PROBABILITY AND STATISTICS

Week 1

Spring-2025

Instructor: Rida Maryam

Today’s Agenda

- Introductory Session
- Course Objectives
- Learning Outcomes
- Administrative Items
- Fundamental Concepts
- Questions ?

Office Hours

- Office: D-Block, 2nd floor, table 5.
- Email: rida.maryam@ucp.edu.pk
- Office Hours:

* will be updated during the semester if required

Overall Learning Outcomes

- Upon completion of this course, you will be able to:
  - Introduce fundamental concepts of statistics, probability, and their applications in software engineering.
  - Develop an understanding of data visualization, descriptive statistics, and probability distributions.
  - Familiarize students with probability rules, Bayes' theorem, and statistical inference techniques.
  - Apply probability models and statistical methods to real world engineering and computing problems.
  - Enhance decision-making skills using hypothesis testing, confidence intervals, and regression analysis

Plagiarism Policy

- All the parties involved in first cheating case will be awarded Zero for that evaluation.
- Afterwards, if any cheating made by any person of the class, will get F in course for all the accused parties.

Tentative Grading Policy *

Class Participation	**10%**
Quiz	**15%**
Assignments	**10%**
Mid Term	**25%**
Final	**40%**

Course Sessions

Highly Interactive !!!

Probability and Statistics

Text Book:

- Probability and Statistics for Computer Science David Forsyth.
- Probability and Stochastic Processes A Friendly Introduction for Electrical and Computer Engineers 3rd Ed – Roy D. Yates and David J. Goodman

Reference Books

1. Introductory Statistics by Barbara Illowsky, De Anza College, Susan Dean, De Anza
2. Probability and Statistics for Engineers and Scientists by Walpole
3. Advanced Level STATISTICS 1 by Steve Dobbs and Jane Miller
4. Discrete Mathematics and its applications, Kenneth H. Rosen.
5. Mathematics for Computer Science, Eric Lehman. (MIT)

Probability and Statistics

Lecture No. 01

- **Basics of Set Theory**
- **Set operations.**
- **Venn Diagrams and Their Applications in Probability**

Set Theory and Set operations

Set

- A set is a well-defined collection of distinct objects or people”
  - Rainbow Colors: { red, orange, yellow, green, blue, indigo, violet }
  - Pakistani Cities: { Islamabad, Lahore, Karachi, Peshawar, Quetta … }
- Sets are notated with curly brackets
- Sets do not have duplicate elements
  - Consider the set of vowels in the alphabet.
    - It makes no sense to list them as {a, a, a, e, i, o, o, o, o, o, u}
    - What we really want is just {a, e, i, o, u}
- Order does not matter
  - We often write them in order because it is easier for humans to understand it that way
  - {1, 2, 3, 4, 5} *is equivalent to* {3, 5, 2, 4, 1}

Specifying a Set – I

- Sets are usually represented by a capital letter (A, B, S, etc.)
- Elements are usually represented by an italic lower-case letter (*a*, *x*, *y*, etc.)
- Easiest way to specify a set is to list all the elements: A = {1, 2, 3, 4, 5}
  - Not always feasible for large or infinite sets
- Can use an ellipsis (…): B = {0, 1, 2, 3, …}
  - Can cause confusion. Consider the set C = {3, 5, 7, …}. What comes next?
  - If the set is all odd integers greater than 2, it is 9
  - If the set is all prime numbers greater than 2, it is 11

Specifying a Set – II

- Can use set-builder notation
  - D = {*x* | *x* is prime and *x* > 2}
  - E = {*x* | *x* is odd and *x* > 2}
  - The vertical bar means “such that”
  - Thus, set D is read (in English) as: “all elements *x* such that *x* is prime and *x* is greater than 2”
- A set is said to “contain” the various “members” or “elements” that make up the set
  - If an element *a* is a member of (or an element of) a set S, we use then notation $a \in S$
    - $4 \in \{1, 2, 3, 4\}$
  - If an element is not a member of (or an element of) a set S, we use the notation $a \notin S$
    - $7 \notin \{1, 2, 3, 4\}$
    - London $\notin \{1, 2, 3, 4\}$

Often used Sets

- **N** = {0, 1, 2, 3, …} is the set of natural numbers
- **Z** = {…, -2, -1, 0, 1, 2, …} is the set of integers
- $\mathbf{Z}^+$ = {1, 2, 3, …} is the set of positive integers (a.k.a whole numbers)
  - Note that people disagree on the exact definitions of whole numbers and natural numbers
- **Q** = $\{p/q \mid p \in \mathbf{Z}, q \in \mathbf{Z}, q \neq 0\}$ is the set of rational numbers
  - Any number that can be expressed as a fraction of two integers (where the bottom one is not zero)
- **R** is the set of real numbers

The Empty Set

- If a set has zero elements, it is called the empty (or null) set
  - Written using the symbol ∅
  - Thus, ∅ = { } ← **VERY IMPORTANT**
  - If you get confused about the empty set in a problem, try replacing ∅ by { }
- As the empty set is a set, it can be an element of other sets
  - { ∅, 1, 2, 3, x } is a valid set
- Note that ∅ ≠ { ∅ }
  - The first is a set of zero elements
  - The second is a set of 1 element (that one element being the empty set)

Set Equality

- Two sets are equal if they have the same elements
  - {1, 2, 3, 4, 5} = {5, 4, 3, 2, 1}
  - {1, 2, 3, 2, 4, 3, 2, 1} = {4, 3, 2, 1}
- Two sets are not equal if they do not have the same elements
  - {1, 2, 3, 4, 5} ≠ {1, 2, 3, 4}

Subsets

- If all the elements of a set S are also elements of a set T, then S is a subset of T
  - For example, if S = {2, 4, 6} and T = {1, 2, 3, 4, 5, 6, 7}, then S is a subset of T
  - This is specified by $S \subseteq T$ **OR** $\{2, 4, 6\} \subseteq \{1, 2, 3, 4, 5, 6, 7\}$
- If S is not a subset of T, it is written as: $S \not\subseteq T$
  - For example, $\{1, 2, 8\} \not\subseteq \{1, 2, 3, 4, 5, 6, 7\}$
- Note that any set is a subset of itself!
  - Given set S = {2, 4, 6}, since all the elements of S are elements of S, S is a subset of itself
  - This is kind of like saying 5 is less than or equal to 5
  - Thus, for any set S, $S \subseteq S$

Proper Subsets

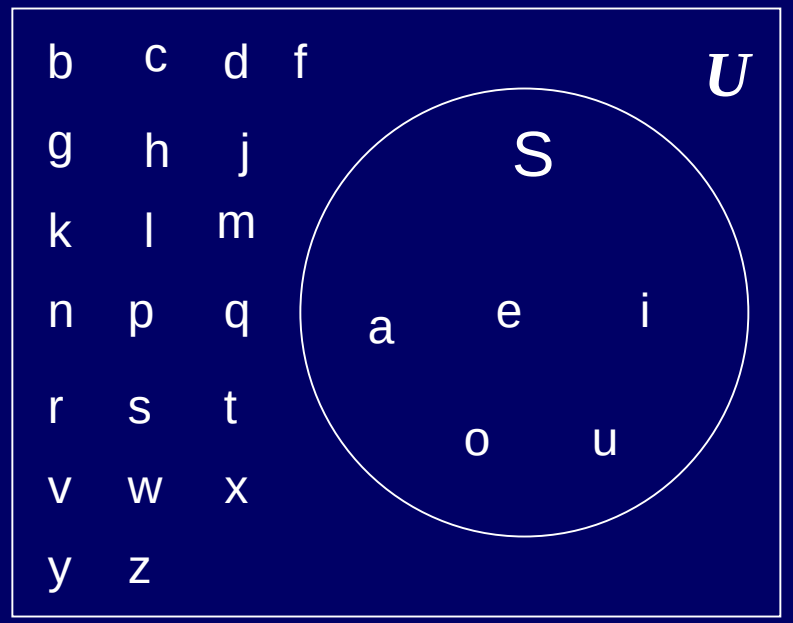
- If S is a subset of T, and S is not equal to T, then S is a proper subset of T
  - Let T = {0, 1, 2, 3, 4, 5}
  - If S = {1, 2, 3}, S is not equal to T, and S is a subset of T
  - A proper subset is written as $S \subset T$
  - Let R = {0, 1, 2, 3, 4, 5}. R is equal to T, and thus is a subset (but not a proper subset) of T
    - Can be written as: $R \subseteq T$ and $R \not\subset T$ (or just R = T)
  - Let Q = {4, 5, 6}. Q is neither a subset of T nor a proper subset of T
- The difference between “subset” and “proper subset” is like the difference between “less than or equal to” and “less than” for numbers

The Universal Set

- ***U*** is the universal set – the set of all of elements (or the "universe") from which given any set is drawn
  - For the set {-2, 0.4, 2}, U would be the real numbers
  - For the set {0, 1, 2}, U could be the natural numbers (zero and up), the integers, the rational numbers, or the real numbers, depending on the context
- For the set of the students in this class, U would be all the students in the University (or perhaps all the people in the world)
- For the set of the vowels of the alphabet, U would be all the letters of the alphabet

Venn diagrams

- Represents sets graphically
  - The box represents the universal set
  - Circles represent the set(s)
- Consider set S, which is the set of all vowels in the alphabet
- The individual elements are usually not written in a Venn diagram



Union

A = {1, 2, 3}

B = {3, 4, 5}

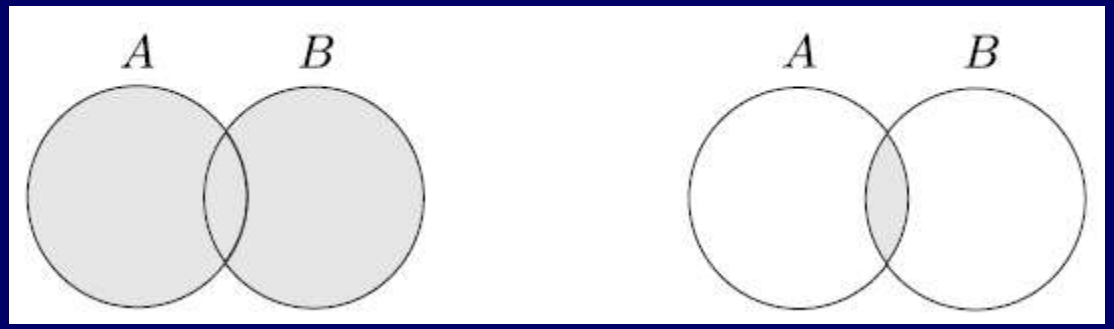
A U B = {1, 2, 3, 4, 5}

Intersection

A = {1, 2, 3}

B = {3, 4, 5}

A ∩ B = {3}



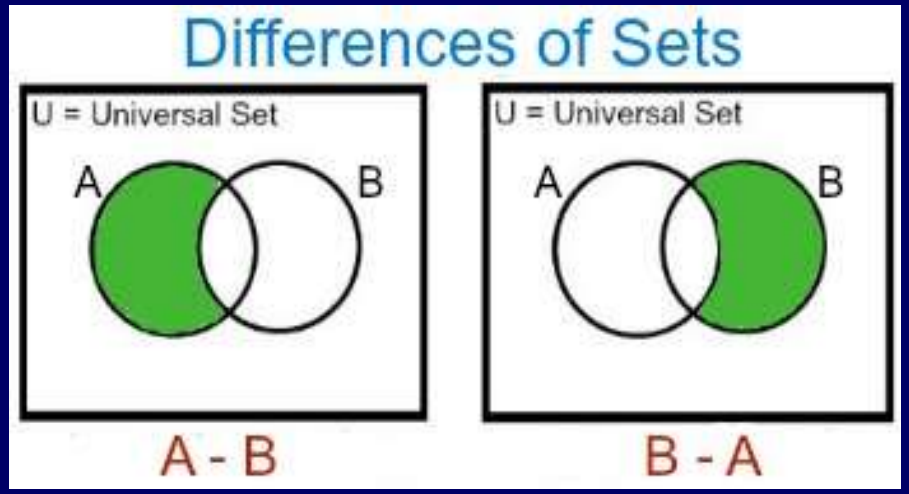
Difference

A = {1, 2, 3}

B = {3, 4, 5}

A - B = {1, 2}

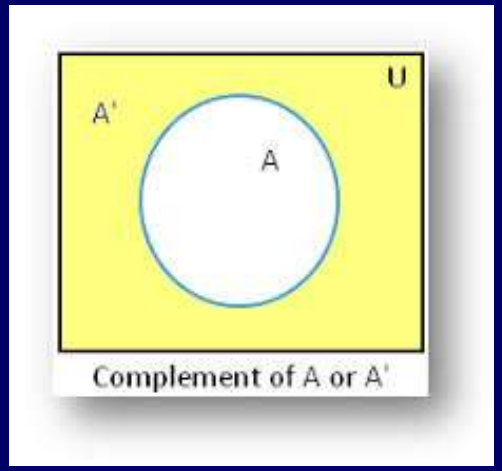
B - A = {4, 5}



Complement

A = A be the set of positive integers greater than 10

A'= {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}



Probability and Statistics

Lecture No. 02+03

- Introduction to Statistics
- Descriptive and inferential Statistics
- Population and Sample
- Variable and its types
  - Qualitative (ordinal , nominal)
  - Quantitative (Discrete and continuous)

Introduction to Statistics

Statistics is a discipline that includes techniques or group of methods used to collect, organize, present as well as analyse and interpret data to make decisions

or

That science which enables us to draw conclusions about various phenomena on the basis of real data collected on sample-basis

Statistics

Branches of Statistics

Descriptive Statistics consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures so as to yield meaningful information

Inferential Statistics consists of methods that use sample results to help make decisions or predictions about a population. Its applying probability theory on descriptive statistics and drawing conclusions

Branches of Statistics

Example

Classify the following statements as descriptive or inferential statistics.

(a) The average age of the students in this class is 21 years.

(b) At least 5% of the killings reported last year in Lahore were due to accidents.

(c) Of the students enrolled in UCP in this year 74% are female and 26% are male.

(d) The chance of winning the Lucky Lottery in any day is 1 out of 167000.

(e) The demand for automobiles may decline next year in Pakistan.

(f) It has been continuously raining in Murree from Monday to Friday. It will continue to rain in the weekend

Statistics

Population vs Sample

Population consists of all elements – individuals, items, or objects – whose characteristics are being studied. The population that is being studied is also called the ***target population***

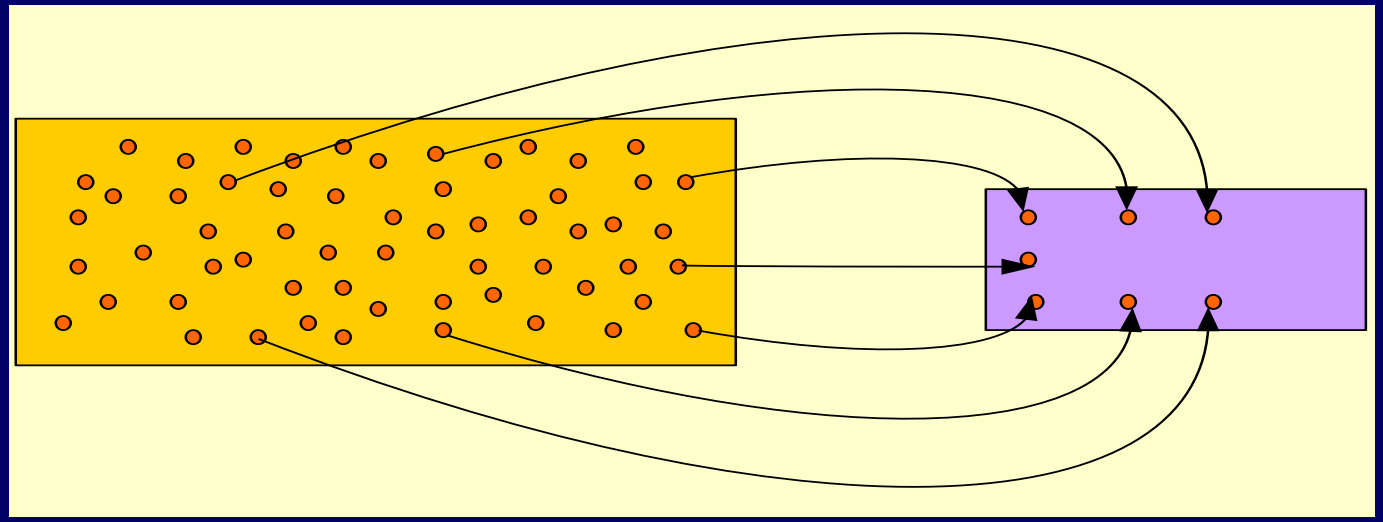
Sample portion of the population selected for study.

The sample must contain the characteristics of the population in order to be a **representative sample.**

Population and Sample

Population **vs** **Sample**

(N) **(n)**



Definitions

Variable characteristic of the individuals of a population or of a sample which varies from individual to individual i.e. height, age, weight etc

Data are the actual values of the variable.

Data observations are the recorded values of a characteristic of every individual of a population or of a sample

Data Set collection of observations on one or more variables

Variable and its Types

Types of Variables

- **Quantitative Variables (Numerical variable)**
  - **Discrete Variables**
  - **Continuous Variables**

- **Qualitative or Categorical Variables**

Variables and its Types

Types of Variables

Quantitative Variable when a characteristic can be expressed numerically such as age, weight, income or number of children.

The data collected on a quantitative variable is called quantitative data

Variables and its Types

Types of Quantitative Variable

Discrete Variable A variable whose values are countable is called a discrete variable. In other words, a discrete variable can assume only certain values with no intermediate values. i. e. A concrete no (no of chairs in classroom)

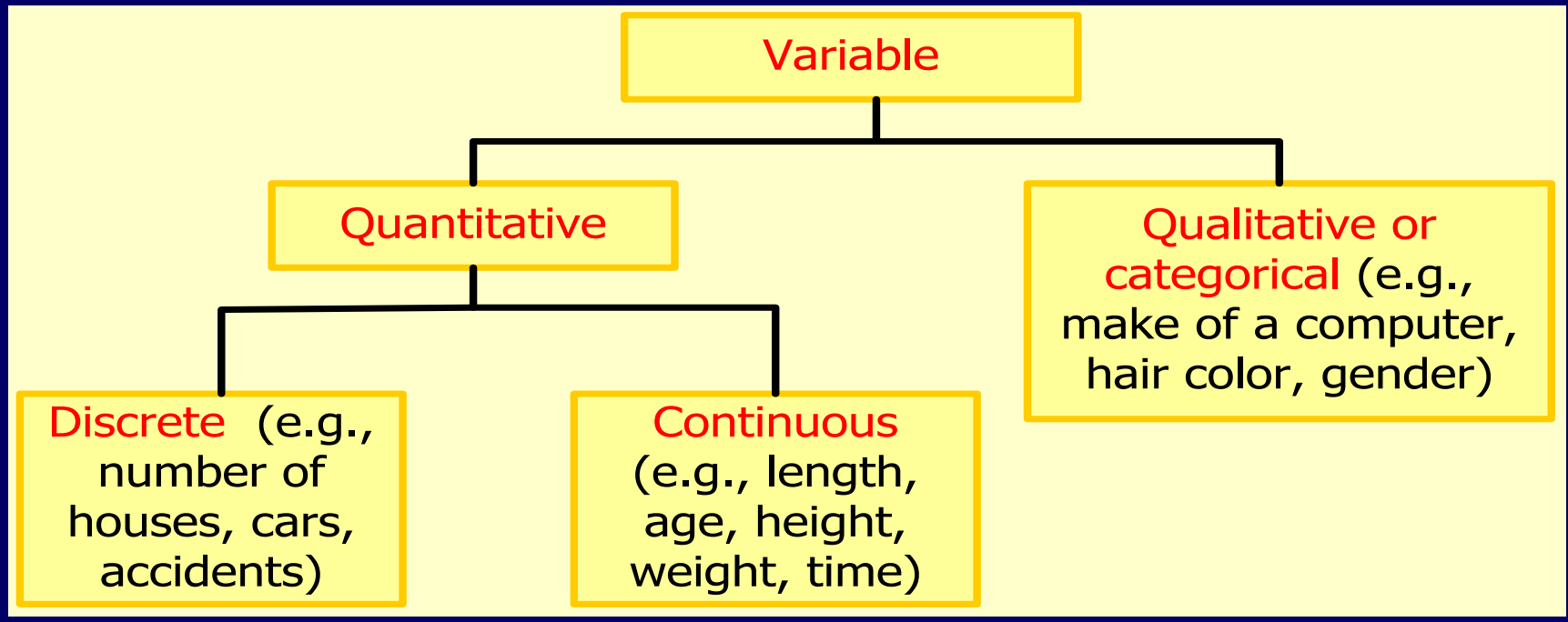
Continuous Variable A variable that can assume any numerical value over a certain interval or intervals is called a continuous variable. i.e.in fraction form (height of students)

Variable and its Types

Types of Variables

Qualitative Variable If the characteristic is non-numerical such as education, eye-colour, quality, intelligence, poverty, satisfaction, etc. the variable is referred to as a qualitative variable. A qualitative characteristic is also called an attribute

Variable and its Types



Identify correct data type (quantitative or qualitative)

a. the number of pairs of shoes you own

b. the type of car you drive

c. the distance it is from your home to the nearest grocery store

d. the number of classes you take per school year.

e. the type of calculator you use

f. weights of sumo wrestlers

g. number of correct answers on a quiz

h. IQ scores

Exercise

You go to the supermarket and purchase three cans of soup (19 ounces tomato bisque, 14.1 ounces lentil, and 19 ounces Italian wedding), two packages of nuts (walnuts and peanuts), four different kinds of vegetable (broccoli, cauliflower, spinach, and carrots), and two desserts (16 ounces pistachio ice cream and 32 ounces chocolate chip cookies).

Name data sets that are quantitative discrete, quantitative continuous, and qualitative.