Q21: Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. It indicates whether the observations in a dataset are skewed to the left (negative skewness) or to the right (positive skewness).

- **Types of Skewness**:
    1. **Positive Skewness** (Right Skewness):
        - The right tail (higher values) is longer or fatter than the left tail.
        - Most data points are concentrated on the left.
        - Mean > Median > Mode.
    2. **Negative Skewness** (Left Skewness):
        - The left tail (lower values) is longer or fatter than the right tail.
        - Most data points are concentrated on the right.
        - Mean < Median < Mode.
    3. **Zero Skewness** (Symmetrical Distribution):
        - The left and right tails are equally balanced.
        - Mean = Median = Mode.

Graphical Representation:

- **Positive Skewness**:
- **Negative Skewness**:

Q22: Probability Mass Function (PMF) and Probability Density Function (PDF)

- **Probability Mass Function (PMF)**:
    - Defines the probability distribution for discrete random variables.
    - The PMF gives the probability that a discrete random variable is exactly equal to some value.
    - Mathematically: P(X=x)P(X = x)P(X=x).
- **Probability Density Function (PDF)**:
    - Defines the probability distribution for continuous random variables.
    - The PDF describes the relative likelihood for this random variable to take on a given value.
    - The area under the PDF curve represents the probability.
    - Mathematically: f(x)f(x)f(x).

Difference:

- PMF is used for discrete variables, while PDF is used for continuous variables.
- PMF gives probabilities directly, while the PDF must be integrated over an interval to obtain probabilities.

Q23: Correlation

Correlation measures the strength and direction of a linear relationship between two variables.

- **Types of Correlation**:
    1. **Positive Correlation**:
        - Both variables move in the same direction.
        - As one variable increases, the other also increases.
    2. **Negative Correlation**:
        - Variables move in opposite directions.
        - As one variable increases, the other decreases.
    3. **No Correlation**:
        - No linear relationship between the variables.

Methods of Determining Correlation:

1. **Pearson's Correlation Coefficient**:
    - Measures the linear relationship between two continuous variables.
    - Values range from -1 to 1.
    - Formula: r=∑(X−X¯)(Y−Y¯)∑(X−X¯)2∑(Y−Y¯)2r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 \sum (Y - \overline{Y})^2}}r=∑(X−X)2∑(Y−Y)2∑(X−X)(Y−Y)
2. **Spearman's Rank Correlation**:
    - Measures the relationship between two ranked variables.
    - Values range from -1 to 1.
    - Formula: rs=1−6∑di2n(n2−1)r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}rs=1−n(n2−1)6∑di2
3. **Kendall's Tau**:
    - Measures the strength of dependence between two variables.
    - Values range from -1 to 1.

Q24: Calculate the Coefficient of Correlation

To calculate the Pearson correlation coefficient between the marks obtained by students in Accountancy and Statistics, we can use the following formula:

r=∑(Xi−X)∑(Yi−Y) / (∑(Xi−X)^2 ∑(Yi−Y)^2)^(1/2)

Where:

- Xi and Yi are the individual sample points indexed with i.
- X and Y are the means of the X and Y datasets, respectively.

Let's calculate it using Python:

```
import numpy as np
```

```
# Marks obtained by students
accountancy_marks = [45, 70, 65, 30, 90, 40, 50, 75, 85, 60]
statistics_marks = [35, 90, 70, 40, 95, 40, 60, 80, 80, 50]

# Calculate Pearson's correlation coefficient
correlation_coefficient = np.corrcoef(accountancy_marks,
statistics_marks)[0, 1]
Correlation_coefficient
```

Q25: Differences between Correlation and Regression

1. **Definition**:
    - **Correlation** measures the strength and direction of a linear relationship between two variables.
    - **Regression** analyzes the nature of the relationship between two variables, specifically how one variable affects another.
2. **Symmetry**:
    - **Correlation** is symmetric, meaning the correlation between XXX and YYY is the same as between YYY and XXX.
    - **Regression** is asymmetric; the regression of YYY on XXX is generally different from the regression of XXX on YYY.
3. **Purpose**:
    - **Correlation** aims to quantify the degree of relationship between variables.
    - **Regression** aims to predict the value of one variable based on the value of another variable.
4. **Units**:
    - **Correlation** is a unitless measure, ranging from -1 to 1.
    - **Regression** coefficients have units and depend on the units of the variables involved.

Q26: Finding the Most Likely Price at Delhi

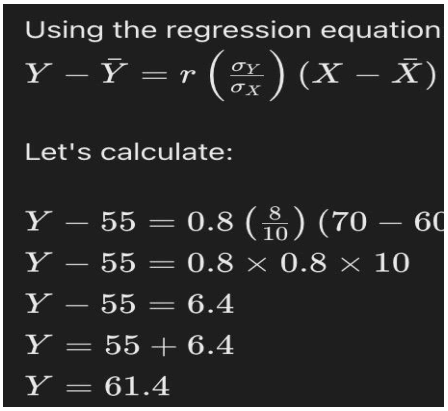
Given:

- Correlation coefficient (rrr) = +0.8
- Price at Agra (XXX) = Rs. 70

Let's assume:

- Mean price at Agra (X^-\bar{X}X^-) = 60

- Mean price at Delhi (Y¯\bar{Y}Y¯) = 55
- Standard deviation at Agra (σX\sigma_XσX) = 10
- Standard deviation at Delhi (σY\sigma_YσY) = 8



Therefore, the most likely price at Delhi corresponding to the price of Rs. 70 at Agra is Rs. 61.4.

Q28: Normal Distribution

Normal Distribution is a continuous probability distribution characterized by a bell-shaped curve, symmetric about the mean.

Assumptions:

1. **Independence**:
    - Each observation is independent of the others.
2. **Sample Size**:
    - Large sample sizes tend to produce more accurate approximations of normality.
3. **Random Sampling**:
    - Data is collected through random sampling methods.
4. **Measurement Scale**:
    - Data should be measured on an interval or ratio scale.

Q29: Characteristics of Normal Distribution Curve

1. **Bell-shaped and Symmetric**:
    - The curve is symmetric about the mean.
2. **Mean, Median, Mode**:
    - All are equal and located at the center of the distribution.
3. **Asymptotic**:

    - The tails approach the horizontal axis but never touch it.
4. **Empirical Rule**:
    - About 68% of data within 1 standard deviation (σ), 95% within 2σ, and 99.7% within 3σ.
5. **Total Area**:
    - The total area under the curve is 1.

Q30: Correct Options about Normal Distribution Curve

- **(a)**: Incorrect. The middle 50% area is covered by mean ± 0.6745σ.
- **(b)**: Correct.
- **(c)**: Correct.
- **(d)**: Correct.
- **(e)**: Correct.