

Assignment 1

Student Name: Youssef Tarek

Part 1: Key Concepts

Q1: Definitions and Examples

- **a) Structured Data:** Data that fits into a fixed format like tables (rows and columns).
 - *Example:* An Excel sheet of hospital patient names and their ages.
- **b) Unstructured Data:** Data that does not have a predefined internal structure.
 - *Example:* Email messages, PDF files, or patient X-ray images.
- **c) Discrete Variable:** Variables that can only take specific, countable values (whole numbers).
 - *Example:* The number of rooms in a hospital.
- **d) Continuous Variable:** Variables that can take any value within a range (including decimals).
 - *Example:* A patient's body temperature (e.g., 37.2°C).
- **e) Categorical Variable:** Data that describes a "category" or group.
 - *Example:* Blood types (A, B, AB, O).

Q2: Statistical Differences

- **a) Population vs. Sample:** Population is the entire group you want to study (e.g., all patients in Egypt). A Sample is a smaller part of that group used for the actual study (e.g., 500 patients from one hospital).
- **b) Parameter vs. Statistic:** A Parameter is a characteristic of the whole population. A Statistic is a characteristic calculated from the sample.
- **c) Descriptive vs. Inferential:** Descriptive statistics summarize the data we have (using means or charts). Inferential statistics use that data to make predictions or "guesses" about the larger population.

Q3: Central Tendency

- **Mean:** The mathematical average.

- **Median:** The middle value when data is ordered.
- **Mode:** The value that appears most frequently.
- **Preference:** The Median is preferred over the Mean when the data has Outliers (extreme values), because it isn't affected by very high or very low numbers.

Q4: Spread and Shape

- **Variance & Standard Deviation:** They measure how much the data "spreads" away from the average.
- **Skewness:** It describes if the data distribution is symmetrical or "leans" to one side.
- **Outliers:** These are values that are significantly different from the rest of the data.
 - *Effect:* Outliers can pull the Mean up or down and increase the Variance, making the data look less reliable.

Part 2: Analytical & Problem-Based Questions

Q5: Exam Scores Analysis

Data: 45, 50, 52, 55, 60, 62, 65, 70, 85, 98

- **a) Compute Mean and Median:**
 - **Mean:** $(45+50+52+55+60+62+65+70+85+98) / 10 = 64.2$
 - **Median:** The middle numbers are 60 and 62. Median = $(60+62) / 2 = 61$
- **b) Symmetry:** The data is Right-Skewed (Positively Skewed) because the Mean (64.2) is larger than the Median (61).
- **c) Best Measure:** The Median is better because the high scores (85 and 98) act as outliers that inflate the Mean.

Q6: Probability

- **a) Event A not occurring:** $1 - 0.3 = 0.7$
- **b) independent formula:** $P(A \text{ and } B) = P(A) * P(B)$
- **c) Importance:** Independence is critical because it simplifies the math in models and ensures that one variable's error doesn't automatically cause an error in another.