Questions from students:

“Can u plz explain coefficients of correlation/ variation/ determination & their values that can explain linearity or non-linearity in variables e.g. if the corr.coef. is 0.7 does it mean nonlinearity & in that case which multivariate analysis to implement etc step by step plz”

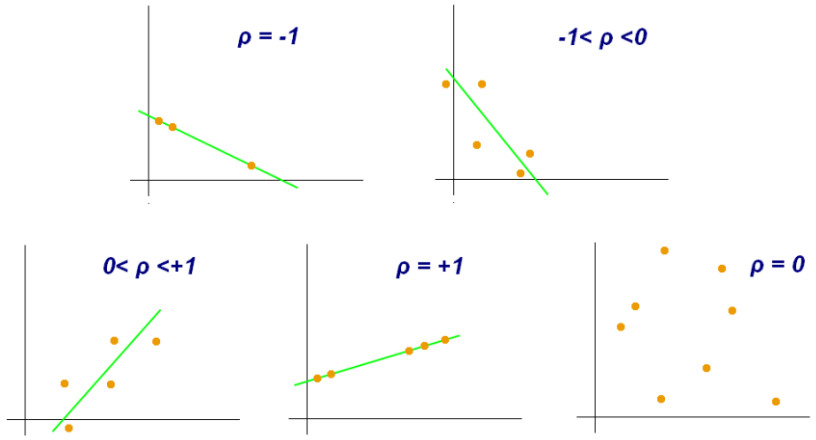
‘can multi variance analyses be next lecture topic?”

Coefficient of Correlation

The coefficient of correlation (usually denoted as *r*) measures the strength and direction of the linear relationship between two variables.

Its value ranges from -1 to +1:

- *r* = +1: Perfect positive linear relationship
- *r* = −1: Perfect negative linear relationship
- *r* = 0: No linear relationship



Source: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#/media/File:Correlation_coefficient.png

Note: The coefficient of correlation alone suggests a strong positive/negative linear relationship between the two variables. However, it does not rule out the possibility of a nonlinear relationship. You should visualise the data using scatter plots to check for linearity visually.

Coefficient of Variation

Coefficient of variation is a type of relative measure of dispersion. It helps to compare two datasets on the basis of the degree of variation. Lower CV indicates less relative variability.

Coefficient of Variation

Population	$\frac{\sigma}{\mu} \times 100$
Sample	$\frac{s}{\bar{x}} \times 100$

Source: https://www.cuemath.com/coefficient-of-variation-formula/

Example usage
Two plants C and D of a factory show the following results:

No. of workers	5000	6000
Average monthly wages	$2500	$2500
Standard deviation	9	10

Plant C CV = (9/2500) x 100 = 0.36
Plant D CV = (10/2500) x 100 = 0.4

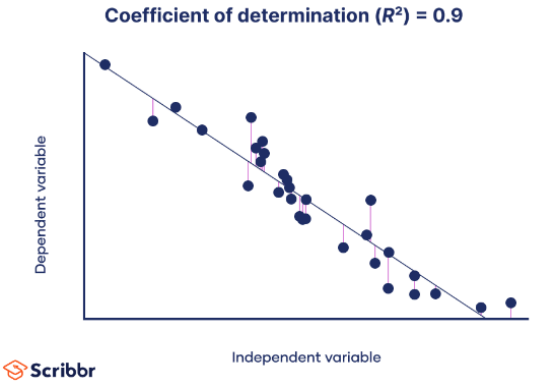
Plant D has greater variability in individual wages.

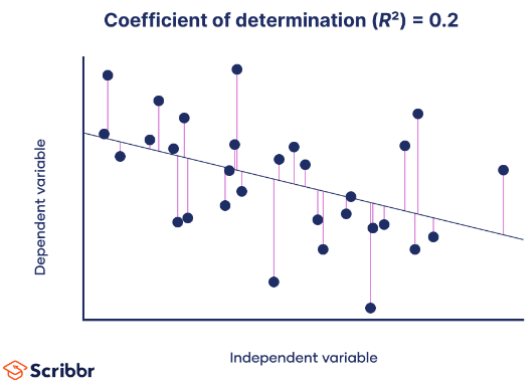
Coefficient of Determination

The coefficient of determination R^2 represents the proportion of variance in the dependent variable that is predictable from the independent variable.

- R^2 = 1: Perfect fit
- R^2 = 0: No fit
- High R^2 means the model explains a large portion of the variance
- Low R^2 means the model explains a small portion of the variance

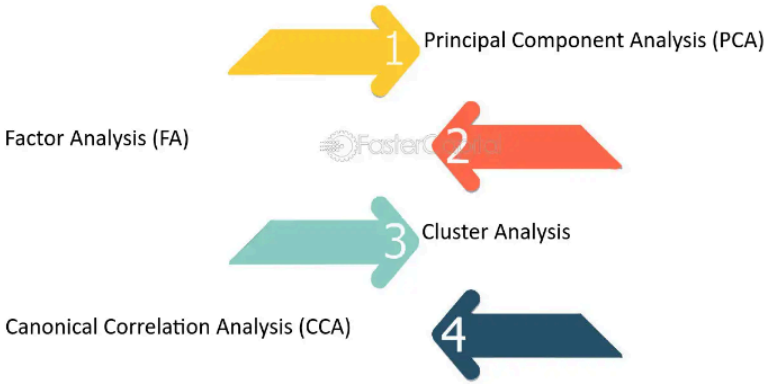
This is one of the performance metrics we used for regression analysis.





Multivariate Analysis

Multivariate analysis refers to a set of statistical techniques used for analysing data that involves **multiple variables simultaneously**. The good news is we have done some multivariate analysis. Multiple Regression Analysis, Discriminant Analysis, and Multivariate Analysis of Variance (MANOVA) are examples of multivariate techniques, and so are the following:



Useful Resources for SQL

An introduction to databases:
https://www.digitalocean.com/community/conceptual-articles/an-introduction-to-databases
Nice examples for beginners:
https://www.programiz.com/sql/database-introduction