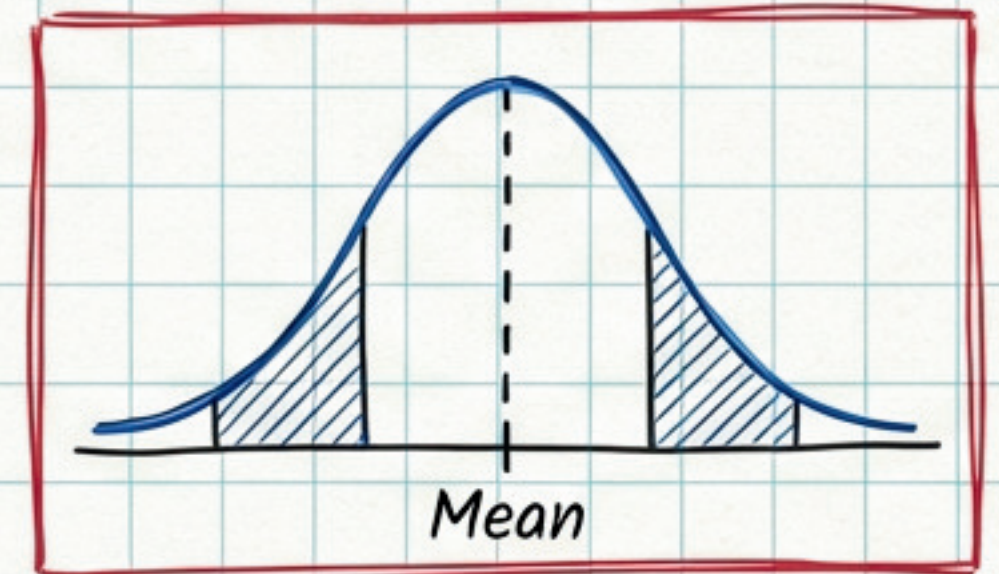


(Notes By ~ Kamal Kishor)

# CSA 6004T – Computer Oriented Statistical Techniques (BCA)

*Detailed Study Notes & Formula Sheet*



*Subject: Statistics | Format: Handwritten Notes*

# UNIT-1: INTRODUCTION TO STATISTICS

## 1. Meaning of Statistics

Statistics is the science of collecting, organizing, presenting, analyzing and interpreting numerical data.

### Uses of Statistics

- ✓ Decision making
- ✓ Research and analysis
- ✓ Business forecasting
- ✓ Computer data analysis
- ✓ Government planning

## 2. Raw Data

Raw data is unorganized original data collected from surveys, experiments, or observations.

### Example

Marks obtained by students:  
45, 60, 72, 50, 90, 38

### 3. Frequency Distribution

#### a) Ungrouped Frequency Distribution

Data arranged with frequencies without class intervals.

Value	Frequency
10	2
20	4

#### b) Grouped Frequency Distribution

Data grouped into class intervals.

Class Interval	Frequency
0-10	5
10-20	8

*Grouped*

## 4. Diagrammatic Presentation

### a) Bar Diagram

- Simple Bar
- Multiple Bar
- Component Bar

(Used for comparison)

### b) Pie Diagram

- Circular diagram
- Shows percentage distribution

$$\text{Angle} = \left( \frac{\text{Value}}{\text{Total}} \right) \times 360^\circ$$

## 5. Graphical Presentation

### a) Histogram – Rectangles

✓ Used for continuous data

### b) Frequency Polygon – Joining midpoints of histogram bars

### c) Frequency Curve – Smooth curve over polygon

### d) Cumulative Frequency Curve (Ogive) – Less than Ogive

✓ More than Ogive

Used to find  
Median & Quartiles

# UNIT-2: MEASURES OF CENTRAL TENDENCY

## 1. Arithmetic Mean (AM)

↳ Ungrouped data:  $\bar{X} = \frac{\sum X}{N}$       Grouped data:  $\bar{X} = \frac{\sum fX}{\sum f}$

## 2. Median ✓

Middle value of ordered data. ✓

For grouped data: Median =  $l + \left[ \frac{\frac{N}{2} - cf}{f} \right] \times h$

## 3. Mode ✓

Most frequent value.

For grouped data: Mode =  $l + \left[ \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right] \times h$

# Measures of Central Tendency (Continued)

## 4. Geometric Mean (GM)

$$GM = (\prod X)^{1/N}$$

Used in growth rates & computer algorithms ✓

## 5. Harmonic Mean (HM)

$$HM = \frac{N}{\sum(1/X)}$$

Used in speed and rate problems.


# B. Measures of Dispersion

## 1. Range

$$\text{Range} = \text{Largest} - \text{Smallest} \quad \checkmark$$

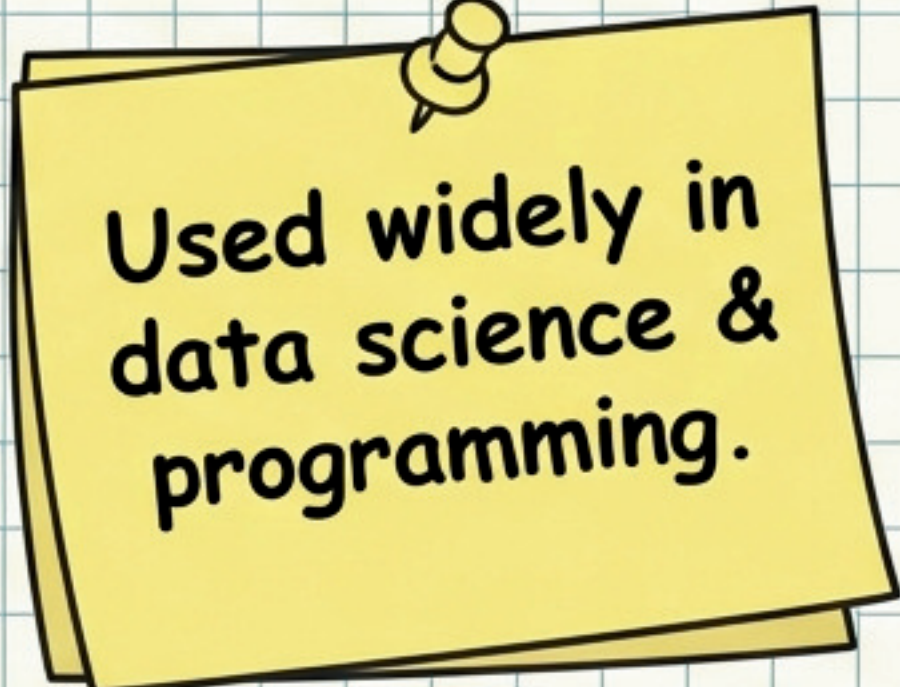
# Dispersion: Detailed Measures

## 2. Mean Deviation


$$MD = \frac{\sum |X - \bar{X}|}{N}$$

## 3. Standard Deviation (SD)


$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$



Used widely in  
data science &  
programming.

# Shape of Distribution

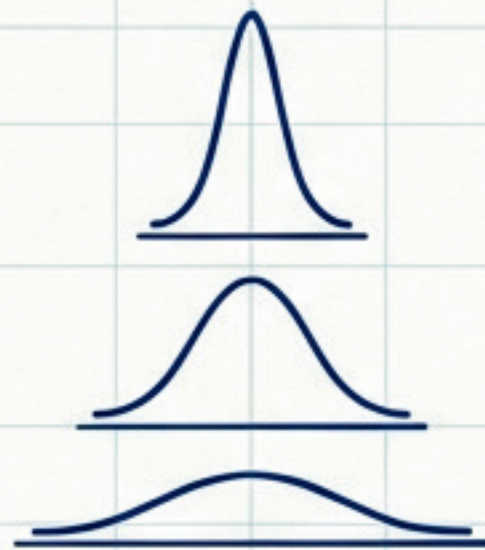
## 4. Skewness

- Positive skew → tail on right
  - Negative skew → tail on left
- ↪ Karl Pearson's coefficient:

$$Sk = \frac{\bar{X} - \text{Mode}}{\sigma}$$

## 5. Kurtosis

- Leptokurtic (peaked)
- Mesokurtic (normal)
- Platykurtic (flat)



# UNIT-3: CORRELATION & REGRESSION

## 1. Correlation

Measures degree of relationship between variables.

- Positive
- Negative
- Zero

## 2. Scatter Diagram

Graphical method to identify correlation.

## 3. Karl Pearson's Coefficient of Correlation

$$r = \frac{\sum(XY)}{\sqrt{\sum X^2 \sum Y^2}}$$

Value of r lies between -1 and +1

## 4. Concurrent Deviation Method

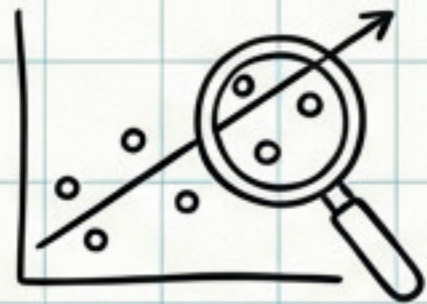
Used for directional relationship only.

## 5. Regression Analysis

Used for prediction.

Regression Lines:

- Y on X
- X on Y



$$Y = a + bX$$

## 6. Least Squares Method

Finds best fitting line by minimizing error.

## 7. Standard Error of Estimate

Measures accuracy of prediction.

## 8. Explained & Unexplained Variation

Total Variation = Explained + Unexplained

# UNIT-4: PROBABILITY & PROBABILITY DISTRIBUTION

## 1. Probability

Chance of occurrence of an event.

$$P(E) = \frac{\text{Favourable}}{\text{Total}}$$

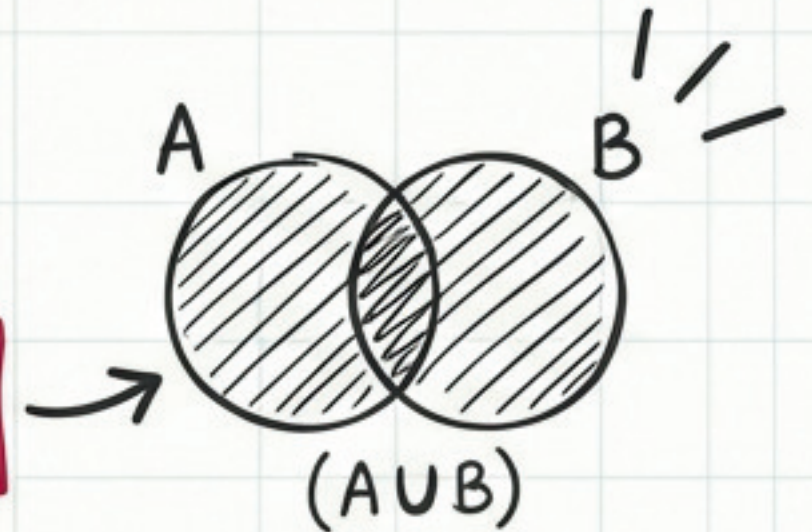


## 2. Types of Events

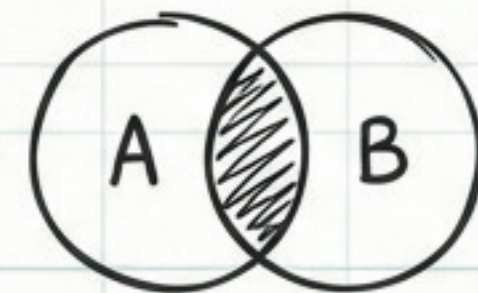
- Independent
- Dependent
- Mutually Exclusive

## 3. Laws of Probability

Addition Law:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Multiplication Law:  $P(A \cap B) = P(A)P(B)$



# Advanced Probability Concepts

## 4. Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

## 5. Bayes' Theorem

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum P(A)P(B|A)}$$

Used in AI, Machine Learning & decision systems.

## 6. Probability Distributions

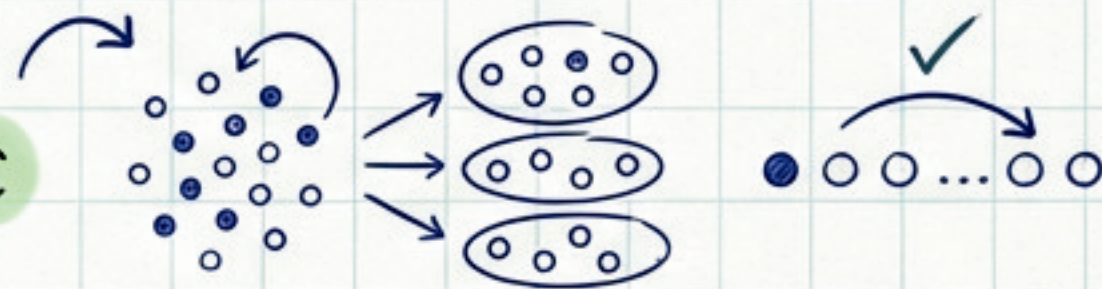
- Discrete
- Continuous //

# ★ UNIT-5: SAMPLING THEORY & TESTING

## 1. Sampling

Selecting subset from population. ★

- Random
- Stratified
- Systematic

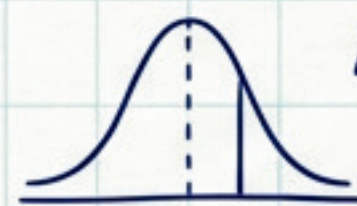


## 2. Sampling With & Without Replacement

Refers to returning or not returning elements after selection.

## 3. Sampling Distributions

- Mean
- Proportion
- Difference & Sums



## 4. Standard Error

$$SE = \frac{\sigma}{\sqrt{n}} \text{ Caveat}$$

Measure of accuracy

## 5. Small Samples?

Used when  $n < 30$

# Estimation & Hypothesis Testing

## 6. Student's t-Distribution

- Small samples
- Unknown population variance

## 7. Confidence Intervals

$$\bar{X} \pm Z \times SE$$

## 8. Tests of Hypothesis

- Null Hypothesis ( $H_0$ )
- Alternative Hypothesis ( $H_1$ )

1. State hypothesis

2. Choose level of significance

3. Test statistic

4. Decision

# Advanced Distributions & Course Conclusion

## 9. Chi-Square Distribution

- Usage:
- Goodness of fit
  - Independence of attributes

$$\chi^2 = \sum \left[ \frac{(O - E)^2}{E} \right]$$

## 10. Degrees of Freedom $df = n - 1$

## 11. F-Distribution

Used in ANOVA

### WHY THIS SUBJECT IS 'COMPUTER ORIENTED'

- Data analysis ✓
- Algorithms & logic ✓
- Machine learning basics ✓
- Statistical programming ✓
- Database analysis ✓