11

Unit 2

Part-A :-

Simple Linear Regression & Correlation

1. Introduction to Linear Regression
2. The Simple Linear Regression Model
3. Least Squares
4. Fitted Model
5. properties of Least Squares estimators
6. Inferences Concerning the Regression Coefficients
7. prediction
8. Simple Linear Regression Case Study.

Unit-2 Part-B

Correlation and Regression:-

Introduction:- We studied in part-A, the Concept of a Single R.V.

i.e → We get An idea about the characteristics of R.V by means of Various statistical averages or Moments like

Mean

Variance

S.d

Skewness

peakedness ....

But in Many Real world problems & many practical problems:- Several Random Variables interact with each other.

Ex:- Doctors may measure many parameters like height, weight, blood pressure, sugar levels ..... etc.

properties:- 1. $0 \le p(x_i, y_j) \le 1$

2. $\sum_i \sum_j p(x_i, y_j) = 1$

3. Joint Cumulative distribution

fn:-

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

$$F_{XY}(2,3) = P(X \le 2, Y \le 3)$$

(1,1) pairs

(1,2)

(1,3)

2,1

2,2

2,3

$$= P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3) + P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3)$$

1

=

What is Multiple R.V:—

Say X & Y are R.V ⇒ 2D R.V

⇒ (X, Y)

The outcome of trials X = x, Y = y

i.e (X, Y) = (x, y).

---- n-dimentional R.V ⇒ n-dimensional Random Vector

(or)

Vector of Random Variables

i.e n-dimensional Eucleadian space.

$$\Rightarrow R^n$$

Say R^2 ⇒ Random Vector (X, Y) i.e Bi Variate

(X, Y) has discrete ⇒ It has only finite number of pairs

↓

Then Joint probability fn ⇒ $f(x,y) = P(X=x, Y=y)$ (PMF)

Joint probability distribution ⇒

P(X=x)@X/Y	1	2	3	x
0	1/2	1/12	1/12	
1	1/12	1/2	1/12	

2

perties:-

1. $0 \le f_{XY}(x,y) \le 1$ @ $0 \le f_{XY} \le 1$

2. $f_{XY} = 1$ @ $f_{XY}(\infty, \infty) = 1$

3. f_{XY} is non decreasing

Marginal probability distribution fn:-

Components of distribution fns are called p.d

$f_X(-\infty, y)$

$f_Y(x, -\infty)$

Joint Probability density fn (CP

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}\, dx\, dy = 1$$

properties:- $f_{XY} \geq 0$

2. $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}\, dx\, dy = 1$$

(3)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx\, dy = 1$$

$Y_i = \beta_0 + \beta_1 x + \epsilon$ — bits — Empirical Model

where e = Random error term.

(2) $E(\epsilon) = 0$

$V(\epsilon) = \sigma^2$

(3) prove $E(\epsilon) = 0$ in Regression ?

W.k.t Regression SLR

$Y = \beta_0 + \beta_1 x + \epsilon$

$E(Y) = E(\beta_0 + \beta_1 x + \epsilon)$

$= E(\beta_0) + E(\beta_1 x) + E(\epsilon)$

$= \beta_0 + \beta_1 x + 0$

$\therefore E(Y) = \beta_0 + \beta_1 x$

(4) Prove $V(\epsilon) = \sigma^2$ in Regression ?

W.k.t SLR $\Rightarrow Y = \beta_0 + \beta_1 x + \epsilon$

Apply Variance

$V(Y) = V(\beta_0 + \beta_1 x + \epsilon)$

$= V(\beta_0) + V(\beta_1 x) + V(\epsilon)$

$= \beta_0 + \beta_1 x + \sigma^2$

$= 0 + \sigma^2$

$V(Y) = \sigma^2$

(a) Y

(1) Regression Analysis (or) Estimation

Introduction:-

→ Many engineering and scientific problems are concerned with determining a relationship between a set of Variables. (especially in Regression)

(or)

✓ Many problems in engineering and Science involves exploring the relationships between two (or) more Variables.

Regression Analysis:- Regression Analysis is a statistical technique that is very useful for these types of problems

→ The Concept of Regression Analysis was 1st introduced by Sir Francis Galton in 19th Century, according to his words

" Regression means "STEPPING BACK TOWARDS AVERAGE "

Now a days Regression mea.

"It is a mathematical measure of aver. relation ship in terms of 2 or more Variables under study."

Meaning:- In Regression Analysis, we can estimate the Value of One Variable with the Value of other Variable, which is known.

Definition:- "The statistical Metted which helps us to estimate the Unknown Value of One Variable from the known Value of other Variable is called Regression."

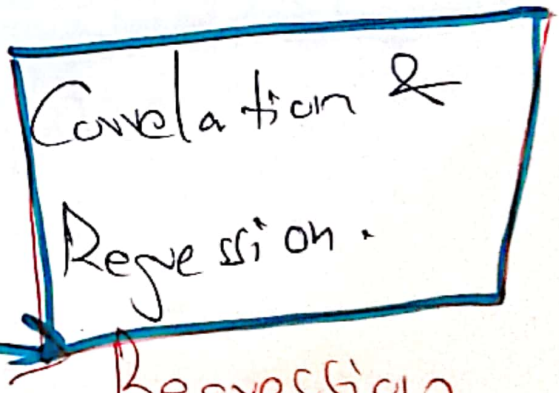
Regression Unit-2 part-A

History & Introduction:-

→ Fitting of Curves to a set of numerical data is of Considerable importance. Hence tical as well as practical.

Theoretically:-

It is useful in study of Curvefitting



Correlation

To measure the strength the linear relationship b/w two R.V say X & Y.

Regression

To describe the relationship b/w the mean of a random Variable Y and One or more other variables.

houses in the same part of the country that are the same square footage of living space will not all be sold for the same price.

Dependant Variables

Responses (y)

Tax Content

Gas mileage (mpg)

price of houses

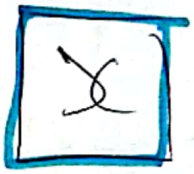
Natural dependant Variables

Single Value (y)

Expected Response.

Independant Variables

Regressors (x)



Inlet temperature

engine Volume

Square feet of living Space

Natural Independant Variables.

Generall, these Variables are Input Variables @ Input Values.

Set of Input Values $(x_1, x_2, \ldots)$

Example:— In Industry, it may be known that "Tar Content" in the outlet stream in a chemical process is related to the inlet temperature.

It may be of interest to develop a method of prediction

ie A procedure for estimating the tar Content for Various levels of the inlet temperature from Experimental information.

Now, of Course, it is highly likely that for many example runs in which the inlet temperature is the same, say 130°C. The out let tar Content will not be the same engine Volume.

They will not all have the same gas mileage.

If the relationship is exact, then it is a deterministic relation between two scientific variables.

ie here there is no random or probabilistic component to it.

But in general, the relationship is not deterministic (ie X does not always give the same value for Y).

But, in real world problems are probalsilistic in nature. So, we cannot expect exact solution.

So, the Concept of Regression Analysis deals with finding the best relationship btw y & x.

But some times multiple regressions also

will be there. Then

$$Y = a + bx_1 + cx_2 + dx_3 + \cdots$$

(or)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$$

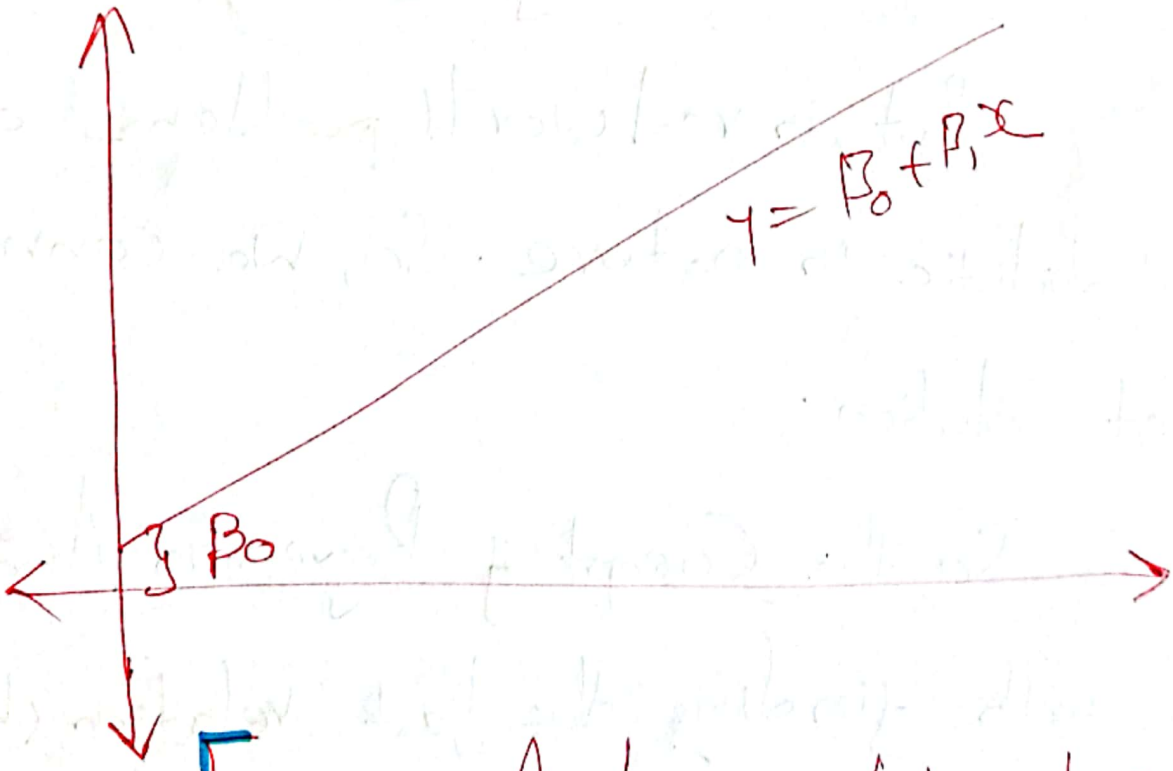


Figure:- A Linear relationship β_0, intercept, β_1: slope.

Uses :-

1. It is used to estimate the relationship b/w economic Variables income & expenditure
2. It is widely used for prediction purpose
3. It is highly valuable in economics & bussiness
4. It is useful in statistical estimation of production function, Cost function, Constant function, - - - etc

Simple Linear Probabilistic Model:

(01) Linear Regression Equation :-

(02) Linear Regression Model :-

(SV) SLR Simple Linear Regression Model

Consider the problem of trying to predict the value of a response Y based on the value of an independent variable x.

Then, the best fitting line $Y = a + bx$.

Where a = intercept

b = slope.

$Y = a + bx_1 + cx_2$

Example:- $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ (or)

(or)

where $Y = \beta_0 + \beta_1 x$ (or) $Y = a + bx$

(or)

$Y = \beta_0 + \beta_1 x + \epsilon$ (or) $Y = a + bx + \epsilon$

where β_0, a = Unknown intercept parameter

β_1 or b = Slope parameter.

ϵ = Random Variable

assumed

$E(\epsilon) = 0$

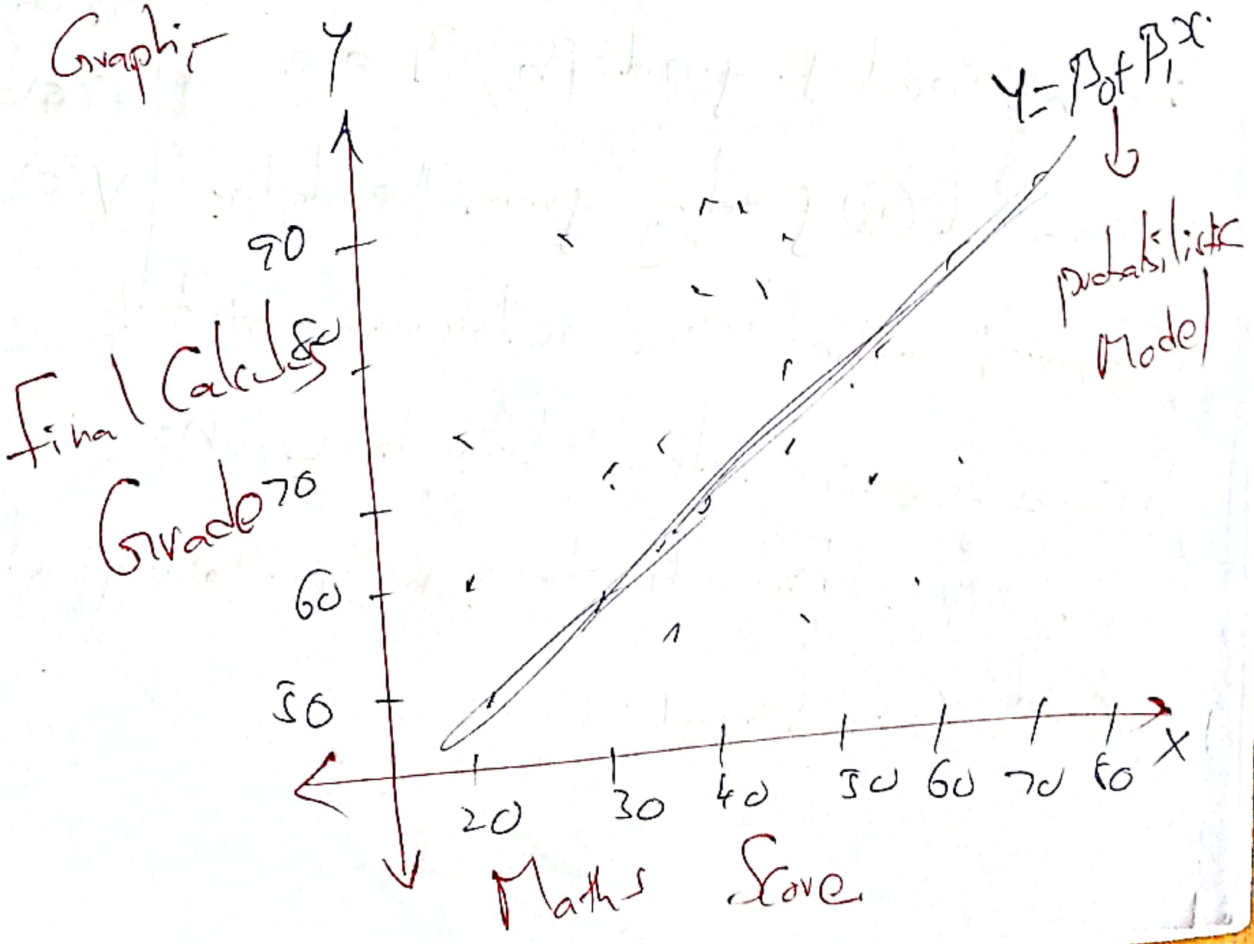
$V(\epsilon) = \sigma^2$

i.e we need to find β_0, β_1 are estimated (or) finding from the data.

Ex:- Mathematics achievement test Scores for a random sample of $n = 5$ in CSE AIML branch along with their final marks in 2nd Year Sem-1.

Now we need to see bivariate.

Student Roll No	Maths achieved	Final Calculus Grade
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92



What is Regression and briefly Explain about Linear Regression Line & Regression & their types?

→ Usually, the primary purpose of a Regression study is prediction.

(estimation, finding the values, approximating - - -)

→ We want to develop an equation by which the value of [Y] can be predicted randomly well,

Based on knowledge of the values assumed by other Variables in equation.

→ Suppose, if the Variables in a bivariate distribution are related. We can find the points in Scatter diagram will

cluster around some Curve called "Curve of Regression"

Small 'x' only (x, x2, x3 ...

Let x = Mathematical Variable

Capital 'Y' Y = Random Variable.

The Curve of regression of Y on X is the graph of mean Value of Y for Various Values of x

$Y = a + bx$ (or) $ax + b$

Linear Regression / Line of Regression :-

If the Curve is a straight line it is called Linear Regression (or) Line of Regression.

Equation

$$Y = a + bx \text{ (or) } Y = ax + b$$

→ First task is attempting to derive a regression equation to decide whether a linear regression curve appears to be appropriate or not.

Ensure that it should be appropriate!

Types of Linear Regression:—
(or)
Methods of Linear Regression:—

There are two approaches to above problem of determining, whether Linear regression is appropriate or not.

1. **Graphic Method:—**
   1. Scatter diagram
2. **Algebraic Method:—**
   2. Method of Least Squares.

1. **Graphic Method:—** Here points representing the pairs of values of the variables are plotted on graph.

Graphic Method :-

(1) Scatter diagram (2) Scatogram

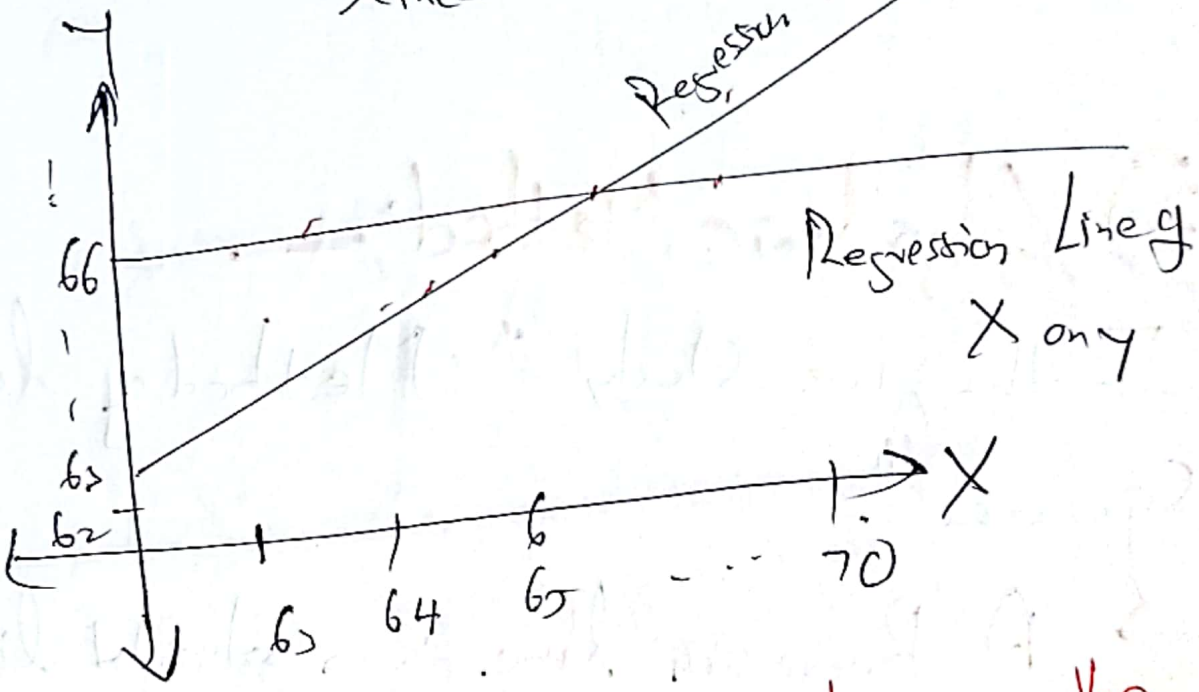
→ This is a picture of data in which the pairs (x, y) are plotted as points in a Coordinate plane.

→ If points tend to form a linear trend, then we use a Simple linear regression method to obtain a prediction equation.

→ Then, our problem is "to estimate the equation for straight line by means of observed data. (2) Theoretical data" finally we will set Expected data (2) Desired data".

i.e Several straight lines can be drawn to include some of data points. & also many lines can be drawn

→ So that points are close to the lines.



Example:— Fit a Regression Line on the Scatter diagram for the following data.

These above points form a Scatter diagram.. A Regression Lines are shown in above graph by free hand..

② Algebraic Method:—

In this, we study "Method of least Squares".

→ A Regression line is a straight line fitted to the data by the method of "least Squares".

→ It indicates the best possible mean value of one Variable Corresponding

(2) Method of least squares:-

to the mean value of the other.

→ There are always two regression lines constructed for the relationship b/w two variables X & Y.

→ Thus,

Thus (SLR) Regression Equations (or) Simple Linear Regression Equations

Y on X

① Straight Line Equations

$Y = a + bx$ (or) $Y = ax + b$

② Second degree parabola :—

$Y = a + bx + cx^2$ (or) $Y = ax^2 + bx + c$

X on Y

$X = a + by$ (or) $X = ay + b$

$x = a + by + cy^2$ (or)

$x = ay^2 + by + c$

③ Regression Coefficient $r = \sqrt{b_{yx} \cdot b_{xy}}$

Y on X

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

$(y - \bar{y}) = \frac{r\sigma_y}{\sigma_x}(x - \bar{x})$

$(y - \bar{y}) = \frac{\Sigma xy}{\Sigma x^2}(x - \bar{x})$

X on Y

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

$(x - \bar{x}) = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$

$(x - \bar{x}) = \frac{\Sigma xy}{\Sigma y^2}(y - \bar{y})$

Principle of Least Squares (2)

Why we are using principle of Least Squares ?

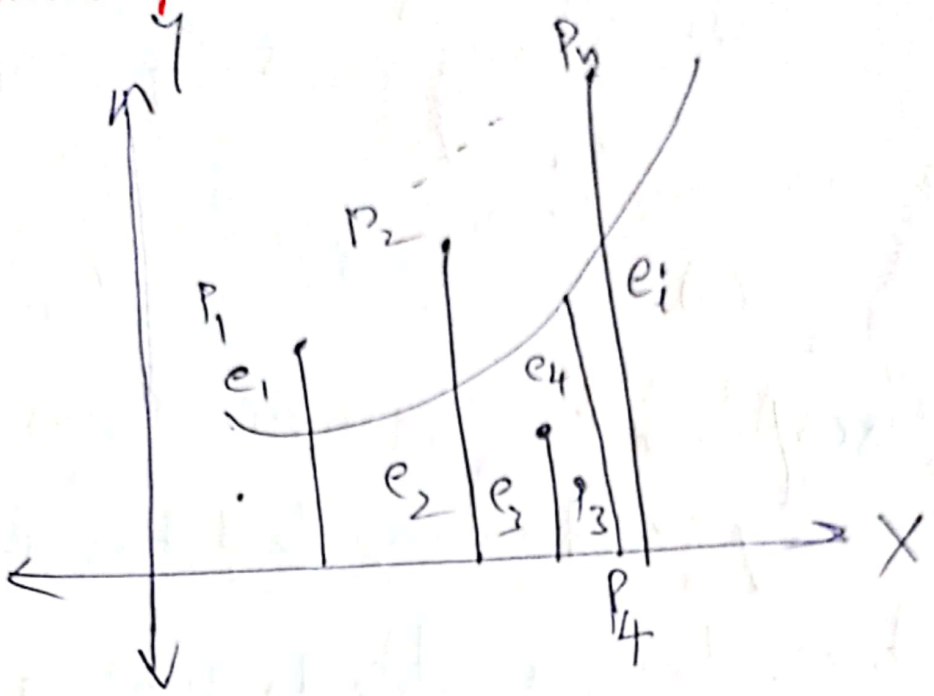
(or)

Disadvantage of Curve of fit (or) Curve fitting?

Legendre in 1806 The graphical method has the obvious drawback of "being unable to give a Unique Curve fit" So, it is necessary to get Unique Curve of fit for any problem. So, we need one platform to get Unique one.

"The principle of least Squares, it provides an elegant procedure for fitting a Unique Curve to a given data."

Explanation:-



Let the Curve $y = a + bx + cx^2 + dx^3 + \dots + kx^n$

n- data points $(x_1, y_1)(x_2, y_2)(x_3, y_3)(y_4, y_4)$ --

Now, our Aim is to determine the Constants

a, b, c, d - - - - .

i.e It represents the Curves of Best fit(

if we found Values a, b, c, d).

→ If $n=m$ in the above equation.

We get Unique set of a-Constants.

→ If more than $n>m$ ⇒ the Cannot Solved for these Constants.

(n: equation, m: Constants)

So, We try to determine these Values of a, b, c ... k, which satisfy all the equations as nearly as possible. So, it gives the best fit.

In these cases, we will apply the principle of Least Squares.

$e_1, e_2, e_3 \ldots$ = Errors (are positive).

We are Considering here, but Some..

times it may -ve also. But, we are only +ve values. So, Thus, to give equal weightage to each error, we square each of these & form their sum.

$$\therefore E = e_1^2 + e_2^2 + \cdots + e_n^2$$

Conclusion:-

→ The curve of best fit is that for which

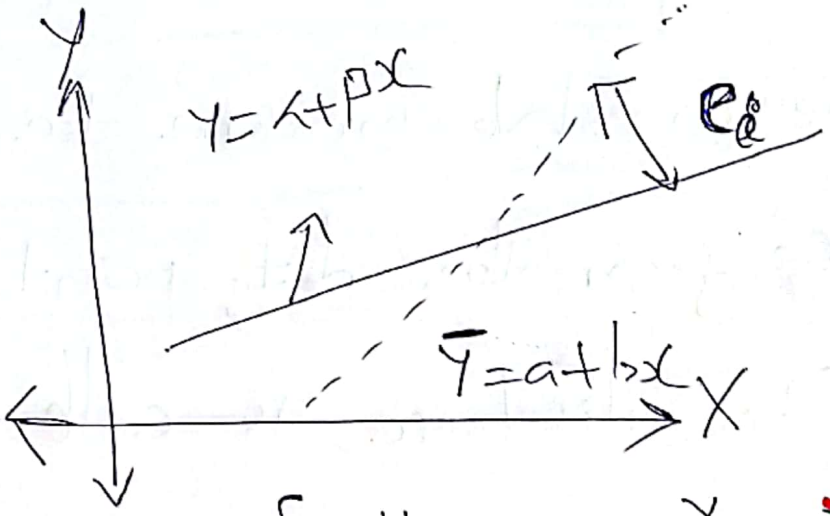
→ es are as small as possible.

→ E, the sum of the squares of the errors is a minimum.

This is called the principle of least squares & suggested by Legendre in 1806.

The reasoning behind the method of least squares is quite simple.

" From among the many straight line drawn through the scatter gram, We pick out the one that is closest to @ best fits the observations. "



Scattergram of X →

@

Scatter diagram of X

where $Y = A + \beta X$ is theoretical Regression line & ideal prediction Curve

$\bar{Y} = a + bx$ is estimated regression line & equation used to make prediction.

Where e_i = Difference b/w estimate Value of Y at x_i,

i.e To determine how close the i^{th} data point is to the estimated line of regression, We measure the Vertical distance from the data point to this line. This distance is called i^{th} residual or i^{th} error

Which is denoted by e_i.

$$Y = a + bx + e$$

$$Y_i = a + b x_i + e_i$$

$$e_i = Y_i - (a + b x_i)$$

e.g. This residual error is +ve

But in general, this is not possible (i.e always e is +ve values). We may have get +ve @ -ve. So, the method of least squares utilizes not the sum of residuals but **Sum of their squares.**

63	47	5
149	150	

st Squa

0·06

+46

+6x

Basic Aim of Least Squares Principle:-

"The technique selects the line through the data that is best in the sense that it minimizes the sum of squares of the residuals."

$$\sum e_i^2 = \sum \left[\left(y_i - (a + b x_i) \right) \right]^2 \text{ as minimum}$$

(or)

We can use Normal Equations.

$$y = ax + b$$

$$\Sigma y = a\Sigma x + nb \quad \text{——— (1)}$$

multiply Σx on both sides

$$\Sigma xy = a\Sigma x^2 + b\Sigma x \quad \text{——— (2)}$$

(1) & (2) are two set of Linear Equations with a & b are two Unknowns.

Solutions of a & b can also find out without solving (1) (2) equations.

$$a = \frac{\Sigma y - b\Sigma x}{n} \quad \& \quad b = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

Disadvantages —

→ The principle of least squares does not help us to determine the form of the appropriate curve, which can fit a given data.

→ It only determines the best possible values of the constants in the equation when the form of the curve is known before hand.

→ The selection of the curve is a matter of experience & practical considerations.

Inferences based on Least Squares Estimator:—

In order to extend point estimates (which are found by least squares), initially to "1" estimate the Common Variance

σ^2

σ^2 is usually estimated in terms of the Vertical deviations of sample points from least squares line.

Sample points = data points

$$S_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(y_i - (a+bx_i)\right)^2$$

or

$$S_e^2 = \frac{1}{n-2} \sum \left(y - (a+bx)\right)^2$$

where S_e = Standard error

$(n-2)S_e^2$ = Residual sum of squares

(or)

Error sum of squares.

Explain Non-Linear Regression Equation

(or)

Explain briefly Curvilinear Regression?

1) Finding the straight line model $Y = a + bx$ is inappropriate, because the true regression Curve is Non-Linear and also called Curvilinear Regression.

Sometimes non linearity is visually determined from scatter diagram. These non Linear Regression Equation can be converted to Linear Regression Equation (SLR). Then, we can find easily for getting the solution.

Non-Linear Model → Transformation (by using log) → Convert to Linear Na

Straight Line

Second degree parabola

We can solve easily

Now.

Non-Linear Model :-

1. $y = a \cdot b^x$

2. $y = a \cdot x^b$

3. $y = \frac{1}{a + bx}$

4. $y = a e^{bx}$

Explain Polynomial Regression?

Briefly & write Normal Equations for Second degree polynomial Regression

If there is no clear indication about functional form of regression of Y on X, we often assume that the underlying relationship is atleast, we believed to the extent that it has a Taylor Series expansion and that the first terms of this expansion will yield a fairly good approximation.

Thus, we fit our data as a polynomial

i.e predicting equation of form.

$$Y = a + bx + cx^2 + dx^3 + \cdots$$

W.r.t Second degree parabola ($y = a + bx + cx^2$)

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

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad \text{---(1)}$$

multiply Σx on B.S

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad \text{---(2)}$$

multiply Σx on B.S

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad \text{---(3)}$$

(1)(2)(3) are called Normal Equations for Second degree Polynomial Regression (Parabola).

Method of Least Squares :-

Procedure :-

To fit the straight line $y = a + bx$

→ Substitute the observed set of x values in this eqn.

→ Write Normal Eqns.

$y = a + bx$

→ Solve these NE as Simultaneous equations for a & b

→ Substitute the values of a & b in $y = a + bx$

→ Which is required Line of best fit

Procedure :-

To fit the parabola $y = a + bx + cx^2$

→ Say

$y = a + bx + cx^2$

→ Same.

(1) Fit a second degree parabola to the following data

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

$a = 1.48$

$b = 1.13$

$c = 0.55$

(2) fit a 2nd degree parabola to the following data

x = 1.0	1.5	2.0	2.5	3.0	3.5	4.0
y = 1.1	1.3	1.6	2.0	2.7	3.4	4.1

$a = 2.07$ $\quad c = 0.061$

$b = 0.511$

5. Fit a 2nd degree parabola

x	1989	1990	1991	1992	1993	1994	1995	1996	1997
y	352	356	357	358	360	361	361	360	359

$X = x - \bar{x}$

$Y = y - \bar{y}$

$Y = -100006.41 + 1034.29X - 0.267X^2$

$a = \frac{694}{231}$

$b = \frac{17}{20}$

$c = -\frac{247}{924}$

H.W problems @ Straight Line

① The following are data on the drying time of a certain paint and the amount of an additive that is intended to reduce the drying time.

Amount of Paint additive x-grams	0	1	2	3	4	5	6	7	8
Drying time (G+S)Y	12	10.5	10	8	7	8	7.5	6.5	9

By using Method of least Squares to predict the drying time of the paint when 6.5 grams of additive is being used

(1) Fit a second degree polynomial (parabola)

$a = 12.2$

$b = -1.4$

$c = 0.143$

final $y = 7.9$ hrs @ $x = 6.5$

(2) Fit a straight Line

(2) The following are data on the number of twists required to break a certain kind of forged alloy bar and percentage of two alloying elements present in metal

Number of twists (Y)	41	49	69	65	40	50	51	57	31
percent of element A (x_1)	1	2	3	4	1	2	3	4	1
percent of element B (x_2)	5	5	5	5	10	10	10	10	15

fit a least squares regression plane and use its equation to estimate the number of twists required to break one of the bars when $x_1 = 2.5$ & $x_2 = 12$?

hint:- $Y = a + bx_1 + cx_2$

$$\Sigma Y = na + b\Sigma x_1 + c\Sigma x_2 \quad \text{---(1)}$$

multiply Σx_1 on B.S

$$\Sigma y x_1 = a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2 \quad \text{---(2)}$$

multiply Σx_2 on B.S

$$\Sigma y x_1 x_2 = a\Sigma x_1 x_2 + b\Sigma x_1^2 \cdot \Sigma x_2 + c\Sigma x_1 x_2^2 \quad \text{---(3)}$$

drop box & solve it by

$x_1 = 9.5$

$x_2 = 12$

(4)

(3) Fit a straight line for the following data. Using (a) $y = a + bx$

(b) $y = ax + b$

(c) $x = a + by$

(d) $x = ay + b$

(e) $y = a + bx + cx^2$

(f) $y = ax^2 + bx + c$

x	53	98	95	81	75	61	59	55
y	47	25	32	37	30	40	39	45

④ In the accompanying table x is the tensile force applied to a steel specimen in thousand of pounds and y is the resulting elongation of an inch.

x	1	2	3	4	5	6
y	14	33	40	63	76	85

Find the equation of least square line and use it to predict the elongation when the tensile force is 3.5 of thousands pounds?

hint:- $y = a + bx$

$b = 14.69$ $a = 1.12$

$x = 3.5$ @ $y = 51.63$

(5) Given the following data

x	10	30	40	60	80	90	110	140
y	10	20	40	40	50	70	80	90

Compute least square Regression Line?

~~hint~~ Here not specific equation

on y on x & x on y.

→ At this case use y on x

$y = a + bx$

$n = 8$, $a = 38.66$ $b = 0.162$

$$y = 38.66 + (0.162)x$$

x	10	30	40	60	80	90	110	140

6. The following table shows the ages X and systolic blood pressures Y of 12 women;

Age(x)	56	42	72	36	63	47	55	49	38	42	68
Blood Pressure(y)	147	125	160	118	149	150	145	115	140	150	160

Determine the least square regression equ of Y on X?

Hint;

$Y = a + bx$

$a = 150.06$

$b = 5.546$

$Y = 150.062 + 5.546x$

Problems

(1) A panel of two judges P & Q graded seven dress performances by independently awarding marks as follows:

Performance	1	2	3	4	5	6	7
Marks by P	46	42	44	40	43	41	45
Marks by Q	40	38	36	35	39	37	41

The eight performance, which judge Q would not attend, was awarded 37 Marks by judge P. If judge Q had also been present. How many marks would be expected to have been awarded by him to the eight performance?

P judge = 37 marks given but 8th performer not attended Q is attend Then having Q will be given

These type problems has been solved by using Regression Co-efficients.

Performance	Marks by P x	$X = x - \bar{x}$	X^2	Marks by Q y	$Y = y - \bar{y}$	XY
1	46	+3	9	46	+2	6
2	42	−1	1	38	0	0
3	44	+1	1	36	−2	2
4	46	−3	1	35	−3	9
5	43	0	0	39	+1	0
6	41	−2	4	37	−1	2
7	45	+2	4	41	+3	6
	30[illegible]		25	266		21

$\bar{x} = \frac{\Sigma x}{n} = \frac{301}{7} = 43$, $\bar{y} = \frac{\Sigma y}{n} = \frac{266}{7} = 38$.

$\therefore \bar{x}, \bar{y}$ are not fractions.

Scanned with CamScanner

Subject: Probability & Statistics

Class Notes

Faculty: Dr.K.Kavitha

Topic:

Unit No:
Lecture No:
Link to Session
Planner (SP): S.No....of SP
Date Conducted:
Page No:

Now we have 2 R-E's so, If we see the data we want to calculate Q i.e (Y).

∴ Regression Equation Y on X.

$$Y - \bar{Y} = b_{YX}(X - \bar{X})$$

where $b_{YX} = \frac{\Sigma XY}{\Sigma X^2} = \frac{21}{28} = 0.75$

$$Y - 38 = 0.75(X - 43)$$

$$Y = 0.89X + 37.36$$

But by data $X = 37$ by D

$$\therefore -Y = 0.89(37) + 37.36$$

$$\boxed{Y = 33.5}$$

Conclusion:- If Q would have been present, he would have awarded 33.5 marks to the eight persons.

(27) In the following table S is weight of Potassium bromide, which will dissolve in 100grams of water at V°C. Fit an equation of the form $S = mT + b$. By the method of Least Squares.

Use this relation to estimate S, when T = 50°C ?

T	0	20	40	60	80
S	54	65	75	85	96

$y = ax + b$ $\quad S = mT + b \Rightarrow$ It's a straight Line.

by $T = 50^{\circ}C$ seeing this we'll have Curtate

or by data T, S

N.E:- $\Sigma S = m\Sigma T + nb$

Multiply ΣT

$\Sigma TS = m\Sigma T^2 + b\Sigma T$ — (1)

T	S	TS	T^2	TS
0	54			
20	65			
40	75			
60	85			
80	96			
200	375		12000	17050

$\bar{T} = \frac{\Sigma T}{n}$ ...

$\bar{S} = \frac{\Sigma S}{n} =$

Substitute These Values in (1).

We will get. $m = 0.52$, $b = 54.2$

But $T = 50^{\circ}$ given by data i.e. $S = 0.52 \times 50^{\circ} + 54.2$

$S = 80.2$

Subject: Probability & Statistics **Class Notes**

Faculty: Dr.K.Kavitha

Topic:

Unit No:
Lecture No:
Link to Sessi
Planner (SP):
Date Conduct
Page No:

(3) Find the most likely production corresponding to a Rainfall 40 from the following data:-

	Rainfall (X)	production Y
Mean = Average	$\bar{x}=30$	$\bar{y}=500$ kgs
σ Standard Deviation	$\sigma_x=5$	$\sigma_y=100$ kgs
r = Coefficient of Correlation	$r=0.8$	-

i.e X = Rainfall = 40 i.e we have to find Y.

i.e R.E Y on X $= (y-\bar{y}) = b_{yx}(x-\bar{x})$

Where $b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$ (or) $r \cdot \frac{\sigma_y}{\sigma_x}$

data related to 2nd one formula

$\therefore$ Data:- $\bar{x}=30$, $\bar{y}=500$

$\sigma_x=5$ $\sigma_y=100$

$(y-500) = 0.8 \cdot \frac{100}{5}(x-30)$

$(y-500) = \frac{4}{100}(40-30)$

$\therefore y = 500.4$

$b_{yx} = 16$

$y = 16x + 20$

When $x=40$ i.e

$y = 660$

(4) From a sample of 200 pairs of observation following quantities were calculated.

$\Sigma X = 11.34$, $\Sigma Y = 20.78$, $\Sigma X^2 = 12.16$,

$\Sigma Y^2 = 84.96$, $\Sigma XY = 22.13$

From the above data. Show how to compute the coefficients of the equation $Y = a + bX$?

$$Y = a + bX$$

N.E.I $\quad \Sigma Y = na + b\Sigma X$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

Substituting above data,

$$20.78 = 200a + b \times 11.34$$
$$22.13 = a\,11.34 + b\,12.16$$

} solving

$$a = 0.0005$$
$$b = 1.82$$

$$\therefore Y = 0.0005 + 1.82X.$$

Topic:

Lecture No:
Link to Session
Planner (SP): S.No....of SP
Date Conducted:
Page No:

(5) Criticise the following

Regression Coefficient of Y on X is 0.7 & X on Y is 1.2

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{0.7 \times 1.2} = \sqrt{2.24}$$

$$= 1.57$$

But Correlation Coeffi can not exceed 1

∴ there is some inconsistency in the information given.

(6)

price	10	12	13	12	16	15
Demand	40	38	43	45	37	43

Calculate Regression Equations of Y on X from the data given below ~~[illegible]~~ from actual Means of X & Y. Estimate the Likely demand when price = 20?

X - price = 20 ∴ we have to find Y.

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

$\Sigma x = 78$

$b_{yx} = \frac{\Sigma xy}{\quad}$ find out we will substitute $x = 20$

Scanned with CamScanner

① Price indices of Cotton and Wool are given for the 12 Months of a year. Obtain the eqn of Lines of Regression b/w the Indices?

Two lines of regression

X (price index of cotton)	78	77	85	88	87	82	81	77	76	83	97	93
Y (price index of wool)	84	82	82	85	89	90	88	92	83	89	96	89

X	Y	$dx = x - \bar{x}$	$dy = y - \bar{y}$	dx^2	dy^2	$dx\,dy$
78	84					
77	82					
85	82					
88	85					

$\bar{x} = 83.9$

$\bar{y} = 87.4$

$r = 0.1$

(1) Price indices of Cotton and Wool are given for the 12 Months of a year. Obtain the eqn of Lines of Regression b/w the Indices?

X (price Index of Cotton)	78	77	85	88	87	82	81	77	76	83	97	93
Y (price index of wool)	84	82	82	85	89	90	88	92	83	89	98	95

X	Y	$dx = x - \bar{x}$	$dy = y - \bar{y}$	dx^2	dy^2	$dxdy$
78	84					

$\bar{x} = 83.9$

$\bar{y} = 87.4$

Subject: Pr
Faculty: Dr.K.
Topic:

Scanned with CamScanner

Angle b/w two Regression Lines:-

$$\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

$\longrightarrow$ θ = Acute Angle:- $\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_y}{\sigma_x}$

$r = -1$ to $+1$

$\rightarrow$ θ = Obtuse Angle:- $\tan\theta = \frac{k^2-1}{r} \cdot \frac{\sigma_y}{\sigma_x}$

$\Rightarrow$ $r=0$:- $\theta = \pi/2$. No-relation b/w x & y.

ie They are Independent.

$\Rightarrow r = \pm 1 \Rightarrow \theta = 0$ (or) π . — Dependent

Two Regression Lines are parallel (or) coincide

ie Correlation b/w Key Variables are perfect.

Problems

(1) If θ is the angle b/w two Regression Lines & S.d. of Y is twice the S.d of X, $r = 0.25$. Find $\tan\theta$?

Sol: $\sigma_y = 2\sigma_x$

$r = 0.25$

W.k.t $\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$

$= \frac{1-r^2}{r} \cdot \frac{\sigma_x \cdot 2\sigma_x}{\sigma_x^2 + (2\sigma_x)^2}$

$= \frac{1-r^2}{r} \cdot \frac{2\sigma_x^2}{\sigma_x^2 + 4\sigma_x^2}$

$= \frac{1-r^2}{r} \cdot \frac{2\sigma_x^2}{\sigma_x^2(1+4)}$

$= \frac{1-r^2}{r} \cdot \frac{2}{5}$

$= \frac{1-(0.25)^2}{(0.25)} \cdot \frac{2}{5}$

$= 1.5$

(2) If $\sigma_x = \sigma_y = \sigma$. The angle b/w two Regression Lines is $\tan^{-1}\left(\frac{4}{3}\right)$?

$\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$

$\sigma_x = \sigma_y = \sigma$

$$\tan\theta = \left[\frac{1-r^2}{r} \cdot \frac{\sigma}{2\sigma}\right]$$

$$\tan\theta = \frac{1-r^2}{2r}$$

$$\theta = \tan^{-1}\left(\frac{1-r^2}{2r}\right) \quad \text{——— (1)}$$

But $\det\theta = \tan^{-1}\left(\frac{4}{3}\right)$ — (2)

(1) = (2)

$$\tan^{-1}\left(\frac{1-r^2}{2r}\right) = \tan^{-1}\left(\frac{4}{3}\right)$$

$$\frac{1-r^2}{2r} = \frac{4}{3}$$

$$3 - 3r^2 = 8r$$

$$3r^2 + 8r - 3 = 0$$

$3r^2 + 8r - 3 = 0$

$r = \frac{1}{3}$ (or) $r = -3$

$$r = \frac{1}{3} \text{ (or) } r = -3$$

But $r = -3$ ✗ not possible since $r = -1$ to $+1$ limit.

$$\therefore r = \frac{1}{3}$$

3) The tangent of angle b/w two Regression Lines is 0.6 & if $\sigma_x = \frac{1}{2}\sigma_y$. Find Correlation Coefficient b/w x & y?

$\sigma_x = \frac{\sigma_y}{2}$

$r = \frac{1}{2}$ (or) $r = -2$ ✗

Finding Mean Values & r :- by using two Regression Equations :-

① Find the Mean Values of Variable x & y & Coeffs Coefficient from the following Regression Equations

$$\left.\begin{aligned} 2y - x - 50 = 0 \\ 3y - 2x - 10 = 0 \end{aligned}\right\} \text{---------- (1)}$$

Sol :-

$\bar{x} = 130, \bar{y} = 90$

r :-

From (1) :-

$2y - x - 50 = 0$

$2y = x + 50$

$y = \frac{x}{2} + 25$ ---------- (2) This is Yon X Reg

$3y - 2x - 10 = 0$

$2x = 3y - 10$

$x = \frac{3y}{2} - 5$ ---------- (3) This is X on Y Reg

$\therefore b_{yx} = \frac{1}{2}, \quad b_{xy} = \frac{3}{2}$

Subject: Probability & Statistics

Class Notes

Faculty: Dr.K.Kavitha

Topic:

Unit No:
Lecture No:
Link to Session
Planner (SP): S.No....of SP
Date Conducted:
Page No:

$$r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{\frac{3}{2} \cdot \frac{1}{2}} = \sqrt{\frac{3}{4}}$$

$$\therefore r = 0.866$$

Test whether the equations $2x+3y=4$ & $x-y=5$ represent Valid Regression Lines?

$r^2 = -3/2$ r^2 = -ve so, the given equations don't represent Valid Regression Lines

For 20 army personal Regression of weight of kidneys (y) on weights of heart X is $Y = 1.99X + 6.394$ & the Regression of weights of heart on weight of kidneys is $X = 1.212Y + 2.461$. find the correlation Coefficient between the two Variables and also their Means?

$\bar{x} = 197720.03$

$\bar{y} = 78890.28$

$r = 0.695$

3) The equation of two R. Lines are $7x-16y+9=0$ & $5y-4x-3=0$ find the Coefficient of Corelation & the Means of x & y?

$\bar{x} = 0.1034, \bar{y} = 0.5172$ $\quad r = \frac{r}{\sqrt{35}}$

31-9 570 37-6 ಪ್ರಶ್ನೆಗಳು

32-6 = part-C

Properties of Regression Coefficients:-

(1) The Co-relation Coefficient r is the geometric mean b/w the two regression Coefficients.

$$r^2 = b_{yx} \cdot b_{xy}$$

(or)

$$r^2 = r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y}$$

(or)

$$r = \sqrt{b_{yx} \cdot b_{xy}} \quad \text{(or)} \quad r = \sqrt{r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y}}$$

Where b_{yx} (or) $r\frac{\sigma_y}{\sigma_x}$ = The regression Coefficient of Y on X

b_{xy} (or) $r\frac{\sigma_x}{\sigma_y}$ = The regression Coefficient of X on Y.

(2) $r = \frac{\Sigma xy}{\Sigma x^2}$ for Y on X

$r = \frac{\Sigma xy}{\Sigma y^2}$ for X on Y.

(3) If θ is the angle b/w two regression lines. Then $\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$

→ $r = 0$, $\tan\theta \to \infty \left(\theta = \pi/2\right)$. Then two lines of regression are $\perp$ to each other

→ $r = \pm 1$, $\tan\theta = 0 \left(\theta = 0 \text{ or } \pi\right)$. Then, two lines of regression are coincide.

ie Here, there is a perfect correlation b/w the two Variables.

(4) $$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

(5) $$r = \frac{\Sigma XY}{n\sigma_x\sigma_y}$$ & $$r = \frac{\Sigma(x-\bar{x})\Sigma(y-\bar{y})}{n\sigma_x\sigma_y}$$

$\bar{x}$ & $\bar{y}$ ave not-fractions

$\bar{x}$ & $\bar{y}$ ave fractions.

(6) $$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$$

(7) If one of the regression coefficients is greater than Unity, the other must be less than Unity.

$$b_{yx} \cdot b_{xy} = r^2 < 1$$

$$b_{xy} \leq \frac{1}{b_{yx}} < 1$$

i.e $b_{xy} > 1$ & $b_{yx} < 1$

(8) $r^2 \leq 1$ & $\sigma_x \sigma_y$ are +ve.

(9) +ve sign gives the acute angle to the lines.

Least Squares Notations in Regression

$$S_{XX} = \Sigma(X-\bar{x})\,\Sigma(X-\bar{x})$$

$$S_{XY} = \Sigma(X-\bar{x})\cdot\Sigma(Y-\bar{y})$$

$$S_{YY} = \Sigma(Y-\bar{y})\cdot\Sigma(Y-\bar{y})$$

$$SS_R = \frac{S_{XX}\,S_{YY} - S_{XY}^2}{S_{XX}}$$

where SS_R = Sum of Squares of the residuals.

Total SS = SSR + SSE

where SSR = Sum of Squares for regression

SSE = Sum of Squares for error.

r^2 Values :- (or) Co-efficient Correlation Values :-

(Q) Correlation Coefficient Values :-

1. $0 < r < 0.5$ → Low degree +vely Correlated

   $0 < r < 0.5$ → " " −vely "

2. $0.5 < r < 0.75$ → Moderately +vely "

   $-0.5 < r < -0.75$ → " −vely "

3. $0.75 < r < 1$ → high degree +vely "

   $-0.75 < r < -1$ → high degree −vely "

4. $0.9 < r < 1$ → Very high degree +vely "

   $-0.9 < r - 1$ → " −vely "

$r = +1$ → perfectly +ve

$r = -1$ → perfectly −ve

$r = 0$ → Un Correlated

Props

Properties of Least Squares Estimators :-

1. The least squares estimators for a straight line $y = a + bx$ are $\boxed{a \& b}$.

2. $Y = a + bx + e$ is a R.V with $E(X) = 0$ & $V(X) = \sigma^2$.

3. The values of a & b depend on the observed y's.

4. The least squares estimators of the Regression Coefficients may be Viewed as Random Variables

Inferences Concerning the Regression Coefficients :-

→ We know that, for estimating the linear relationship b/w x & Y for purposes of testing applications & prediction.

→ The experiments may also be interested in drawing certain inferences about the slope and intercept.

→ In order to allow for the testing of hypothesis and the construction of confidence interval on a & b.

→ Now, here we need to assume that e_i is normally distributed

$$Y = a + bx + e_i$$

e_i = residual error $i = 1, 2 - - - n$

→ In this, Interval Confidence Regression Coefficients, we use χ^2 test. & also t-test

---

t test

(Test statistic)

$$T = \frac{b - \beta}{\frac{s}{\sqrt{\Sigma(x-\bar{x})}}} \quad \text{(or)} \quad \frac{b-\beta}{\frac{s}{\sqrt{S_{xx}}}}$$

&

Confidence Interval :-

$$\left(\beta - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{S_{xx}}}\right) < b < \left(\beta + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{S_{xx}}}\right)$$

where $t_{\frac{\alpha}{2}}$ = t- distribution with $(n-2)$ degrees of freedom.

Simple Linear Regression SM

Cases Study:—

Write briefly Simple Linear Regression Case Study? (SLR – Case Study):—

→ In the manufacture of Commercial wood products, it is important to estimate relationship b/w the density of a wood product and its stiffness.

→ A relation near type of particleboard is being considered that can be formed with Considerations, more ease than the accepted Commercial product.

→ It is necessary to know at value density the stiffness is Comparable to that of the well-known, well documented product (Commercial product)

→ A study was done by Terrance ↗ E. Conners, they investigate of certain a Mechanical properties of a wood foam: Composite.

→ This particle boards were produced at densities ranging from roughly 8 to 26 pounds & the stiffness was measured in pounds per Square inch.

→ Now, it is necessary for the data analyst to focus on an appropriate fit to the data & use inferential methods discussed in this.

→ Hypothesis testing on the slope of the regression, as well as Confidence (CI) prediction interval estimation may well be appropriate

→ Now, we will begin by demonstrating a simple scatter plot of the data with a simple linear regression superimposed.

For example,- $y = -25433.739 + 31849x$

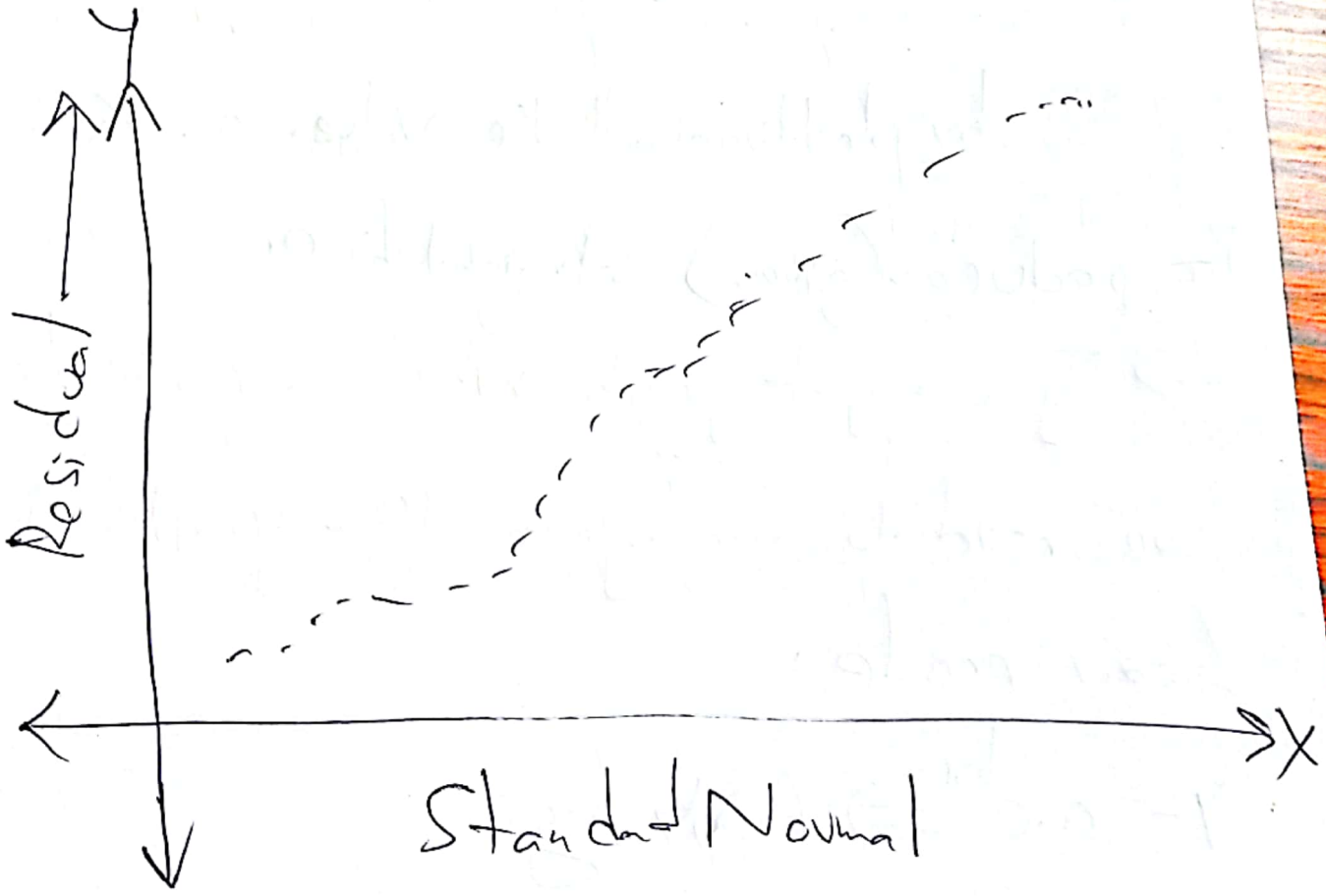


Figure:- Normal probability plot of residuals for wood density data.

X-axis :- Which is horizontal axis to represents the empirical normal distribution

for

Y-axis :- which is Vertical axis represents the Residuals.

by plotting all the values on X & Y axis it produces (gives) Straight Line.

→ If we get graph which is non-Linear we need to transform the graphs into Linear mode.

$y = a \cdot e^{bx}$ ⟹ Apply log

$\Rightarrow \log y = \log a + bx \log e$

$\boxed{\Rightarrow Y = A + bx}$ Then we have

to apply normal procedure of Regression equations (SLR).

Case-study-2 :-

Case-study-3 :-

} We can write from Real time applications.

This is Left for HW.

Fitted Model :—

Explain fitted Model in Regression?

→ write about SLR (which is Simple Linear Regression Equation)

+

and also write Least Squares Methods stating that ($y = a + bx + \epsilon$) ve sidual error also.

+

→ Write about few lines of Curvilinear Regression.

~~$y = a + bx$~~

→ $y = a \cdot e^{bx}$, $y = a \cdot b^{x}$ -- all these Non-linear equations is to Linear Regression Equations by transforming (i.e Apply [log] for

Curvilinear Equations:

Prediction :-

There are several reasons for building a linear regression. One, of course is predict response values at one or more values of the independent variable. In this section, the focus is on errors associated with prediction.

The eqn $Y = \alpha + \beta x$, may be need to predict or estimate the mean response or it may be used to predict a single value. We would expect the error of prediction to be higher in the case of a single predict value than in the case where mean is predicted. Then, we can use here confidence interval also by applying t- distribution.

15M

Unit-3

15M

Part-A

(or)

Part-B

Continuous Probability Distributions

1. Normal distributions
2. Areas under the Normal Curve
3. Applications of N.D
4. Normal Approximation to B.D

Fundamental Sampling distributions

1. Random Sampling
2. Sampling distributions
3. Sampling distributions of Means
4. Central Limit theorem
5. Sampling distributions of S^2
6. t- Distributions
7. F- distribution.

of X on Y & Y on X.

Regression Equation :- (or) Equation of a Straight Line :-

Regression Equation is an Algebra expression of the Regression Line. It can be classified into Regression Equation, Regression Coeff Individual Observation & Group Discussion

Standard form of Regression Equation

$$Y = a + bX.$$

Where a, b = Constants

Subject: Probabil
Faculty: Dr.K.Kavitha
Topic:

a :- Value of y, when $x=0$ - It is called y - Intercept.

b :- Value of slope of Regression Lines called Regression Coefficient y on x.

To find a & b :- With help of Normal Equations

(1) Regression Equation of y on x :-

$y = a + bx$

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

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

(2) R.E of x on y :-

$x = a + by$

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

$\Sigma xy = a\Sigma y + b\Sigma y^2$

Problems

(1) Determine the Equation of a straight Line which best fits the data. (i.e Regression Equation.)

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

St. Line $y = a + bx$

Normal Equations $\Sigma y = b\Sigma x + na$

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

Faculty: Dr.K.Kavitha

Topic:

Lecture No:
Link to Session
Planner (SP): S.No....of SP
Date Conducted:
Page No:

What is Meant by Curve fitting? Explain the principle of Least Squares on Legendre principle of Least Squares?

Let (x_i, y_j) where $i = 1$ to n
$j = 1$ to n.

Given set of n-pairs of values.

x = Independent Variable
y = Dependent Variable.

Analytic expression of the form $y = f(x)$. It is useful in study of Correlation & Regression. In practice statistics enables us to represent the relationship b/w two Variables by Simple Algebric expressions

Ex:- polynomial & exponential functions
- - - etc.

It is used to estimate the value of first Variable which would Correspond to specified Values of the Variable.

The Legendre principle of Least Square consists of minimising the sum of square deviation of the actual values from the estimated values are given by the difference of fitting curve of line.

Non Linear Regression & Curvilinear Regression

In this Method, we have to use Method of Least Square to fit a straight line

1. $y = a \cdot b^x$ — power curve. $\quad y = a \cdot e^{bx}$ = use log base e
2. $y = a \cdot e^{bx}$ — exponential curve $\quad y = a \cdot b^x$ = use log base 10.
3. $y = a x^b$ — power curve
4. $y = a + bx + cx^2$ — Second degree parabola

---

(1) $y = a \cdot b^x$ —

apply log

$\log y = \log a + x \log b$

$Y = A + xB$

$\sum Y = nA + b\sum x$

x	y	$Y = \log y$	x^2	xY

(2) $y = a e^{bx}$

$\log y = \log a + bx$

$Y = A + bx$

$\sum Y = nA + b\sum x$

$\sum xY = A\sum x + b\sum x^2$

x	y	$Y = \log y$	xY	x^2

(3) $y = a \cdot x^b$

$\log y = \log a + b \log x$

$Y = A + bX$

X	Y	$X = \log x$	$Y = \log y$	X^2

(4) $y = a + bx + cx^2$

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

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

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

x	y	x^2	x^3	x^4	xy	x^2y

$y = a + bx$

$\therefore y = a + bx$

$= 0.82 + 1.56x$

(3) The following data pertain to the number of jobs per day and the actual processes unit time required. Fit a straight line. Estimate the mean CPU time at $x = 3.5$ (2) fit a straight line by method of least squares.

No of jobs	1	2	3	4	5
CPU time	2	5	4	9	10

(1) x and y

$y = a + bx$

$30 = 5a + 15b$

$110 = 15a + 55b$

$a = 0, b = 2$

$\therefore y = 0 + 2x \Rightarrow y = 2x$

$x = 3.5 \Rightarrow y = 2(3.5)$

$y = 7$

(4) fit a least square regression equation of X_3 on X_1 & X_2?

$\Sigma X_3 = a + bX_1 + cX_2 \Rightarrow$ Normal equation write

$\Sigma X_1 X_3 =$ Mult X_1 & X_2

$\Sigma X_2 X_3 =$

X_1	X_2	X_3	X_1^2	X_2^2	X_1X_2	X_1X_3	X_2X_3

substitution & solve

$a = 61.40, b = -3.65$

$c = 2.54$

$\therefore X_3 = 61.40 - 3.65X_1 + 2.54X_2$

2014

1.

Quartile Deviation : Mean Deviation : Standard Deviation = 10 : 12 : 15

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{2}{3}\sigma$$

$$\text{Mean Deviation} = \frac{4}{5}\sigma$$

2. Define probable error :— (1)

" It is such that the probability of an error falling within the limits $(\mu - \lambda$ & $\mu + \lambda)$ is exactly equal to the chance of an error falling outside these limits.

i.e. chance of an error lying within $\mu - \lambda$ & $\mu + \lambda$ $= \frac{1}{2}$

Example :- In a Industry, lot of articles manufactured to Certain specifications is subject to small errors. In fact, measurement of any physical quantity shows slight error.

1. Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$

2. Kurtosis :-

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{(\sigma^2)^2} = 3$$

3. Recurrence relation for Moments of N-D :-

$$\mu_{2n} = \sigma^2 (2n-1)\, \mu_{2n-2}$$

all odd moments = Varishes

all Even moments = Exists.

4. Moments about Origin	Moments about Mean
1. 1^{st} Mont = Mean = $\mu_1' = b$	1. Mean = $\mu_1 = b$ (2) Variance = σ^2

Table of functions

3 types of R.V = X = a function.

towards we have to Consider R.V = function

P.D.f	P.m.f	P.d.f
distribution function	Probability mass function	probability density function
Let X be a R.V. Then function F is defined for all x $F(x) = P(X \le x) = \int_{-\infty}^{+\infty} f(x)\,dx$ is called P.D.F.	Let X = discrete RV with distinct values $x_1, x_2 \dots x_n$. Then function p(x) is $P(x) = \begin{cases} P(X=x), & \text{if } x = x_i \\ 0, & \text{if } x \neq x_i \end{cases}$ $i = 1, 2, \dots$ is called p.m.f.	Let X = Continuous R.V with p.d.f f(x). Then the function The pdf is the derivative of the probability distribution function.
P.D.F Properties:-	**P.m.f. properties:-**	**P.d.f properties:-**
1. If F is d.f of 1D R.V is X, Then ① $0 \le F(x) \le 1$ ② $F(x_1) \le F(y)$ * or All distribution functions are monotonically non-decreasing & lie b/w 0 & 1.	1. $p(x_i) \ge 0$ $\forall i$ 2. $\sum p(x) = 1$ or 3. The set of values, which X takes is	1. $f(x) \ge 0$ 2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

Explain Mean, Median & Mode?

Mean:- If X is a Continuous R.V with probability $f(x)$. Then

$$M = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx = \int_{a}^{b} x \cdot f(x)\, dx$$

Median:- Median is the point, which devides the entire distribution into two equal parts. Then

$$\int_{a}^{M} f(x)\, dx = \int_{M}^{b} f(x)\, dx = \frac{1}{2}$$

Mode:- Mode is the value of X for which $F(x)$ is Maximum. Mode is given by

$$f'(x) = 0$$
$$f''(x) < 0 \quad \text{for } a < X < b$$