Statistics Essentials for Data Science

Applications of Statistics in Business

Learning Objectives

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

- ◉ Apply statistical methods to enhance daily business operations

- ◉ Interpret statistical data accurately to prevent misconceptions

- ◉ Evaluate visualizations for their truthfulness and effectiveness in conveying information

- ◉ Identify instances of false implications about relationships in data

Business Scenario

ABC is an e-commerce platform that sells multiple products online. As user numbers and needs increase, it needs to scale its business. However, before proceeding, it must understand the business in depth, as it currently lacks meaningful insight into the business outcomes.

To do this, the platform needs to analyze data using statistics and incorporate this analysis into its daily business operations. This approach will help it understand the data in detail and improve its business.

Additionally, it should explore how statistics can integrate into daily business and comprehend its key concepts.

How to Use Statistics in Day-to-Day Business?

Discussion

Discussion

Duration: 15 minutes

Does statistics play a vital role in daily business operations?

- What areas could benefit from the application of statistics?
- What advantages can businesses gain from using statistics?

Using Statistics in Day-to-Day Business

Statistics offers potential for business growth and plays a crucial role in decision-making.

Statistics can be applied in the following areas:

Quality control

Market research

Managing business risk

Design of psychometric tests

Statistical Quality Control (SQC)

Statistical Quality Control (SQC) is the term used to describe the set of statistical tools employed by quality professionals.

Here are the features of SQC:

It analyzes and solves quality problems.

It uses statistical methods to monitor and maintain product quality.

Statistical Quality Control (SQC)

Decisions about whether to accept or reject materials or components rely on the inspection of a sample.

The following factors are incorporated into inspection plans and decisions:

- The cost of inspection
- The impact of accepting poor quality
- The impact of rejecting good quality

The possibility of errors in decision-making is based on the results of a sample instead of an entire population.

Statistical Process Control

It preempts operations when processes do not meet expected standards, using the following tools:
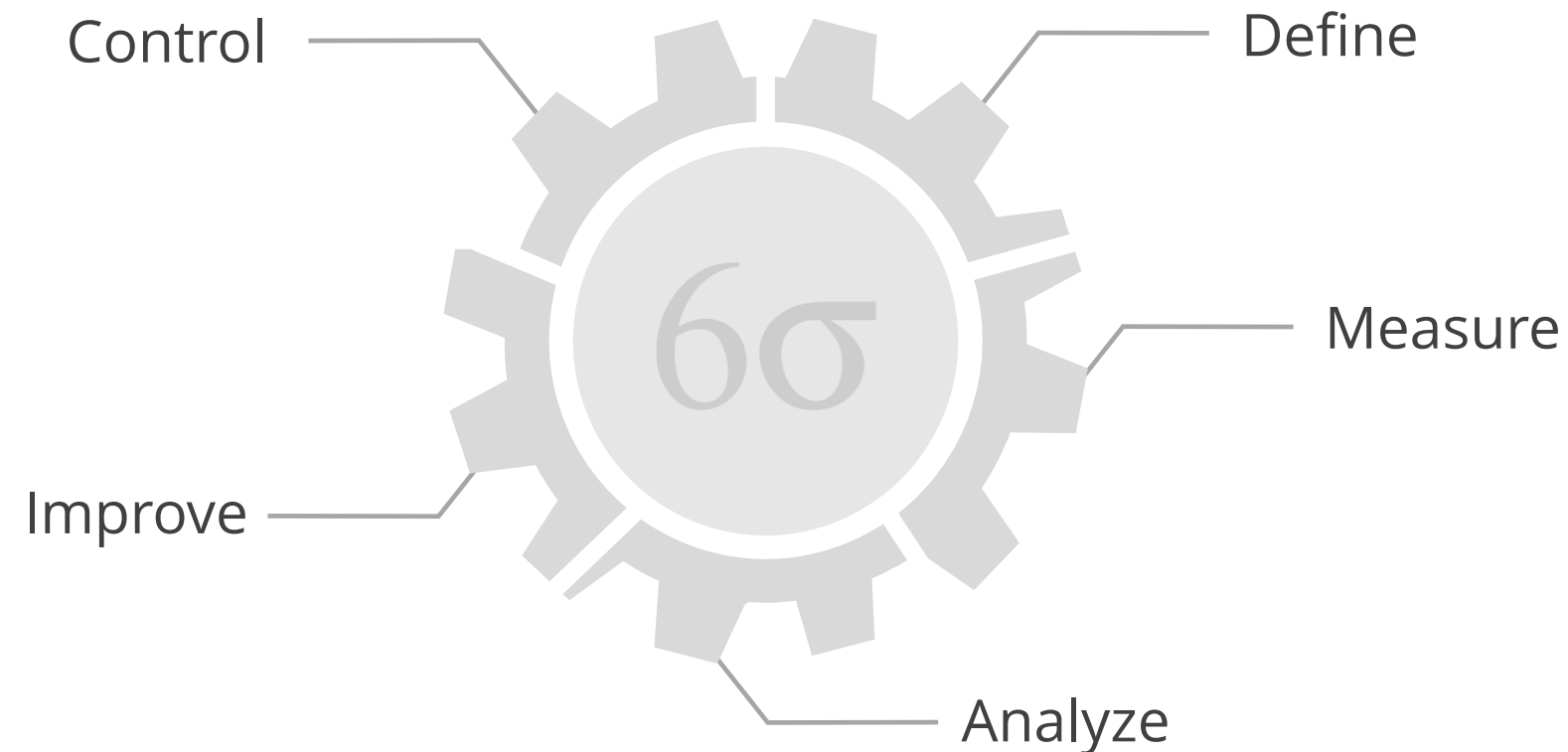
Charts

Incorporated statistical concepts

Six Sigma Philosophy

It aims to identify and eliminate the root causes of variations in product quality.

The five phases of Six Sigma are:

Control

Define

Measure

Improve

Analyze

The term Six Sigma signifies a level of quality that corresponds to only 3.4 defects per million opportunities.

Six Sigma Philosophy

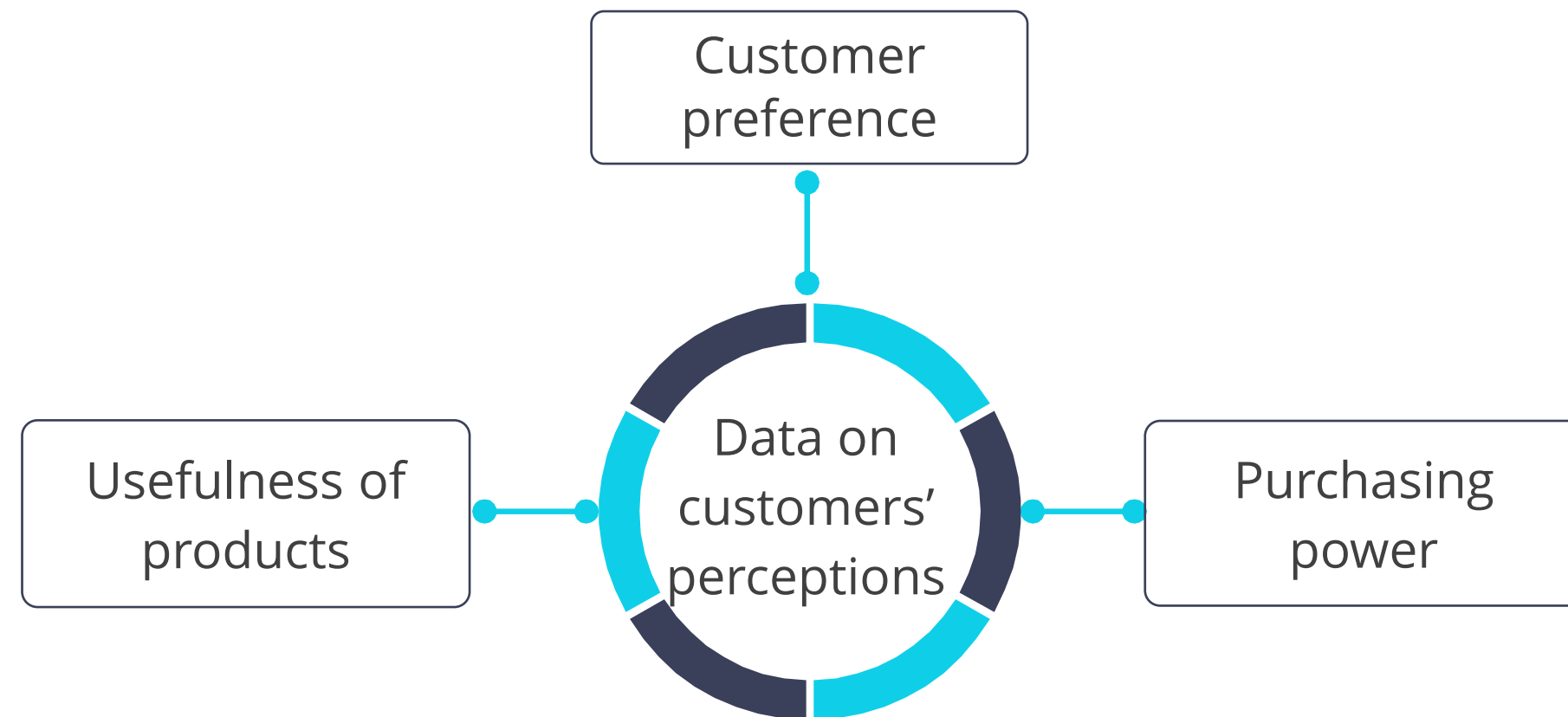In addition to product quality, it addresses factors such as:

6σ

Lead time or waiting time

Service durations

Market Research Studies

Market research involves evaluating the potential of a new product or service by reaching out to potential customers.

Customer preference

Usefulness of products

Data on customers' perceptions

Purchasing power

This collected data is then subjected to statistical analysis to form business plans.

Market Research Studies

During new product development, use market and technical data such as:

Productibility

Successful product development

Cost considerations

These factors enable decision-making.

Managing Potential Risks

Business decisions regarding the capacity of manufacturing and service facilities are fraught with uncertainties and potential risks.

Surplus capacity

Surplus capacity describes a situation where a facility's available capacity exceeds the current demand or utilization.

Deficit capacity

Conversely, deficit capacity refers to a situation where demand or utilization surpasses a facility's available capacity.

Coping with Potential Risks

The following are the uncertainties in capacity requirements:

Market demands

Productive capability of the operating facilities
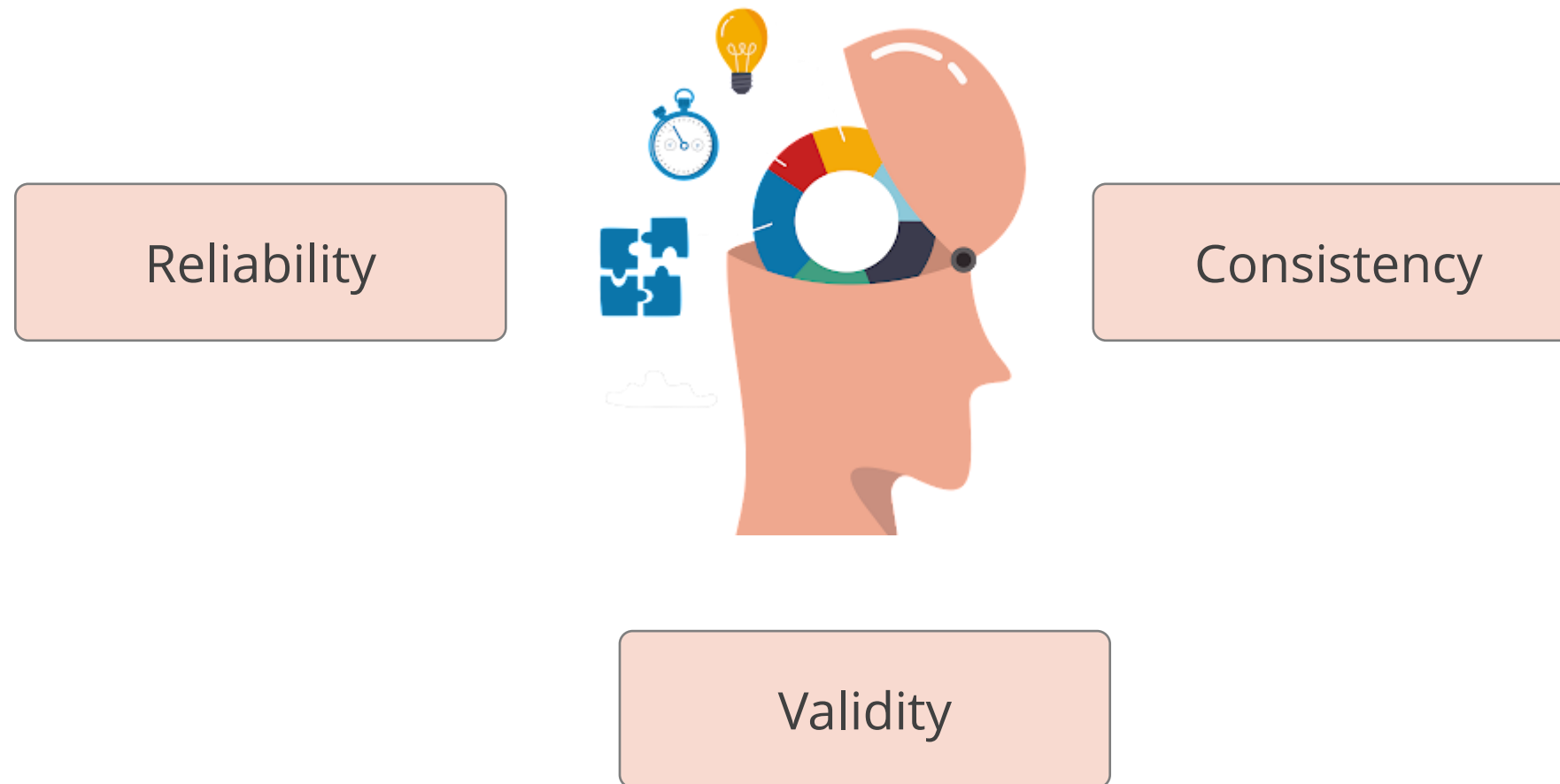
Performance of competitors

Government policies

The statistical concept of probability helps to address uncertainty.

Design of Psychometric Tests

Psychometric tests are employed during the selection and training of professional personnel.

The following tests are analyzed using statistical concepts such as mean, median, and mode:

Reliability

Consistency

Validity

Discussion

Does statistics play a vital role in daily business operations?

- What areas could benefit from the application of statistics?
  Answer: Statistics provides substantial potential for businesses. It can be applied in areas such as quality control, market research, business risk management, and the design of psychometric tests.

- What advantages can businesses gain from using statistics?
  Answer: It aids in managing statistical quality control, implementing controls during production, conducting reliability studies, and more.

How Not to Lie with Statistics?

Discussion

Discussion

As someone new to using statistics in daily business operations, you have decided to explore and understand its utility in your business.

To do this, carry out the following tasks:

- Identify incorrect approaches to data analysis
- Establish the importance of using appropriate and relevant data

How Not to Lie with Statistics

Statistical tools and techniques play a pivotal role in numerous businesses.

They enable organizations to:

- Analyze data

- Make informed decisions

- Develop strategies

Many businesses have significantly leveraged the use of these statistical tools and techniques.

Incorrect Approaches

Strong conclusions, as opposed to inferences, from hypothesis testing become crucial when dealing with small sample sizes.

Incorrect Approaches: Example

Assume a hypothesis claims that the new machine's average hourly output significantly surpasses 40 units

Exceed a threshold of 40 units

Reject the null hypothesis that the current rate is valid only when the average number of units per hour, based on sample observations, surpasses a certain threshold limit

Incorrect Approaches

The threshold limit tends to be high when using a small sample size.

Exceed a threshold

When the null hypothesis cannot be rejected, it does not disprove the claim of the new machine.
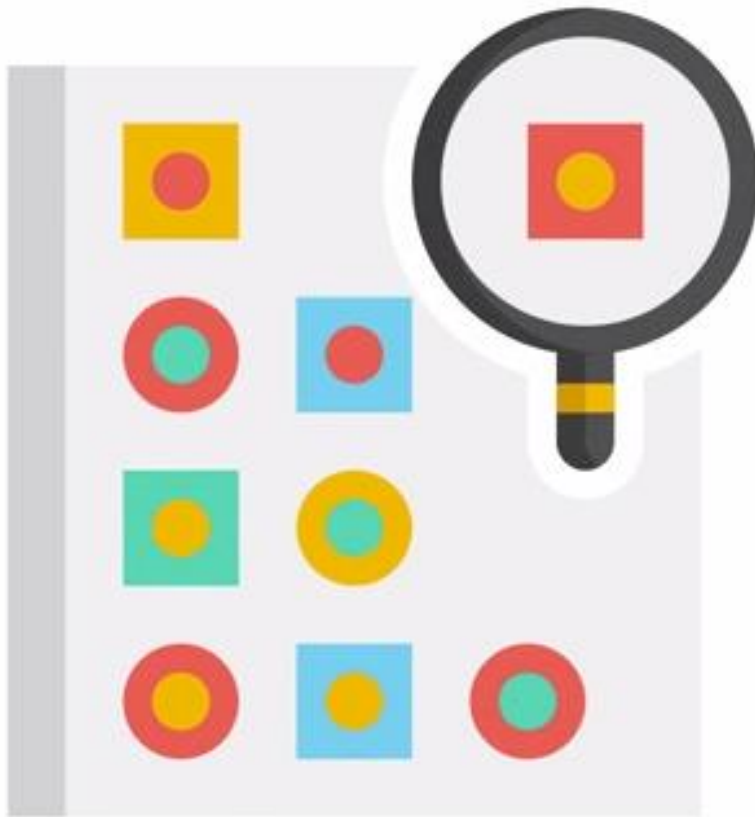
Incorrect Approaches

A hypothesis doesn't validate a claim; it implies that the available evidence is insufficient.

The findings from the analysis must be carefully communicated.

Purposive vs. Random Sampling

Purposive sampling is a non-probability sampling technique where individuals or groups are selected based on specific criteria or characteristics relevant to the study.

Conversely, random sampling involves the selection of individuals or groups without any specific criteria, purely at random.

Purposive vs. Random Sampling: Example

If information about a company's product or service is exclusively collected from repeat buyers for convenience or other reasons, it could lead to misleading findings.
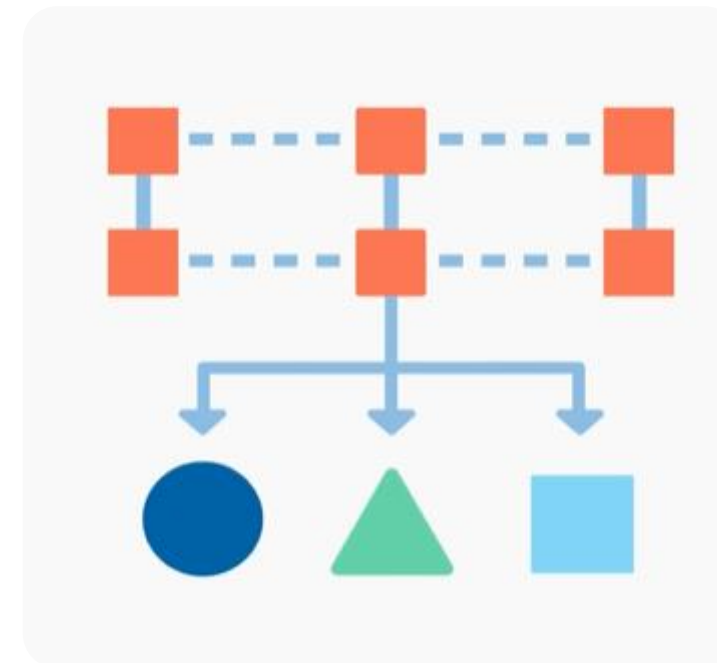
The opinions of dissatisfied customers might be overlooked in such a scenario.

Example: How Not to Lie with Statistics?

Example 1

A factory experienced delays in executing certain operations.

A machine breakdown was suspected as the cause.

Data concerning the frequency of machine breakdowns were gathered to pinpoint factors contributing to these breakdowns.

Example 1

However, it was discovered that operations on machines did not contribute to the delays.

Collecting data solely on the frequency of breakdowns proved insufficient, as it neglected other factors such as restoration times and the availability of alternative machines.

Example 1

To circumvent such pitfalls, companies typically conduct brainstorming sessions to identify various potential causes and gather necessary data.

Failure to collect data may occur in surveys when respondents misunderstand the questions, underscoring the importance of clear communication.

Example 1

Pilot surveys are employed to assess the quality of questionnaires.

Incorporating interview data alongside questionnaire data is crucial to guarantee a comprehensive analysis. This approach ensures diverse perspectives are considered and offers deeper insights into the research topic.

Example 1

The possibility of errors in data is high when data collection is delegated.

This also happens when there is a lack of clarity on the process involved.

Example 2

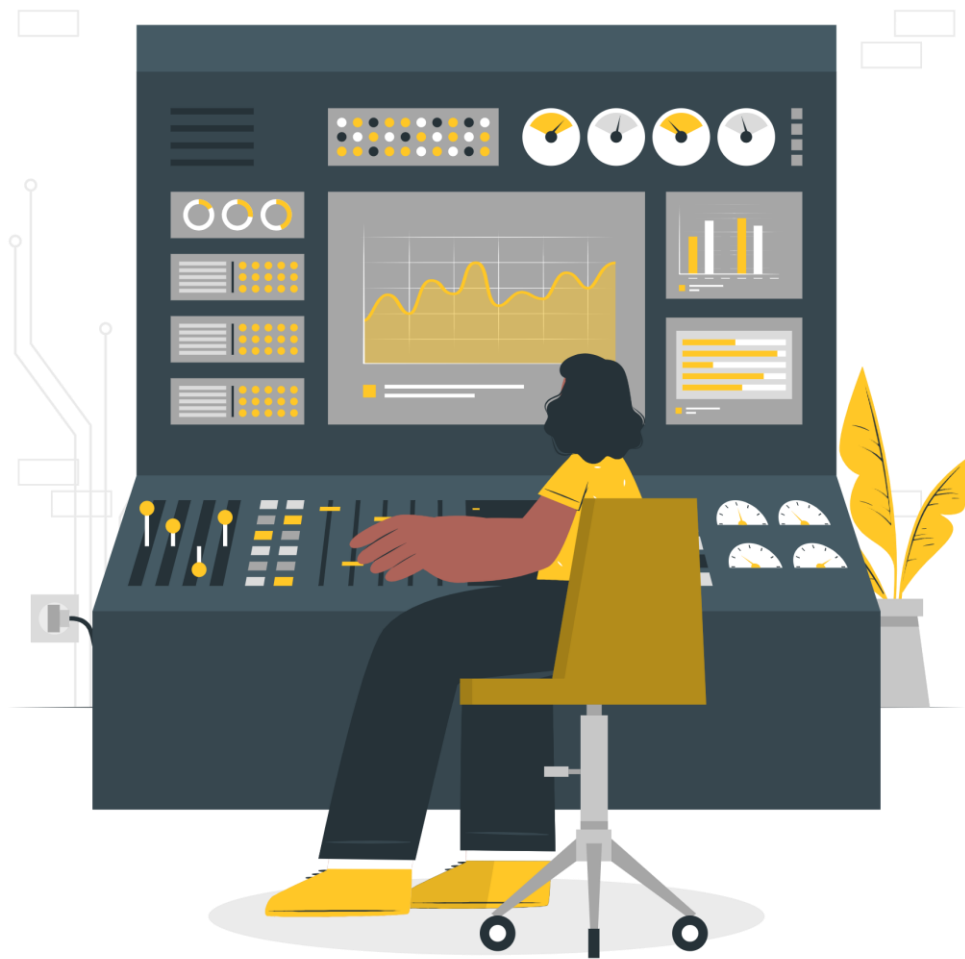A company aimed to identify the reasons for low production capacity.

An activity of sampling study was undertaken where:

- Machines were observed at random intervals, and
- Causes of non-utilization were recorded

Example 2

The observed causes include:

- ✓ Operator unavailability
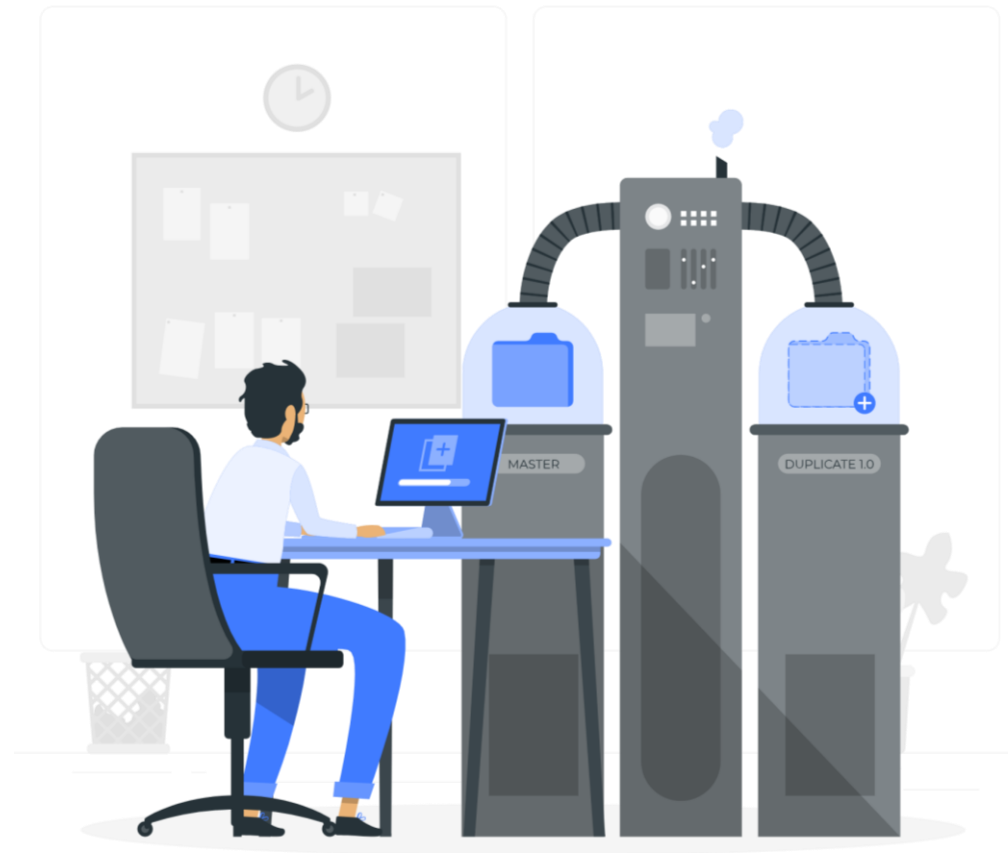- ✓ Operator engagement
- ✓ Machine breakdown
- ✓ Wait time

Example 2

When operators were occupied with other tasks, which were meant to signify other operations, it created a bottleneck.

As a solution, the operator was reassigned to that machine.

Example 2

The situation also occurred when the necessary materials were not available, which resulted in operators being occupied with other tasks.

Since data collection was delegated and there was a communication gap, it was concluded that the data collected might not be accurate.

Statistical Study

A statistical study should be carefully planned to ensure:

Collection of quality data

Systematic and scientific analysis

The data collection process should be carefully organized.

How Not to Lie with Visualizations?

Importance of Charts

Charts are powerful tools for communicating information on data sets.

Importance of Using Appropriate Data

Appropriate data is essential to ensure the accuracy and reliability of any analysis or decision-making process.
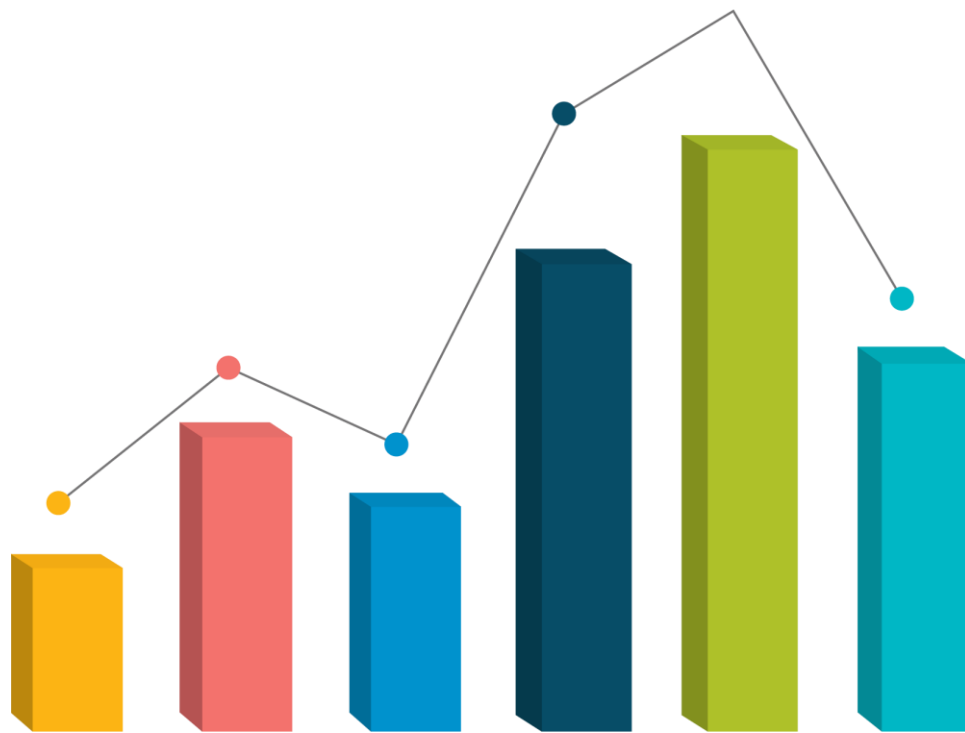

The usage of correct information is vital when data is presented visually.

Representing Qualitative Data

Bar charts are effective tools for representing qualitative data.
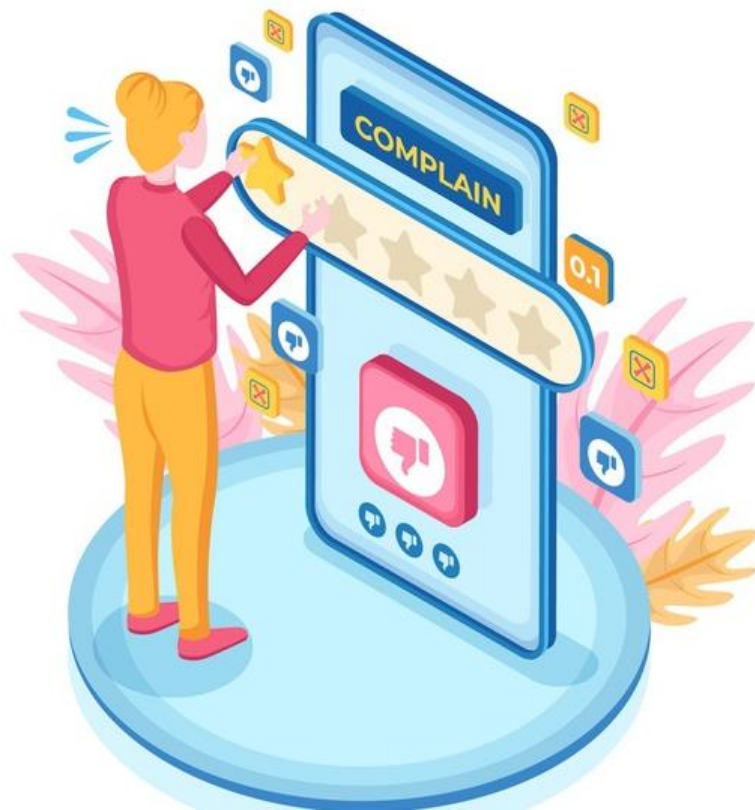
A bar chart consists of:

Rectangles of equal width

Heights proportional to the frequencies of attributes

Representing Qualitative Data: Example

Visualizing customer complaint records and highlighting the impact of the quality drive:

Bar charts will help illustrate how the quality drive has affected things visually.

Representing Qualitative Data: Example

Each type of customer complaint can be represented using two adjacent bars, resulting in several pairs of bars.

The differences in their heights will highlight the impact of the quality drive. In such cases, using frequencies rather than relative frequencies is advised.
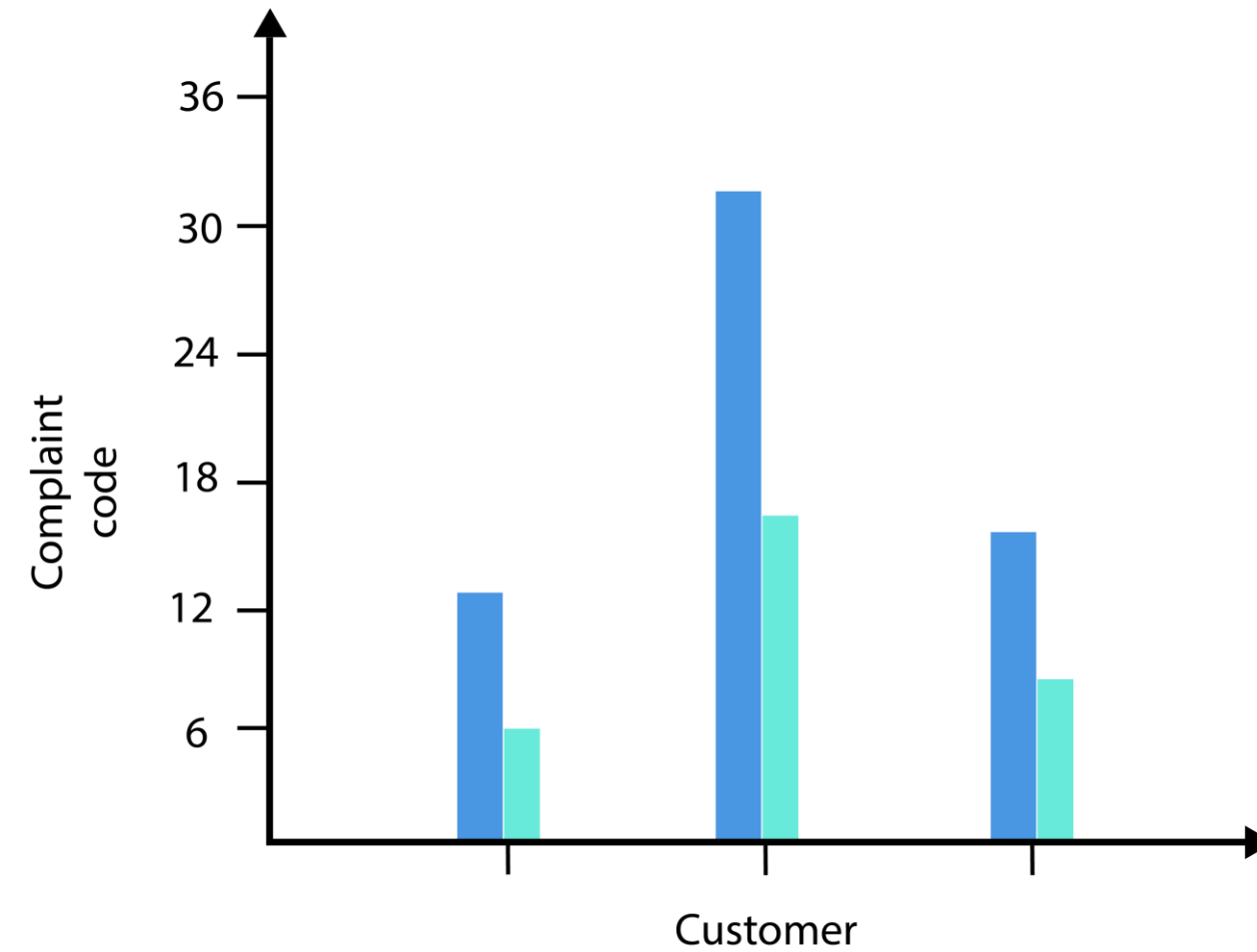
Representing Qualitative Data: Example

Consider an example to demonstrate the influence of a quality drive:

| Complaint Code	I	II	III	Total
Frequency before the quality drive	12	32	16	60
Frequency after the quality drive	6	16	8	30

Representing Qualitative Data: Example

The previous data can be represented as shown below:

A bar graph provides a visual summary of categorical information that allows for easy comparison and identification of patterns or trends.

Using Relative Frequencies: Example

Let's take an example of creating a histogram for room occupancy in hotels.
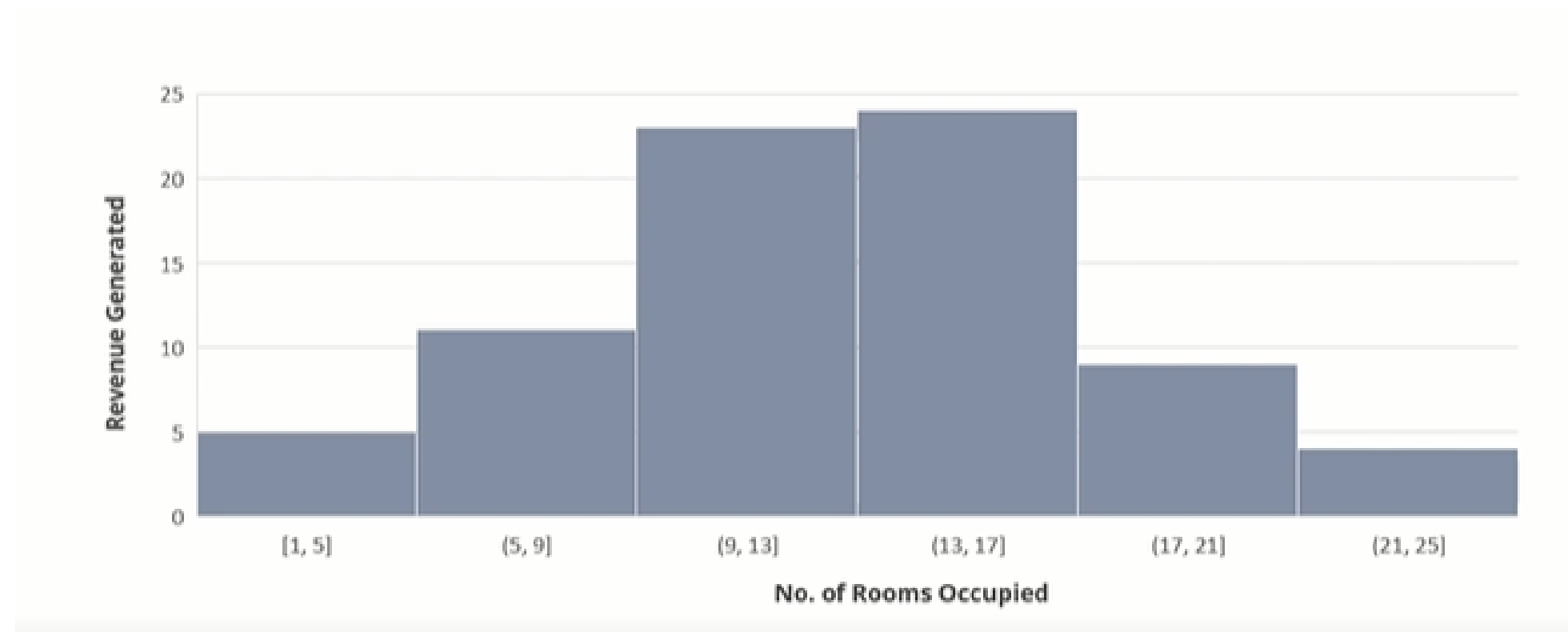
The example uses the number of occupied rooms and revenue generated; the use of either set of data is inadequate as it fails to consider other important factors that can affect the overall performance and profitability of a business.

Importance of Data in a Statistical Investigation

In a statistical investigation, it is important to plan what to collect and use.

Incorrect projections can be a result of:

Biases of the analyst

Planned distortion

Importance of Data in a Statistical Investigation

Incorrect projections can be prevented when key stakeholders are involved in the data analysis process.

Discussion

Duration: 15 minutes

As someone new to using statistics in daily business operations, you have decided to explore and understand its utility in your business.

To do this, carry out the following tasks:

- Identify incorrect approaches to data analysis
  Answer: The following are the incorrect approaches:

  - Drawing strong conclusions rather than making inferences from hypothesis testing

  - Using purposive sampling over random sampling

- Establish the importance of using appropriate and relevant data
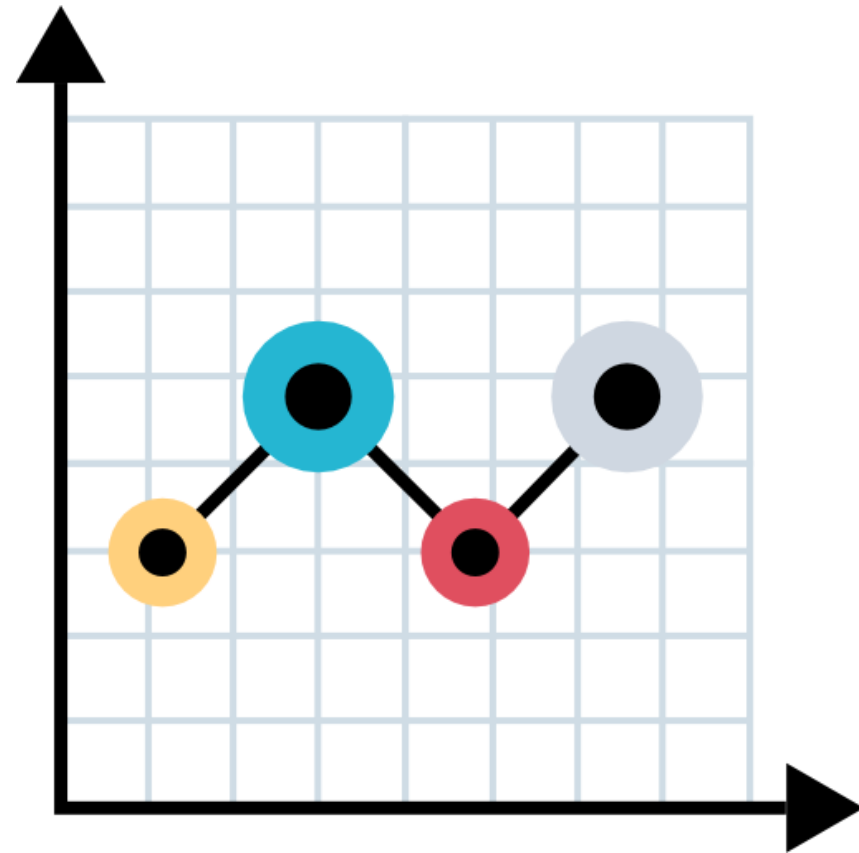 **Answer:** The following are the importance of using appropriate data:

  - Representing quality data

  - Using relative frequencies

  - Statistical investigation

How Not to Lie About Relationships in the Data?

Correlations

Correlation refers to the statistical relationship between the two entities. It measures the extent to which two variables are linearly related.

- All correlations are not due to a cause-and-effect relationship.

- Correlation indicates a relationship between variables, but it may not imply causation.

- The prediction of one variable knowing the value of the other should be avoided.

Identifying the cause-and-effect variables is vital when a cause-and-effect relationship is established. Use the appropriate regression equation

Cause-and-Effect Relationship

Refers to the notion that a change in one variable is directly responsible for a change in another variable

Cause X — Y on X ✓ — Y Effect

When variable X is the cause and variable Y is the effect, the regression is Y on X and not X on Y.

Cause-and-Effect Relationship

The cause should be properly determined and used for prediction.

Cause X — Y on X ✓ — Y Effect

Two regression equations differ unless the correlation is +1 or -1.

When a sample is used, a relationship between two variables cannot be inferred only based on the numerical value of the correlation coefficient.

Correlation Coefficient

The correlation coefficient is a statistical measure of the strength of a linear relationship between two variables. Its values can range from -1 to 1.

- It is important to perform a test of significance before making an inference.

- The test of significance is related to the correlation coefficient in determining whether the observed correlation between two variables is statistically significant or if it could have occurred by chance.

The test must be based on the sample size, specifically the pair of observations, the numerical value of the correlation coefficient, and the level of significance.

Correlation Coefficient

For a smaller sample size, a large threshold value of the correlation coefficient is required to infer that a strong relationship exists.
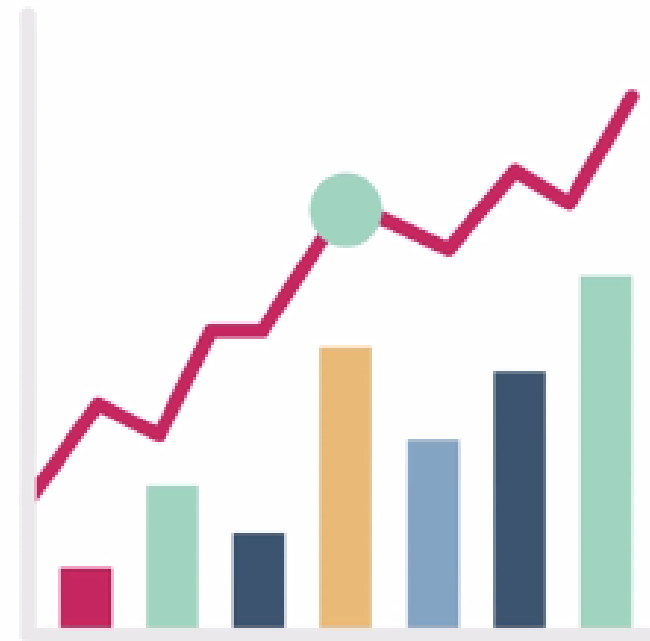
Coefficient of Determination

The coefficient of determination is a statistical measurement that examines how differences in one variable can be explained by the difference in a second variable when predicting the outcome.

The coefficient of determination is an index.

This index quantifies the quality of a regression equation for prediction purposes.

Coefficient of Determination

Statistical significance is a determination made by an analyst that the results in the data are not explainable by chance alone.

HYPOTHESIS

CONCLUSION

Statistical hypothesis testing is the method by which the analyst makes this determination.

Coefficient of Determination

The test of significance will incorporate:

- Sample size
- Regression sum of squares
- Error sum of squares
- Level of significance

- If the null hypothesis is not rejected, then the equation does not predict the variables in the regression equation.
- It indicates the extent to which the regression equation explains the magnitude of the variation.

Analytics

Analytics

Analytics highlights the importance of scientifically and systematically planned and executed analysis for high-quality results.

Analysts

Vital role to play

Users

Thomas Davenport (American author, researcher, and thought leader in the field of business analytics and information technology) states that the "*CEO leading the analytics change requires both an appreciation and a familiarity with the subject*."

Analytics

A background in statistics is not necessary, but leaders must understand the theory behind various quantitative methods, including mean, standard deviation, regression, hypothesis testing, and sample size determination.

Leaders should also recognize the limitations of these methods, specifically in terms of the factors that are considered and those that are not.

Importance of Analytics

CEOs often seek assistance from experts when it comes to utilizing quantitative techniques.

These experts possess a deep understanding of how analytics can be effectively applied to various aspects of business operations.

Importance of Analytics

A good analyst must:

Be able to express complex ideas in simple terms

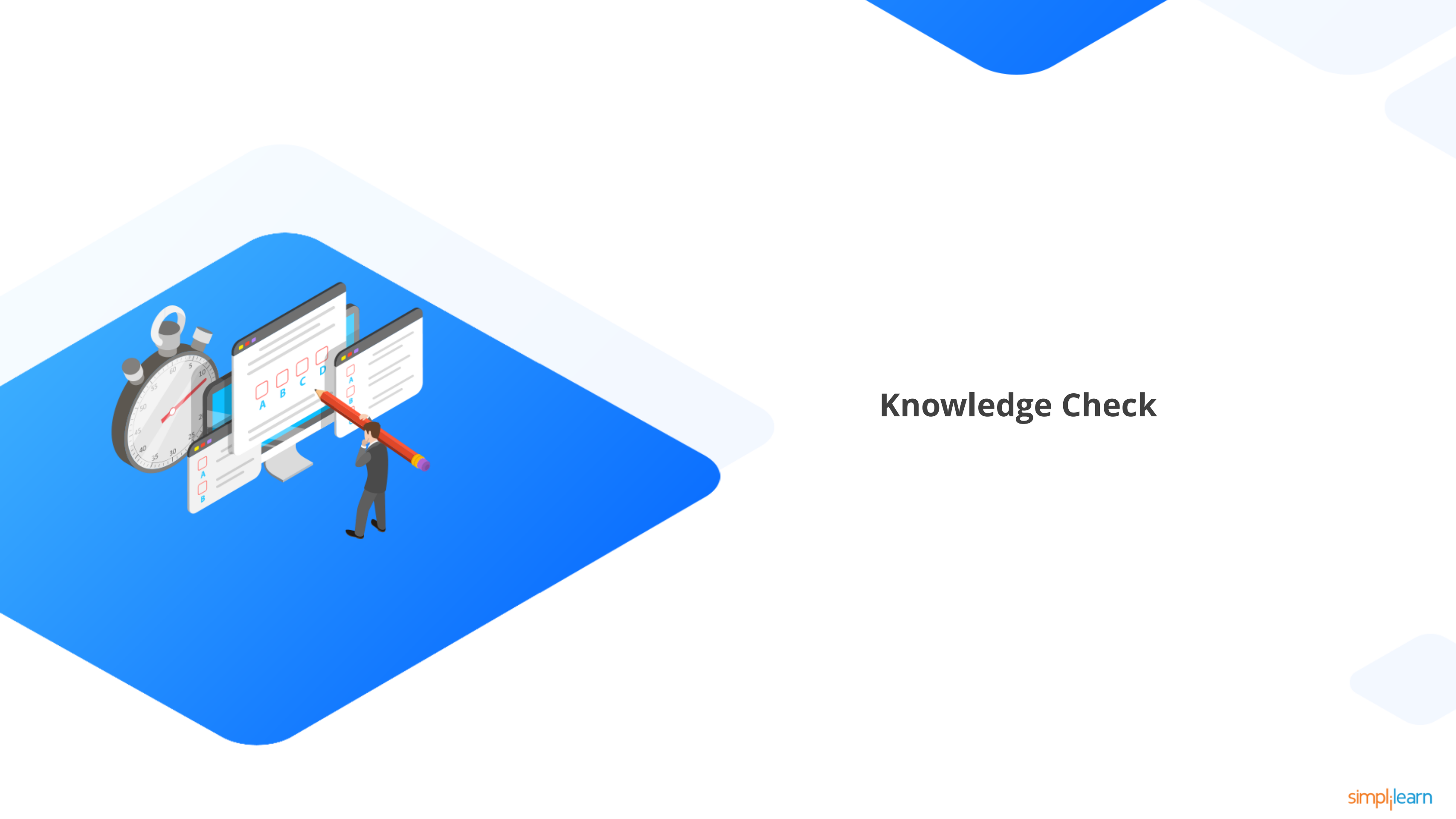Have interpersonal skills to interact with decision-makers

Have good business acumen

Help market their work internally and sometimes externally

Key Takeaways

- Statistical process control serves as a preemptive measure in operations where processes do not meet the expected standards.

- Market research enables businesses to assess the potential of new products or services by engaging with potential customers.

- Bar charts are effective tools for visually representing qualitative data.

- Identifying the cause-and-effect variables is vital when a cause-and-effect relationship is established.

Knowledge Check

Which of the following helps in deciding to accept or reject materials or components by inspecting a sample?

A.   Statistical control quality

B.   Exercising controls during production

C.   Building quality into products

D.   Reliability studies

Which of the following helps in deciding to accept or reject materials or components by inspecting a sample?

A.   Statistical control quality

B.   Exercising controls during production

C.   Building quality into products

D.   Reliability studies

The correct answer is **A**

Statistical control quality helps in deciding to accept or reject materials or components by inspecting a sample.

When a _____ holds, it is important to identify the cause-and-effect variables and use the appropriate regression equation.

A. Correlations

B. Cause-Effect Relationships

C. Correlation Coefficient

D. Coefficient of Determination

When a _____ holds, it is important to identify the cause-and-effect variables and use the appropriate regression equation.

A.   Correlations

B.   Cause-Effect Relationships

C.   Correlation Coefficient

D.   Coefficient of Determination

The correct answer is **B**

When a cause-effect relationship holds, it is important to identify the cause-and-effect variables and use the appropriate regression equation.

Which of the following might provide false information?

A.   Correlations

B.   Cause-Effect Relationships

C.   Correlation Coefficient

D.   Coefficient of Determination

Which of the following might provide false information?

A. Correlations

B. Cause-Effect Relationships

C. Correlation Coefficient

D. Coefficient of Determination

The correct answer is **A**

Correlations sometimes give false information.

Thank You