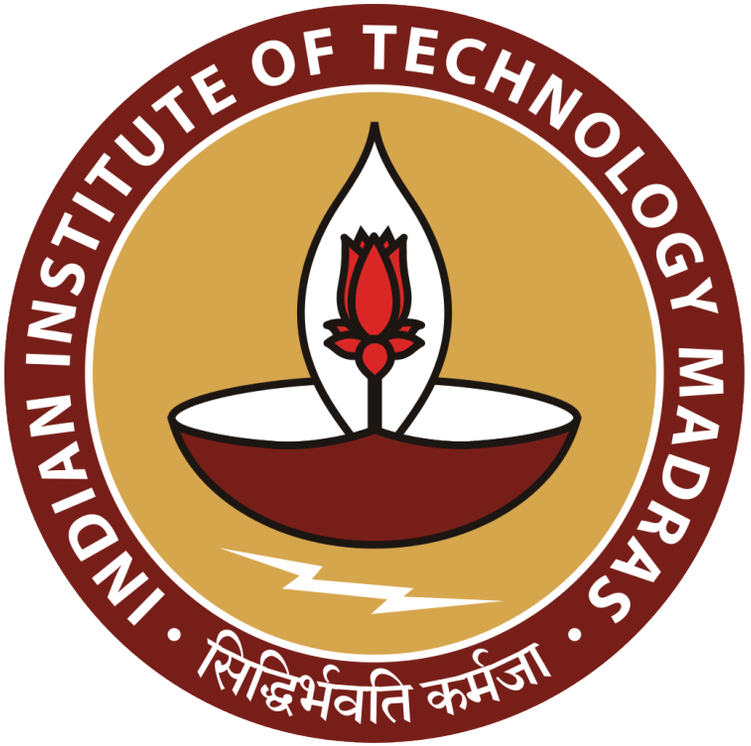


IIT Madras

ONLINE DEGREE

Statistics for Data Science -1

Course Overview

Usha Mohan

Indian Institute of Technology Madras

Introduction and course overview

Week wise schedule and learning objectives

Introduction

Statistics for Data Science-1 is an introductory course in statistics intended for beginners.

Introduction

Statistics for Data Science-1 is an introductory course in statistics intended for beginners. Students learn to create handle data sets and summarize them using both graphical techniques and numerical techniques.

Introduction

Statistics for Data Science-1 is an introductory course in statistics intended for beginners. Students learn to create handle data sets and summarize them using both graphical techniques and numerical techniques. Further, the notion of uncertainty is introduced and probability as a tool to handle uncertainty is discussed in detail.

Introduction

Statistics for Data Science-1 is an introductory course in statistics intended for beginners. Students learn to create handle data sets and summarize them using both graphical techniques and numerical techniques. Further, the notion of uncertainty is introduced and probability as a tool to handle uncertainty is discussed in detail. The concept of a random variable is introduced with a detailed discussion on the Binomial distribution and Normal distribution.

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

1. To create, download, and manipulate datasets.

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

1. To create, download, and manipulate datasets.
2. To learn methods for presenting and describing sets of data. Select an appropriate graphical technique for a given scenario.

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

1. To create, download, and manipulate datasets.
2. To learn methods for presenting and describing sets of data. Select an appropriate graphical technique for a given scenario.
3. To learn measures that can be used to summarize a data set. Use of appropriate numerical summaries for a given scenario/question.

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

1. To create, download, and manipulate datasets.
2. To learn methods for presenting and describing sets of data. Select an appropriate graphical technique for a given scenario.
3. To learn measures that can be used to summarize a data set. Use of appropriate numerical summaries for a given scenario/question.
4. To understand uncertainty through probability.

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

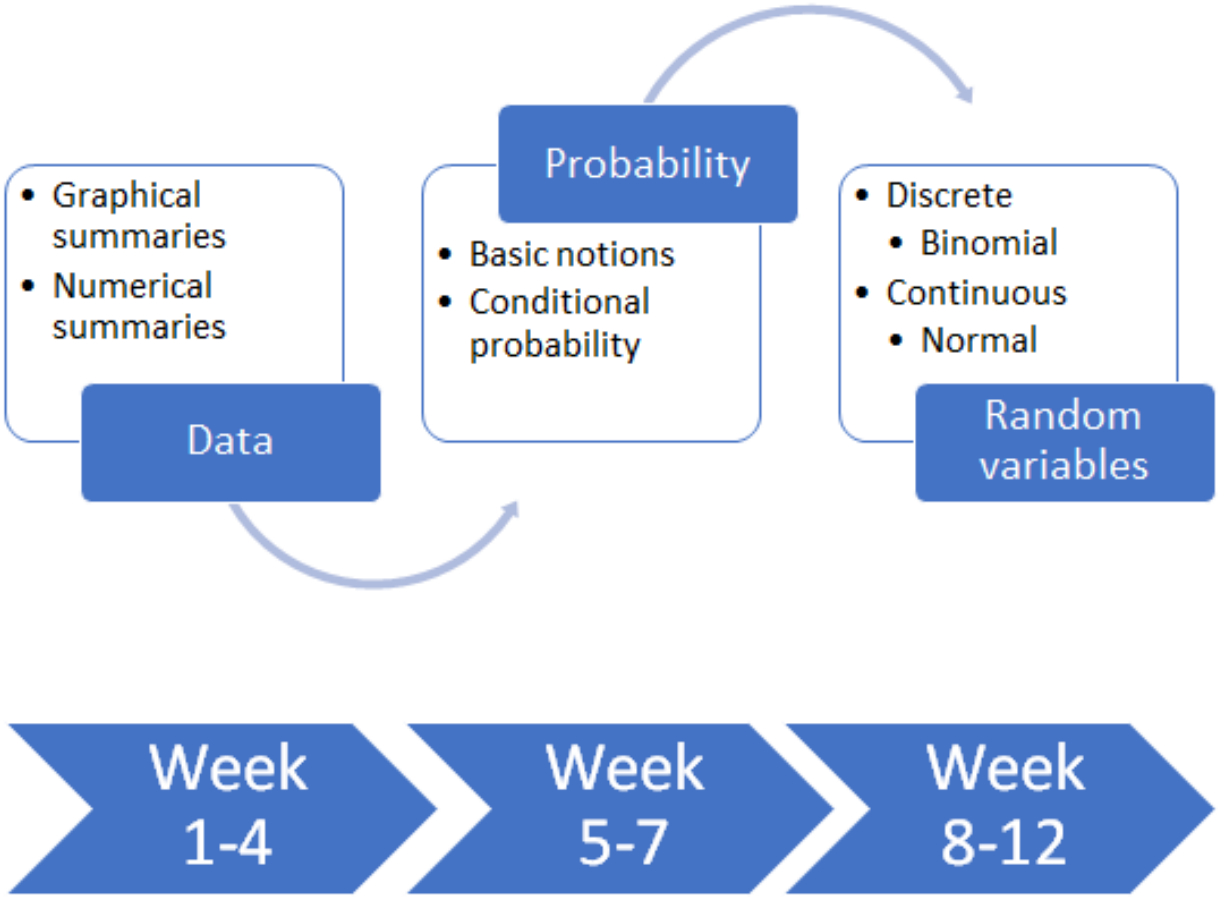
1. To create, download, and manipulate datasets.
2. To learn methods for presenting and describing sets of data. Select an appropriate graphical technique for a given scenario.
3. To learn measures that can be used to summarize a data set. Use of appropriate numerical summaries for a given scenario/question.
4. To understand uncertainty through probability.
   - 4.1 Understand notions of random experiment, events, probability and conditional probability.

Course objectives

To provide students an understanding of statistics at a conceptual level to achieve the following objectives:

1. To create, download, and manipulate datasets.
2. To learn methods for presenting and describing sets of data. Select an appropriate graphical technique for a given scenario.
3. To learn measures that can be used to summarize a data set. Use of appropriate numerical summaries for a given scenario/question.
4. To understand uncertainty through probability.
   - 4.1 Understand notions of random experiment, events, probability and conditional probability.
   - 4.2 Understand use of random variables, both discrete (in particular, Binomial) and continuous (in particular, Normal).

Road map


• Graphical summaries
• Numerical summaries
Data
Probability
• Basic notions
• Conditional probability
• Discrete
• Binomial
• Continuous
• Normal
Random variables
Week 1-4
Week 5-7
Week 8-12

Example

XYZ university has just completed admissions to their undergraduate program. Every admitted student fills up a form and the information is tabulated.

`https://docs.google.com/spreadsheets/d/15nJvZ-xBZDGb0oii-NCvSIY4fETotXcJdm5pV1Fq2aI/edit?usp=sharing`

Example

XYZ university has just completed admissions to their undergraduate program. Every admitted student fills up a form and the information is tabulated.

`https://docs.google.com/spreadsheets/d/15nJvZ-xBZDGb0oii-NCvSIY4fETotXcJdm5pV1Fq2aI/edit?usp=sharing`

A portion of the data obtained by the admissions office is given below:

Example

XYZ university has just completed admissions to their undergraduate program. Every admitted student fills up a form and the information is tabulated.

`https://docs.google.com/spreadsheets/d/15nJvZ-xBZDGb0oii-NCvSIY4fETotXcJdm5pV1Fq2aI/edit?usp=sharing`

A portion of the data obtained by the admissions office is given below:

S.No	Name	Gender	Date of Birth	Marks in Class 10	Board (Board)	Marks in Class 12	Board (Class 12)	Mobile Number
1	Anjali	F	17-Feb-03	484	State Board	394	CBSE	xxx7252826
2	Pradeep	M	03-Jun-02	514	ICSE	437	ICSE	xxx5243748
3	Varsha	F	02-Mar-01	527	CBSE	442	CBSE	xxx5242824
4	Divya	F	22-Mar-03	397	State Board	401	State Board	xxx6546889
5	Thomas	M	19-Dec-02	562	CBSE	451	CBSE	xxx4242736
6	Sarita	F	19-May-02	533	ICSE	462	ICSE	xxx5242577
7	Prashant	M	30-Oct-01	496	CBSE	413	CBSE	xxx3352630
8	Harsha	M	11-Feb-01	436	CBSE	375	CBSE	xxx1702736
9	Rafiq	M	31-Jul-02	501	ICSE	423	CBSE	xxx0026248

Questions

1. Identify variables, observations

Week 1: Introduction

Week 1: Introduction

1. Understand how data are collected.

Week 1: Introduction

1. Understand how data are collected.
   - Identify variables and cases (observations) in a data set

Week 1: Introduction

1. Understand how data are collected.
   - Identify variables and cases (observations) in a data set
2. Types of data- classify data as categorical(qualitative) or numerical(quantitative) data.

Week 1: Introduction

1. Understand how data are collected.
   - Identify variables and cases (observations) in a data set
2. Types of data- classify data as categorical(qualitative) or numerical(quantitative) data.
3. Understand cross-sectional versus time-series data.

Week 1: Introduction

1. Understand how data are collected.
   - Identify variables and cases (observations) in a data set
2. Types of data- classify data as categorical(qualitative) or numerical(quantitative) data.
3. Understand cross-sectional versus time-series data.
4. Creating data sets; Downloading and manipulating data sets; working on subsets of data.

Week 1: Introduction

1. Understand how data are collected.
   - Identify variables and cases (observations) in a data set
2. Types of data- classify data as categorical(qualitative) or numerical(quantitative) data.
3. Understand cross-sectional versus time-series data.
4. Creating data sets; Downloading and manipulating data sets; working on subsets of data.
5. Framing questions that can be answered from data.

Week 1: Introduction

1. Understand how data are collected.
   - Identify variables and cases (observations) in a data set
2. Types of data- classify data as categorical(qualitative) or numerical(quantitative) data.
3. Understand cross-sectional versus time-series data.
4. Creating data sets; Downloading and manipulating data sets; working on subsets of data.
5. Framing questions that can be answered from data.
6. Distinguish between a sample and a population.

Questions

1. What is the gender diversity, in other words, what is the proportion of women students and proportion of male students?
2. How many students come from each board?

Week 2: Describing categorical data- one variable

Week 2: Describing categorical data- one variable

1. Organizing and graphing categorical data.

Week 2: Describing categorical data- one variable

1. Organizing and graphing categorical data.
2. Create frequency tables for tabulated data.

Week 2: Describing categorical data- one variable

1. Organizing and graphing categorical data.
2. Create frequency tables for tabulated data.
3. Choosing an appropriate graphical technique for displaying data.

Week 2: Describing categorical data- one variable

1. Organizing and graphing categorical data.
2. Create frequency tables for tabulated data.
3. Choosing an appropriate graphical technique for displaying data.
4. Discuss about misleading graphs.

Questions

1. What are the average marks obtained by students in Class 10/Class 12?
2. Is there a lot of variability in the marks obtained?
3. What is the least mark obtained? Highest marks obtained?
4. What is the average age of students admitted?

Week 3: Describing numerical data- one variable

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution
2. Compute and interpret numerical summaries of data

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution
2. Compute and interpret numerical summaries of data
   2.1 Compute and interpret measures of central tendency: mean, median, mode.

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution
2. Compute and interpret numerical summaries of data
   - 2.1 Compute and interpret measures of central tendency: mean, median, mode.
   - 2.2 Compute and interpret measures of dispersion: range, variance, standard deviaiton.

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution
2. Compute and interpret numerical summaries of data
   - 2.1 Compute and interpret measures of central tendency: mean, median, mode.
   - 2.2 Compute and interpret measures of dispersion: range, variance, standard deviaiton.
   - 2.3 Compute and interpret percentiles, Interquartile Range (IQR).

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution
2. Compute and interpret numerical summaries of data
   - 2.1 Compute and interpret measures of central tendency: mean, median, mode.
   - 2.2 Compute and interpret measures of dispersion: range, variance, standard deviaiton.
   - 2.3 Compute and interpret percentiles, Interquartile Range (IQR).
3. Compute and interpret five-number summary

Week 3: Describing numerical data- one variable

1. Visual representation of numerical data and interpret shape of distribution
2. Compute and interpret numerical summaries of data
   - 2.1 Compute and interpret measures of central tendency: mean, median, mode.
   - 2.2 Compute and interpret measures of dispersion: range, variance, standard deviaiton.
   - 2.3 Compute and interpret percentiles, Interquartile Range (IQR).
3. Compute and interpret five-number summary
4. Use histogram and box-plot to identify outliers in a dataset.

Questions

1. Are there more women from state board when compared to men from state board?
2. Do students who have scored high marks in Class 10 score high marks in class 12 also?
3. Do students from State board score higher marks than those from other boards?

Week 4: Association between two variables

Week 4: Association between two variables

1. Use of two-way contingency tables to understand association between two categorical variables.

Week 4: Association between two variables

1. Use of two-way contingency tables to understand association between two categorical variables.
2. Understand association between numerical variables through scatter plot; compute and interpret correlation.

Week 4: Association between two variables

1. Use of two-way contingency tables to understand association between two categorical variables.
2. Understand association between numerical variables through scatter plot; compute and interpret correlation.
3. Understand relationship between a categorical and numerical variable.

Questions

After joining a college, the students want to form committees.

1. How many ways can a committee of 3 be formed from 10 people?
2. How many ways can a committee of 3 (President, Vice-president, and secretary) be formed from 10 people?
3. Basic principle of counting.

Week 5: Permutations and combinations

Week 5: Permutations and combinations

1. Understand the basic principle of counting.

Week 5: Permutations and combinations

1. Understand the basic principle of counting.
2. Concept of factorials.

Week 5: Permutations and combinations

1. Understand the basic principle of counting.
2. Concept of factorials.
3. Understand differences between counting with order (permutation) and counting without regard to order (combination).

Week 5: Permutations and combinations

1. Understand the basic principle of counting.
2. Concept of factorials.
3. Understand differences between counting with order (permutation) and counting without regard to order (combination).
4. Use permutations and combinations to answer real life applications.

Questions

1. What are the chances of a student getting a top grade?
2. What are the chances of a student getting a top grade given the student is from a particular board?
3. Key word is "chance"

Week 6-7: Probability

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.
3. Understand the notion of simple event and compound events.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.
3. Understand the notion of simple event and compound events.
4. Basic laws of probability.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.
3. Understand the notion of simple event and compound events.
4. Basic laws of probability.
5. Calculate probabilities of events and use a tree diagram to compute probabilities.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.
3. Understand the notion of simple event and compound events.
4. Basic laws of probability.
5. Calculate probabilities of events and use a tree diagram to compute probabilities.
6. Understand notion of conditional probability, i.e find the probability of an event given another event has occurred.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.
3. Understand the notion of simple event and compound events.
4. Basic laws of probability.
5. Calculate probabilities of events and use a tree diagram to compute probabilities.
6. Understand notion of conditional probability, i.e find the probability of an event given another event has occurred.
7. Distinguish between independent and dependent events.

Week 6-7: Probability

1. Understand uncertainty and concept of a random experiment.
2. Describe sample spaces, events of random experiments.
3. Understand the notion of simple event and compound events.
4. Basic laws of probability.
5. Calculate probabilities of events and use a tree diagram to compute probabilities.
6. Understand notion of conditional probability, i.e find the probability of an event given another event has occurred.
7. Distinguish between independent and dependent events.
8. Solve applications of probability.

Questions

Suppose one of the questions asked in the questionnaire asked students to report the number of siblings(sisters and brothers) they have.

1. What is the chance that a randomly selected student has 2 siblings?

Week 8-9: Discrete random variables

Week 8-9: Discrete random variables

1. Define what is a random variable.

Week 8-9: Discrete random variables

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.

Week 8-9: Discrete random variables

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.
3. Probability mass function, graph, and examples.

Week 8-9: Discrete random variables

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.
3. Probability mass function, graph, and examples.
4. Cumulative distribution function, graphs, and examples.

Week 8-9: Discrete random variables

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.
3. Probability mass function, graph, and examples.
4. Cumulative distribution function, graphs, and examples.
5. Expectation and variance of a random variable.

Questions

A multiple-choice examination has 4 possible answers for each of 25 questions.:

1. What is the chance of getting exactly 5 questions correct just by guessing?
2. What is the chance of getting more than 5 questions correct just by guessing?

Week 10: Binomial distribution

Week 10: Binomial distribution

1. Understand the binomial distribution.

Week 10: Binomial distribution

1. Understand the binomial distribution.
2. Applications of binomial distribution.

Questions

The time taken to write a test is recorded for each student. What is the chance that

1. the student requires more than 45 minutes to complete the test?
2. The student requires between 30 to 45 minutes to complete the test?

Week 11-12: Continuous distributions and Normal Distribution

Week 11-12: Continuous distributions and Normal Distribution

1. Concept of probabaility density function

Week 11-12: Continuous distributions and Normal Distribution

1. Concept of probabaility density function
2. The empirical rule of Normal distribution

Week 11-12: Continuous distributions and Normal Distribution

1. Concept of probabaility density function
2. The empirical rule of Normal distribution
3. Standard Normal distribution.

Week 11-12: Continuous distributions and Normal Distribution

1. Concept of probabaility density function
2. The empirical rule of Normal distribution
3. Standard Normal distribution.
4. Applications of Normal distributions.