Basics of Statistics

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

- Types of Data: Qualitative and Quantitative
- Qualitative Data: Describes qualities or characteristics.
    - Example: Eye color (blue, brown), types of fruits (apple, orange).
- Quantitative Data: Represents numerical values and measurements.
    - Example: Height (160 cm, 170 cm), temperature (25°C, 30°C).

Scales of Measurement:

```
1. Nominal: Categorical data without order.
```

> Example: Gender (Male/Female), colors (Red, Blue).

```
2. Ordinal: Ordered categories without equal spacing.
```

> Example: Ratings (Good, Better, Best).

```
3. Interval: Numerical data with equal intervals, but no true zero.
```

> Example: Temperature in Celsius.

```
4. Ratio: Numerical data with equal intervals and a true zero.
```

> Example: Weight (0 kg means no weight), distance.

---

2. **What are the measures of central tendency, and when should you use each? Discuss the mean, median, and mode with examples and situations where each is appropriate.**

- **Measures of Central Tendency**
    1. Mean: The average value.

- Use: When data has no extreme outliers.

> Example: Test scores (70, 80, 90): Mean = (70+80+90)/3 = 80.

- 2. Median: The middle value when data is ordered.
  - Use: When data has outliers.

> Example: Salaries (30k, 50k, 1M): Median = 50k.

- 3. Mode: The most frequent value.
  - Use: For categorical data or when identifying common values.

> Example: Shirt sizes (M, L, L, XL): Mode = L.

---

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

- **Concept of Dispersion**
  - Dispersion shows how data is spread around the central value.
  - Variance: Measures average squared deviations from the mean.
  - Standard Deviation (SD): Square root of variance, providing spread in the same units as the data.

> Example: Heights (160, 170, 180 cm): Mean = 170, Variance = 66.67, SD ≈ 8.16.

```
Larger SD indicates greater spread.
```

---

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

Box Plot

- A box plot visually shows data distribution:
  - **Box:** Interquartile range (IQR: Q3 - Q1).
  - **Line inside:** Median.
  - **Whiskers:** Range (excluding outliers).
  - **Outliers:** Points outside 1.5 × IQR.

```
**Use:** Identify spread, symmetry, and outliers.
```

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

- **Role of Random Sampling**
    - Random sampling ensures every individual in a population has an equal chance of selection, reducing bias and enabling accurate population inferences.

```
Example: To survey voting preferences, randomly select 500 people from a voter list.
```

---

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

Skewness

- Skewness measures asymmetry in data distribution:
    - **Positive Skew:** Long tail on the right.

    > Example: Income data.

    - **Negative Skew:** Long tail on the left.

    > Example: Students' test scores where most scored high.

```
Skewed data affects mean and interpretation.
```

---

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

Interquartile Range (IQR)

- IQR = Q3 - Q1 (middle 50% of data).
- Use: Detect outliers.
- Outlier if:
    - Below Q1 - 1.5 × IQR.
    - Above Q3 + 1.5 × IQR.

> Example: Data = [1, 2, 3, 4, 5, 6, 7]. Q1 = 2, Q3 = 6, IQR = 4.

---

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

Binomial Distribution

Conditions:

- Fixed number of trials.
- Each trial has two outcomes (success/failure).
- Probability of success is constant.

> Example: Flipping a coin 10 times for heads.

---

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

- Normal Distribution & Empirical Rule
- Normal Distribution: Bell-shaped curve; mean = median = mode.

Empirical Rule:

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

> Example: Heights (mean = 170 cm, SD = 10 cm): 68% are between 160–180 cm.

---

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

Poisson Process Example

> Example: Calls received at a helpline = 3/hour. Find probability of 2 calls in an hour:

- Probability = 22.4%.

---

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

Random Variables

- Numerical outcome of an experiment.

- **Discrete:** Countable values (e.g., dice roll).
- **Continuous:** Infinite values (e.g., height).

---

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

- **Covariance & Correlation**

  Example Dataset: X = [1, 2, 3], Y = [2, 4, 6].

- Covariance: Measures joint variability.

  > Cov(X, Y) = 2.

- Positive value = X and Y increase together.
- Correlation: Standardized measure (-1 to 1).

  > Corr(X, Y) = 1 (perfect positive relationship).
