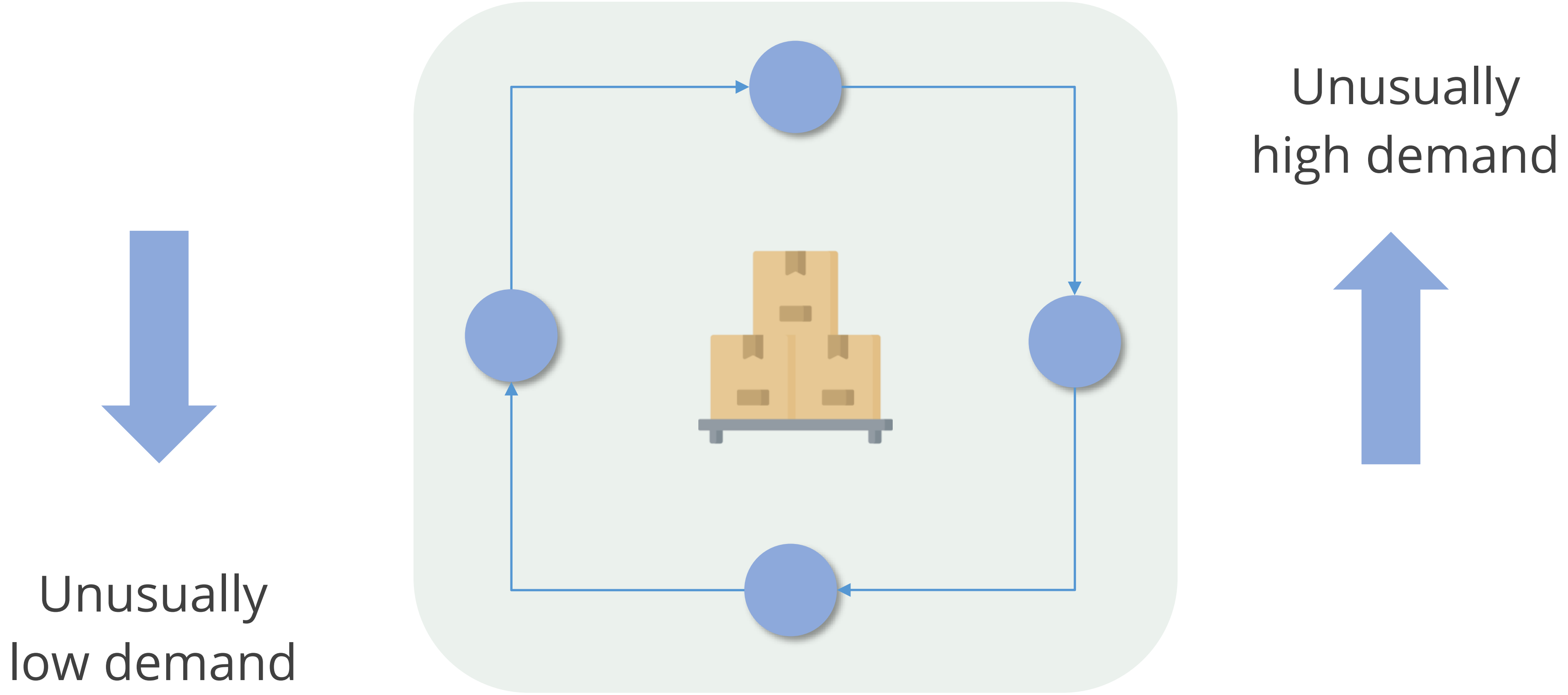
When Z-scores lie outside the range (-3, 3), there is a strong probability of an outlier.



However, all large deviations from the mean cannot be considered outliers.

Z-Score Example

Example: If there is an unusually large or small value of demand for a product observed in some period, it can be due to a transition from one phase to another in the product's life cycle.

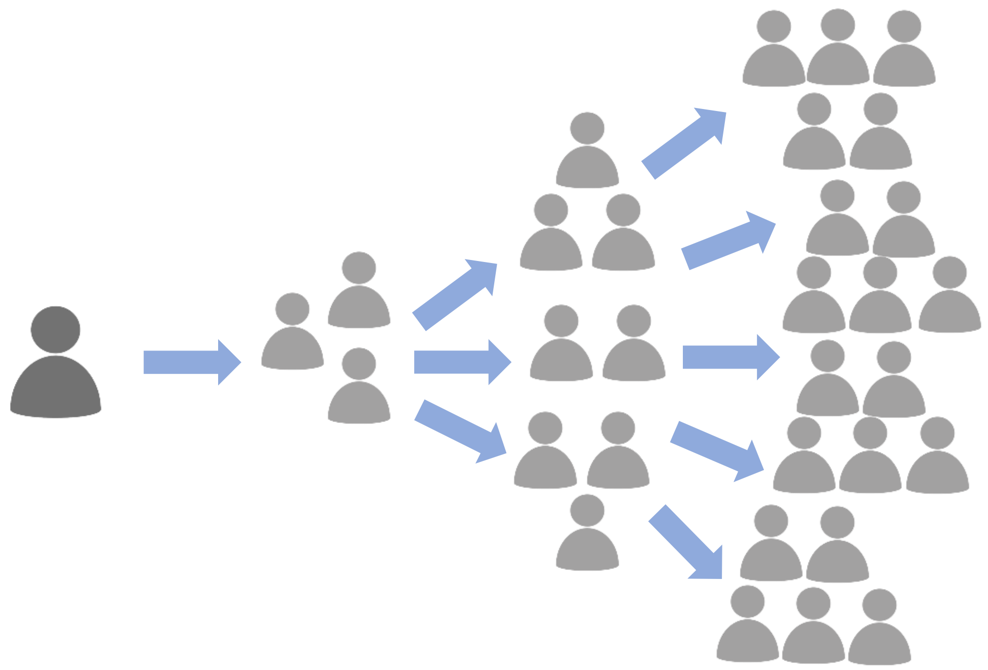


It is important to understand the change in demand patterns to plan further.

Coefficient of Variation and Its Application

Coefficient of Variation

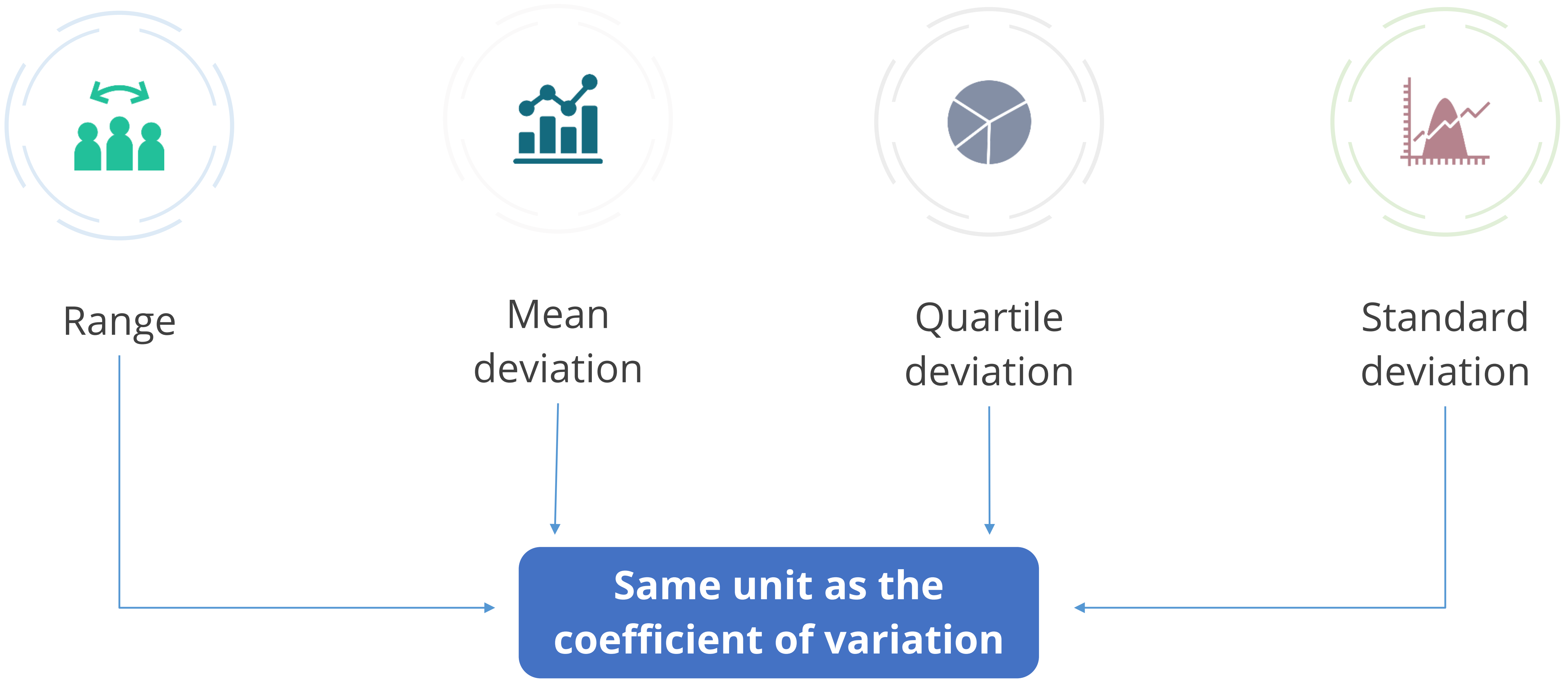
The coefficient of variation is a statistical measure of a data series' relative dispersion around the mean.



Coefficient of Variation

Unit of Coefficient of Variation

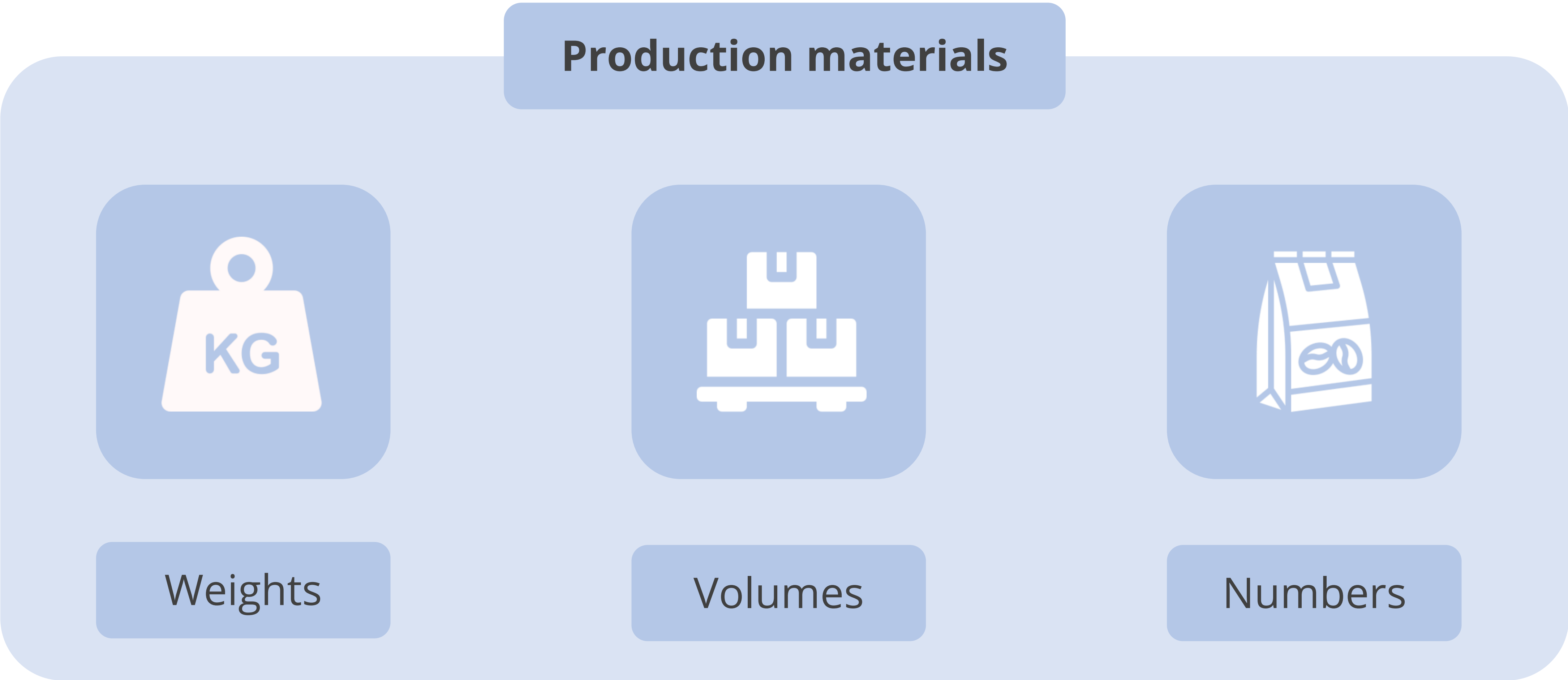
The coefficient of variation is the same unit for the dispersion measures listed below.



It is also important to compare the dispersion of two datasets whose units differ.

Coefficient of Variation in Comparison

Example: In a factory, the units of different materials used in production vary.



The consumption of certain materials is quantified in weights or volumes, while the consumption quantities of packing bags are measured in numbers.

Coefficient of Variation in Comparison

To compare the dispersion in consumption, one must use a coefficient of variation.

Coefficient of variation = (Standard Deviation / Mean) * 100

It is the ratio of the standard deviation to the arithmetic mean expressed as a percentage.

Measure of Consistency

The coefficient of variation can be used to assess consistency.

When two datasets represent the scores of two different people, the person with the lower coefficient of variation is thought to be more consistent.

Coefficient of Variation as Measure of Consistency

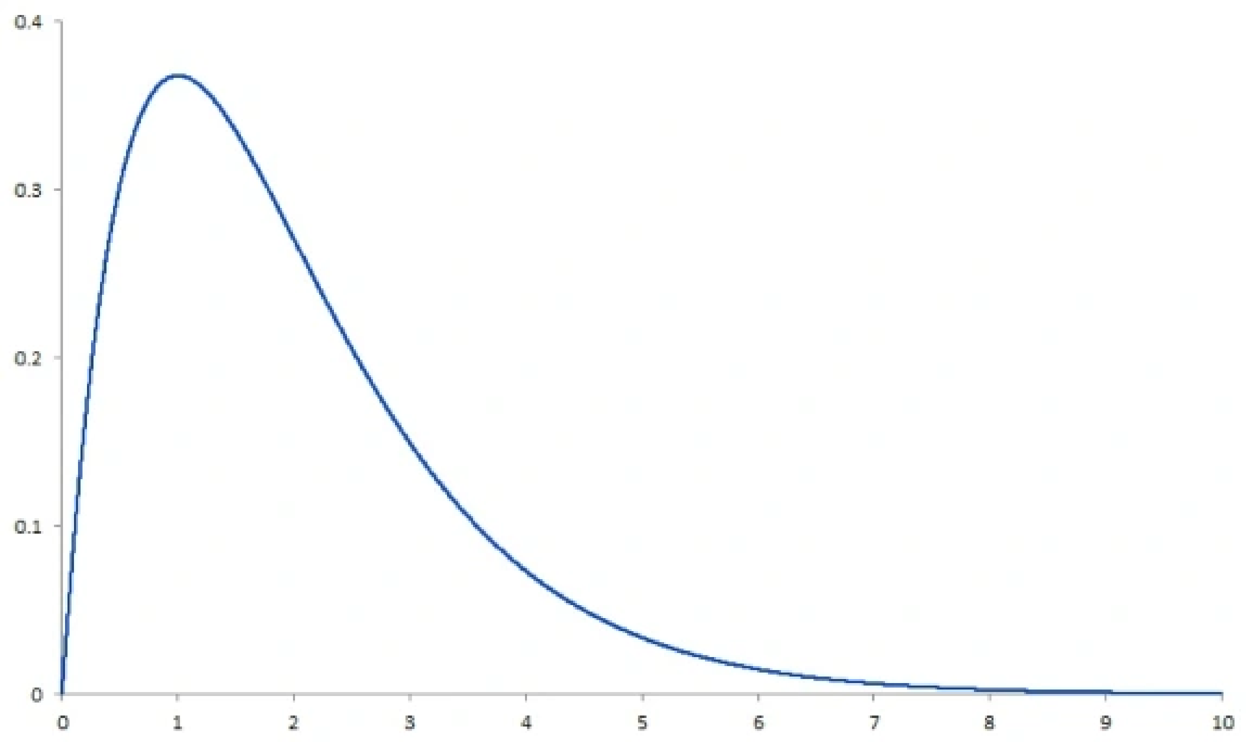
The dataset with a smaller coefficient of variation is less dispersed than the dataset with a larger coefficient of variation.

Coefficient of variation = 47.13	>	Coefficient of variation = 23.53
Less consistent		More consistent
More dispersed		Less dispersed

Measures of Shape

Skewness

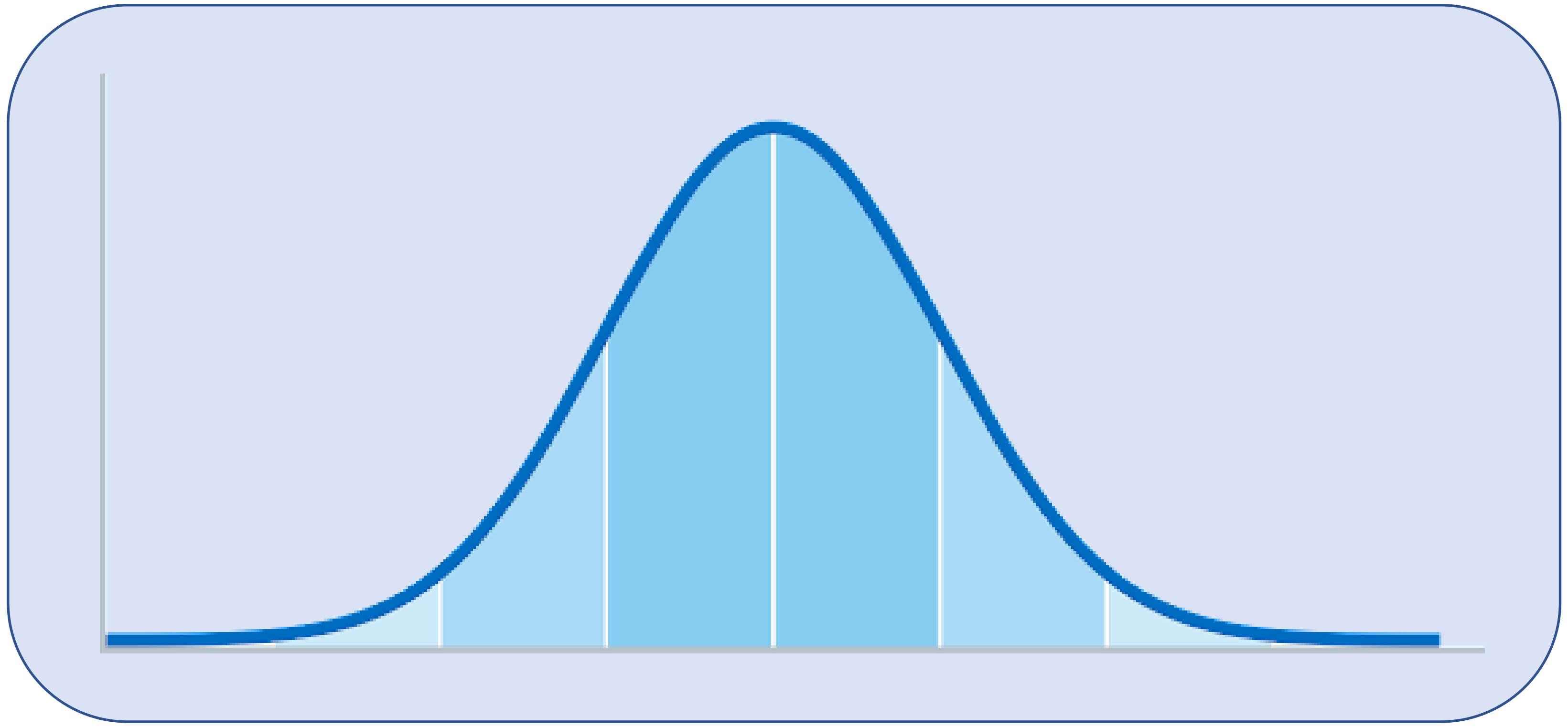
Data can be illustrated graphically, as shown below:



Skewness is a measure used to describe the shape of a dataset when presented graphically.

Normal Distribution

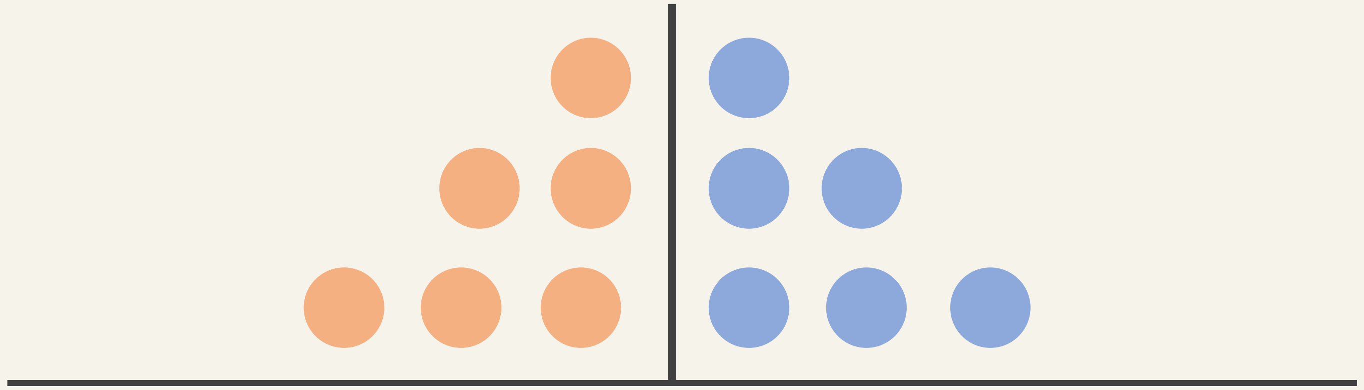
The process of reorganizing data within a database so that users can use it for further queries and analysis is known as data normalization.



It is the procedure for generating clean data.

Skewness in Normal Distribution

Data is symmetrically distributed in a normal distribution, and the skewness is zero because all measures of central tendency are in the middle.

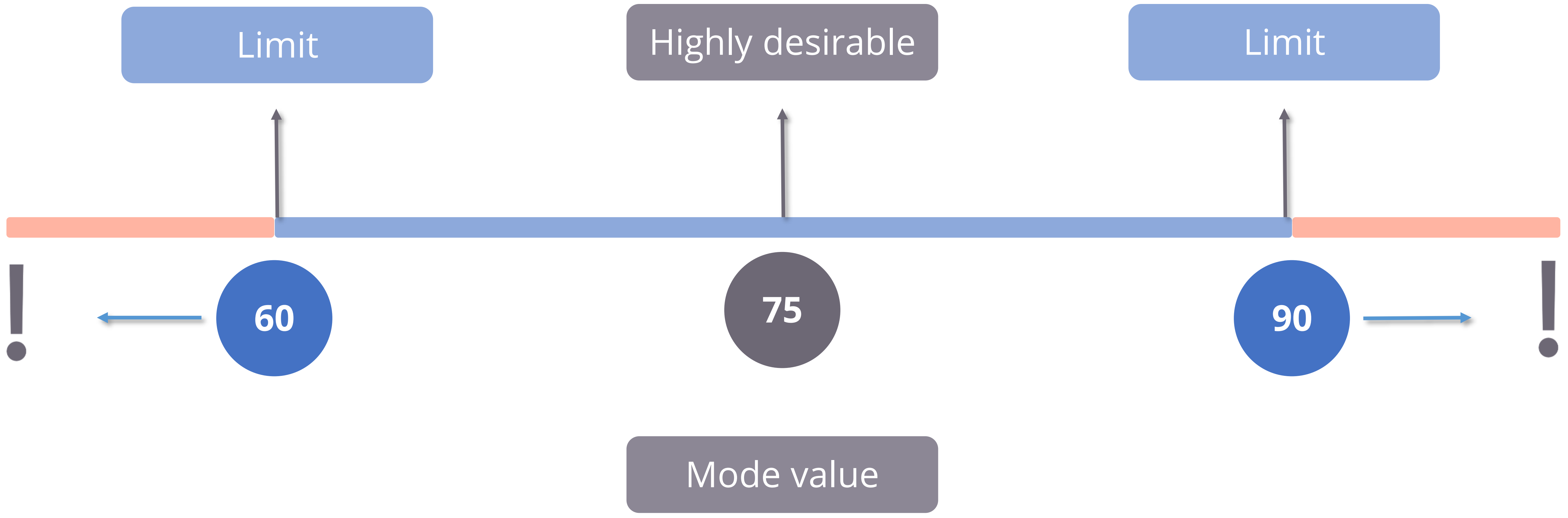


Skewness = 0

Mean = Median = Mode

Skewness: Example

Quality is a characteristic of specification limits of 60-90.



The average of these limits is 75, and it is highly desirable to produce units with that quality characteristic value of 75. Exceeding 90 or falling below 60 can be a concern.

Skewness: Example

Example: Quality as a characteristic to measure

The manufacturing process should ideally be designed so that values below and above 75 by any magnitude are equally likely to occur.

75

-K

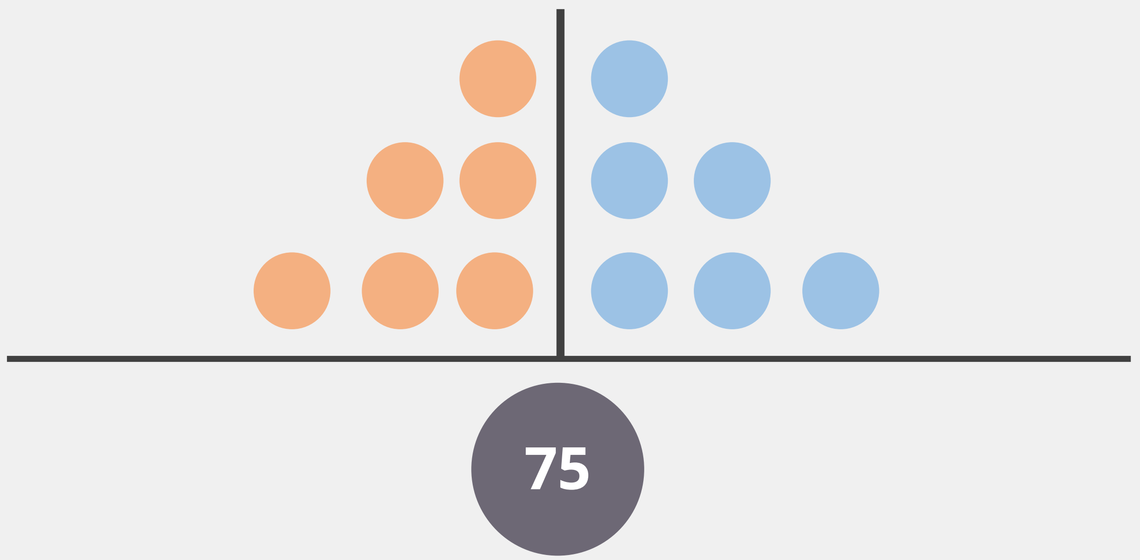
+K

m(75 + K) = m(75 – K)

Skewness: Example

When the values of variables appear at regular frequencies or intervals around the mean, this is referred to as symmetric data.

The frequency curve will then be symmetric around 75.



The arithmetic mean will also be 75.

For symmetric datasets, the three values are identical.

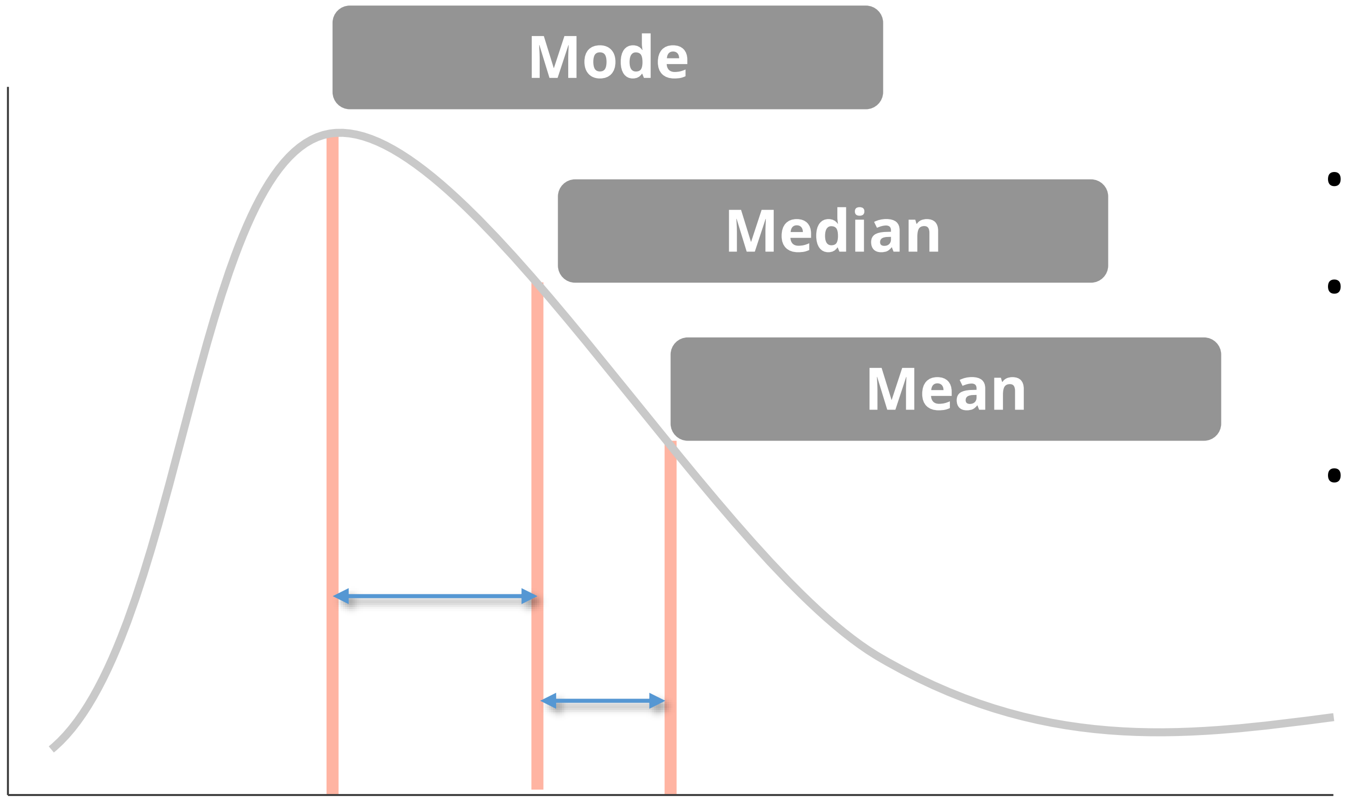
Empirical relation

Mean – Mode = 3*(Mean – Median)

The median is 75.

Positively Skewed

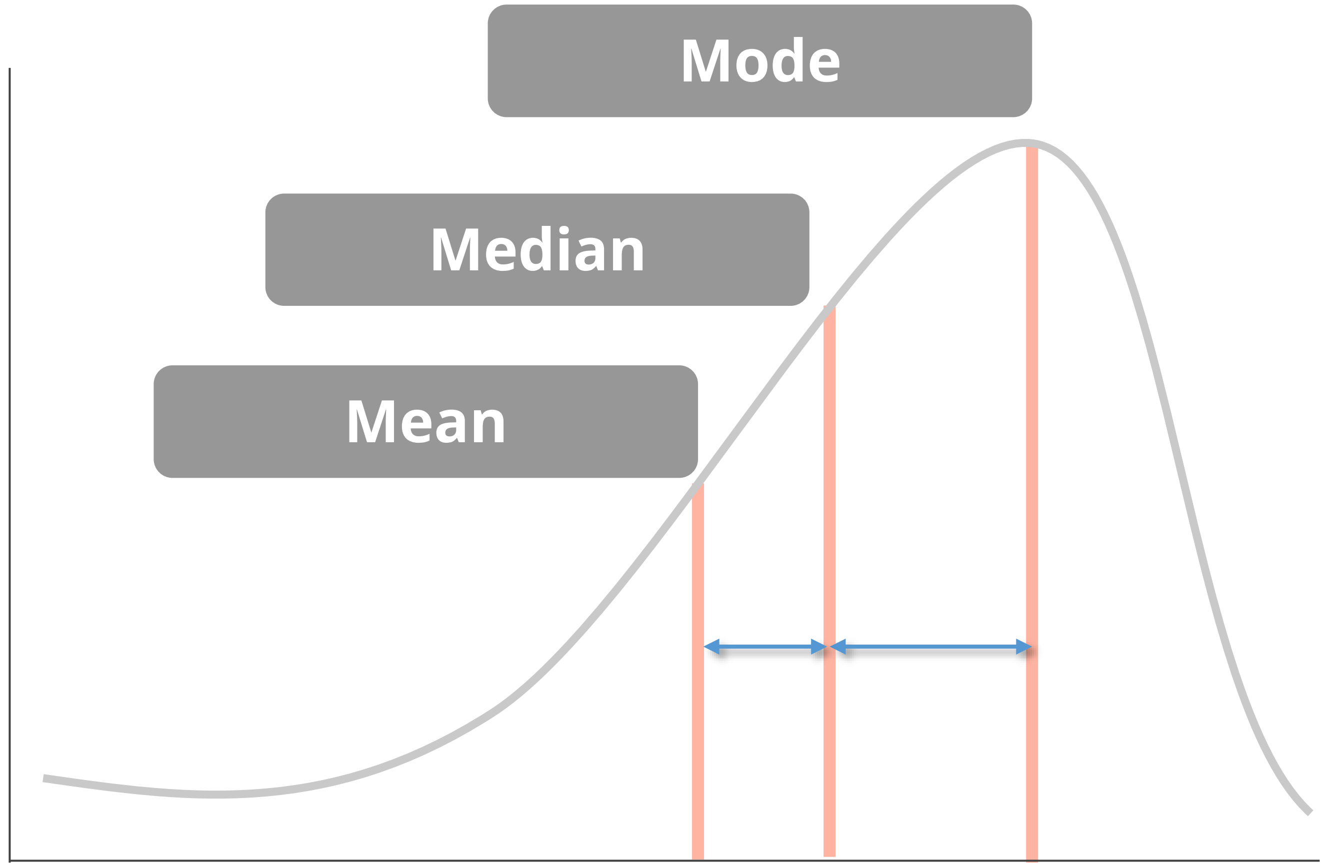
Given that the dataset is positively skewed, the frequency curve has a long tail to the right.



- The mode value occurs to the left.
- The arithmetic mean is impacted by large values, making it larger than the mode.
- The median, however, will lie in between and be closer to the mean.

Negatively Skewed

The difference between the mean and mode indicates the direction and magnitude of skewness.



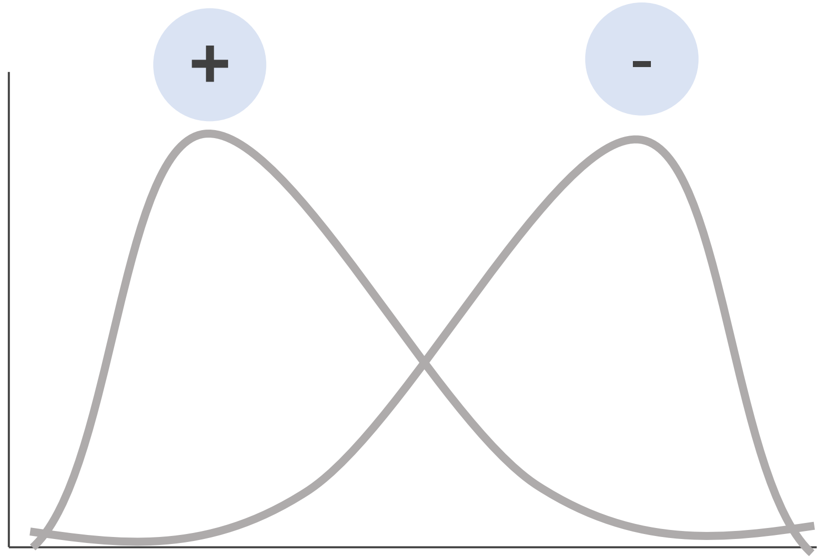
Measures of Shape of a Dataset

Bowley formulated a way to obtain dimensionless measures.

Bowley's coefficient of skewness

= (Mean – Mode)/Standard Deviation

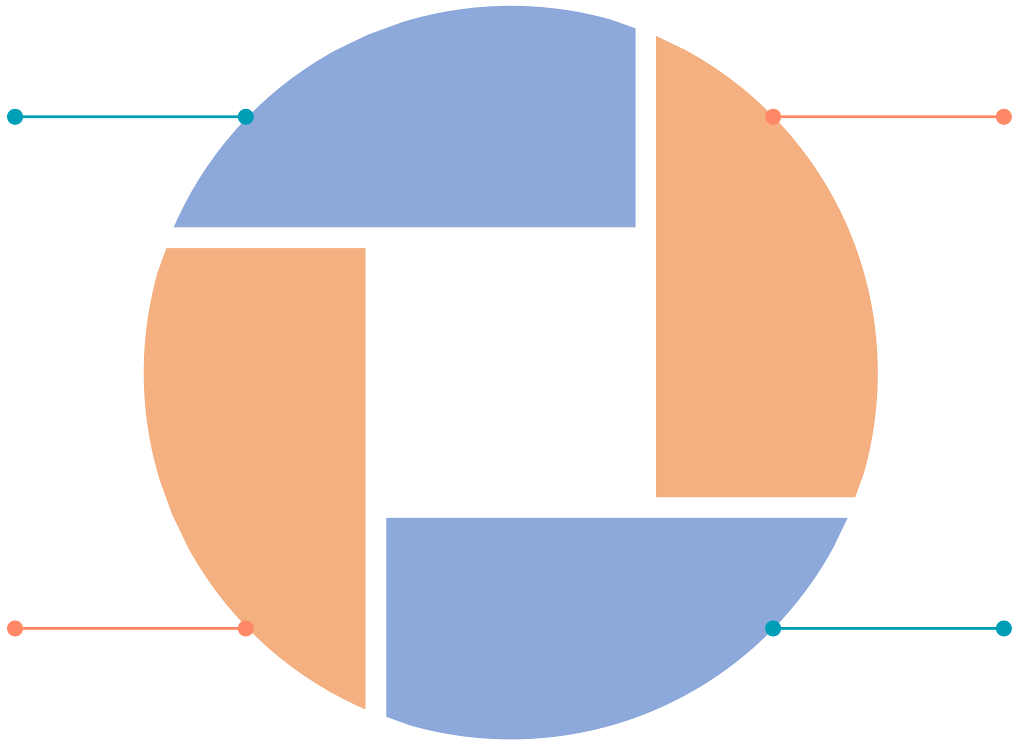
= 3 * (Mean – Median)/Standard Deviation



Depending on the negative or positive deviation, the long tail will be to the left or right.

Summary Measures

The important summary measures are:



Measures of central tendency

Measures of dispersion

Measures of skewness

Quartiles

Kurtosis

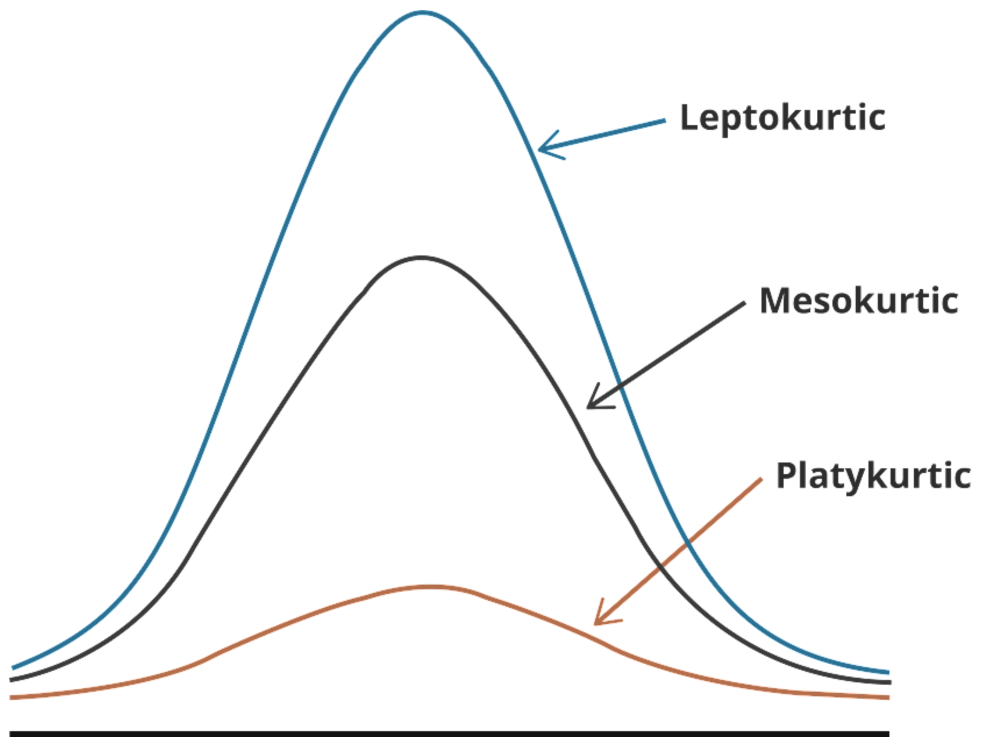
It is a statistical measure used to describe the distribution of observed data around the mean. It indicates the heaviness of the tails of a distribution.

Kurtosis identifies the tails and sharpness of a distribution.

- If the distribution is tall and thin, it is said to have a high kurtosis.
- A distribution is said to have low kurtosis if it is short and broad.

Kurtosis

There are three types of kurtosis: leptokurtic, mesokurtic, and platykurtic.

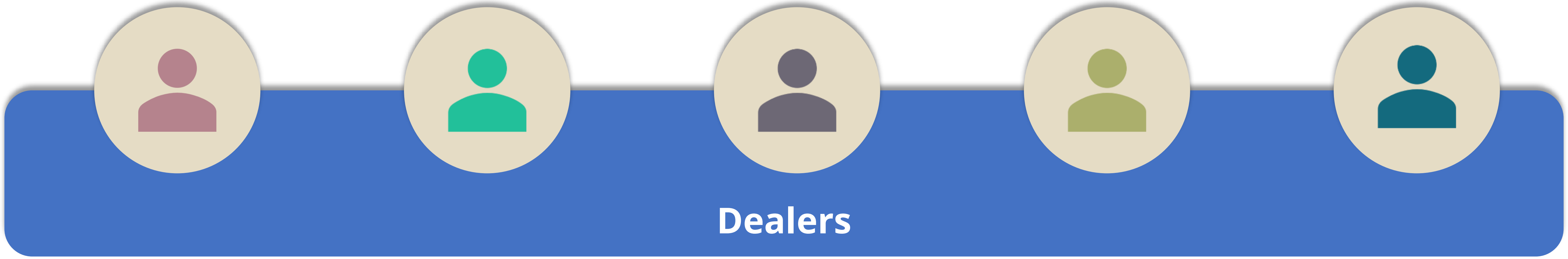


- **Leptokurtic distributions** have a positive kurtosis value and exhibit heavy tails on either side compared to a normal distribution.
- **Mesokurtic distributions** are like a normal distribution and exhibit neither heavy nor light tails.
- **Platykurtic distributions** have a negative kurtosis value and exhibit lighter tails than a normal distribution.

Case Study: Descriptive Statistics

Problem Statement

A factory dispatches its spare products to five dealers each day.



Problem Statement

Quantities requested by a dealer are received and delivered on the same day.

Request

Delivery

Problem Statement

The following must be determined:

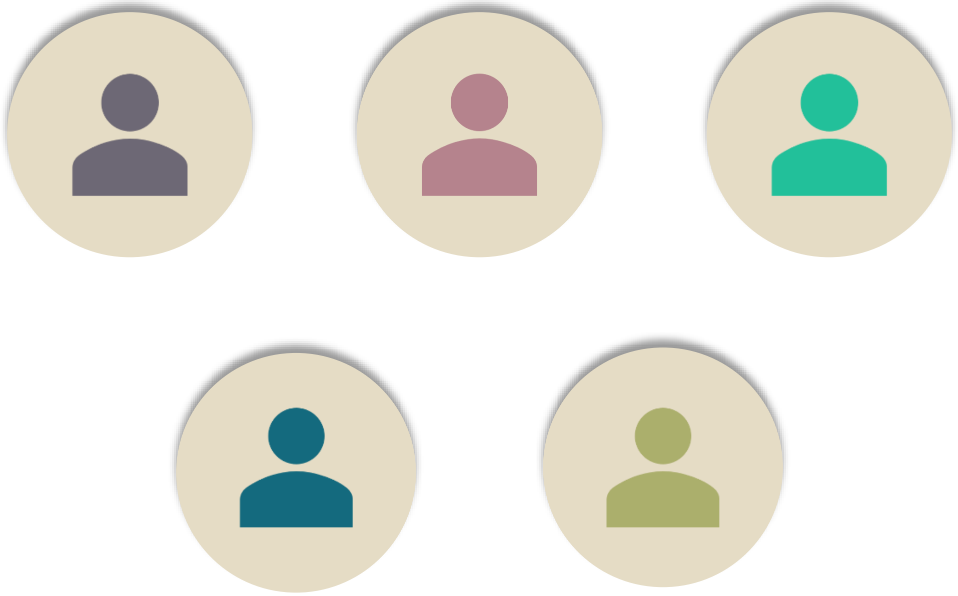
	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6
Dealer 1	83	67	85	74	62	82
Dealer 2	83	85	82	75	69	69
Dealer 3	85	66	77	84	69	75
Dealer 4	73	91	82	85	76	83
Dealer 5	81	82	76	74	70	61
Factory	405	391	402	392	346	370

The table shows the daily dispatches that each dealer gets in a span of six days and the total dispatches by the factory.

Problem Statement

The following must be determined:

Requirement variation at the dealer level

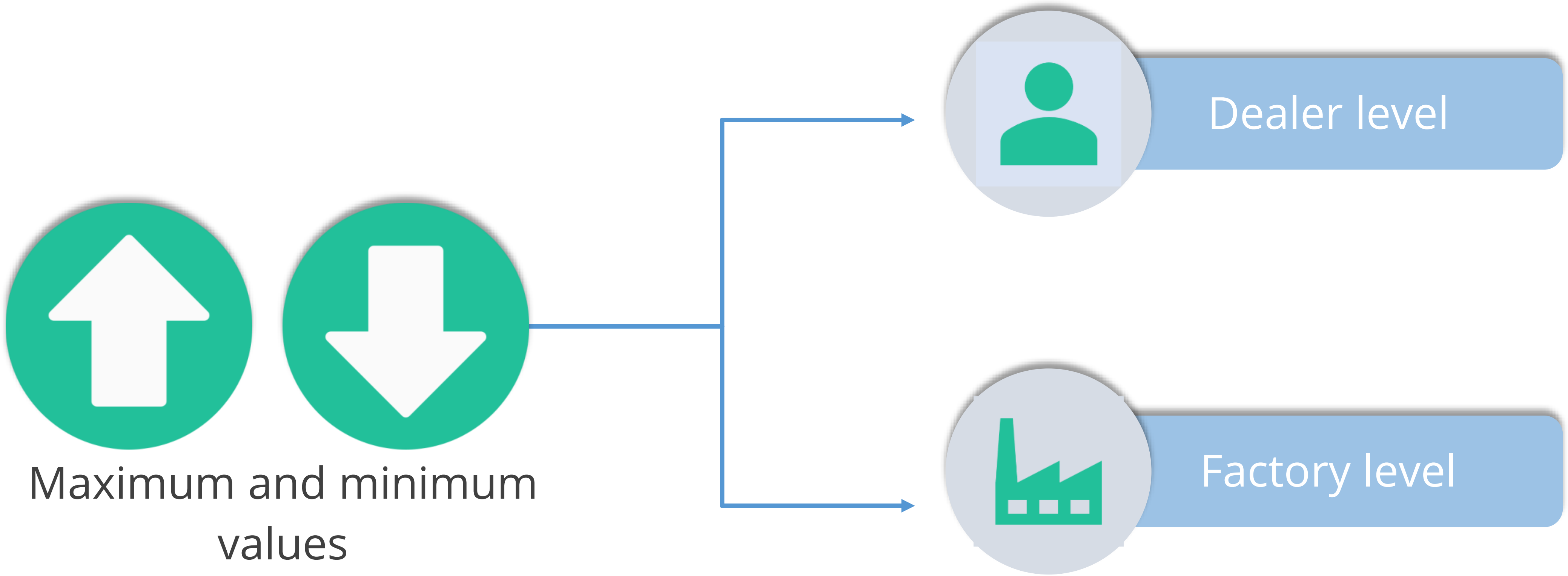


Production variation at the factory level

The extent to which the requirements vary at the dealer level and subsequently affect the production at the factory level

Problem Statement

Determine findings and Z-value



The Z-value will help one plan the production at the factory.

Solution

Calculate the mean, standard deviation, and coefficient of variation for dealer 1.

Mean

Standard deviation

Coefficient of variation

Dealer 1

Solution

The mean for dealer one is calculated as shown below:

Mean

Dealer 1

(1/6) * {83+67+85+74+62+82}

= 453/6

= 75.5

Note

If there are N values

$6 \longrightarrow N$

Summation range: 1 to x

Solution

Standard deviation for dealer 1 is calculated as shown below:

	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Mean	Square root
Values	83	67	85	74	62	82	75.5	
Deviation from mean	7.5	-8.5	9.5	-1.5	-13.5	6.5		
Squares of deviation	56.25	72.25	90.25	2.25	182.25	42.25	74.25	8.61684397

Standard deviation

s=√1/n−1∑ni=1(xi−x̄)2

Solution

The coefficient of variation for dealer one is calculated as shown below:

Dealer 1

Coefficient of variation

(Standard deviation/Mean) * 100

= 8.61684/75.5 * 100

= 11.413

Solution

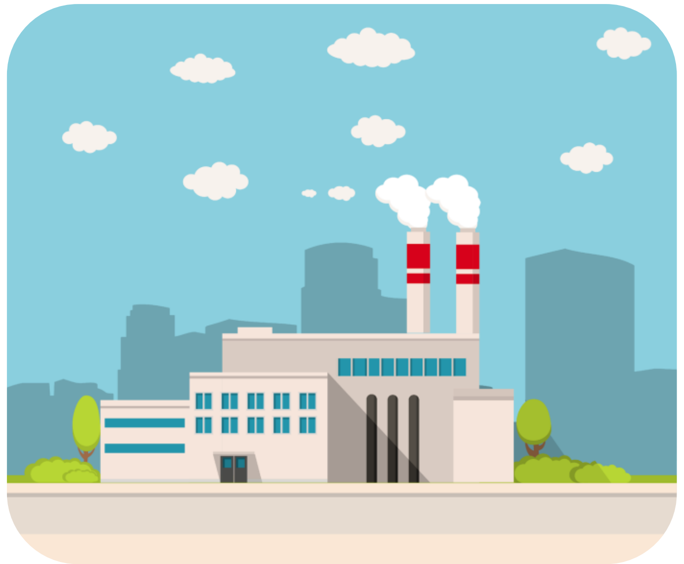
To find the degree variations, the coefficient of variations (C.V) for both the dealers and the factory must be calculated.

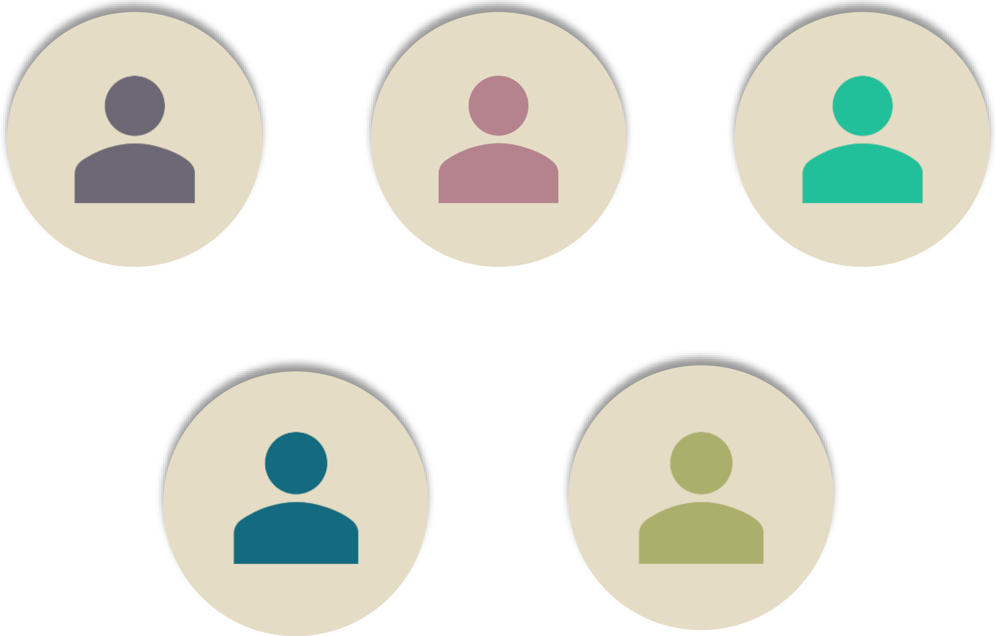
	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6	MEAN	S.D	C.V
Dealer 1	83	67	85	74	62	82	75.5	8.61684	11.413
Dealer 2	83	85	82	75	69	69	77.1667	6.5426	8.47853
Dealer 3	85	66	77	84	69	75	76	7.02377	9.2418
Dealer 4	73	91	82	85	76	83	81.6667	5.8784	7.19804
Dealer 5	81	82	76	74	70	61	74	7.0946	9.5873
Factory	405	391	402	392	346	370	384.333	20.4831	5.3295

Solution

After calculating the coefficient of variation for all the dealers and the factory, the results indicate that:

Coefficient of variation at the factory level < Coefficient of variation at the dealers' level





Solution

The variation is relatively higher for the first dealer as compared to the others.

	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6	MEAN	S.D	C.V
Dealer 1	83	67	85	74	62	82	75.5	8.61684	11.413
Dealer 2	83	85	82	75	69	69	77.1667	6.5426	8.47853
Dealer 3	85	66	77	84	69	75	76	7.02377	9.2418
Dealer 4	73	91	82	85	76	83	81.6667	5.8784	7.19804
Dealer 5	81	82	76	74	70	61	74	7.0946	9.5873
Factory	405	391	402	392	346	370	384.333	20.4831	5.3295

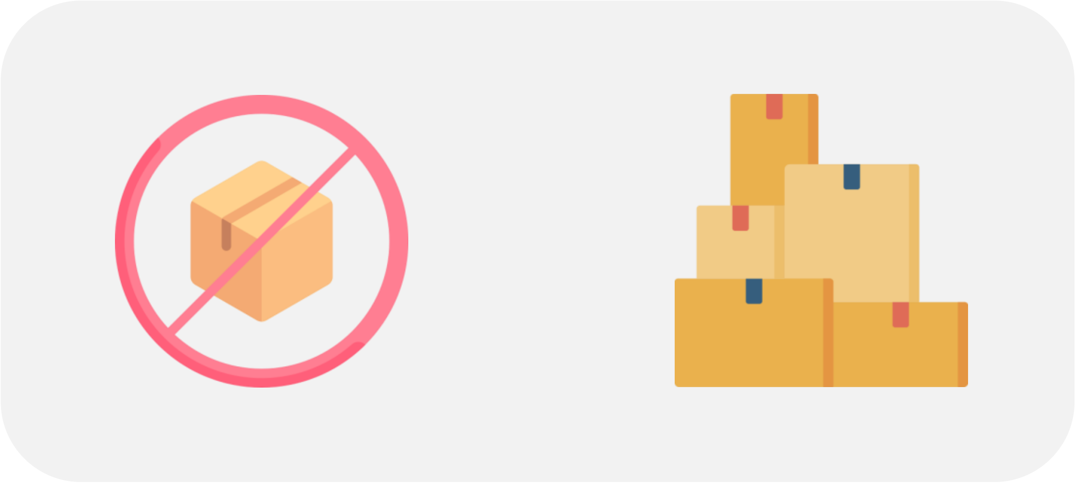
Solution

In general, high values of the coefficient of variation indicate that the level of uncertainty in demand is high.

High coefficient of variation

High level of uncertainty in demand

Risk of stock shortages or surplus stocks



In such cases, the dealer or factory could run the risk of shortages or surplus stocks.

Solution

When the value of the coefficient of variation is low, such risks are less.

Low coefficient of variation → Low risk

Determination of Z-Value

The table lists the maximum (MAX) and minimum (MIN) values and the corresponding Z-values (ZMAX and ZMIN).

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Mean	S.D	MAX	MIN	ZMAX	ZMIN
Dealer 1	83	67	85	74	62	82	75.5	8.61684	85	62	1.10249	-1.5667
Dealer 2	83	85	82	75	69	69	77.1667	6.5426	85	69	1.19728	-1.2482
Dealer 3	85	66	77	84	69	75	76	7.02377	85	66	1.28136	-1.4237
Dealer 4	73	91	82	85	76	83	81.6667	5.8784	91	73	1.58773	-1.4743
Dealer 5	81	82	76	74	70	61	74	7.0946	82	61	1.12762	-1.8324
Factory	405	391	402	392	346	370	384.333	20.4831	405	346	1.00896	-1.8715

Determination of Z-Value

Calculation of Z-value:

Max	Min
85	62
85	69
85	66
91	73
82	61
405	346

The highest of the six values for a dealer

The lowest of the six values for a dealer

The column **Max** in the table indicates the highest of the six values for a dealer, and the column **Min** indicates the lowest value.

Determination of Z-Value

Dealer 1 has values 83, 67, 85, 74, 62, and 82.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Max	Min
Dealer 1	83	67	85	74	62	82	**85**	**62**
Dealer 2	83	85	82	75	69	69	**85**	**69**
Dealer 3	85	66	77	84	69	75	**85**	**66**
Dealer 4	73	91	82	85	76	83	**91**	**73**
Dealer 5	81	82	76	74	70	61	**82**	**61**
Factory	405	391	402	392	346	370	**405**	**346**

The maximum and minimum values for the other dealers are also obtained.

Determination of Z-Value

Calculate the Z-score:

Z-Score

(Value - Mean)/Standard Deviation

For Dealer 1:

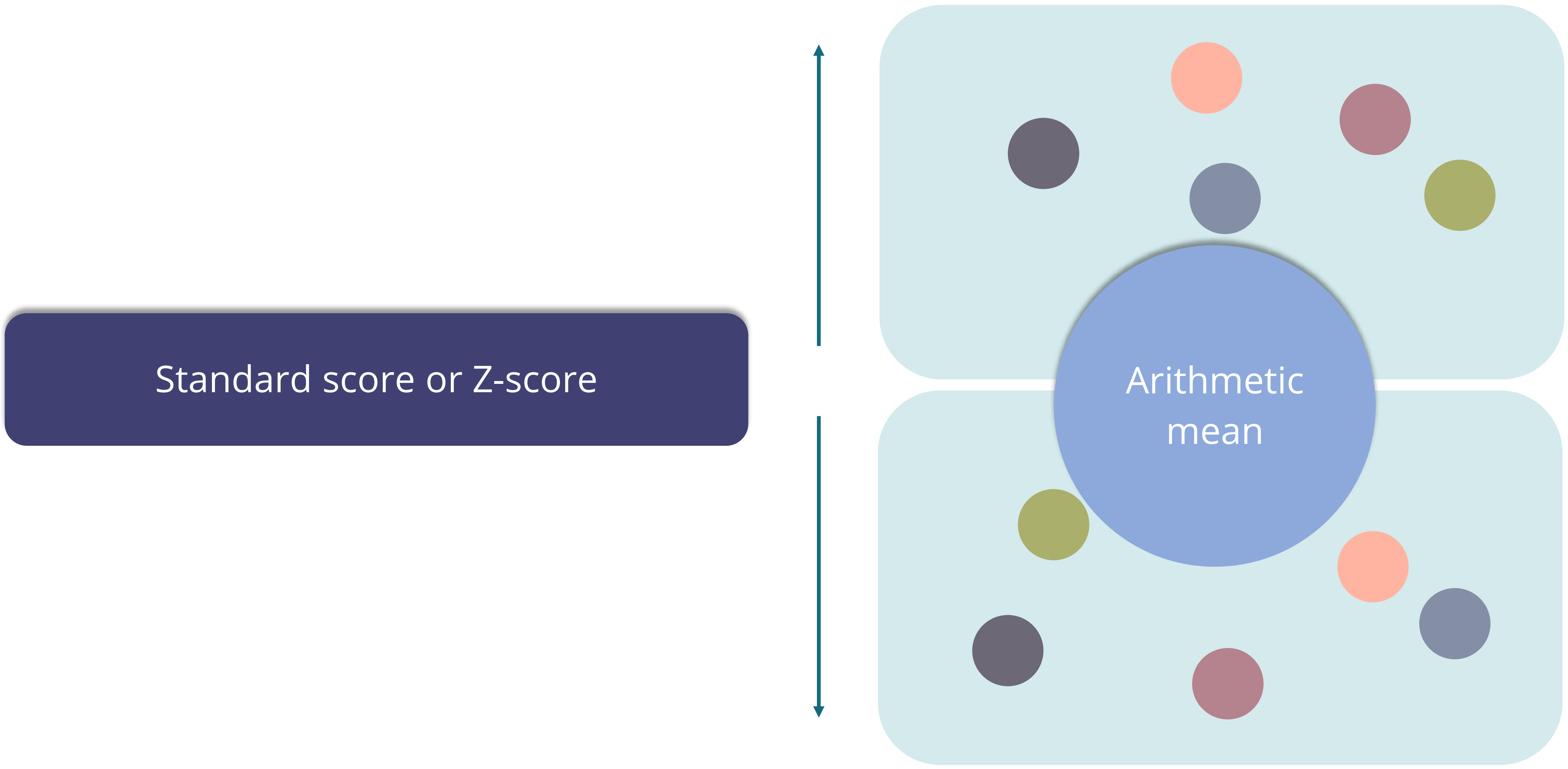
Max value = 85

ZMAX = (85 – 75.5)/8.61684

= 1.10249

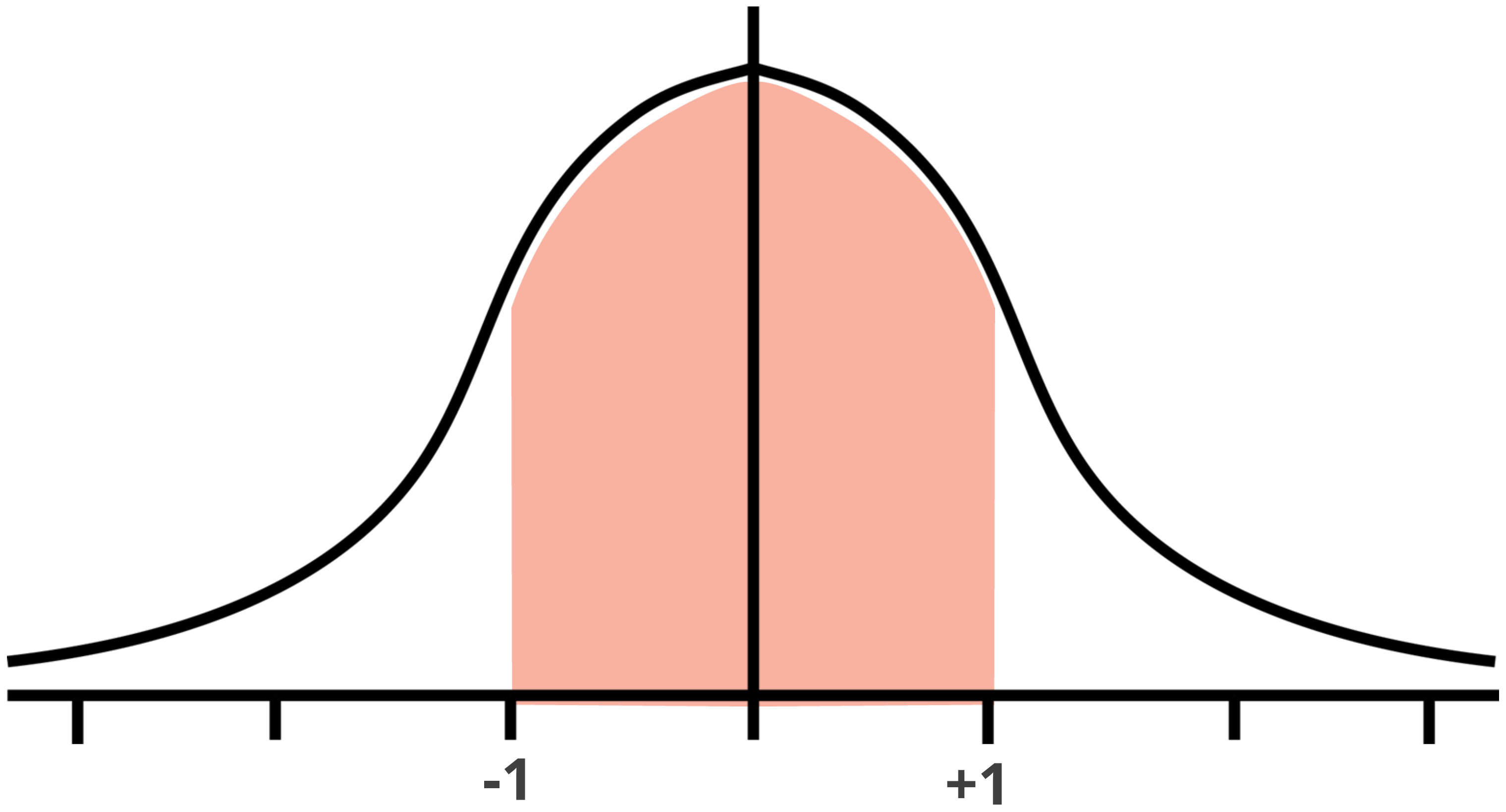
Determination of Z-Value

The Z-score of a value is the number of standard deviations between the value and the set mean.



Determination of Z-Value

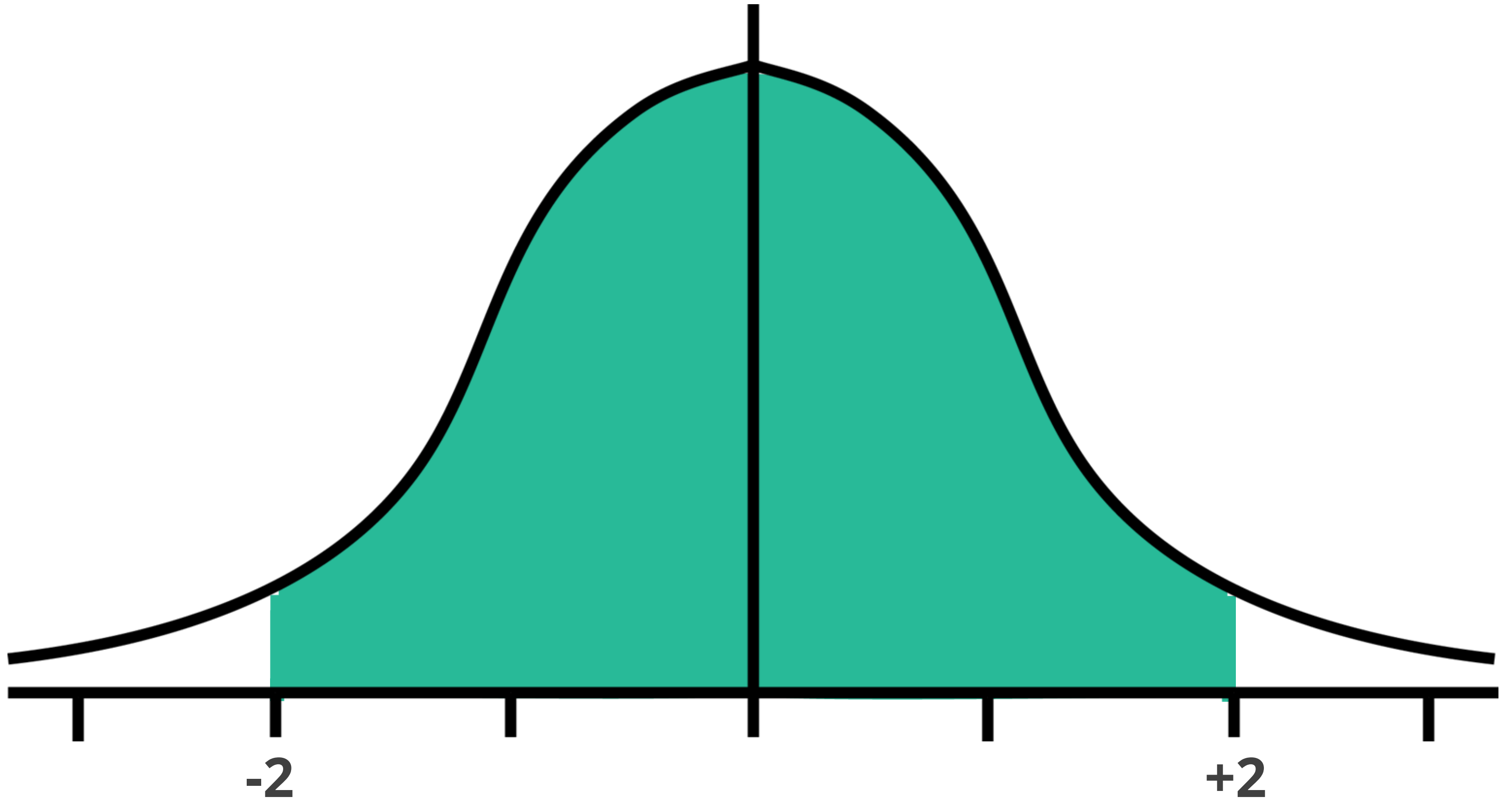
In many business situations:



About 65% of the data have Z-scores between -1 and +1

Determination of Z-Value

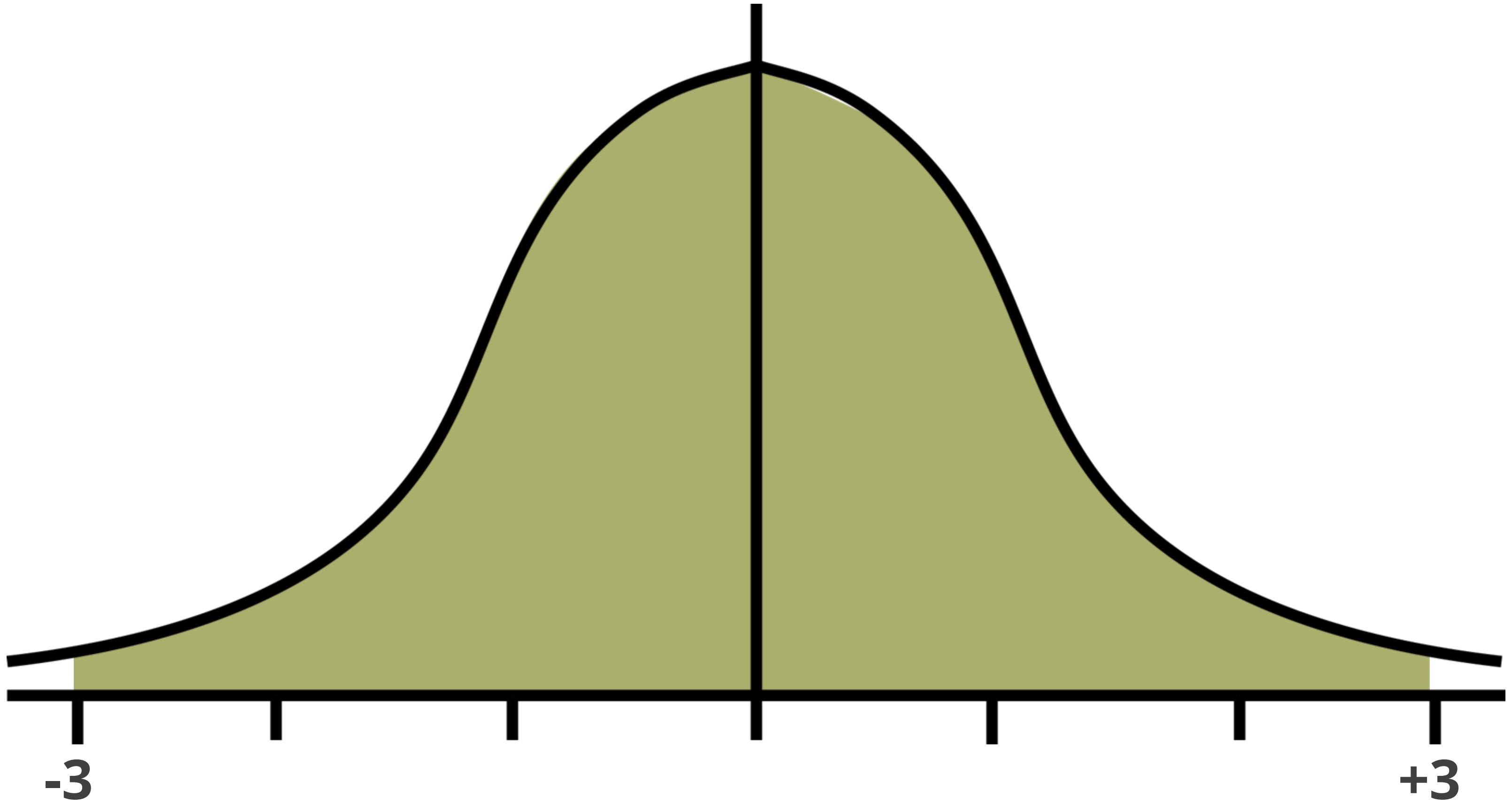
In many business situations:



About 95% of the data have Z-scores between -2 and +2

Determination of Z-Value

In many business situations:



About 99.7% of the data have Z-scores between -3 to +3

Analysis

In this specific example, low levels of variations are observed from the coefficient of variation.

ZMAX	ZMIN
1.10249	-1.5667
1.19728	-1.2482
1.28136	-1.4237
1.58773	-1.4743
1.12762	-1.8324
1.00896	-1.8715

Level of variations

Z-scores vary over a narrow range

Analysis

It shows that the levels of uncertainty both for the factory and the dealers, and consequently the risks, are quite low.

Level of uncertainty:

Factory and dealers

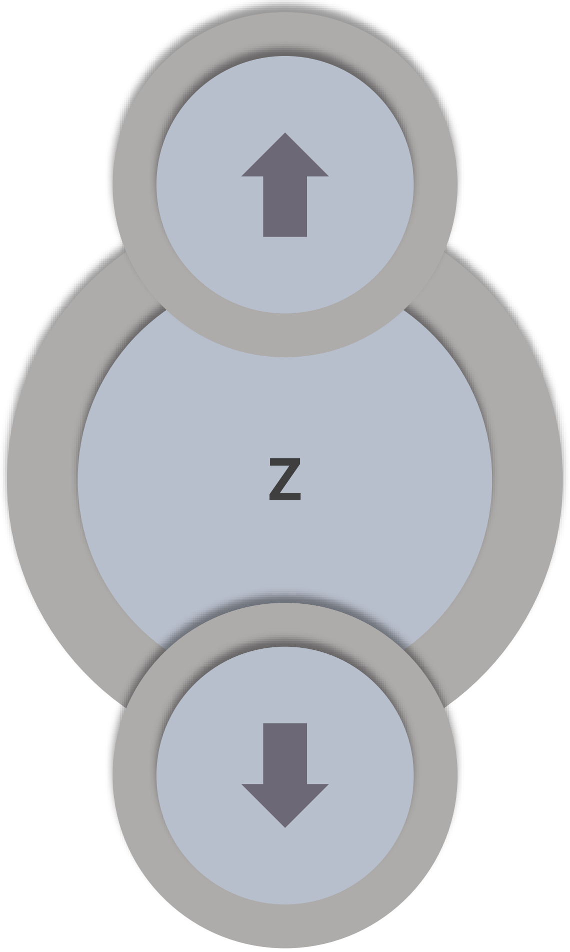
Level of uncertainty for the factory

Level of uncertainty for the dealers

The level of uncertainty is lower for the factory than for the dealers.

Analysis

The Z-score value is higher when the risk of production is also high, and the Z-score value is lower when the risk of production is lower.



Higher the risk on production

Lower the risk on production

This enables the factory to choose the best dealer who can dispatch the goods on a regular basis.

Key Takeaways

- Quartile deviation is half the difference between the first and third quartiles.

- Outliers are values in a dataset that are unusually large or small and are not representative of values in that set.

- The standard deviation is the positive square root of the average of the squares of deviations. Its square is called the variance.

- The standard score or Z-Score of a value in a dataset is the number of standard deviations by which that value is above or below its arithmetic mean.

- Skewness is another measure used to describe the shape of a dataset when presented graphically.

Knowledge Check

Knowledge Check

1

_______________ is the middle value or observation of a given set of data.

A. Mean

B. Median

C. Mode

D. Quartile

Knowledge Check

1

______________ is the middle value or observation of a given set of data.

A. Mean

B. Median

C. Mode

D. Quartile

The correct answer is **B**

Median is the middle value or observation of a given set of data.

Knowledge Check

2

Which of the following is one of the nine values that divide a dataset into ten equal parts?

A. Quartile

B. Decile

C. Percentile

D. Mode

Knowledge Check 2

Which of the following is one of the nine values that divide a dataset into ten equal parts?

A. Quartile

B. Decile

C. Percentile

D. Mode

The correct answer is **B**

Decile is one of the nine values that divide a dataset into ten equal parts.

Knowledge Check

3

Which of the following is a measure used to describe the shape of a dataset when presented graphically?

A. Skewness

B. Coefficient of variation

C. Measure of consistency

D. Normal distribution

Knowledge Check 3

Which of the following is a measure used to describe the shape of a dataset when presented graphically?

A. Skewness

B. Coefficient of variation

C. Measure of consistency

D. Normal distribution

The correct answer is **A**

Skewness is a measure used to describe the shape of a dataset when presented graphically.

Thank You