

Statistics Concepts Summary

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

- Qualitative Data: Descriptive data that categorizes or labels attributes.
Example: Gender, Eye color.
- Quantitative Data: Data that can be measured numerically.
Example: Height, Weight.
- Nominal: Categories without order (e.g., Blood type).
- Ordinal: Ordered categories (e.g., Satisfaction level).
- Interval: Equal intervals, no true zero (e.g., Temperature).
- Ratio: Equal intervals, true zero (e.g., Income).

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.

- Mean: Use with symmetrical data; Example: Average marks.
- Median: Use with skewed data; Example: Median salary.
- Mode: Use with categorical data; Example: Most common age group.

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

- Dispersion indicates how spread out data is.
- Variance: Average of squared deviations.
- Standard Deviation: Root of variance; shows spread around the mean.

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

- Box Plot is a graphical summary showing Q1, median, Q3, and outliers.
- It helps visualize spread, skewness, and potential outliers.

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

- Random sampling gives all individuals equal chance.

Statistics Concepts Summary

- Ensures representation, reduces bias, supports generalization to the population.

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

- Skewness measures asymmetry of data.
- Positive Skew: Tail on right, Mean > Median.
- Negative Skew: Tail on left, Mean < Median.
- Affects choice of central tendency measure.

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

- IQR = $Q3 - Q1$; measures middle 50% spread.
- Outliers: Below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$.

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

- Fixed number of trials.
- Two outcomes: success/failure.
- Constant success probability.
- Independent trials.

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

- Normal distribution is symmetric, bell-shaped.
- Empirical rule: 68% ($\pm 1\sigma$), 95% ($\pm 2\sigma$), 99.7% ($\pm 3\sigma$).

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

- Example: 5 calls/hour. Find $P(X=3)$:

$$P(X=3) = (e^{-5} * 5^3) / 3! \text{ approx. } 0.14$$

- So, 14% chance of 3 calls in an hour.

Statistics Concepts Summary

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

- Random Variable: Value from random experiment.
- Discrete: Countable (e.g., number of cars).
- Continuous: Any value in range (e.g., height).

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

- Dataset: $X=[2,3,4,5]$, $Y=[4,5,6,8]$
- $\text{Mean}(X)=3.5$, $\text{Mean}(Y)=5.75$
- $\text{Cov}(X,Y)=1.67$, $\text{Corr}(X,Y)=0.76$
- Interpretation: Strong positive relation.