Class X

Lesson Plan

Data Science

1. Use of Statistics in Data Science

Teaching Objectives

Students will learn about

- What are Subsets?
- One-Way Frequency Table
- Central Tendency
- Some Methods of Subsetting
- Two-way frequency Table
- Standard Deviation

Number of Periods	
Theory	Practical
x	x

Teaching Plan

While teaching this chapter, tells the students statistic is a quantitative feature of a sample that frequently helps estimate or test the population constraint.

Tell the students population, sample, parameter, and statistic are four main terms in statistics.

Discuss with the students Data science is a field that works with and examines large amounts of data to give meaningful information that can be used for making decisions and solving problems.

Discuss with the students the division of a small set of data from a large set of data is known as a Subset.

Tell the students Data Frame is a two-dimensional data structure in which data is arranged in tabular form and the procedure of selecting a set of desired rows and columns from a data frame is known as subsetting.

Share with the students that the number of subsets can be determined from the number of elements in the set.

Tell the students that the subsetting the data is a useful indexing feature for accessing object elements. It can be used for selecting and filtering variables and observations.

Share with the students there are many methods of subsetting the data, such as Row-based subsetting, Column-based subsetting, and Data-based subsetting.

Tell the students that in row-based subsetting, we take some rows from the top or bottom of the table; in column-based subsetting, you may choose specific columns from the dataset; and the data-based subsetting is used to subset the data based on specific data.

Introduce the frequency as the number of times an event or a value occurs to the students.

Explain frequency table as a table that lists items and shows the number of times the items occur to the students.

Tell the students that two-way table is a statistical table that determines the observed number or frequency for two variables, the rows indicate one category and the columns indicate the other category.

Ask the students to perform **Activity 1** given on Page 120 of the main course book.

Tell the students that central tendency means the value derived from the random variables in the set of data that reflects the midpoint of the data distribution.

Tell the students that mean is a measure of central tendency. The mean is a value in the dataset around which the entire data is spread out.

Explain the three different types of mean covering Arithmetic mean, Geometric mean and Harmonic mean.

Ask the students to perform **Activity 2** given on Page 122 of the main course book.

Explain the students that median is a form of central tendency. To find the median, your numbers have to be arranged in ascending or descending order.

Tell the students that Mean Absolute Deviation (MAD) of a dataset is the average distance between each data value and the mean.

Ask the students to perform **Activity 3** given on Page 124 of the main course book.

Tell the students that mode is the number that appears the most in a dataset.

Introduce the relationship between mean, median, and mode by using a frequency distribution graph to the students.

Explain the students that standard deviation is a measurement of the data's dispersion from the mean. A low standard deviation means that the data are grouped around the mean, whereas a high standard deviation means that the data are more dispersed.

Ask the students to perform **Activity 4** given on Page 128 of the main course book.

Extension

Ask the students some oral questions based on this chapter.

Q. Explain the concept of a subset in statistics.

Q. What is a Data Frame in programming languages like R and Python?

Q. Why is subsetting data useful in data science?

Q. Explain row-based subsetting with an example.

Q. Define a one-way frequency table.

Q. How do you interpret a two-way frequency table?

Q. Explain three measures of central tendency.

Q. Describe the process of calculating standard deviation.

Q. Why is it important to understand subsets in data science?

Q. How do frequency tables help in data analysis?

Q. How are two-way relative frequency tables different from two-way frequency tables?

Q. Why is understanding standard deviation crucial for interpreting data trends?

Encourage the students to walk through the chapter and ask them to explain any one topic from the chapter.

Evaluation

After explaining the chapter, let the students do the exercises given on pages 131 to 136 in the main course book as **Exercise (Solved and Unsolved Questions)**.

Tell them to solve the critical and thinking skill developing exercises as **Higher Order Thinking Skills** given on pages 136.

Ask the students to practice the project in given in **Applied Project** section given on Page 136 in the main course book. This will help the students understand how concept of Data Science can assist them in their daily lives.

Suggested Activity

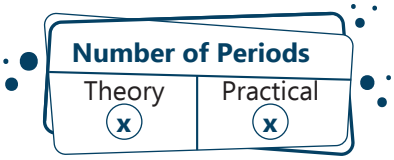
Ask the students to calculate the central tendency (mean, median, mode) of the scores on a recent exam in their class, then discuss whether the measure they chose reflects the typical performance of their peers.

2. Distributions in Data Science

Teaching Objectives

Students will learn about

- ☞ What is Distribution in Data Science?
- ☞ Types of Distributions
- ☞ Statistical Problem-Solving Process
- ☞ Activity - Choosing Groups for School Dance Programs



Teaching Plan

While teaching this chapter, tells the students distribution is a simple way to visualise a set of data. Tell the students in data science, the word "distribution" typically refers to a probability distribution. Probability distribution is a mathematical approach that displays the likely values for a variable and how frequently they occur.

Discuss with the students that discrete data refers to the data that accepts only particular values.

Tell the students continuous data refers to the data that can accept any value within a range. This range of values will have a lower bound and an upper bound, which we call the minimum and the maximum possible values.

Explain discrete probability distribution as a distribution where observation can take only a finite number of values. It is expressed with a formula (Density Function) describing the shape of the distribution.

Tell the students density functions are mathematical functions that describe the probability distribution of a random variable X.

Tell the students that Probability Mass Functions (PMF) describe the probability of a random variable X taking on a particular value x, and it is only applicable for discrete distributions.

Tell the students that discrete probability distribution is of three types: Bernoulli distribution, Binomial distribution and Poisson distribution.

Discuss with the students the Bernoulli distribution describes events that have exactly two outcomes.

Introduce binomial distribution is a common probability distribution that models the probability of obtaining one of two outcomes under a given number of parameters to the students.

Ask the students to perform **Activity 1** given on Page 142 of the main course book.

Tell the students poisson distribution is a probability distribution that is used to display the frequency of an event over a given time period.

Share with the students that all outcomes greater than 0 (which would include numbers whose decimals continue indefinitely; such as pi = 3.14159265) are used to create a continuous distribution.

Tell the students that continuous probability distributions are two types: Uniform distribution and Normal distribution.

Tell the students that a statistical distribution where the probability of outcomes is equally likely and with finite values is called the discrete uniform distribution.

Tell the students that an infinite number of equally likely measurable values make up a continuous uniform distribution.

Share with the students about characteristics of a normal distribution.

Tell the students that statistical problem solving process is a way to gather and analyse data in order to respond to statistical inquiry questions is the goal.

Ask the students to perform Activity - choosing Groups for School Dance Programs.

Extension

Ask the students some oral questions based on this chapter.

Q. How is distribution typically visualised in data science?

Q. Explain the concept of probability using the example of flipping a coin.

Q. What is a uniform distribution and how is it characterised?

Q. Describe the characteristics of discrete data. Also, provide examples of discrete data.

Q. What distinguishes continuous data from discrete data? Give examples of continuous data.

Q. Describe the Bernoulli distribution and provide an example.

Q. What is the probability mass function (PMF) in the context of discrete probability distributions?

Q. What criteria must be met for a distribution to be considered binomial?

Q. How is the probability density function (PDF) related to the cumulative distribution function (CDF)?

Q. Outline the steps of the statistical problem-solving process.

Q. What considerations are necessary when collecting data for statistical analysis?

Q. How do you interpret the results of statistical analysis?

Encourage the students to walk through the chapter and ask them to explain any one topic from the chapter.

Evaluation

After explaining the chapter, let the students do the exercises given on pages 154 to 159 in the main course book as **Exercise (Solved and Unsolved Questions)**.

Tell them to solve the critical and thinking skill developing exercises as **Higher Order Thinking Skills** given on pages 159.

Ask the students to practice the project in given in **Applied Project** section given on Page 159 in the main course book. This will help the students by analysing and interpreting the data to make informed decision-making.

Suggested Activity

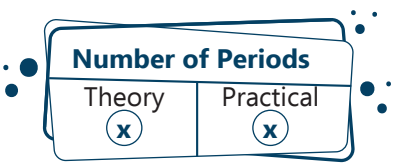
Ask the students to perform a dice rolling experiment. Based on your observations and calculations, what type of probability distribution does this experiment represent? Justify your answer.

3. Identifying Patterns

Teaching Objectives

Students will learn about

- ☞ Facts and Opinion
- ☞ What is Partiality, Preference and Prejudice?
- ☞ How to Identify the Partiality, Preference and Prejudice?
- ☞ Probability for Statistics
- ☞ What Is a Sampling Error?
- ☞ The Central Limit Theorem
- ☞ Why is the Central Limit Theorem Important?



Teaching Plan

While teaching this chapter, tells the students facts are true and cannot be argued with, because they can be proven and are supported by evidence.

Tells the students opinions vary according to the attitudes of the writer.

Discuss with the students that partiality is tendency to favour unfairly.

Tells the students preference is the act of selecting or having a particular preference for one person or thing rather than another or others.

Discuss with the students that prejudice means preconceived opinion that is not based on reason or actual experience.

Tells the students bias is the tendency to favour one thought over another and maybe to ignore competing ideas.

Ask the students to perform **Activity 1** given on Page 162 of the main course book.

Tell the students that statistical and cognitive bias can be grouped into Selection Bias, Linearity Bias, Confirmation Bias, Recall Bias, Survivor Bias, and Availability Bias categories.

Ask the students to perform **Activity 2** given on Page 165 of the main course book.

Tells the students sample is an unbiased, objective group of people chosen to represent the entire population; parameter is a characteristic of a population; standard deviation is the variation in the population that is inferred from the variation in the sample.

Tells the students a sampling error is a deviation in the sampled value versus the true population value.

Ask the students that sampling is an analysis performed by selecting a number of observations from a larger population.

Introduce Central Limit Theorem as a statistical theory stating that given a significantly large sample size from a population with finite variance all samples taken from the same population will have a mean that is nearly equal to the population mean.

Ask the students to perform **Activity 3** given on Page 171 of the main course book.

Extension

Ask the students some oral questions based on this chapter.

Q. What are some strategies for identifying bias in a text?

Q. How does preference differ from partiality?

Q. Define prejudice and explain its origins.

Q. Describe the impact of prejudice on society.

Q. What is bias in data science applications?

Q. What are the consequences of biased data in predictive models?

Q. How can selection bias be mitigated?

Q. What is recall bias and how does it affect data labelling?

Q. What are the dangers of relying solely on available data?

Q. How is probability used to predict events?

Q. Differentiate between populations, samples, parameters, and statistics.

Q. Why is the Central Limit Theorem important in statistics?

Encourage the students to walk through the chapter and ask them to explain any one topic from the chapter.

Evaluation

After explaining the chapter, let the students do the exercises given on pages 172 to 177 in the main course book as **Exercise (Solved and Unsolved Questions)**.

Tell them to solve the critical and thinking skill developing exercises as **Higher Order Thinking Skills** given on pages 177.

Ask the students to practice the project in given in **Applied Project** section given on Page 177 in the main course book.

Suggested Activity

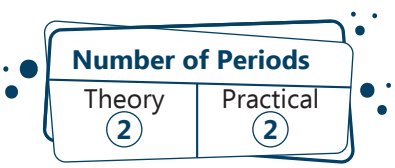
Ask the students to conduct a survey to gather information about student preferences for a new school cafeteria menu. How can you ensure your survey questions are unbiased and will produce reliable data?

4. Data Merging

Teaching Objectives

Students will learn about

- ☞ Overview of Data Merging
- ☞ What is Z-Score?
- ☞ Quartiles
- ☞ Data Joins
- ☞ Concept of Percentile
- ☞ Decilestant?



Teaching Plan

While teaching this chapter, tells the students data merging is the procedure of combining two or more datasets into a single data frame.

Tells the students there are three types of data joins which are: One to one Join, One to many Join and Many to many Join.

Discuss with the students that z-score defines the position of a point in terms of its distance from the mean when it is calculated in standard deviation units. If the value is above the mean, the z-score is positive; if it is below the mean, it is negative.

Ask the students to perform **Activity 1** given on Page 182 of the main course book.

Tells the students percentile is a number that represents the percentage of scores that fall at or below that number.

Discuss with the students that quartiles are a type of percentile. They are a set of descriptive statistics. They summarise the central tendency and variability of a dataset or distribution.

Tells the students interquartile range is defined as the measure of the middle 50% of the values when ordered from lowest to highest.

Tell the students that decile is a technique that is used to divide a distribution into ten equivalent parts. When data is divided into deciles, a decile rank is allotted to each data point in order to sort the data into an ascending or descending order.

Extension

Ask the students some oral questions based on this chapter.

Q. Why is data merging necessary in data analysis?

Q. What are some common challenges faced during data merging?

Q. Give an example of a scenario where data merging would be beneficial.

Q. What are the three types of data joins?

Q. When is a one-to-one join typically used?

Q. How does the choice of join type affect the resulting dataset?

Q. What is the relationship between z-scores and standard deviation?

Q. Discuss the utility of z-scores in probability calculations.

Q. Define percentile and its significance in statistical analysis.

Q. Discuss the relationship between quartiles and percentiles.

Q. How is the interquartile range calculated using quartile values?

Q. What is the purpose of assigning decile ranks to data points?

Encourage the students to walk through the chapter and ask them to explain any one topic from the chapter.

Evaluation

After explaining the chapter, let the students do the exercises given on pages 188 to 192 in the main course book as **Exercise (Solved and Unsolved Questions)**.

Tell them to solve the critical and thinking skill developing exercises as **Higher Order Thinking Skills** given on pages 192.

Ask the students to practice the project in given in **Applied Project** section given on Page 192 in the main course book.

Suggested Activity

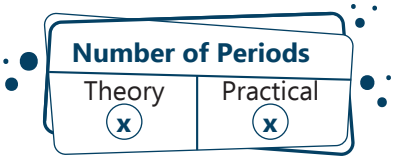
Ask the students to describe a situation in your own field of interest where data merging could be beneficial for analysis or decision-making.

5. Ethics in Data Science

Teaching Objectives

Students will learn about

- ☞ Data Governance Framework
- ☞ Ethical Guidelines Around Data Analysis
- ☞ Discarding the Data



Teaching Plan

While teaching this chapter, tells the students data governance framework is a requirement for all business organisations and governments.

Tells the students data governance framework aims at creating methods, set of responsibilities and processes to standardise, integrate, protect and store data.

Discuss with the students that data governance is a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information to help an organisation reach its objectives.

Tells the students data governance tool is a tool that helps in the process of creating and maintaining a structured set of policies, procedures, and protocols that control how an organisation's data is stored, used, and managed. In today's landscape, data is at the core of every business.

Discuss with the students that Data Governance Quality Index (DGQI) toolset provides a unique mechanism for self-evaluation of data readiness levels throughout the Government of India.

Tells the students privacy does not always mean confidentiality because private data may need to be audited based on the pertinent criteria.

Tell the students that disposition refers to the wide range of actions undertaken to manage records over time, which may include retention, destruction, or transfer to archival storage.

Tells the students destruction is the act of disposing of records permanently by obliterating records so that the information in them can no longer be physically or electronically reconstructed or recovered.

Discuss with the students that there are two possible ways you may have kept the information– in the digital format or as a physical copy.

Extension

Ask the students some oral questions based on this chapter.

Q. What is the purpose of a data governance framework in business organisations and governments?

Q. How does a data governance framework contribute to standardising, integrating, and protecting data?

Q. Discuss the importance of effective information security in data governance.

Q. Discuss the distinction between confidentiality, privacy, and personal data according to Indian laws.

Q. Explain why it is essential to protect the confidentiality of shared private information.

Q. Discuss the ethical implications of ensuring that data does not interfere with human will.

Q. What are some considerations for disposing of digital records securely?

Q. How does data destruction differ from data disposition?

Q. When is cutting up documents a suitable method for data disposal?

Q. Discuss the advantages and disadvantages of burning documents as a disposal method.

Q. What are the environmental considerations associated with document incineration as a disposal method?

Encourage the students to walk through the chapter and ask them to explain any one topic from the chapter.

Evaluation

After explaining the chapter, let the students do the exercises given on pages 198 to 202 in the main course book as **Exercise (Solved and Unsolved Questions)**.

Tell them to solve the critical and thinking skill developing exercises as **Higher Order Thinking Skills** given on pages 202.

Ask the students to practice the project in given in **Applied Project** section given on Page 202 in the main course book.

Suggested Activity

Ask the students to if you're a member of the school's yearbook committee. How would you manage student photos and personal information to create a memorable yearbook while also respecting everyone's privacy?