# <u>Statistics Basics</u>

# Assignment Questions

**1. Explain the different types of data (qualitative and quantitative) and provide examples of each. Discuss nominal, ordinal, interval, and ratio scales.**

- **Qualitative Data** (Categorical): Describes qualities or characteristics.
    - *Example:* Colors (Red, Blue), Gender (Male, Female)
    - **Nominal Scale:** Categories with no logical order (e.g., Hair color).
    - **Ordinal Scale:** Categories with a meaningful order but no equal intervals (e.g., Rankings: 1st, 2nd).
- **Quantitative Data** (Numerical): Represents measurable quantities.
    - *Example:* Age, Height, Temperature
    - **Interval Scale:** Numeric with equal intervals but no true zero (e.g., Temperature in Celsius).
    - **Ratio Scale:** Numeric with a true zero (e.g., Weight, Height).

**2. What are the measures of central tendency, and when should you use each?**

- **Mean:** Average value; best for symmetrical distributions.
  *Example:* Average test score.
- **Median:** Middle value; preferred when data has outliers or is skewed.
  *Example:* Median income.
- **Mode:** Most frequent value; used for categorical data.
  *Example:* Most common blood type.

**3. Explain the concept of dispersion. How do variance and standard deviation measure the spread of data?**

- **Dispersion** shows how much the data varies.
- **Variance:** Average of squared deviations from the mean.
- **Standard Deviation:** Square root of variance; indicates how spread out data is.
    - Low SD: Data points close to mean.
    - High SD: Data points are spread out.

**4. What is a box plot, and what can it tell you about the distribution of data?**

A **box plot** displays the median, quartiles, and possible outliers of a dataset.
It helps visualize:

- Central tendency (median)
- Spread (interquartile range)
- Skewness (asymmetry)
- Outliers (individual extreme values)

**5. Discuss the role of random sampling in making inferences about populations.**

**Random sampling** ensures each member of a population has an equal chance of being selected.
It reduces bias and allows valid **statistical inferences**, helping researchers generalize results to the whole population.

**6. Explain the concept of skewness and its types. How does skewness affect the interpretation of data?**

**Skewness** measures the asymmetry of a distribution:

- **Positive Skew:** Tail on the right; mean > median.
- **Negative Skew:** Tail on the left; mean < median.

Skewed data affects choice of central tendency and interpretation of results.

**7. What is the interquartile range (IQR), and how is it used to detect outliers?**

**IQR = Q3 - Q1**
It measures the spread of the middle 50% of data.

- **Outliers** are values below Q1 - 1.5×IQR or above Q3 + 1.5×IQR.
- Helps detect extreme values in the dataset.

**8. Discuss the conditions under which the binomial distribution is used.**

Use the **binomial distribution** when:

- Fixed number of trials (n)
- Only two outcomes (Success/Failure)
- Probability (p) is constant
- Trials are independent

*Example:* Flipping a coin 10 times and counting heads.

**9. Explain the properties of the normal distribution and the empirical rule (68–95–99.7 rule).**

- **Normal Distribution:** Bell-shaped, symmetric curve.
- **Mean = Median = Mode**

**Empirical Rule:**

- 68% of data within 1 SD
- 95% within 2 SD
- 99.7% within 3 SD

Used to estimate probability in normal data.

**10. Provide a real-life example of a Poisson process and calculate the probability for a specific event.**

*Example:* Number of calls at a call center per hour.

If average = 3 calls/hour,

probability of exactly 2 calls:

$P(X=2)=(e^{-3} \cdot 3^2)/2! = (e^{-3} \cdot 9)/2 \approx 0.224$

**11. Explain what a random variable is and differentiate between discrete and continuous random variables.**

- **Random Variable:** A variable that takes numerical outcomes of a random process.
- **Discrete:** Countable values (e.g., number of students)
- **Continuous:** Infinite values in a range (e.g., height, weight)

**12. Provide an example dataset, calculate both covariance and correlation, and interpret the results.**

**Dataset:**
X = [1, 2, 3, 4, 5]
Y = [2, 4, 6, 8, 10]

- **Covariance:** Positive (values increase together)
- **Correlation (r):** = +1 (perfect positive linear relationship)

**Interpretation:** As X increases, Y increases at a consistent rate.