Statistics Notes (With Definitions and Equations)

1. Statistics

Statistics is the science of **collecting, organizing, summarizing, visualizing, and analyzing data**.

---

2. Types of Statistics

2.1 Descriptive Statistics

Used to **organize, summarize, and visualize** information.

2.2 Inferential Statistics

Used to **make conclusions** about a population using sample data.

---

3. Descriptive Statistics

3.1 Measures of Central Tendency

These represent the **center** or typical value of a dataset.

Mean (Average)

$$\text{Mean}(\mu) = \frac{\sum X}{N}$$

Median

Middle value of the sorted dataset.

Mode

Most frequently occurring value of a dataset.

3.2 Measures of Dispersion

These show how much the data points **spread out** from the mean.

Variance

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

Standard Deviation

$$\sigma = \sqrt{\sigma^2}$$

Shows how spread out the data is.

---

4. Sampling Techniques

1. **Simple Random Sampling** – Every sample has an equal chance of selection.
2. **Stratified Sampling** – Population divided into subgroups (e.g., gender), and samples taken from each.
3. **Systematic Sampling** – Select every *kth* item.
4. **Convenience Sampling** – Sample selected based on availability.
5. **Purposive Sampling** – Select sample intentionally for a purpose.
6. **Cluster Sampling** – Select entire subgroups that share characteristics.

---

5. Covariance & Correlation

5.1 Covariance – Measures relationship between two variables.

$$\text{Cov}(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N}$$

5.2 Correlation – Standardized measure (ranges from -1 to +1).

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

---

6. Probability Distribution Function (PDF)

Describes probability of all possible outcomes.

Types:

6.1 Discrete Distribution

Finite distinct possible outcomes.

6.2 Continuous Distribution

Infinite consecutive outcomes.

---

7. Probability Functions

7.1 PDF (Probability Density Function)

For continuous variables.

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

7.2 PMF (Probability Mass Function)

For discrete variables.

$$P(X = x) \geq 0, \quad \sum P(X = x) = 1$$

7.3 CDF (Cumulative Distribution Function)

$$F(x) = P(X \leq x)$$

---

8. Types of Probability Distributions

8.1 Bernoulli Distribution

One trial, two outcomes (0 or 1).

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

8.2 Binomial Distribution

Multiple independent trials.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

8.3 Poisson Distribution

Events in a fixed time interval.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

8.4 Normal (Gaussian) Distribution

Bell-shaped & symmetric.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

8.5 Uniform Distribution

$$f(x) = \frac{1}{b-a}$$

8.6 Standard Normal Distribution

$$Z = \frac{X - \mu}{\sigma}$$

Properties: 1. Symmetric 2. Bell-shaped curve 3. Mean = Median = 0, Standard deviation = 1

---

9. Scaling Methods

9.1 Normalization (Min-Max Scaling)

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Scale: **0 to 1** or **-1 to 1**

9.2 Standardization (Z-score Scaling)

$$Z = \frac{X - \mu}{\sigma}$$

Scale: **Mean = 0, SD = 1**

---

10. Inferential Statistics

10.1 t-test

Used for comparing means of two groups (**n < 30**).

10.2 z-test

Used when sample size is large (**n > 30**).

10.3 Chi-Square Test

Used for **categorical data**.

10.4 ANOVA

Compares means among **3 or more groups**.

10.5 Hypothesis Testing

Decide whether to accept or reject a hypothesis.

Steps: 1. Define H0 & H1 2. Select significance level 3. Calculate test statistic 4. Compare p-value with α

10.6 Estimation

Used to estimate population parameters.

Point Estimate – Single numeric value

Interval Estimate – Range of values (confidence interval)

If you want, I can convert this into **PDF, Word, or PPT** with styling. Let me know!