23MA301: MATHEMATICAL STATISTICAL FOUNDATION

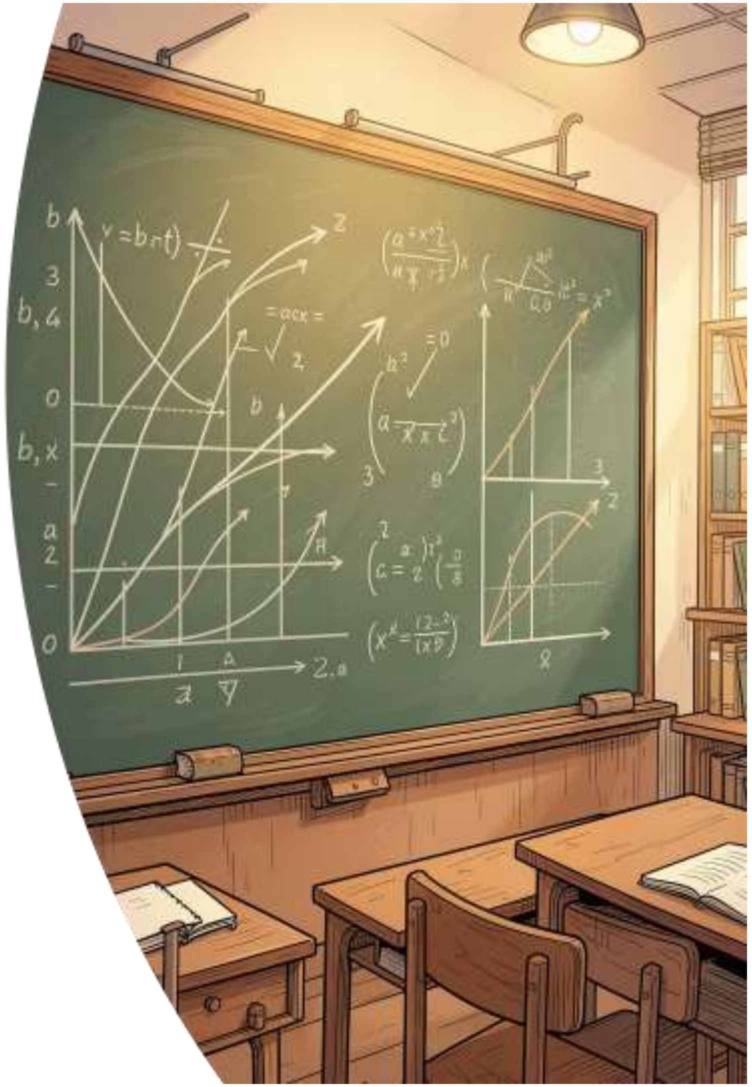
Topic: APPLIED STATISTICS

Mrs. K.Srividya

Assistant Professor, Freshman engineering
Narsimha Reddy Engineering College (Autonomous)
Secunderabad, Telangana, India- 500100.

Applied Statistics

Curve fitting, correlation, and regression — the mathematical tools for discovering patterns and relationships hidden in data.



OVERVIEW

What We'll Cover

01

Least Squares Method

Principle and derivation for best-fit curves

02

Fitting Straight Lines

Linear regression and normal equations

03

Second Degree Parabolas

Quadratic curve fitting for non-linear trends

04

Correlation & Regression

Measuring and modeling variable relationships

05

Rank Correlation

Spearman's coefficient for ordinal data

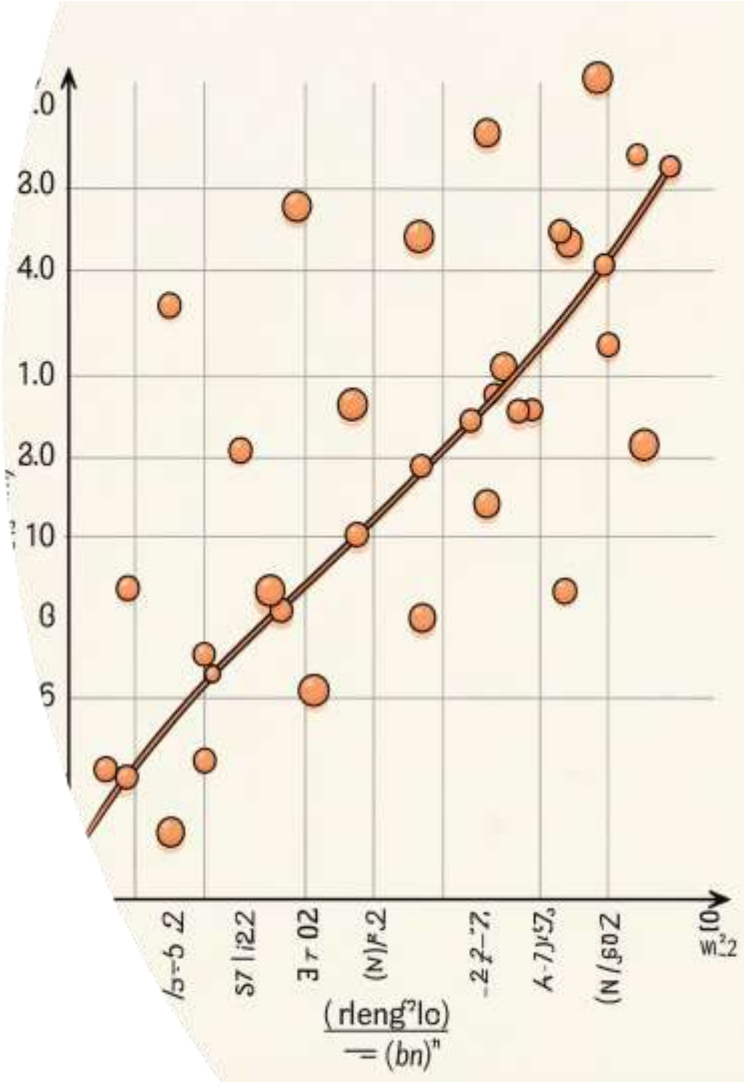
CORE CONCEPT

The Method of Least Squares

Given data points *(x_i, y_i)*, we seek a curve *y* = *f(x)* that **minimizes the sum of squared residuals**:

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Squaring ensures positive contributions and penalizes larger errors more heavily. Setting *\ frac{\ partial S}{\ partial a_k}* = *0* for each parameter yields the **normal equations** — a solvable linear system.



Fitting a Straight Line :

The Model

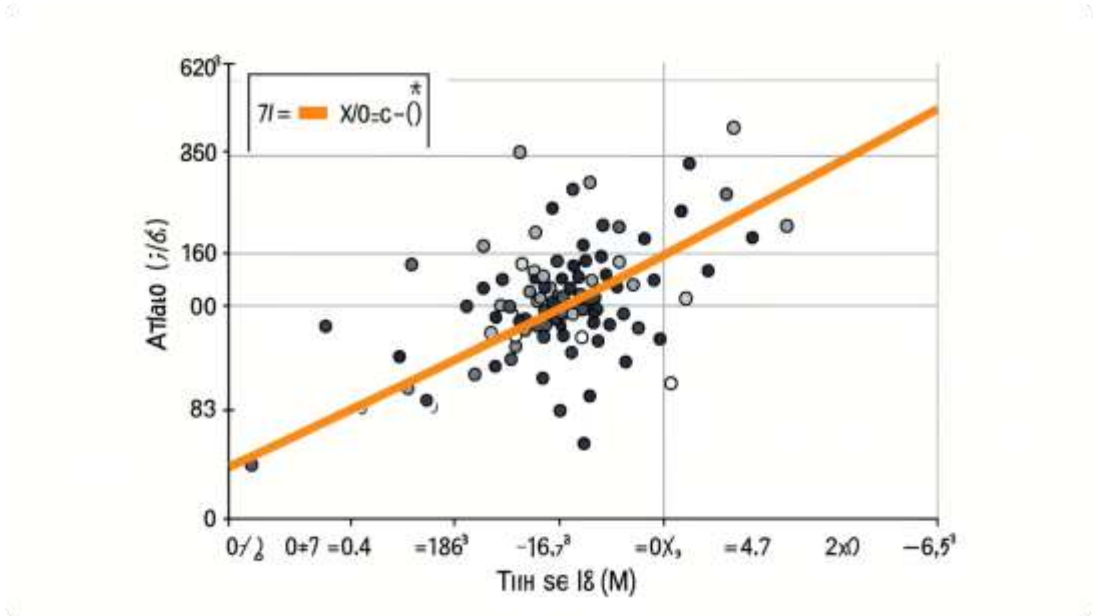
We fit $y = a + bx$ by minimizing $S = \sum(y_i - a - bx_i)^2$.

Normal Equations

$$\sum y = na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

Solving these simultaneously gives the best-fit slope *b* and intercept *a*.



Fitting of a Second Degree Polynomial (Parabola) by the Method of Least Squares

Definition

Curve fitting is the process of finding a mathematical equation that best represents the relationship between two variables. When the relationship is parabolic, a **second-degree polynomial** is fitted.

The required equation is $y = a + bx + cx^2$

where:

- a= intercept ,b= coefficient of x ,c= coefficient of x^2

The constants a, b, and $c$are determined using the **Least Squares Method**, which minimizes the sum of the squares of the errors.

Normal Equations

For fitting $y = a + bx + cx^2$

the normal equations are

$$\sum y = na + b\sum x + c\sum x^2$$

$$\sum x\,y = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

where $n$is the number of observations.

RELATIONS HIPS

Correlation

Correlation quantifies the **strength and direction** of a linear relationship between two variables.

Pearson's r

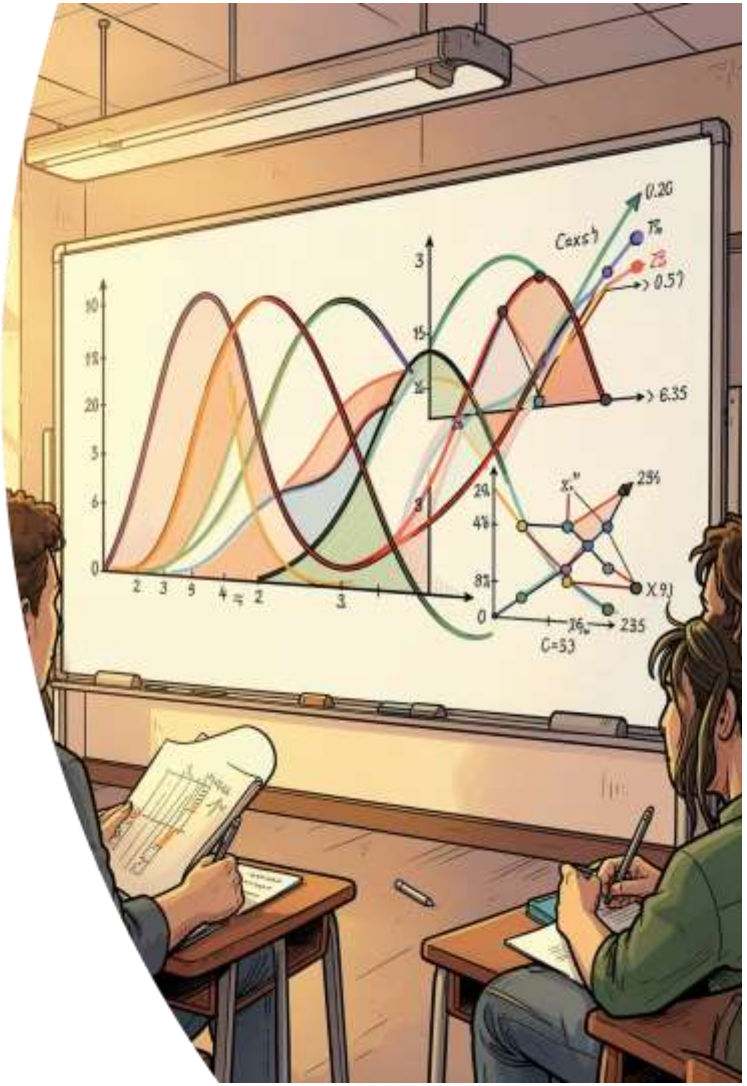
$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Interpretation

$r = +1$: perfect positive
$r = 0$: no linear correlation
$r = -1$: perfect negative

Properties

Unitless, symmetric, and bounded: $-1 \le r \le 1$.
Sensitive to outliers.



For **Spearman's Rank Correlation Coefficient (ρ or r_s)** with **equal number of observations (n)**, these are the key formulas you should remember.

1. When there are no repeated (tied) ranks

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Where:

• n= number of paired observations

• d= difference between the two ranks

• $\sum d^2$= sum of squared rank differences

When repeated (tied) ranks are present

Assign the **average rank** to tied observations, then use the tie-corrected formula:

$$r_s = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}\sum(t^3 - t) + \frac{1}{12}\sum(u^3 - u)\right]}{n(n^2 - 1)}$$

Where:

t= size of each tied group in the first variable

u= size of each tied group in the second variable

Add $\frac{t^3-t}{12}$ for every tied group in the first ranking.

Add $\frac{u^3-u}{12}$ for every tied group in the second ranking.

Regression Analysis

Regression of Y on X

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

where:

X, Y = Variables

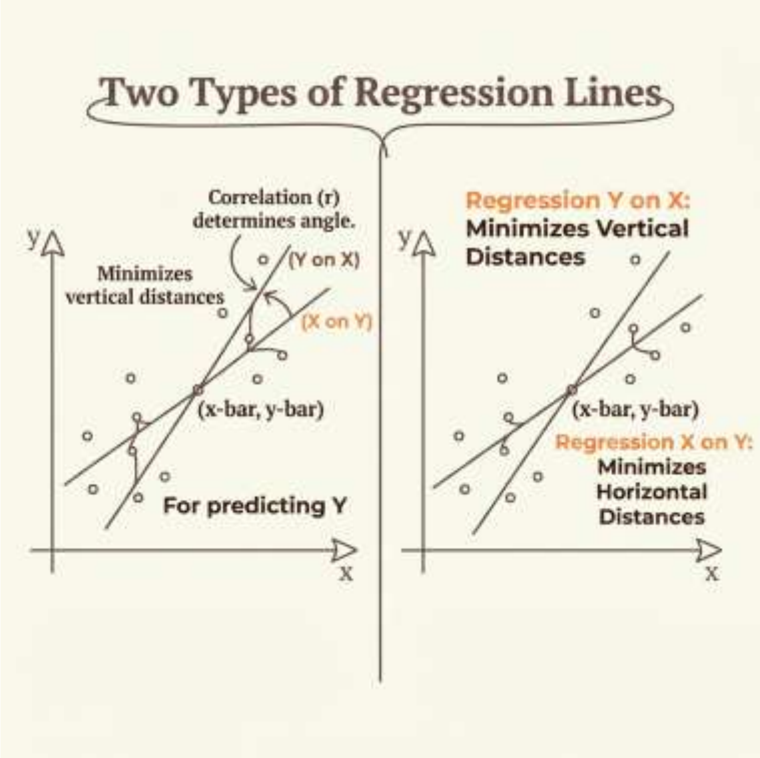
$\bar{X}$ = Mean of X

$\bar{Y}$ = Mean of Y

b_{yx} = Regression coefficient of Y on X

Regression of X on Y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$



SUMMARY

Key Takeaways

Least Squares

Minimize $\sum(y_i - f(x_i))^2$ to derive normal equations for any curve linear in its parameters.

Linear & Quadratic Fitting

Straight lines use 2 normal equations; parabolas use 3. Higher-degree polynomials follow the same pattern.

Correlation vs. Regression

Correlation measures association strength; regression models the predictive relationship between variables.

Rank Correlation

Spearman's ρ handles ordinal data and non-normal distributions where Pearson's r may mislead.

Thank You..