

Department of AI & DS

CSE and CS&IT

COURSE NAME: PROBABILITY, STATISTICS AND QUEUING THEORY

COURSE CODE: 23MT2005

Topic

Measures of Central Tendency

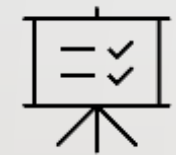
Session - 10

AIM OF THE SESSION



To familiarize students with the basic concept of measures of central tendency

INSTRUCTIONAL OBJECTIVES



This Session is designed to:

1. Demonstrate the examples of mean, median and mode
2. List out the importance of average in data analysis
3. Describe the important characteristics of measures of central tendency

LEARNING OUTCOMES



At the end of this session, you should be able to:

1. Define mean, median and mode
2. Describe various measures of central tendency
3. Summarize the role of average.

CONTENTS

❖ Measures of Central Tendency

Mean

Median

Mode

MEASURES OF CENTRAL TENDENCY

- A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location.
- There are five different types of averages:
 - (i) Mean
 - (ii) Median
 - (iii) Mode.
 - (iv) Geometric Mean
 - (v) Harmonic Mean
- Measure of central tendency provides a very convenient way of describing a set of scores with a single number that describes the **PERFORMANCE** of the group.

Arithmetic Mean

Arithmetic mean is also called arithmetic average. It is most commonly used measure of central tendency.

Arithmetic average of a series is the value obtained by dividing the total value of various item by its number.

Mean For Ungrouped data:

Let x be the variable which takes values $x_1, x_2, x_3, \dots, x_n$ over 'n' items, then arithmetic mean is given by \bar{X} .

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Example: Marks obtained by 6 students are given below: 20, 15, 23, 22, 25, 20. Calculate the mean.

Solution:

$$\bar{X} = \frac{20+15+23+22+25+20}{6} = 125/6 = 20.83$$

Properties

- It measures stability. Mean is the most stable among other measures of central tendency because every score contributes to the value of the mean.
- It may easily be affected by the extreme scores.
- The sum of each score's distance from the mean is zero.
- It can be applied to interval level of measurement
- It may not be an actual score in the distribution
- It is very easy to compute.

Median

- The score that **divides the distribution into two equal parts**, so that half the cases are above it and half below it.
- You compute the median value by following one of two rules:
 - Rule 1** If there are an *odd* number of values in the data set, the median is the middle-ranked value.
 - Rule 2** If there are an *even* number of values in the data set, then the median is the *average* of the two middle ranked values.
- The median is the **middle score**, or average of middle scores in a distribution.
 - Fifty percent (50%) lies below the median value and 50% lies above the median value.
 - It is also known as the middle score or the 50th percentile.

Median

- The median is the middle number in a sorted, ascending or descending, list of numbers and can be more descriptive of that data set than the average.
- The median is sometimes used as opposed to the mean when there are outliers in the sequence that might skew the average of the values.
- So Median is only measure of central tendency that can be used only for qualitative (ordinal) data.

Example-1: The three-year annualized returns for the seven small-cap growth funds with low risk are ranked from the smallest to the largest:

The Median is the 4th position value i.e. 22.4.

Example-2: Consider the time taken for reaching office from home in minutes arranged from lowest to highest

29 31 35 39 39 40 43 44 44 52

The number of observations are 10. Hence the median is the average of values at the 5th and 6th Position, i.e. average of 39 and 40 = **39.5**

Mode for ungrouped data

- Mode is that value of the observation which occurs maximum number of times. It is denoted by 'z'.
- The category or score with the largest frequency (or percentage) in the distribution.
- The mode can be calculated for variables with levels of measurement that are: nominal, ordinal, or interval-ratio.

Properties

- It can be used when the data are qualitative as well as quantitative.
- It may not be unique.
- It is not affected by extreme values.

When to Use the Mode

- When the data set is measured on a nominal scale

- **Example:** A systems manager in charge of a company's network keeps track of the number of server failures that occur in a day. Compute the mode for the following data, which represents the number of server failures in a day for the past two weeks:..

• 1 3 0 3 26 2 7 4 0 2 3 3 6 3
- Because 3 appears five times, more times than any other value, the mode is 3. Thus, the systems manager can say that the most common occurrence is having three server failures in a day.

Advantages and disadvantages of each measure of central tendency

Mean

Advantages:

- It is simple and easy to compute.
- It is rigidly defined.
- It is amenable for algebraic manipulations.
- It is based on all observations in the series.
- It helps for direct comparison.

Disadvantages:

- It is unduly affected by extreme items.
- It is sometimes un-realistic.
- It may leads to confusion.
- Suitable only for quantitative data (for variables).
- It cannot be located by graphical method or by observations.

Advantages and disadvantages of each measure of central tendency

Median

Advantages:

- It is simple, easy to compute and understandable.
- It's value is not affected by extreme variables.
- It can be determined by inspection for arrayed data.
- It can be found graphically also.

Disadvantages:

- It may not be representative value as it ignores extreme values.
- It can't be determined precisely when its size falls between the two values.
- It is not useful in cases where large weights are to be given to extreme values.
- It is capable not for further algebraic treatment.

Advantages and disadvantages of each measure of central tendency

Mode

Advantages:

Its result will not be affected by extreme values and openend classes

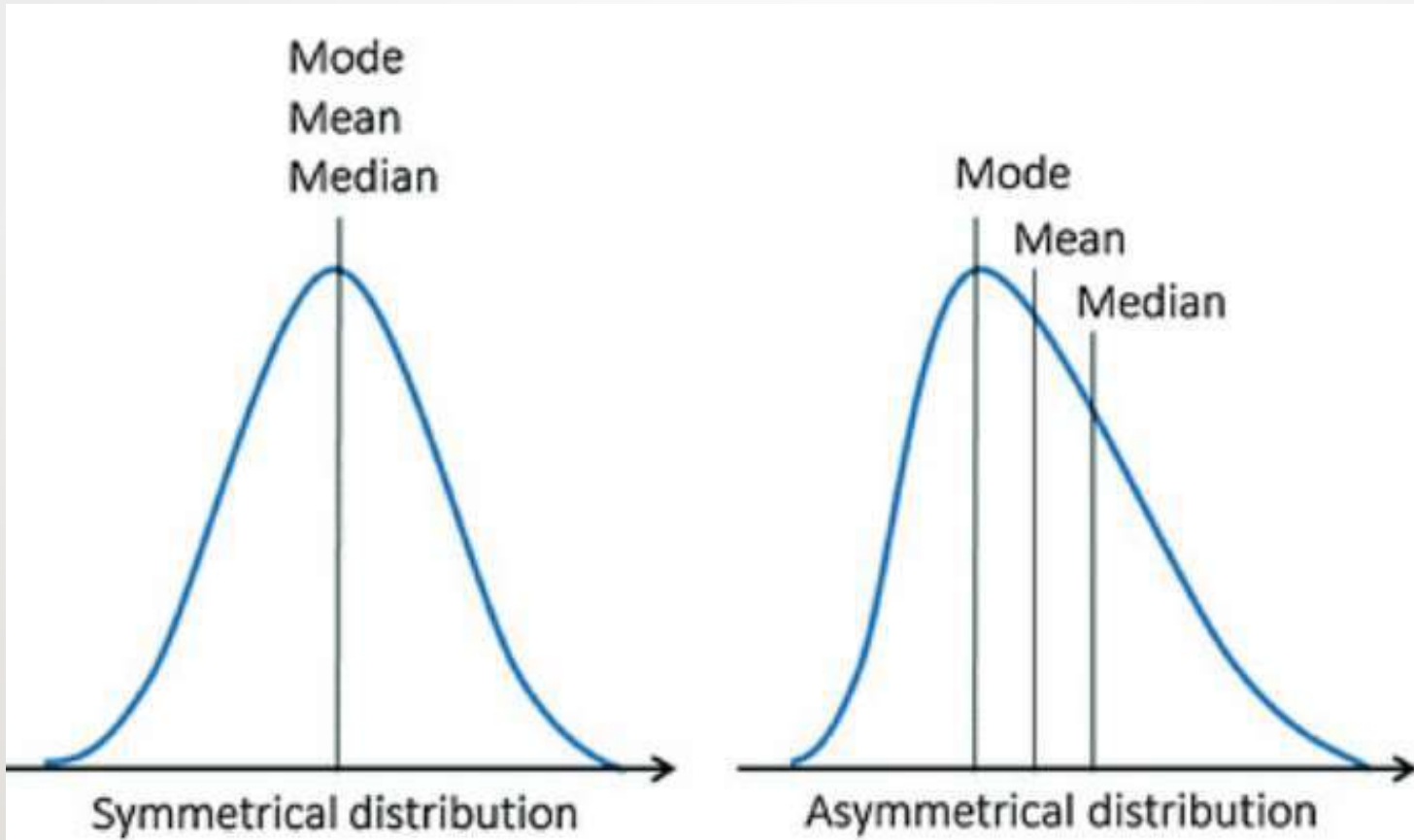
If data are not grouped, it can be determined easily

Disadvantages:

It is to be supplemented by other statistics

It is difficult to obtain an accurate estimate of the mode if the values are classified into a frequency distribution.

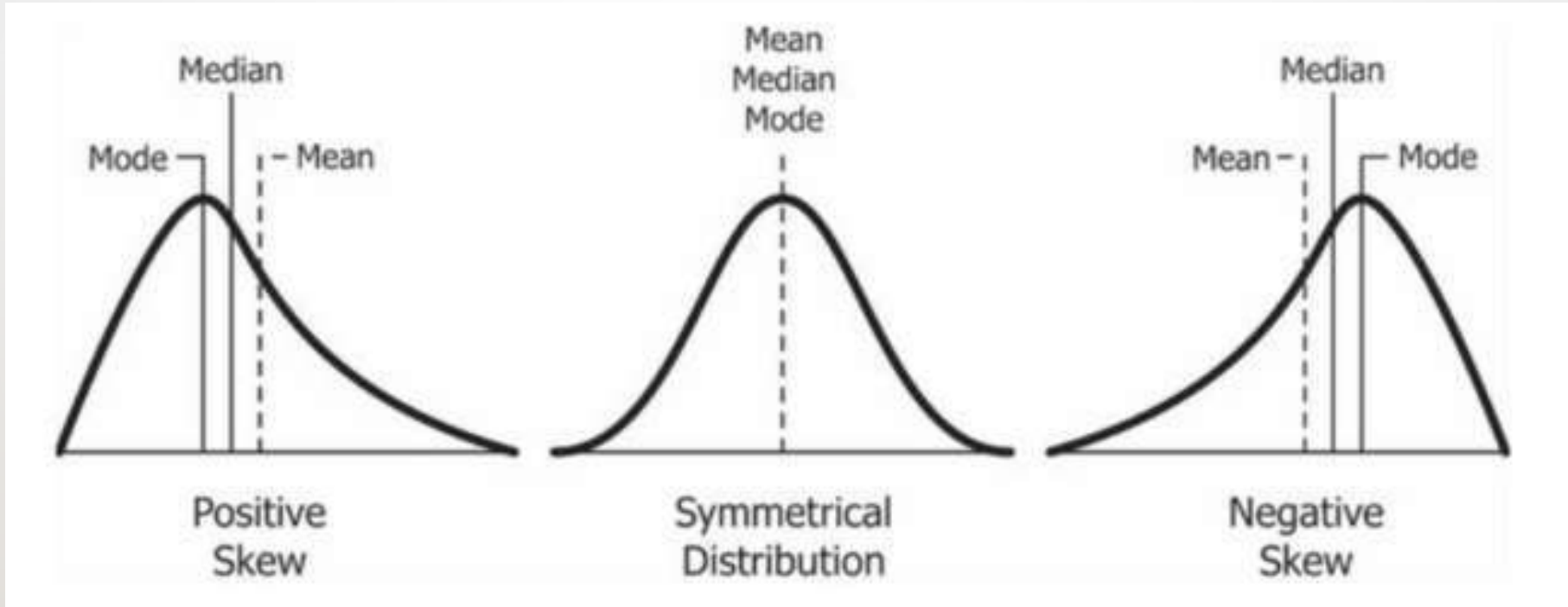
Data : Symmetrical and Asymmetrical



SHAPE OF THE DISTRIBUTION OF DATA

- Symmetrical : Mean is equal to median
- Skewed
 - Negatively (Left): $\text{mean} < \text{median}$
 - Distributions with a left skew have long left tails;
 - Positively (Right) : $\text{mean} > \text{median}$
 - Distributions with a right skew have long right tails.
- Bimodal : has two distinct modes
- Multi-modal : has more than 2 distinct modes

TYPES OF DISTRIBUTION



EXAMPLES

Example: The runs scored by Sachin in 5 test matches are 140, 153, 148, 150 and 154 respectively. Find the mean.

Runs scored by Sachin in 5 test matches: 140, 153, 148, 150 and 154

Means of the runs = total runs number of matches/5

$$\text{Mean} = 140+153+148+150+154/5 = 745/5 = 149.$$

Example: Suppose a restaurant collects the cans for two weeks and sends it to a recycling plant. The number of cans collected each day are: 84, 97, 77, 31, 84, 58, 63, 72, 47, 84, 64, 94, 43 and 68.

Now we need to find the median of these numbers. The first step to find the median is to arrange the numbers either in ascending order or descending order.

So arranging the data in the ascending order, 31, 43, 47, 58, 63, 64, 68, 72, 77, 84, 84, 84, 94 and 97. Here the total numbers are even. So using the formula

$$\text{Median} = ((n/2)^{\text{th}} + (n/2+1)^{\text{th}} \text{ terms})/2$$

Since $n = 14$, 7th term = 68 and 8th term = 72. Now we have our two middle terms as 68 and 72, hence Median = $(68+72)/2 = 70$.

Why Average is not Engouh?

Sl. No.	X_1	X_2
1	2	1
2	8	15
3	5	5
4	3	5
5	7	6
6	8	3
7	5	5
8	2	2
9	5	3
Total	45	45

Statistical measures	Group 1 & 2
Mean	5
Median	5
Mode	5

We need to go to measure of variability

Described measures of central tendency

Mean, Median and Mode

TERMINAL QUESTIONS

1. Describe various measures of central tendency and dispersion
2. List out the characteristics of good average
3. Summarize the role of descriptive statistics in data analysis
4. For the following list, $n = 19$. Find the median.

24, 25, 28, 31, 33, 33, 36, 42, 42, 48, 51, 57, 57, 68, 75, 79, 79, 79, 85

5. Five people play golf and at one hole their scores are 3, 4, 4, 5, 7 For these scores, find (a) the mean (b) the median (c) the mode.

Reference Books:

1. Chapter 1 of TP1: William Feller, An Introduction to Probability Theory and Its Applications: Volume 1, Third Edition, 1968 by John Wiley & Sons, Inc.
2. Richard A Johnson, Miller & Freund's Probability and statistics for Engineers, PHI, New Delhi, 11th Edition (2011).

Sites and Web links:

3. Section 3.1.1 of TS1: Alex Tsun, Probability & Statistics with Applications to Computing (Available at: http://www.alextsun.com/files/Prob_Stat_for_CS_Book.pdf)

Video:

<https://www.youtube.com/watch?v=5sOBWV0qH8&list=PLB45KifGiuHesi4PALNZSYZFhViVGQJK&index=19>

THANK YOU



Team – PSQT EVEN SEMESTER 2024-25