# Linear Regression Q&A

## What is Simple Linear Regression

A model that shows a linear relationship between one independent variable (X) and one dependent variable (Y).

## What are the key assumptions of Simple Linear Regression

1. Linearity between X and Y.  
2. Independence of residuals.  
3. Homoscedasticity (constant variance of errors).  
4. Normality of residuals.

## What does the coefficient m represent in the equation Y=mX+c

It is the slope of the line, representing how much Y changes for a one-unit change in X.

## What does the intercept c represent in the equation Y=mX+c

The value of Y when X = 0.

## How do we calculate the slope m in Simple Linear Regression

m = [n(ΣXY) - (ΣX)(ΣY)] / [n(ΣX²) - (ΣX)²]

## What is the purpose of the least squares method in Simple Linear Regression

To minimize the sum of squared residuals and find the best-fitting line.

## How is the coefficient of determination (R²) interpreted in Simple Linear Regression

It represents the proportion of variance in Y explained by X. For example, R² = 0.9 means 90% of Y's variance is explained by X.

## What is Multiple Linear Regression

A regression model with two or more independent variables.

## What is the main difference between Simple and Multiple Linear Regression

Simple has one predictor; Multiple has two or more predictors.

## What are the key assumptions of Multiple Linear Regression

1. Linearity  
2. Independence  
3. Homoscedasticity  
4. Normality of residuals  
5. No multicollinearity

## What is heteroscedasticity, and how does it affect the results of a Multiple Linear Regression model

It occurs when residuals have non-constant variance, leading to inefficient estimates and biased standard errors.

## How can you improve a Multiple Linear Regression model with high multicollinearity

Remove correlated variables, use PCA, or apply Ridge/Lasso regression.

## What are some common techniques for transforming categorical variables for use in regression models

One-hot encoding, label encoding, and ordinal encoding.

## What is the role of interaction terms in Multiple Linear Regression

They capture the combined effect of two variables on the outcome.

## How can the interpretation of intercept differ between Simple and Multiple Linear Regression

Simple: Y when X=0. Multiple: Y when all Xs=0.

## What is the significance of the slope in regression analysis, and how does it affect predictions

It indicates how much Y changes with a one-unit increase in X.

## What are the limitations of using R² as a sole measure of model performance

R² doesn't detect overfitting or bias and increases with more variables regardless of relevance.

## How would you interpret a large standard error for a regression coefficient

It indicates high variability in the estimate, possibly due to multicollinearity or model issues.

## What is polynomial regression

A type of regression that uses an nth-degree polynomial to model non-linear relationships.

## When is polynomial regression used

When data shows a curved trend that can't be captured by linear regression.

## How does the intercept in a regression model provide context for the relationship between variables

It represents the expected value of Y when all predictors are 0.

## How can heteroscedasticity be identified in residual plots, and why is it important to address it

By looking for patterns like a funnel shape. It affects the reliability of inference.

## What does it mean if a Multiple Linear Regression model has a high R² but low adjusted R²

The model includes irrelevant predictors that do not improve performance.

## Why is it important to scale variables in Multiple Linear Regression

To ensure all features contribute equally and coefficients are comparable.

## How does polynomial regression differ from linear regression

Polynomial regression includes higher-degree terms for non-linearity.

## What is the general equation for polynomial regression

Y = β0 + β1X + β2X² + ... + βnXⁿ + ε

## Can polynomial regression be applied to multiple variables

Yes, including cross-product and higher-order terms of multiple predictors.

## What are the limitations of polynomial regression

It can overfit, be less interpretable, and is sensitive to outliers.

## What methods can be used to evaluate model fit when selecting the degree of a polynomial

Cross-validation, learning curves, AIC/BIC, and adjusted R².

## Why is visualization important in polynomial regression

To visually inspect if the curve fits the data and check for overfitting.

## How is polynomial regression implemented in Python?

Using `PolynomialFeatures` with `LinearRegression` in sklearn: