# Department of AI & DS
# CSE and CS&IT

## COURSE NAME: PROBABILITY, STATISTICS AND QUEUING THEORY

## COURSE CODE: 23MT2005

### Topic

Measures of Dispersion

**Session - 11**

## AIM OF THE SESSION

To familiarize students with the basic concept of measures of dispersion

## INSTRUCTIONAL OBJECTIVES

This Session is designed to:

1. Describe various measures of dispersion
2. List out the importance of average and variation in data analysis
3. Describe the important characteristics of measures of dispersion

## LEARNING OUTCOMES

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

1. Describe various measures of variation
2. Summarize the role of average and variance.

# CONTENTS

- Measures of dispersion

  - Range
  - Standard deviation
  - Variance

Five Point Summary

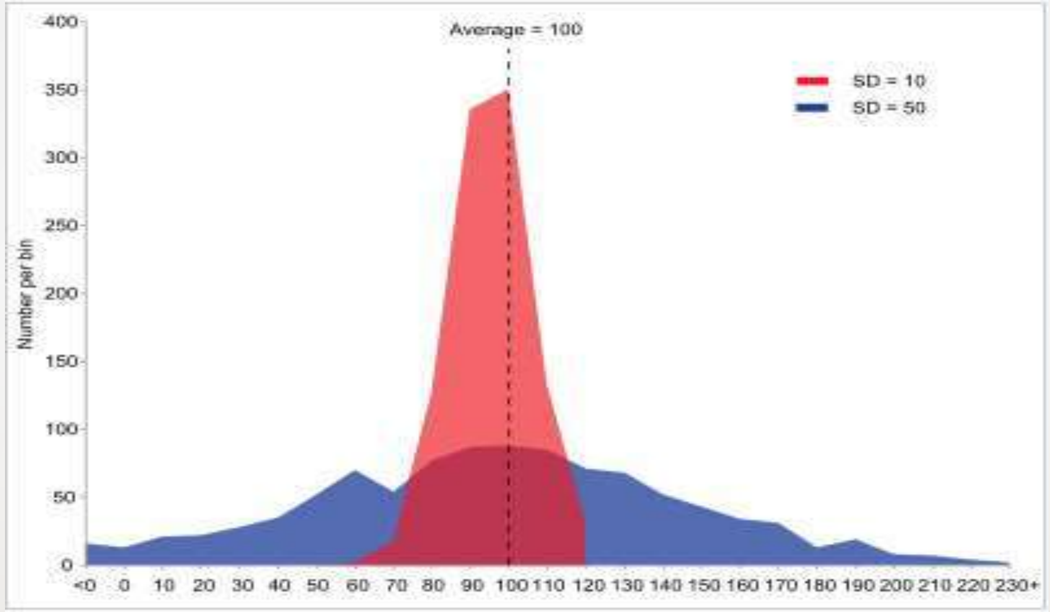
# MEASURES OF Dispersion

- Dispersion is the measure of the variation of the items".----A. L. Bowely
- "The term dispersion is used to indicate the facts that within a given group, the items differ from one another in size or in other words, there is lack of uniformity in their sizes".---W. I. King

Example:

| Sections | | | | | | | | | | Total | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 135 | 15 |
| B | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 135 | 15 |
| C | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 135 | 15 |

# MEASURES OF Dispersion




Example of samples from two populations with the same mean but different dispersion. The blue population is much more dispersed than the red population.

Purpose of dispersion:

1. To compare two or more series with regard to their variability.

2. To serve as a basis for the control of the variability.

3. To determine the reliability of average.

4. To facilitate the computation of other statistical measures.

**.Properties of Good measure of dispersion**

1. It should be simple to understand.

2. It is should be easy to compute.

3. It is should rigidly defined.

4. It should be based on all observations.

5. It should measure the sampling fluctuations.

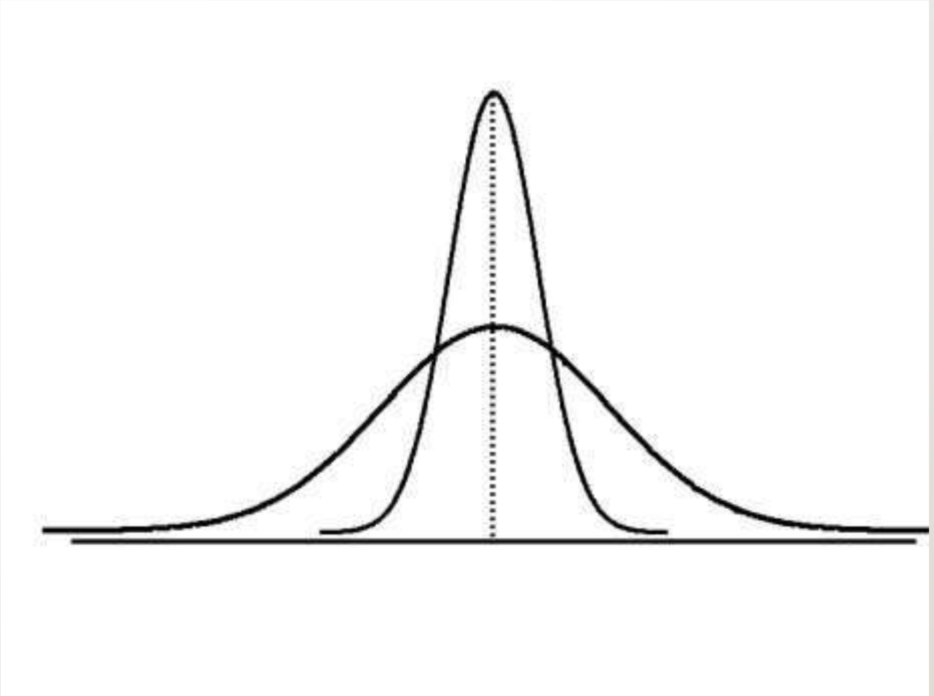
6. It should be suitable for further algebraic treatment.

7. It should be not be affected by extreme observations.

# MEASURES OF Dispersion

- Variation
  - Range
  - Variance
  - Standard Deviation
  - Coefficient of Variation

▪Measures of variation give information on the **spread** or **variability** or **dispersion** of the data values.



# MEASURES OF Dispersion

## Range

- Simplest measure of variation

- Difference between the largest and the smallest values:

$$\text{Range} = X_{\text{largest}} - X_{\text{smallest}}$$

## Variance

Average (approximately) of squared deviations of values from the mean

Sample variance:

$\overline{X}$ = arithmetic mean

n = sample size

$X_i$ = $i^{th}$ value of the variable X

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - x)^2}{n} \quad or$$

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - x)^2}{n-1}$$

# MEASURES OF Dispersion

**Example**

**Sample Data ($X_i$) :** 10 12 14 15 17 18 18 24

$$S = \sqrt{\frac{(10-\overline{X})^2 + (12-\overline{X})^2 + (14-\overline{X})^2 + \cdots + (24-\overline{X})^2}{n-1}}$$

$$= \sqrt{\frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + \cdots + (24-16)^2}{8-1}}$$

$$= \sqrt{\frac{130}{7}} = 4.3095$$

**Case Study 1:**

Find the Variance and Standard Deviation of the Following Numbers: 1, 3, 5, 5, 6, 7, 9, 10.

The mean = 46/ 8 = 5.75

**Step 1:** (1 – 5.75), (3 – 5.75), (5 – 5.75), (5 – 5.75), (6 – 5.75), (7 – 5.75), (9 – 5.75), (10 – 5.75)

= -4.75, -2.75, -0.75, -0.75, 0.25, 1.25, 3.25, 4.25

**Step 2:** Squaring the above values we get, 22.563, 7.563, 0.563, 0.563, 0.063, 1.563, 10.563, 18.063

**Step 3:** 22.563 + 7.563 + 0.563 + 0.563 + 0.063 + 1.563 + 10.563 + 18.063
= 61.504

**Step 4:** n = 8, therefore variance ($\sigma^2$) = 61.504/ 8 = 7.69 (3sf)

Now, Standard deviation ($\sigma$) = 2.77 (3sf)

# ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

**Case Study 2:**

The length of 20 similar crystals is measured (in mm) in a chemistry experiment. Calculate the standard deviation and the coefficient of variation for the observations taken.

| Crystal no. | Length (mm) | Crystal no. | Length (mm) |
|---|---|---|---|
| 1 | 9 | 11 | 7 |
| 2 | 2 | 12 | 4 |
| 3 | 5 | 13 | 12 |
| 4 | 4 | 14 | 5 |
| 5 | 12 | 15 | 4 |
| 6 | 7 | 16 | 10 |
| 7 | 8 | 17 | 9 |
| 8 | 11 | 18 | 6 |
| 9 | 9 | 19 | 9 |
| 10 | 3 | 20 | 4 |

# ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

**Solution** : We can construct the table as given below

| Crystal no. | Length (mm) | Crystal no. | Length (mm) |
|---|---|---|---|
| 1 | 9 | 2 | 4 |
| 2 | 2 | -5 | 25 |
| 3 | 5 | -2 | 4 |
| 4 | 4 | -3 | 9 |
| 5 | 12 | 5 | 25 |
| 6 | 7 | 0 | 0 |
| 7 | 8 | 1 | 1 |
| 8 | 11 | 4 | 16 |
| 9 | 9 | 2 | 4 |
| 10 | 3 | -4 | 16 |
| 11 | 7 | 0 | 0 |
| 12 | 4 | -3 | 9 |
| 13 | 12 | 5 | 25 |
| 14 | 5 | -2 | 4 |
| 15 | 4 | -3 | 9 |
| 16 | 10 | 3 | 9 |
| 17 | 9 | 2 | 4 |
| 18 | 6 | -1 | 1 |
| 19 | 9 | 2 | 4 |
| 20 | 4 | -3 | 9 |
| N = 20 | $\sum x_i = 140$ | | $\sum (x_i - A)^2 = 178$ |
| $A = \sum x_i / N$ $= 140/20 = 7$ mm | | | |

Now, we may give the Standard Deviation as –

$$S.D. = \sqrt{\frac{\sum (X_i - A)^2}{N}} = \sqrt{\frac{178}{20}}$$

$$= 2.9832 (mm)$$

We can calculate the coefficient of variation as –

$$C.V. = (S.D./Mean) \times 100$$

$$= (2.9832/7) \times 100$$

$$= 42.62 \text{ percent}$$

# 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

$$\text{Means of the runs} = \text{total runs number of matches}/5$$
$$\text{Mean} = 140+153+148+150+1545 = 7455/5 = 149.$$

**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 of 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

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

Since n = 14, $7^{th}$ term = 68 and $8^{th}$term = 72. Now we have our two middle terms as 68 and 72, hence Median = 68+722 = 70.

# Five Point  Summary

- The five number summary of data includes 5 items:

❖ **Minimum**.

❖ **Q1** (the first quartile, or the 25% mark).

❖ **Median**.

❖ **Q3** (the third quartile, or the 75% mark).

❖ **Maximum**.

# INTERQUARTILE RANGE (IQR)

- It is measure of Variation
- Also Known as Mid-spread : Spread in the Middle 50%
- Difference Between Third & First Quartiles:
- Not Affected by Extreme Values

Interquartile Range = IQR = $Q_3 - Q_1$

**Data in Ordered Array: 11 12 13 16 16 17 17 18 21**

**Position of $Q_1$** $= \frac{1.(9+1)}{4} = 2.50$

**Position of** $\boldsymbol{Q_1} = \frac{1.(9+1)}{4} = 2.50$

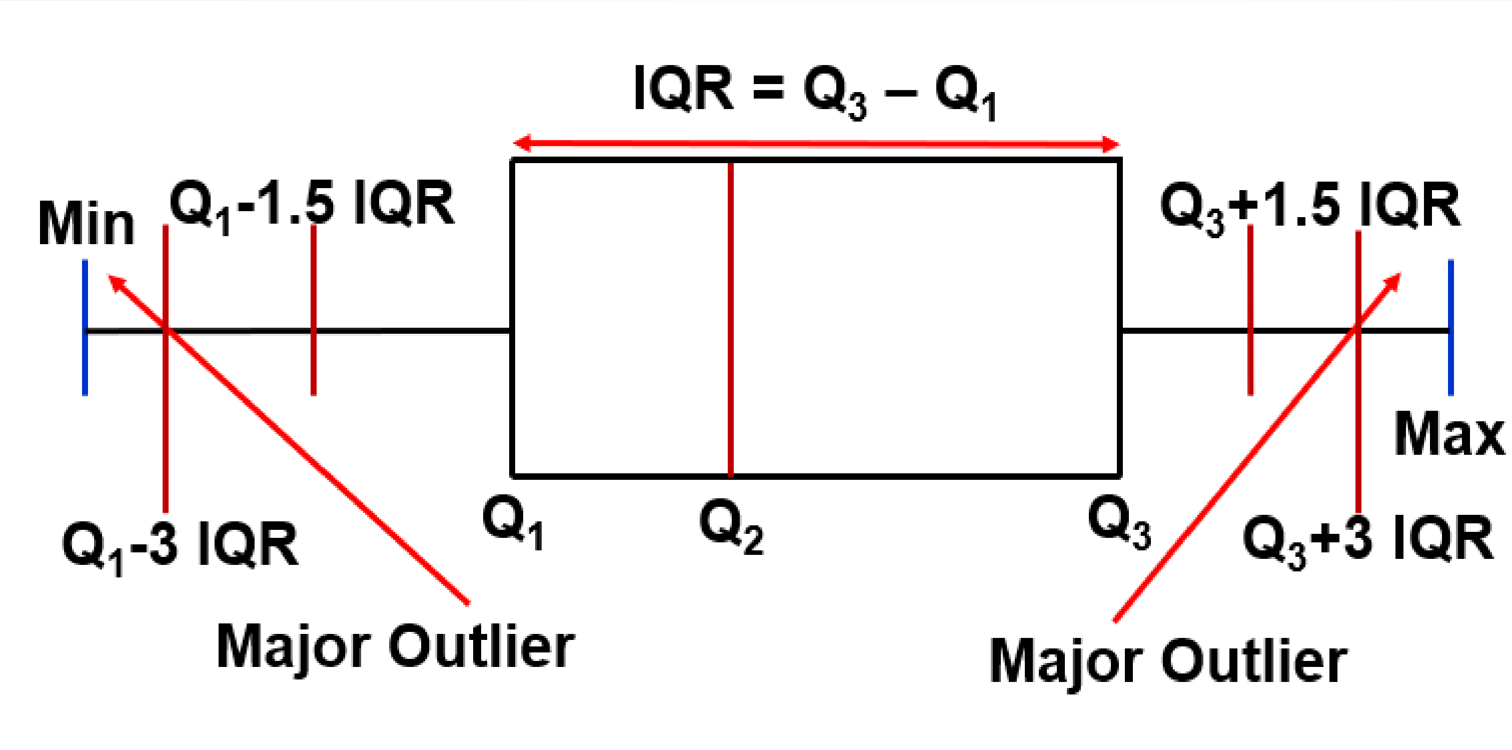
$\therefore \boldsymbol{Q_1} = 12.5$

**Position of** $\boldsymbol{Q_3} = \frac{3.(9+1)}{4} = 7.50$

$\therefore \boldsymbol{Q_1} = 12.5 = 17.5$

**Interquartile Range = IQR= $Q_3 - Q_1$ = 17.5 - 12.5 = 5**

# Box and Whisker plot




- The lower limit and upper limit of a data set are given by:

  Lower limit = $Q_1$ - 1.5 x IQR, Upper limit = $Q_3$ + 1.5 x IQR

- Data points that lie below the lower limit or above the upper limit are potential outliers.

# SUMMARY

Described measures of central tendency

Mean, Median and Mode

Described measures of variation

Range, interquartile range, variance and standard deviation, coefficient of variation,

# SELF-ASSESSMENT QUESTIONS

The measurements of spread or scatter of the individual values around the central point is called

a) measures of dispersion
b) Measures of central tendency
c) Measures of skewness
d) measures of kurtosis

The mean of an examination is 69, the median is 68, the mode is 67, and the standard deviation is 3. The measures of variation for this examination is

(a) 67

(b) 68

(c) 69

(d) 3

# SELF-ASSESSMENT QUESTIONS

3. If all the scores in a data set are the same, the Standard Deviation is equal to 1.00

i) True
ii) False

4. The standard deviation measures

(1) Sum of squared deviation scores

(2) Standard distance of a score from the mean

(3) Average deviation of a score from the mean

(4) Average squared distance of a score from the mean

# TERMINAL QUESTIONS

**1**. The following data represent the battery life (in shots) for three pixel digital cameras:

300 180 85   170 380 460 260 35   380 120 110 240

List the Five-point summary.

2. For the data set below:

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 82 | 45 | 64 | 80 | 82 | 74 | 79 | 80 | 80 | 78 | 80 | 80 | 48 | 73 | 80 | 79 | 81 | 70 | 78 | 73 |

a. Obtain and interpret the quartiles.
**b.** Determine and interpret the interquartile range.
**c.** Find and interpret the five-number(point) summary.
**d.** Identify potential outliers, if any.
**e.** Construct and interpret a boxplot.

# REFERENCES FOR FURTHER LEARNING OF THE SESSION


**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=PLeB45KifGiuHesi4PALNZSYZFhViVGQJK&index=19

# THANK YOU

**Team –PSQT EVEN SEMESTER 2024-25**