Descriptive Statistics Project – Theory & Definitions

Part A: Theory & Definitions

1. Types of Data: Numerical Data: Quantitative measurable values such as Age, Household Income, and Family Size. Categorical Data: Qualitative categories such as Education Level and Urban/Rural classification. 2. Types of Statistics: Descriptive Statistics: Used to summarize and describe features of a dataset using numerical measures and graphs. Inferential Statistics: Used to draw conclusions or make predictions about a population based on sample data. 3. What is Descriptive Statistics? Descriptive statistics involves organizing, summarizing, and presenting data in meaningful ways using measures of central tendency, dispersion, and graphical tools.

Measures of Central Tendency

Mean: Arithmetic average. Formula: Mean = (Sum of all values) / (Number of observations). Median: Middle value when data is arranged in ascending order. Mode: Most frequently occurring value.

Measures of Dispersion

Range: Difference between maximum and minimum values. Variance: Average of squared deviations from the mean. Standard Deviation: Square root of variance; measures spread of data.

Important Statistical Concepts

Gaussian (Normal) Distribution: A symmetric bell-shaped distribution defined by mean (μ) and standard deviation (σ). Formula: f(x) = (1/(σ√2π)) * e^(-(x-μ)^2 / 2σ^2) Log-Normal Distribution: A distribution where the logarithm of the variable follows a normal distribution. 3-Sigma Rule (Empirical Rule): 68% of data lies within ±1σ 95% within ±2σ 99.7% within ±3σ Percentiles: A percentile indicates the value below which a given percentage of observations fall. Quartiles: Q1 (25th percentile), Q2 (Median), Q3 (75th percentile). Five Number Summary: Minimum, Q1, Median, Q3, Maximum. Skewness: Measures asymmetry of a distribution. Kurtosis: Measures heaviness of tails relative to normal distribution.

Age vs Household Income – Scatter Plot Interpretation

A scatter plot with regression line was used to analyze the relationship between Age and Household Income. The plot shows a mild positive relationship, indicating that income tends to increase slightly with age. The regression line represents the overall trend in the data. The spread of data points suggests moderate variability among individuals within similar age groups. No extreme outliers were observed in the distribution.