**[Que-23] - What is correlation. Explain its type in details. what are the methods of determining correlation**

**Correlation** refers to the statistical relationship between two or more random variables. It describes how changes in one variable are associated with changes in another variable. Correlation does not imply causation, meaning that a correlation between variables does not necessarily indicate a cause-and-effect relationship.

### **Types of Correlation**

1. **Pearson Correlation Coefficient (Pearson's r)**:
   * Measures the linear relationship between two continuous variables.
   * **Range**: *−1-1*−1 to *+1+1*+1
     + *+1+1*+1: Perfect positive correlation (as one variable increases, the other also increases linearly).
     + *−1-1*−1: Perfect negative correlation (as one variable increases, the other decreases linearly).
     + *00*0: No linear correlation (variables are not linearly related).
   * **Formula**: *r=∑i=1n(Xi−Xˉ)(Yi−Yˉ)∑i=1n(Xi−Xˉ)2∑i=1n(Yi−Yˉ)2r = \frac{\sum\_{i=1}^{n} (X\_i - \bar{X})(Y\_i - \bar{Y})}{\sqrt{\sum\_{i=1}^{n} (X\_i - \bar{X})^2} \sqrt{\sum\_{i=1}^{n} (Y\_i - \bar{Y})^2}}*r=∑i=1n (Xi −Xˉ)2 ∑i=1n (Yi −Yˉ)2 ∑i=1n (Xi −Xˉ)(Yi −Yˉ) where *XiX\_i*Xi and *YiY\_i*Yi are the individual data points, *Xˉ\bar{X}*Xˉ and *Yˉ\bar{Y}*Yˉ are the means of *XX*X and *YY*Y, respectively.
2. **Spearman's Rank Correlation Coefficient**:
   * Measures the strength and direction of association between two ranked variables.
   * Appropriate for both continuous and ordinal variables.
   * Non-parametric measure (based on ranks rather than actual values).
   * **Range**: *−1-1*−1 to *+1+1*+1
   * Calculated similarly to Pearson's r but using ranks of data points.
3. **Kendall's Tau Coefficient**:
   * Measures the strength and direction of association between two variables.
   * Non-parametric measure based on concordant and discordant pairs of data.
   * Appropriate for ordinal data.
   * **Range**: *−1-1*−1 to *+1+1*+1

### **Methods of Determining Correlation**

1. **Scatter Plots**:
   * Visual representation of the relationship between two variables.
   * Helps visualize the direction and strength of correlation.
2. **Correlation Coefficient**:
   * Numerical measure of the strength and direction of the linear relationship between two variables.
   * Pearson's r, Spearman's rho, and Kendall's tau are commonly used correlation coefficients.
3. **Covariance**:
   * Measures the joint variability of two random variables.
   * Positive covariance indicates a positive relationship, while negative covariance indicates a negative relationship.
   * However, covariance alone does not standardize the scale and hence is less interpretable than correlation coefficients.
4. **Coefficient of Determination (*R2R^2*R2)**:
   * Square of the Pearson correlation coefficient (r).
   * Represents the proportion of the variance in one variable that is predictable from the other variable.
   * *R2R^2*R2 values range from 0 to 1, where 0 indicates no relationship and 1 indicates a perfect relationship.
5. **Hypothesis Testing**:
   * Tests such as Pearson's correlation test, Spearman's rank correlation test, and Kendall's tau test determine if the correlation coefficient is significantly different from zero.
   * Helps assess whether the observed correlation is statistically significant.
6. **Machine Learning Models**:
   * Techniques like feature importance in regression models or decision trees can indirectly indicate the strength of correlation between predictors and outcomes.