

Statistics Essentials for Data Science



Introduction to Statistics



Learning Objectives

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

- 🕒 Distinguish among various types of statistics
- 🕒 Understand the significance of statistical concepts in data science
- 🕒 Identify the applications of statistics across diverse business sectors
- 🕒 Investigate the discipline of statistics and its methods
- 🕒 Explain the fundamentals of statistics



Business Scenario

ABC operates an over-the-top (OTT) media platform. Lately, the company has seen a decrease in viewership and has struggled to onboard new viewers. To resolve these issues, ABC needs to delve into the viewer's data. The company must investigate which shows should be broadcast and how to win back viewers.

Analysis of usage data from the OTT platform can yield insights into the most popular shows and user preferences. Using this data, ABC can create and broadcast viewer-preferred shows while discontinuing less popular ones. This strategy is likely to attract more viewers and expand the business.

To carry out this plan, the organization needs a comprehensive understanding of data statistics and its concepts.





What Is Statistics?



Discussion

What Is Statistics?

Duration: 15 minutes



Is data storage important?

- What does data mean?
- What does data statistics mean?

Data

People handle a significant amount of data daily.

Example: Newspaper



Data

Data is examined to gain possible insights, especially when the stakes associated with decisions are high.



Data plays a key role in running businesses.

Data Collection and Maintenance

Data has been collected and maintained since the dawn of ancient civilizations.

Data on population and resources were maintained by the governments of:



Babylonia



Egypt



Rome

Data Collection and Maintenance

Religious institutions kept records of births, deaths, and marriages.



Data Usage

Data collection and use in various fields lead to the generation of insights. Such analysis has evolved into a discipline known as statistics.



Statistics is a discipline that involves the collection, organization, analysis, interpretation, and presentation of data to make informed decisions or predictions.

Statistics

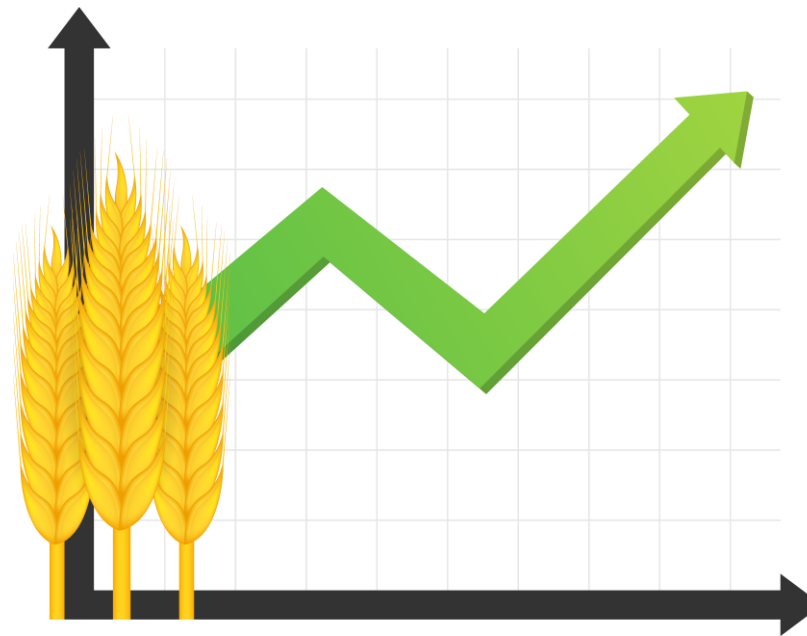
Example: Study of relative consumer preference for various brands



Statistical data is influenced by numerous factors because each variable is affected by several economic and social factors. This is known as the multiplicity of causes.

Example: Multiplicity of Causes

Factors such as seed variety, soil fertility, and the quantity and quality of manure have an impact on crop production.



Example: Multiplicity of Causes

The time of arrival of a doctor affects the duration of a patient's visit to the clinic, the waiting time for other patients, and their consultation time.



What Is Statistics?

Duration: 15 minutes



Is data storage important?

- What does data mean?

Answer: Data is information that has been transformed into a format that is efficient for storage or processing. Individuals encounter vast amounts of data daily. Data can be analyzed to gain valuable insights, particularly when significant decisions are at stake.

- What does data statistics mean?

Answer: Data statistics refers to the study and manipulation of data and encompasses techniques for its collection, examination, analysis, and drawing of conclusions.



Why Statistics?



Discussion

Why Statistics?

Duration: 10 minutes



How are the statistics of the data used?

- Why is it important?
- What are the benefits of data statistics?

Why Statistics?

Statistics helps in:



Analyzing quantitative and qualitative data



Utilizing statistical tools and techniques

Benefits of Statistics

The following are the benefits of statistics:

Make comparison

Clear and concise
presentation of data

Increase efficiency

Generalization of large-volume
of data using samples

Enables forecasting

Improve decision-making

Makes problem-solving
easier

Clear and Concise Presentation of Data

Data can be presented in various forms, such as graphs, charts, or tables, enabling the extraction of valuable insights.



Generalize on Large Volume of Data Using Samples

Statistical analysis often requires the examination of numerous items or units, and analyzing a large volume of data can be time-consuming and inefficient.



For example, a bank cannot collect feedback from every customer to improve customer service.

Generalize on Large Volume of Data Using Samples

Statistical theory offers logical methods to select samples and draw inferences based on the study of those samples.



Make Predictions or Forecasts

If death rates are considered in a particular city or country, one may be able to predict the number of future deaths.



Statistics supports making predictions and forecasting based on available data.

Make Comparisons

In an organization, it is essential to compare the performance of different machines, assess the profitability of various products, and analyze other factors.



Analyze conversions from past campaigns to get the best ROI

For example, the performance of two marketing campaigns is compared.

Why Statistics?

Duration: 15 minutes

How are the statistics of the data used?

- Why is it important?

Answer: Statistics enables individuals to conduct technical analyses of data. In most organizations, managers need to make periodic decisions that impact performance based on the available quantitative and qualitative data.

- What are the benefits of data statistics?

Answer: Statistics assists in the clear and concise presentation of data, generalizing from a large volume of data using samples, making predictions or forecasts, and facilitating comparisons.

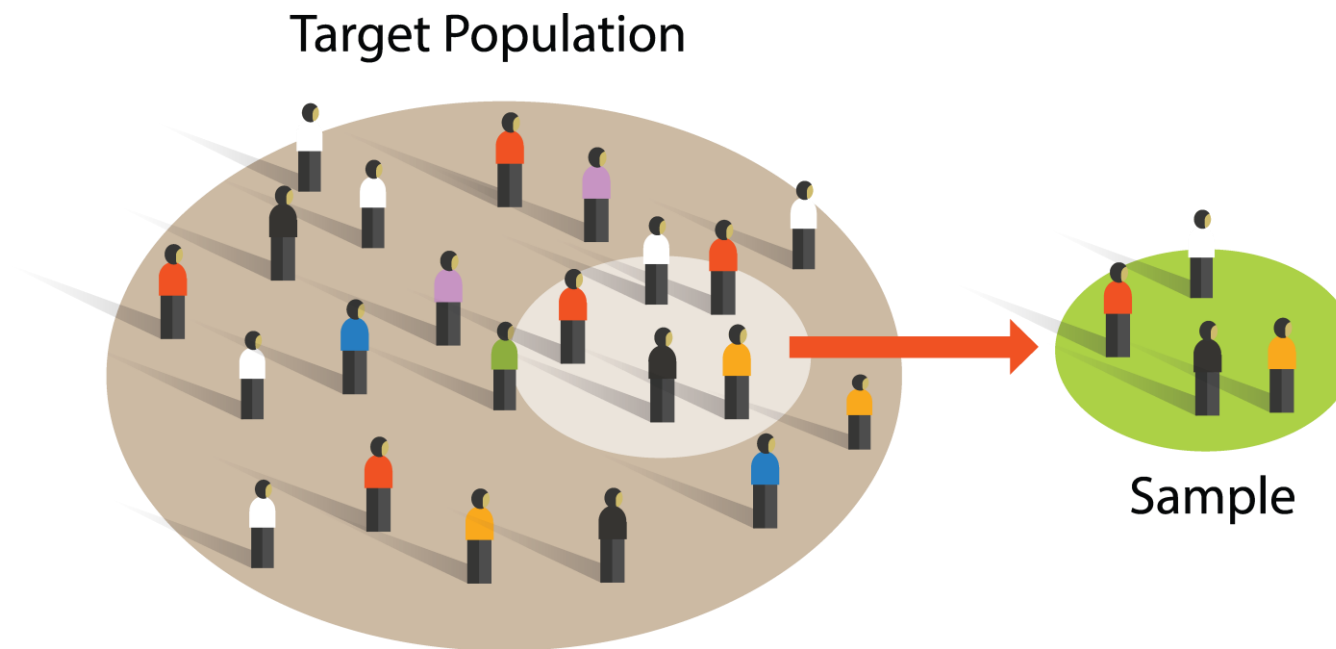




Difference Between Population and Sample

Population and Sample: Example

In a school, the population comprises all the enrolled students



Among the students at the school, a sample can consist of those who speak French.

Population Characteristics

The characteristics of a population based on a sample are:



Population Characteristics: Example

In the example cited earlier:

Average height of all students = **Parameter**

Average height of students in the sample = **Statistic**

A statistic can serve as an estimate of the parameter.



Different Types of Statistics



Discussion

Types of Statistics

Duration: 15 minutes

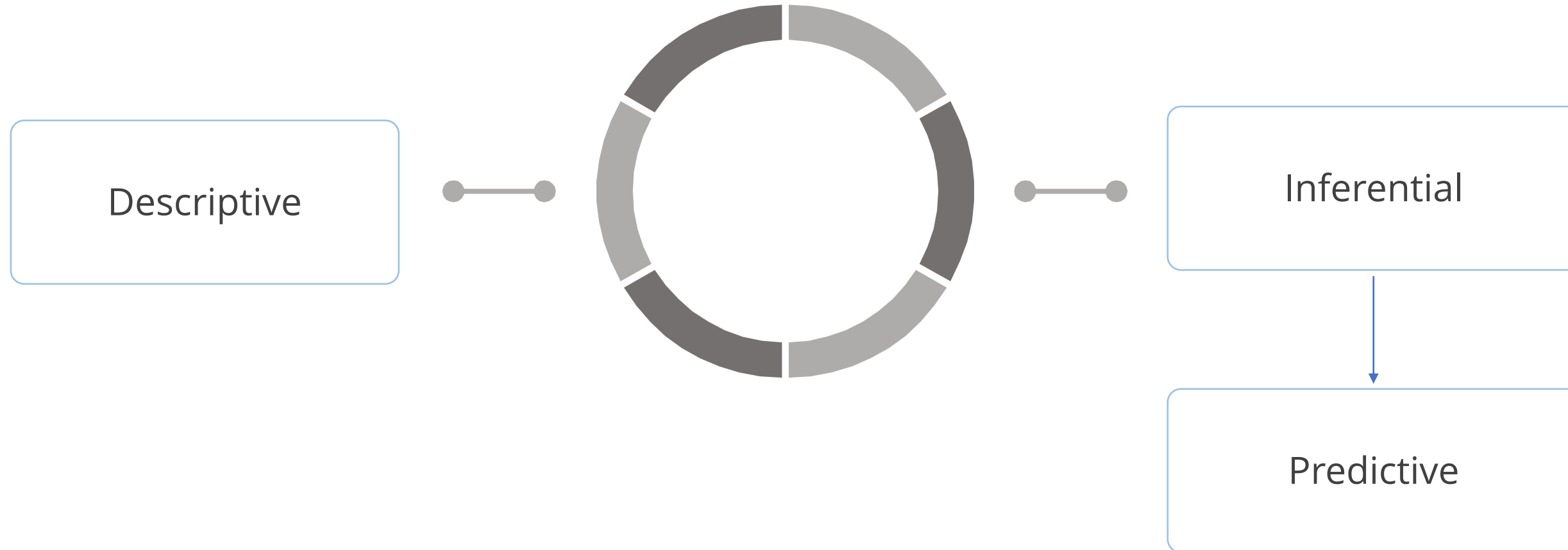


You are working as a data scientist for a government body, and you have been asked to analyze the population of the country and determine what the population will be five years from now.

Which type of statistics will be used here?

Types of Statistics

Statistic theory is broadly categorized into:



Descriptive Statistics

Descriptive statistics involves the collection, presentation, and characterization of data.



It also defines the sampling techniques that will be used to collect the data.

Descriptive Statistics

For example, summarization of social media engagement data, such as Instagram or Facebook likes



Descriptive Statistics: Example

Example: The average scores of students in classes A and B

Grades of class A students are:
 $\{73, 95, 76, 69\}$

Average grade of class A is:
78.25

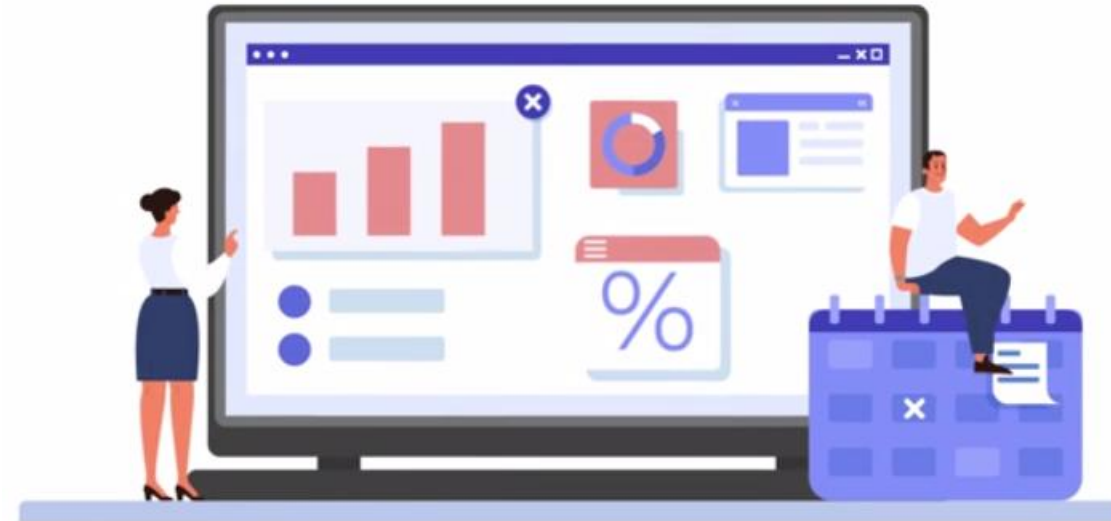
Grades of class B students are:
 $\{69, 75, 98, 86\}$

Average grade of class B is:
82

This denotes that the average of class B is more than class A.

Inferential Statistics

Inferential statistics is used to estimate the characteristics of a population.



It enables decision-making by drawing conclusions from random samples.

Inferential Statistics

The main functions of inferential statistics are to estimate population characteristics, test hypotheses, and draw conclusions about the data.

Population or universe

Sample

Parameter

Statistic

Inferential Statistics: Example

Example: The average height of 500 adults in a city



The population is the adult population of the city.

Sample

The selected 500 adults from the city

Parameter

The average height of adults in the city

Statistics

The average height of selected 500 adults

Uses of Inferential Statistics

Inferential statistics have two main uses:

01

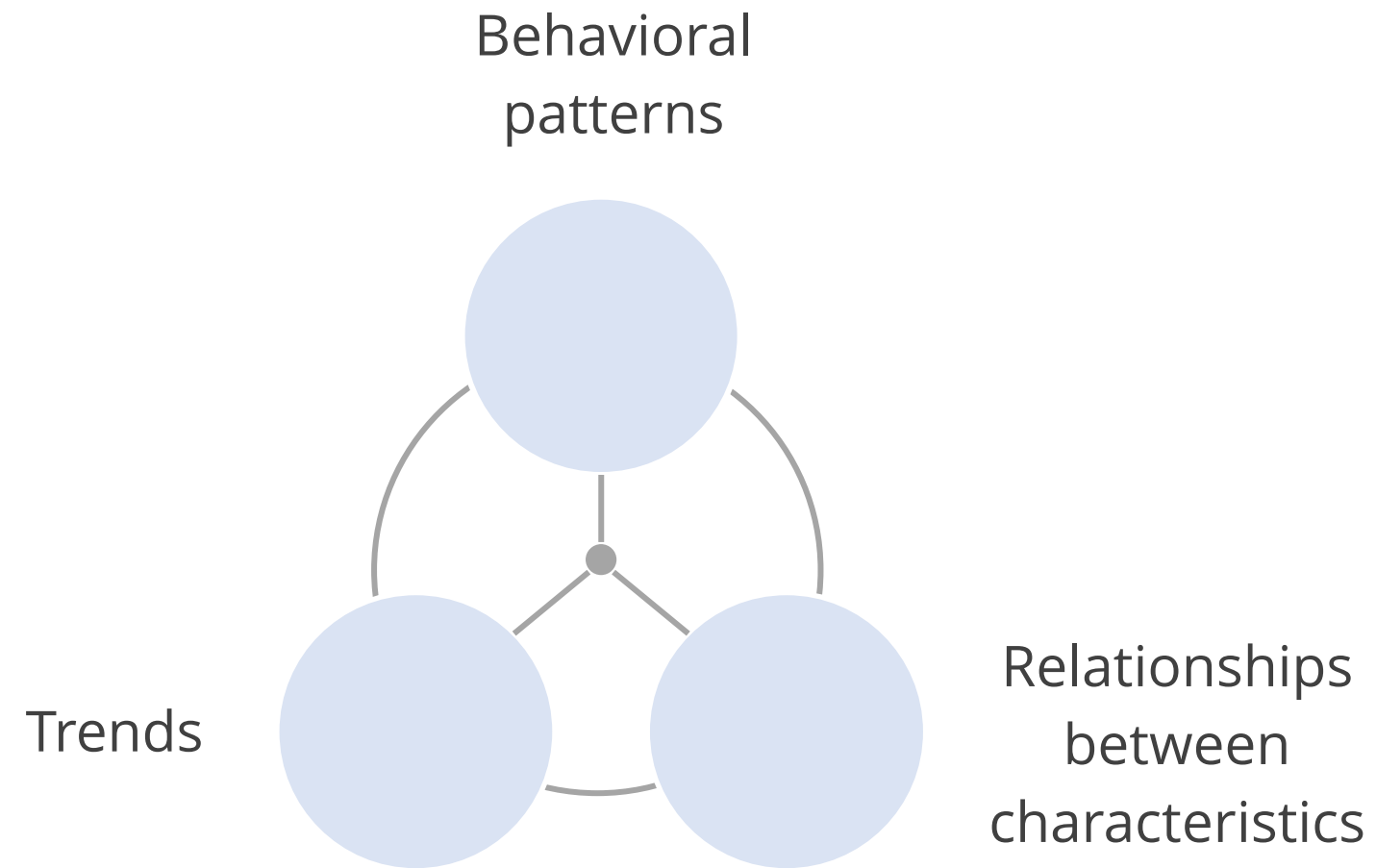
Make estimates about populations (for example, the mean GMAT score of all junior year students in the United States)

02

Test hypotheses to draw conclusions about a population (for example, the relationship between GMAT scores and family income)

Predictive Statistics

Predictive statistics is the science of extracting information from data and using it to predict:



Predictive Statistics Example: Demand Estimation

For example, data on the number of residents in a city recorded over the years can be used to predict its future population.



Examples

Below are examples of data in predictive statistics:



The demand for spares in the future is predicted based on the sales of the car.



Job performance scores can be predicted based on the performance of candidates in selection tests.

Such data can be used to predict the values of one characteristic by knowing the values of others.

Types of Statistics

Duration: 15 minutes



You are working as a data scientist for a government body, and you have been asked to analyze the population of the country and determine what the population will be five years from now.

- Which type of statistics will be used here?

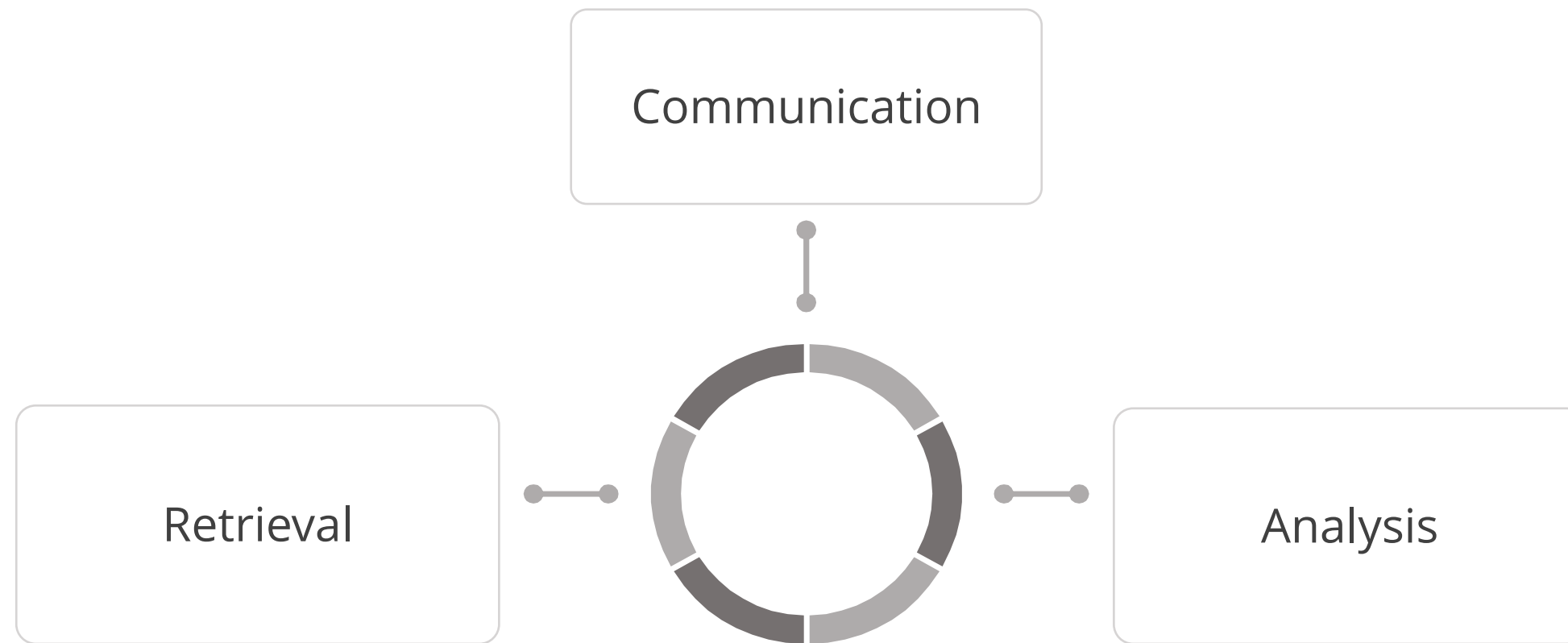
Answer: Predictive statistics will be used. It is the science of extracting information from data and using it to predict the population.



Importance of Statistical Concepts in Data Science

Statistics with Data Science

It provides insights to help understand large amounts of data.



Statistics has given rise to a scientific approach to capitalize on the power of data science.

Data Science

It is a combination of practices, especially in statistics, to understand data.



It uses scientific methods, processes, algorithms, and systems to extract knowledge and derive insights from data.

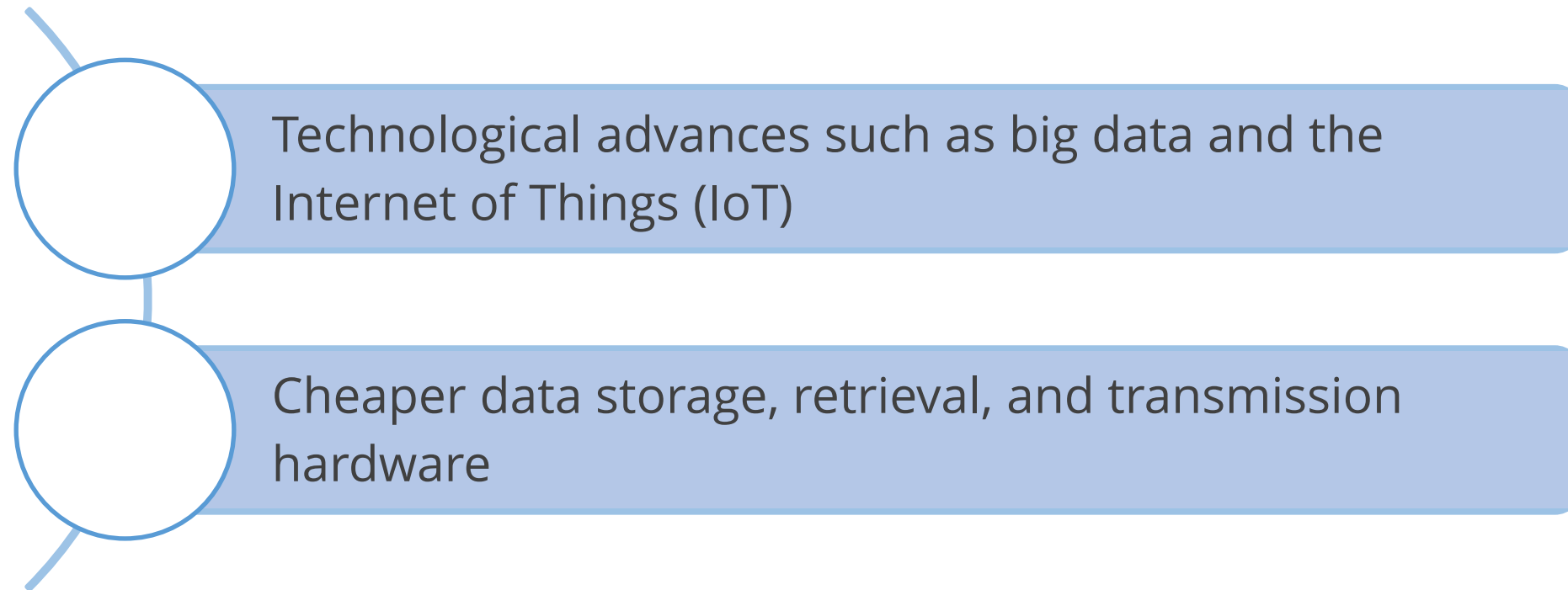
Use of Data Science

Using data science techniques, unstructured and unclean data can be organized and summarized.



Generation of Data in Businesses

Today, businesses are flooded with a sea of data. The primary reasons are as follows:



Generation of Data in Businesses

Every day, 2.5 quintillion bytes of data are generated by humans through social media and IoT.



Making sense of this data is essential for business success.

Example: Managing Intensive Care Units (ICU)

For example, managing a shortfall in an ICU's capacity



This problem has been studied using data science.

Example: Managing Intensive Care Units (ICU)

Discharged ICU patients are readmitted when the level of risk is severe.



More readmissions impose constraints on ICU capacity.

Problem Solved

Patients who were less likely to be readmitted were recommended for discharge by incorporating the following:

Demand from incoming patients

Current level of severity



Service quality improved and the problem of a deficit in ICU capacity was solved.

Analyzing the Approach Used

Some of the methods employed to manage the ICU capacity problem are as follows:

Intuitive approach

Powerful methodology

Interrelated factors

Significant improvements can be anticipated with the inclusion of more variables and complex interconnections.

Importance of Statistical Techniques

The factor determining readmission possesses a statistical likelihood.



Even after systematically assessing the severity, the need for readmission is uncertain.

Importance of Statistical Techniques

Randomness and uncertainty impact several factors. Demand and supply are typical business factors that involve uncertainty.



Therefore, statistics is very useful for such cases.

Vital Role of Data Science

Data science greatly benefits from the use of statistical techniques and methods. It is essential to a company's success.





Why Statistics Is Vital for Business Success



Discussion

Statistics Is Vital for Business Success

Duration: 20 minutes

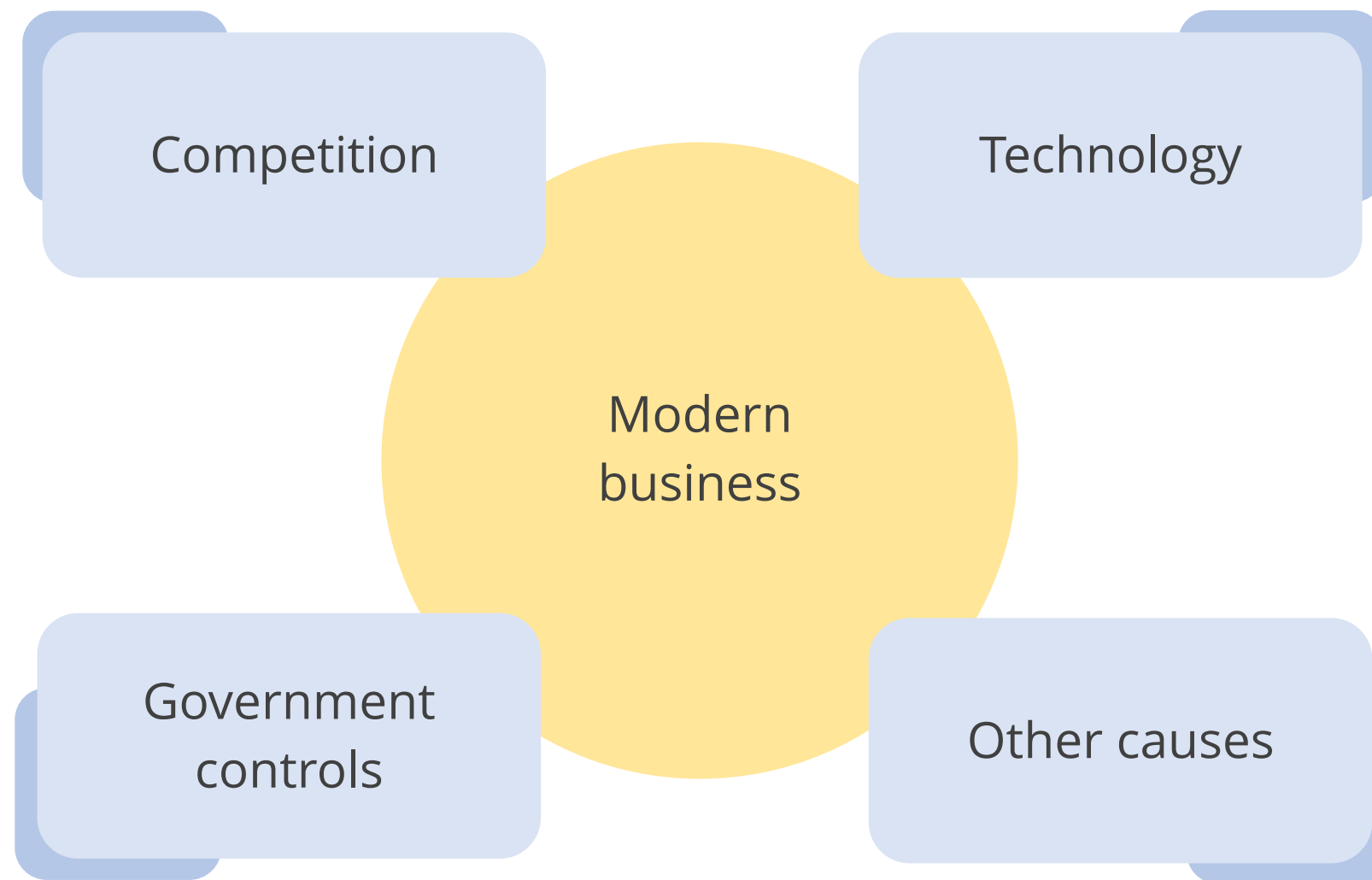


Does data statistics improve decision-making?

- How does it help improve a business?
- Which techniques are used in statistics?

Statistics Is Vital for Business Success

Business decision-making is a complex process.



Role of Computer and Information Technology

The scope of statistical work has expanded due to the development of computers and information technology.



Automation of Data Analysis

The use of software has automated data analysis.



Minitab[®] 19

Such software can process vast amounts of data and quickly generate results on a computer.

Use of Statistics

Statistical techniques are used extensively in:



Market research



Quality
management



Psychometric
studies

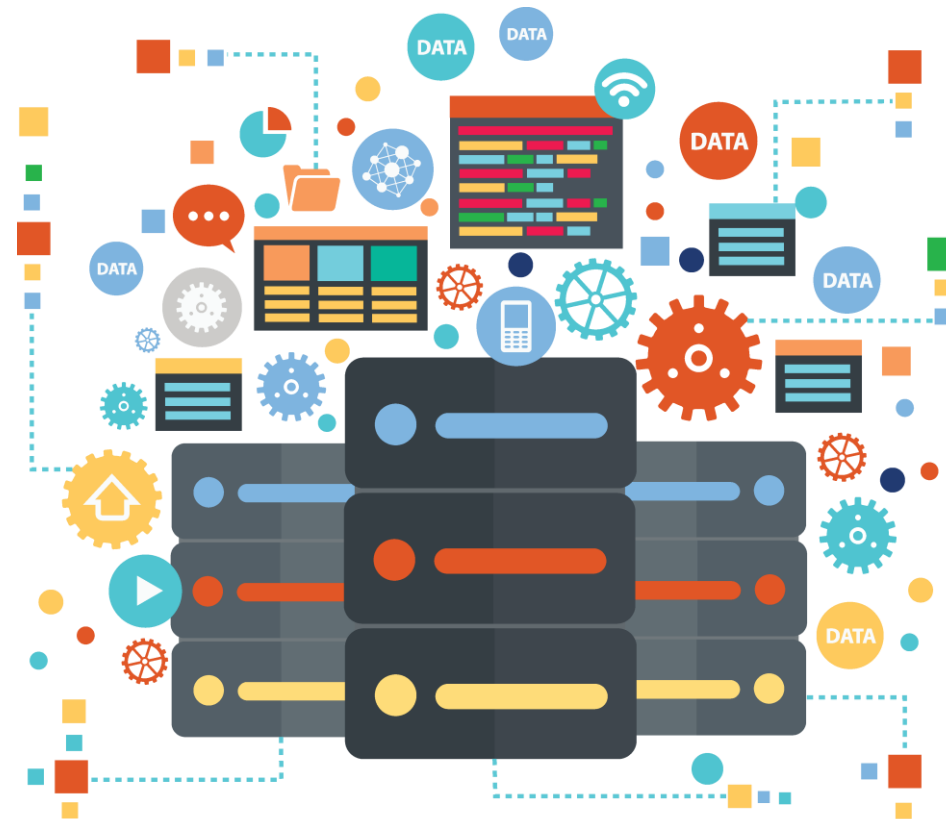
Use of Statistics

Given the prevailing confidence in big data and business analytics, the use of statistics is anticipated to increase in the coming years.



Rise of Big Data

IT development has enabled the use and real-time analysis of large volumes of data.



BIG DATA

Rise of Big Data

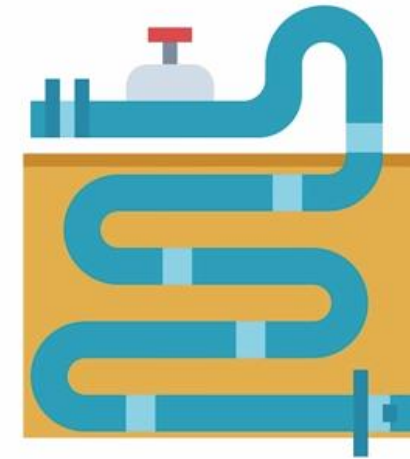
Big data is indicative of a wide-ranging variety of data, which is delivered at a faster pace and exhibits a significant increase in volume.



Refinery



Sensors



Stress on pipes

For example, a refinery installed sensors throughout the plant to collect real-time data on the stress exerted on its pipes.

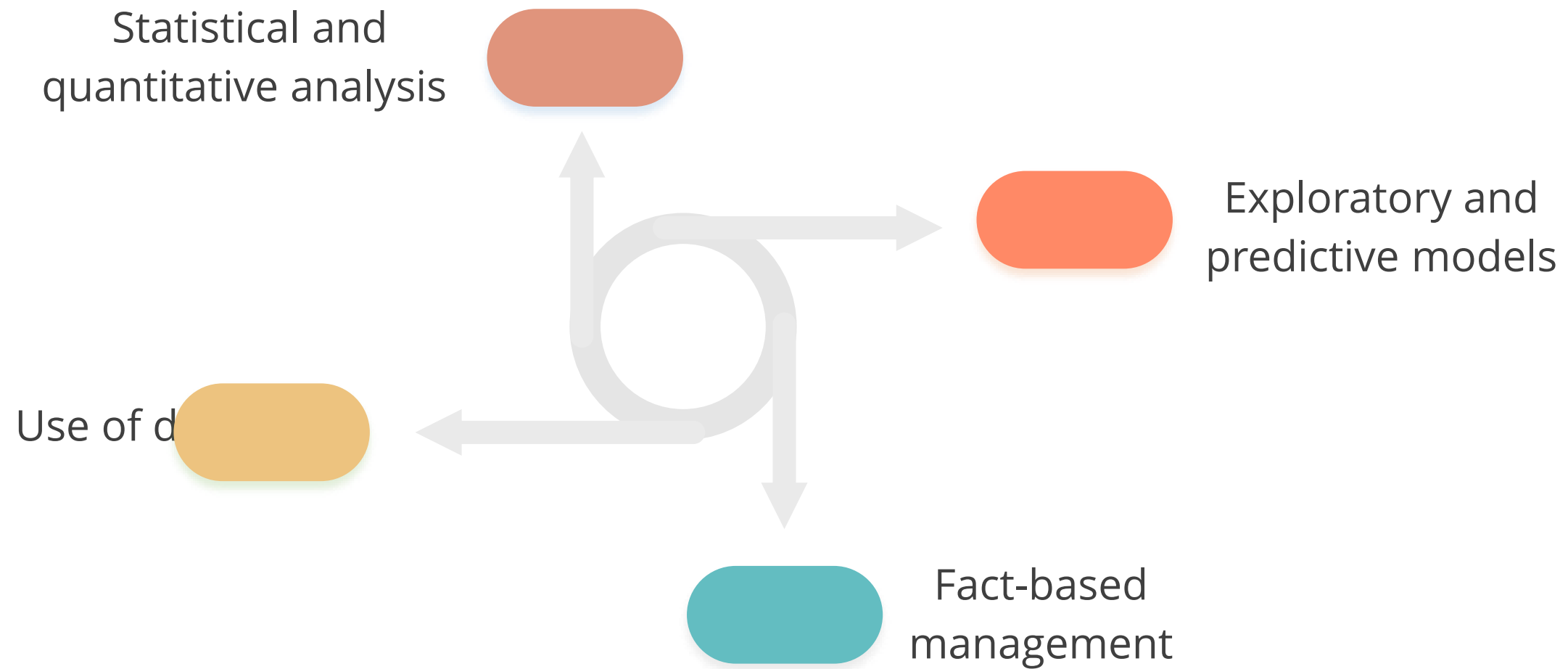
Rise of Big Data

Big data can address business-related decisions when it is thoroughly analyzed and studied.



Analytics

Analytics involves the following:



Statistics Is Vital for Business Success

Duration: 20 minutes



Does data statistics improve decision-making?

- How does it help improve a business?

Answer: Business decision-making is a complex process. Having access to accurate data when needed assists in making superior decisions and enhancing the business.

- Which techniques are used in statistics?

Answer: The techniques used for decision-making include market research, quality management, and psychometric studies.



Case Studies of Statistics Use in Business



Discussion

Case Studies of Statistics

Duration: 15 minutes



Today's healthcare sector is extensive, consistently generating, collecting, and storing vast amounts of data. This data comprises patient information, details about doctors, medication records, and bed availability. Analyzing such data allows healthcare institutions to make informed decisions about medication use, medical staff performance, and patient welfare. Provide additional examples of instances where statistical analysis can improve healthcare outcomes.

To explore this topic further, let's examine examples from various industries.

Global Retail Chain

A retail giant analyzed the purchasing patterns of its customers.



Items that were regularly bought during a specific period were identified.

Global Retail Chain

The store rearranged the layout to place frequently purchased items closer together for the convenience of customers.



As a result, customer satisfaction and sales increased.

Global Retail Chain

A life insurance company identified existing customers who were likely to repurchase or renew their policy.

The company used the following:

Advanced propensity models to identify existing customers

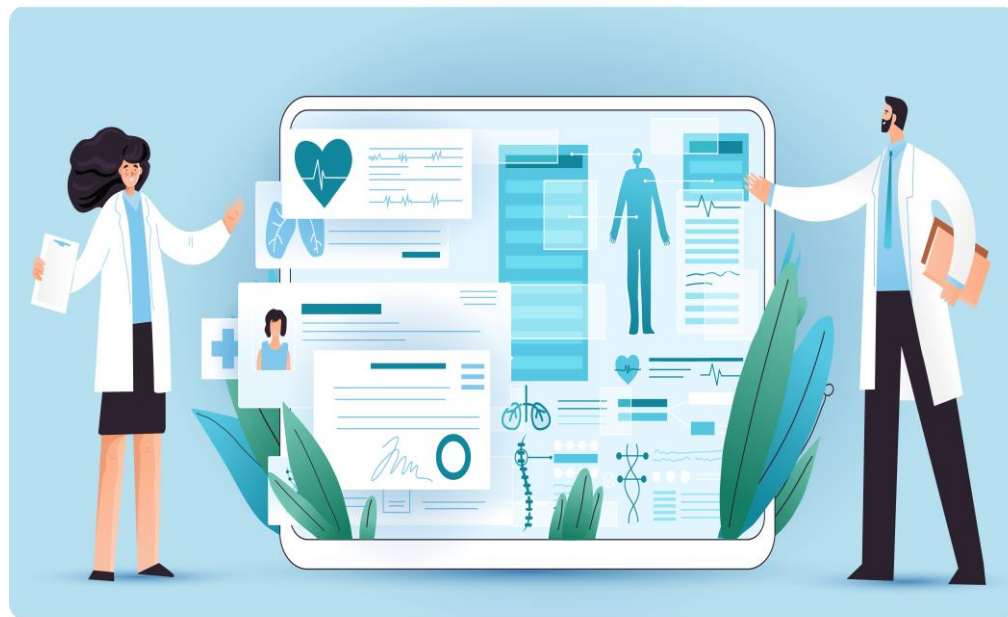
Analytics-based customer selection methodology, which enhanced the company's revenue by 80%

Statistics in Healthcare

Healthcare professionals in the United States have established a network to promote the use of data in managing healthcare systems.



Statistics in Healthcare



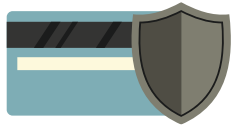
2019 HITECH Act

A huge capital was allocated to encourage the adoption and managerial use of health information systems.

Governments have also supported such initiatives.

Statistics for Machine Learning

Machine learning models deploy statistical methods for:



Credit card fraud detection



Face detection and recognition



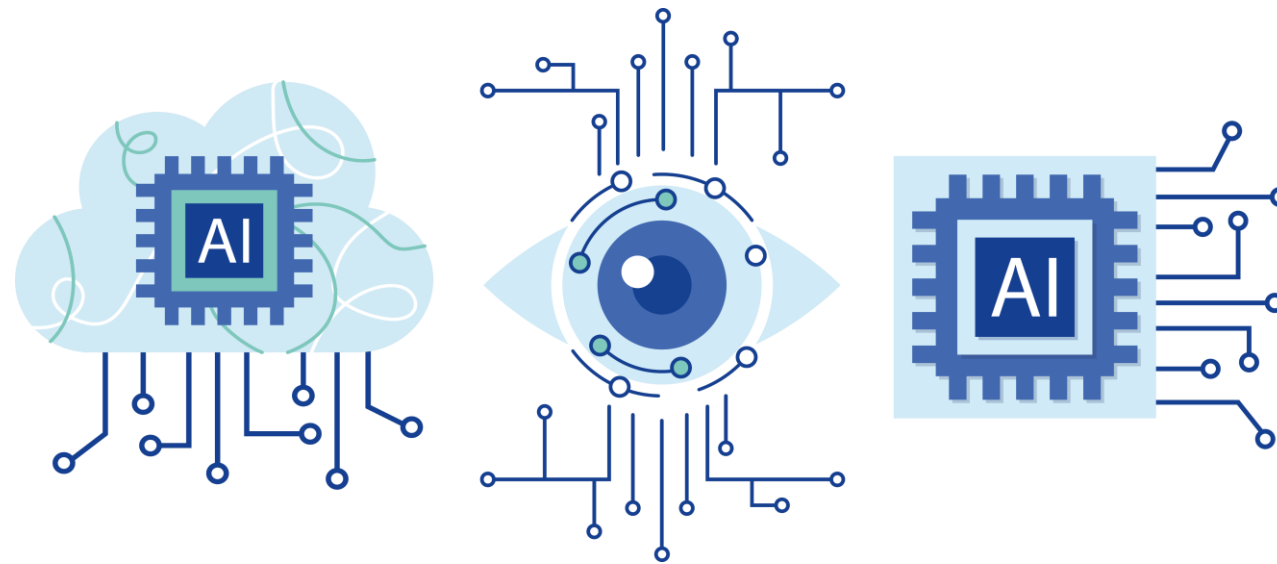
Self-driving cars



Identification of high-risk patients

Statistics in Computer Vision

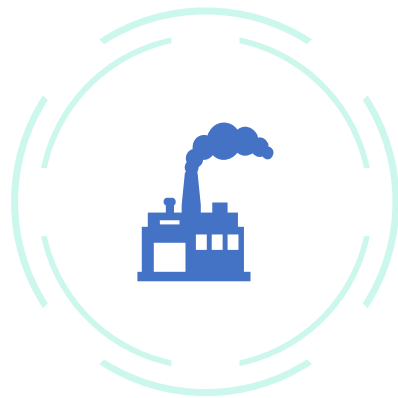
A branch of artificial intelligence, known as computer vision, has evolved with the application of basic statistics.



Software programs automatically extract information from images and videos by emulating the human eye.

Uses of Statistics

Statistics is widely used by governments and enterprises throughout the world for strategic planning.



Industrial
production



Energy
usage



Infrastructure
building

Statistics in Finance

Major banking and financial organizations, such as the Federal Reserve in New York, conduct:



Statistics for Decision-Making and Planning

Statistically derived insights in trade journals and publications of governmental bodies are widely used by:

CEOs

Policymakers

Bankers

Chambers of Commerce

Industries

Nielsen Holdings

Nielsen Holdings, a market research firm, uses statistics to build its offerings and sell research information and analysis.



Other data analytics vendors include the Datastream Group, Versium, Reklam, and AnalyticsIQ.

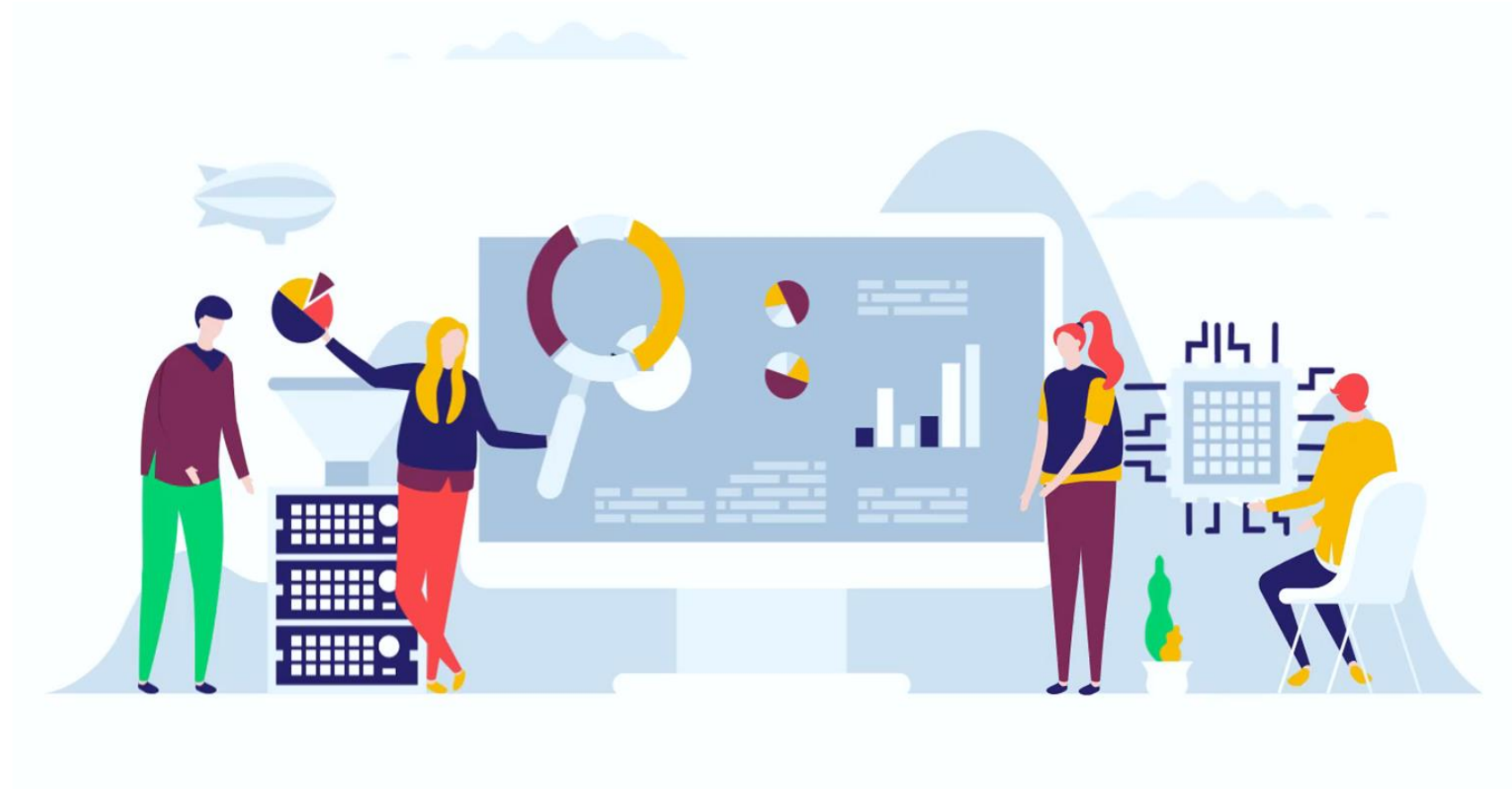
Cytel

The statistical consulting services offered by Cytel Statistical Software and Services can assist pharmaceutical companies in optimizing the clinical development process of a new drug and measuring its efficacy.



Importance of Statistics

Business, governmental, and economic decisions are driven by statistical insights and analyses.



H.G. Wells once said, "Statistical thinking will one day be as important as the ability to read and write."

Case Studies of Statistics

Duration: 15 minutes



Today's healthcare sector is extensive, consistently generating, collecting, and storing vast amounts of data. This data comprises patient information, details about doctors, medication records, and bed availability. Analyzing such data allows healthcare institutions to make informed decisions about medication use, medical staff performance, and patient welfare. Apart from this, there can be many uses for data in other fields of health care.

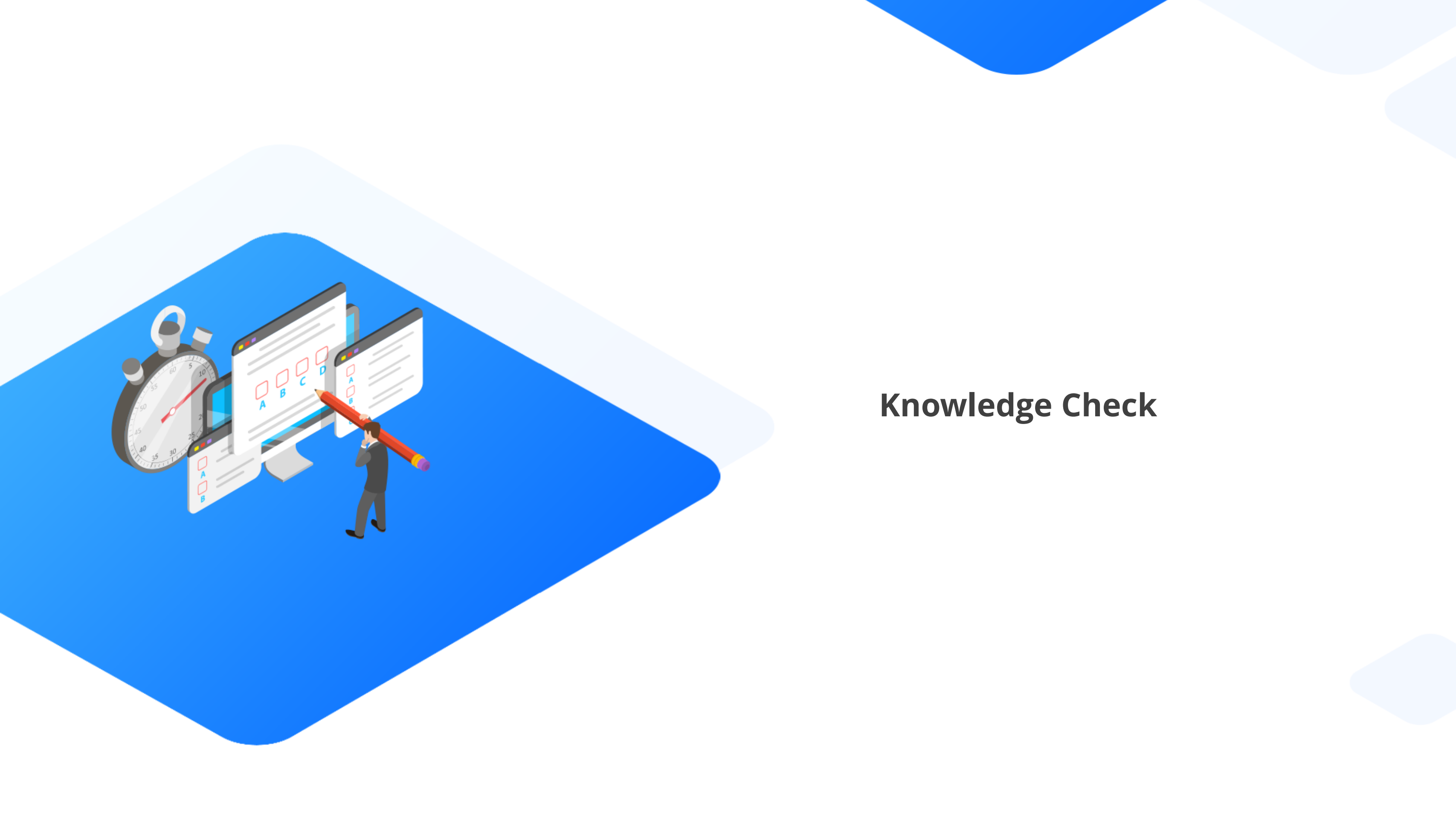
In which domains of the healthcare industry, apart from hospitals, are data statistics used?

Answer: Data statistics is also used in other healthcare domains such as research institutions, universities, and consulting firms.

Key Takeaways

- Statistics is the collection, presentation, analysis, and interpretation of numerical data.
- Statistics gives us the scope to present data concisely, generalize from vast amounts of data, and make predictions and comparisons.
- Statistical methods assist data scientists in extracting knowledge and insights from data.





Knowledge Check

Knowledge Check

1

Which of the following is not a statistical technique?

- A. Market research
- B. Quality management
- C. Psychometric studies
- D. None of the above



Knowledge Check

1

Which of the following is not a statistical technique?

- A. Market research
- B. Quality management
- C. Psychometric studies
- D. None of the above

The correct answer is **D**

Market research, quality management, and psychometric studies are all statistics techniques.



**Knowledge
Check**

2

Which of the following types of statistics is used to estimate the characteristics of a population and enable decision-making based on the sample results?

- A. Descriptive
- B. Inferential
- C. Predictive
- D. All of the above



**Knowledge
Check**

2

Which of the following types of statistics is used to estimate the characteristics of a population and enable decision-making based on the sample results?

- A. Descriptive
- B. Inferential
- C. Predictive
- D. All of the above

The correct answer is **B**

Inferential statistics is used to estimate the characteristics of a population and enable decision-making based on the sample results.





Thank You