Complete Statistics for Data Science - Notes

1. Types of Data / Variables

1.1 Nominal Variable

- Represents categories with no order.

- Examples: Colors: Red, Blue, Green; Gender: Male, Female

1.2 Ordinal Variable

- Represents ordered categories but distance not measurable.

- Examples: Poor, Average, Excellent

1.3 Interval Variable

- Numerical, meaningful differences, no true zero.

- Examples: Temperature, IQ Scores

1.4 Ratio Variable

- Numerical, true zero, all arithmetic allowed.

- Examples: Age, Height, Weight, Salary

2. Special Categorical Types

2.1 Binary Variable: Only two categories (Yes/No, True/False)

2.2 Multinomial Variable: More than two categories without order (Red, Blue, Green)

3. Descriptive Statistics

3.1 Central Tendency: Mean, Median, Mode

3.2 Dispersion: Range, Variance, Standard Deviation (sqrt(Variance))

3.3 Percentiles/Quartiles: Q1, Q2 (Median), Q3

4. Probability Concepts

4.1 Random Variable: depends on chance

4.2 Probability Distribution: Discrete vs Continuous

4.3 Common Distributions: Normal, Binomial, Poisson, Uniform

5. Inferential Statistics

5.1 Hypothesis Testing: H0, H1, p-value

5.2 Confidence Intervals: 95% CI example

5.3 Correlation & Regression: r, linear prediction

6. Sampling Techniques

Simple Random, Stratified, Systematic, Cluster Sampling

7. Data Visualization Basics

Bar Chart, Histogram, Box Plot, Scatter Plot, Pie Chart

8. Key Notes

Nominal, Ordinal, Interval, Ratio, Binary, Multinomial

9. Data Cleaning & Preparation

- Handling Missing Values: mean, median, drop
- Detecting Outliers: IQR, Z-score
- Normalization/Standardization: Min-Max, Z-score

10. Advanced Descriptive Measures

- Skewness: measure of asymmetry
- Kurtosis: measure of peakness
- Coefficient of Variation (CV)

11. Probability Theorems

- Law of Total Probability
- Bayes Theorem
- Independence & Conditional Probability

12. Advanced Distributions

- t-distribution: small sample stats
- Chi-square: goodness of fit
- F-distribution: ANOVA

13. More Inferential Tests

- Z-test, t-test (1-sample, 2-sample, paired)
- Chi-square test (independence & goodness of fit)
- ANOVA, Non-parametric tests (Mann-Whitney, Kruskal-Wallis)

14. Effect Size & Statistical Power

- Cohen's d, Eta squared
- Importance of effect size beyond p-value

15. Time Series Basics

- Components: Trend, Seasonality, Residuals
- Moving Averages, Exponential Smoothing

16. Multivariate Analysis Basics

- Multiple Regression
- Introduction to Principal Component Analysis (PCA)