Linear Regression - Complete Questions & Answers

## Q: What is linear regression?

A: Linear regression is a statistical technique used to model the relationship between a dependent variable (target) and one or more independent variables (features) by fitting a straight line.

## Q: What are the assumptions of linear regression?

A: 1. Linearity: Relationship between predictors and target is linear.  
2. Independence: Observations are independent.  
3. Homoscedasticity: Constant variance of residuals.  
4. Normality: Residuals follow a normal distribution.  
5. No multicollinearity: Predictors are not highly correlated.

## Q: Difference between simple and multiple linear regression?

A: Simple linear regression has one predictor, while multiple linear regression uses two or more predictors.

## Q: What is the equation of a linear regression model?

A: y = β₀ + β₁x₁ + β₂x₂ + ... + βnxn + ε, where β are coefficients and ε is the error term.

## Q: What does the slope (β₁) represent?

A: The slope represents the expected change in the dependent variable for a one-unit change in the independent variable, keeping others constant.

## Q: What does the intercept (β₀) represent?

A: It is the predicted value of the dependent variable when all predictors are zero.

## Q: What is the difference between correlation and regression?

A: Correlation measures strength/direction of a linear relationship, while regression builds a predictive model.

## Q: When should you use linear regression?

A: When the target variable is continuous, assumptions hold, and the relationship is approximately linear.

## Q: What is the difference between linear and logistic regression?

A: Linear regression predicts continuous outcomes; logistic regression predicts categorical probabilities.

## Q: What is the purpose of the error term (ε)?

A: It represents unexplained variation in the dependent variable.

## Q: How do we estimate the coefficients (β)?

A: Using Ordinary Least Squares (OLS) or iterative optimization methods like gradient descent.

## Q: Explain the Ordinary Least Squares (OLS) method.

A: OLS minimizes the sum of squared differences between actual and predicted values.

## Q: Derive the formula for slope (β₁) in simple regression.

A: β₁ = Σ((x - mean(x)) \* (y - mean(y))) / Σ((x - mean(x))²).

## Q: Derive the formula for intercept (β₀).

A: β₀ = mean(y) - β₁ \* mean(x).

## Q: What is the cost function used in linear regression?

A: Mean Squared Error (MSE) = (1/n) Σ(y\_pred - y\_actual)².

## Q: Why do we minimize the sum of squared errors?

A: Squaring penalizes larger errors more and ensures non-negative values.

## Q: Explain MSE and RMSE.

A: MSE is average squared error; RMSE is its square root, providing error in original units.

## Q: What is R²?

A: R² measures proportion of variance explained by the model (0 ≤ R² ≤ 1).

## Q: How is Adjusted R² different from R²?

A: Adjusted R² accounts for number of predictors, penalizing irrelevant features.

## Q: What is the F-statistic?

A: It tests overall model significance (whether at least one predictor is useful).

## Q: What are key assumptions of linear regression?

A: Linearity, independence, homoscedasticity, normality, no multicollinearity.

## Q: What happens if assumptions are violated?

A: Model estimates may be biased, inefficient, or misleading.

## Q: How do you check for multicollinearity?

A: By using correlation matrices or Variance Inflation Factor (VIF).

## Q: What is VIF?

A: Variance Inflation Factor measures how much variance is inflated due to multicollinearity. VIF > 10 indicates severe multicollinearity.

## Q: How do you check for homoscedasticity?

A: Plot residuals vs. fitted values; constant spread suggests homoscedasticity.

## Q: What is heteroscedasticity?

A: Unequal variance of residuals across values of predictors, which violates assumptions.

## Q: How do you detect autocorrelation?

A: Use residual plots or statistical tests like Durbin-Watson.

## Q: What is the Durbin-Watson test?

A: It checks for autocorrelation in residuals. Value ~2 means no autocorrelation.

## Q: How do you check if residuals are normal?

A: By using Q-Q plots, histograms, or Shapiro-Wilk test.

## Q: How do you handle outliers?

A: By removing, transforming, or using robust regression techniques.

## Q: How does gradient descent work in linear regression?

A: It iteratively updates coefficients in the opposite direction of the gradient of the cost function.

## Q: What is the gradient of the cost function?

A: For MSE: ∂/∂β₀ = -(2/n) Σ(y - y\_pred), ∂/∂β₁ = -(2/n) Σx(y - y\_pred).

## Q: Why partial derivatives for β₀ and β₁?

A: They indicate how much to adjust coefficients to reduce error.

## Q: Difference between batch, stochastic, and mini-batch gradient descent?

A: Batch uses all data, SGD uses one sample at a time, mini-batch uses subsets.

## Q: How does learning rate affect convergence?

A: Too high → divergence, too low → slow convergence.

## Q: Why might gradient descent fail?

A: Poor learning rate choice or numerical instability.

## Q: Closed-form OLS vs Gradient Descent?

A: OLS solves directly; gradient descent iteratively optimizes.

## Q: Which is faster for large datasets?

A: Gradient descent scales better; OLS is better for small datasets.

## Q: Can gradient descent get stuck in local minima?

A: No, for linear regression cost function is convex; only global minimum exists.

## Q: How do you implement regression in Python NumPy?

A: By manually applying OLS formulas or using gradient descent.

## Q: How to implement in scikit-learn?

A: Use LinearRegression() from sklearn.linear\_model.

## Q: How do you split data?

A: Using train\_test\_split from sklearn.model\_selection.

## Q: How do you evaluate model performance?

A: Using metrics such as R², MSE, RMSE, MAE.

## Q: What metrics are commonly used?

A: MSE, RMSE, MAE, R², Adjusted R².

## Q: What is polynomial regression?

A: Regression where predictors are polynomial terms of original variables.

## Q: What is interaction term?

A: Product of two variables, capturing combined effect.

## Q: How do you handle categorical variables?

A: By encoding them (One-Hot, Label Encoding, etc.).

## Q: What is regularization in regression?

A: Adding penalty terms (L1, L2) to prevent overfitting.

## Q: What is ridge regression?

A: Linear regression with L2 penalty to shrink coefficients.

## Q: What is lasso regression?

A: Linear regression with L1 penalty, which can shrink some coefficients to zero.

## Q: What is elastic net?

A: Combination of L1 and L2 penalties.

## Q: How does regularization prevent overfitting?

A: By penalizing large coefficients, reducing variance.

## Q: What is bias-variance tradeoff?

A: Balance between underfitting (high bias) and overfitting (high variance).

## Q: What is cross-validation?

A: Technique to evaluate model performance using different training/test splits.

## Q: How do you detect overfitting?

A: High training accuracy but low test accuracy.

## Q: How do you handle missing values?

A: By imputation (mean/median/mode) or removing rows/columns.

## Q: Can regression be used for time series forecasting?

A: Yes, but assumptions must be carefully checked, and features engineered.

## Q: Real-world example?

A: Predicting house prices based on features like area, rooms, location.

## Q: Can linear regression model non-linear relationships?

A: Not directly, but polynomial regression or transformations can help.

## Q: What happens if predictors are highly correlated?

A: Multicollinearity inflates variance of coefficients and weakens reliability.

## Q: Why is regression sensitive to outliers?

A: Because OLS minimizes squared errors, outliers disproportionately influence results.

## Q: Can regression be used for categorical targets?

A: No, logistic regression or classification models are better.

## Q: Difference between parametric and non-parametric models?

A: Parametric assumes fixed functional form; non-parametric is more flexible.

## Q: Is regression parametric?

A: Yes, because it assumes a linear functional form between variables.

## Q: What are leverage points?

A: Data points with extreme predictor values that influence model fit.

## Q: What is Cook’s distance?

A: Measure of influence of a data point on regression results.

## Q: R² = 0.9 vs R² = 0.5 meaning?

A: R² = 0.9 means 90% variance explained; R² = 0.5 means only 50% explained.

## Q: Why prefer a simpler model even with lower R²?

A: For better generalization, interpretability, and avoiding overfitting.