Simple Linear Regression - Viva Questions

Conceptual Questions

1. What is simple linear regression?

- Explain it as a statistical method to model the relationship between a single independent variable (predictor) and a dependent variable (outcome) by fitting a linear equation.

2. What is the equation of a simple linear regression model?

- $y = b0 + b1x + e$

- Where y is the dependent variable, x is the independent variable, b0 is the intercept, b1 is the slope, and e is the error term.

3. What do the slope and intercept represent in the model?

- Intercept (b0): The predicted value of y when x = 0.

- Slope (b1): The change in y for a one-unit increase in x.

4. What are the assumptions of simple linear regression?

- Linearity: Relationship between x and y is linear.

- Independence: Observations are independent.

- Homoscedasticity: Constant variance of residuals.

- Normality: Residuals are normally distributed.

5. What is the difference between correlation and regression?

- Correlation measures the strength and direction of a linear relationship between two variables, while regression predicts the dependent variable based on the independent variable.

Simple Linear Regression - Viva Questions

6. What is the purpose of residuals in regression?

- Residuals (errors) represent the difference between the observed and predicted values of y. They help assess model fit and assumptions.

Interpretation Questions

7. What does an R-squared value indicate?

- It indicates the proportion of variance in the dependent variable explained by the independent variable (ranges from 0 to 1).

8. How do you interpret the p-value in regression analysis?

- A p-value tests the null hypothesis that the coefficient of the predictor is zero. A small p-value (e.g., < 0.05) indicates that the predictor is statistically significant.

9. What if the slope coefficient is negative?

- A negative slope indicates that as the independent variable increases, the dependent variable decreases.

10. What does it mean if the residuals show a pattern when plotted?

- It suggests that one or more assumptions (e.g., linearity or homoscedasticity) are violated.

Practical/Implementation Questions

Simple Linear Regression - Viva Questions

11. How do you evaluate the performance of a regression model?

- Metrics include R-squared, adjusted R-squared, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

12. What are outliers, and how do they affect a regression model?

- Outliers are data points that deviate significantly from the pattern. They can skew the regression line and affect the slope and intercept.

13. What steps would you take if the assumptions of linear regression are violated?

- Transform variables, remove outliers, use polynomial regression, or switch to a non-linear model.

14. What is multicollinearity, and does it apply to simple linear regression?

- Multicollinearity is when independent variables are highly correlated. It does not apply to simple linear regression, as there is only one predictor.

15. How would you implement simple linear regression in Python?

- Steps:
    - Import libraries: `import numpy as np`, `import pandas as pd`, `from sklearn.linear_model import LinearRegression`.
    - Fit the model using `LinearRegression().fit(X, y)`.
    - Evaluate using `.score(X, y)` and other metrics.

16. How do you check if the model assumptions hold?

- Residual plots, Q-Q plots, Shapiro-Wilk test for normality, and Breusch-Pagan test for homoscedasticity.

Applied Scenario Questions

17. How would you decide if the model fits well?

- Check R-squared, residual plots, and the significance of the predictor using p-values.

18. What happens if there is no relationship between x and y?

- The slope b1 would be close to zero, and the R-squared value would also be near zero.

19. What is overfitting in the context of regression?

- Overfitting occurs when the model captures noise rather than the underlying relationship, which can be more common in complex models but less so in simple linear regression.

20. Can you use simple linear regression with categorical data?

- Not directly; categorical data must first be encoded (e.g., one-hot encoding) into numerical values.