Assignment 1: Statistics & Data Analytics Foundations

Part 1: Key Concepts

Q1: Define the following terms and provide one example for each:

a) Structured data

b) Unstructured data

c) Discrete variable

d) Continuous variable

e) Categorical variable

Q2: Differentiate between:

a) Population vs Sample

b) Parameter vs Statistic

c) Descriptive statistics vs Inferential statistics

Q3: Explain the difference between the following measures of central tendency:

- Mean
- Median
- Mode

Under what circumstances is the median preferred over the mean?

Q4: Define and explain:

- Variance
- Standard deviation
- Skewness
- Outliers

How do outliers affect statistical measures?

Part 2: Analytical & Problem-Based Questions

Q5: A dataset contains exam scores of 10 students:

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

a) Compute the mean and median.

b) Comment on whether the data appears symmetric or skewed.

c) Which measure better represents the dataset and why?

Q6: Explain the concept of probability and answer:

a) If the probability of an event A is 0.3, what is the probability that A does not occur?

b) If two events A and B are independent, write the formula for P(A and B).

c) Why is independence important in data analytics?

Part 3: Reflection (Optional)

Q7: Why is understanding statistics essential before studying Machine Learning?

Relate your answer to concepts such as data distribution, variance, and probability.

Q8: Imagine you are given a dataset with 10,000 rows and 50 features.

Discuss:

- How you would describe the dataset statistically before building a model
- What potential data issues you would look for
- Why dataset understanding is critical in AI applications