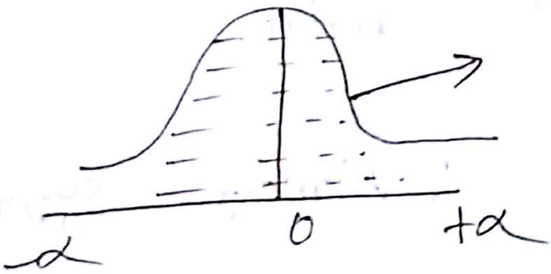
Final

Skawness

1

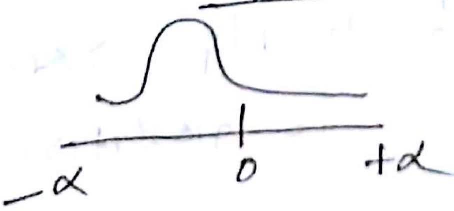


Symmetric distribubion

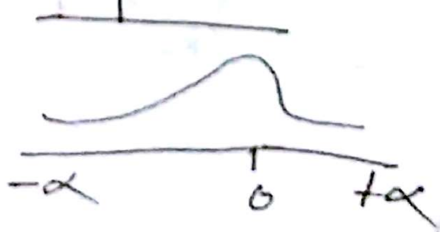
Equal "

skenied distuibution

Postively



Negatively



Coefficient of skewness

$$S_K = \frac{3(\text{mean} - \text{median})}{S.D}$$

$S_K (-3 \text{ to } +3)$

2

if,

1. $S_K = 0$; symmetric distn.
2. $S_K = (-0.5 \text{ to } 0.5)$; Almost symmetric
3. $S_K > (0.5 \text{ to } 1)$; Moderately skewed which is positive.
4. $S_K (-1 \text{ to } < -0.5)$; ~~which i~~ moderately skawed which is Negative.
5. $S_K < -1$; Highly skewed which is negative.

Scanned by CamScanner

6 $S_K > 1$; Highly skowed which is positive.

* 10, 2, 11, 20, 1

3

Step 1: 1, 2, 10, 11, 20.

Calculated coeff of skewence and comment interprete on your result. or Comment on the shape of the distribution

Med: 10

mean: 8.8

$$S.D = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{(10-8.8)^2+(2-8.8)^2+(11-8.8)^2+(20-8.8)^2 \quad (1-8.8)^2}{5-1}}$$

$$= \sqrt{\frac{1.44 + 46.24 + 4.84 + 125.44 + 60.84}{4}}$$

$$= \sqrt{59.7} = 7.73.$$

$$S_k = \frac{3(\text{mean} - \text{median})}{S.D}$$

$$= \frac{3(8.8 - 10)}{7.73}$$

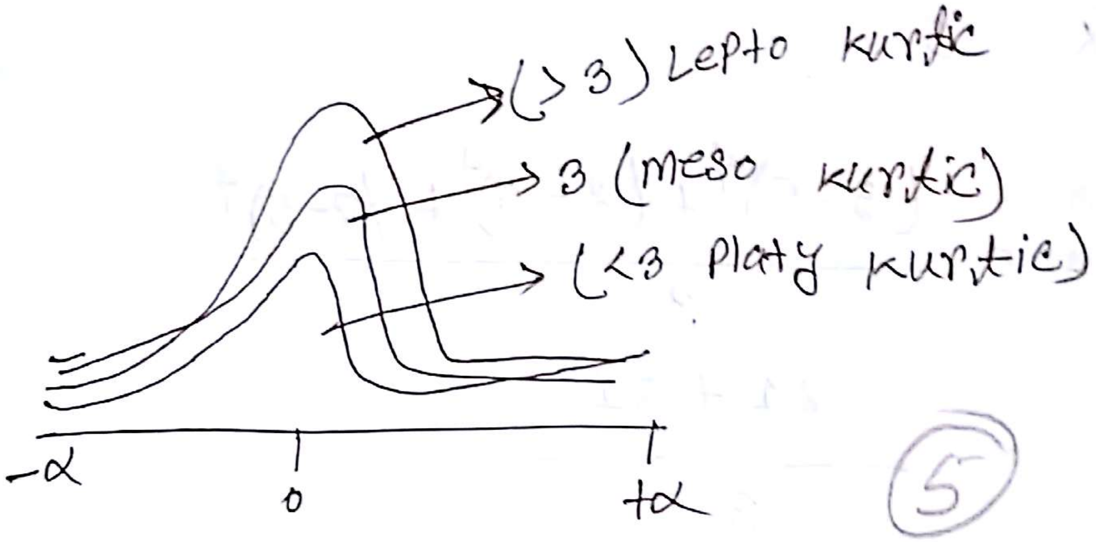
$$= -0.47$$

4

Almost symmetic

kurtosis

1. Lepto kurtic :
2. meso kurtic ;
3. Platy kurtic ;



5

Cofficient of kurtosis :

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

μ = Moments.

* $$\mu_r = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^r}{N}$$

$(5, 2, 8)$

$\bar{X} = 5$

6

$$\mu_4 = \frac{(5-5)^4 + (2-5)^4 + (8-5)^4}{3}$$

$$= \frac{81 + 81}{3}$$

$$= 54$$

$$\mu_2 = \frac{(5-5)^2 + (2-5)^2 + (8-5)^2}{3}$$

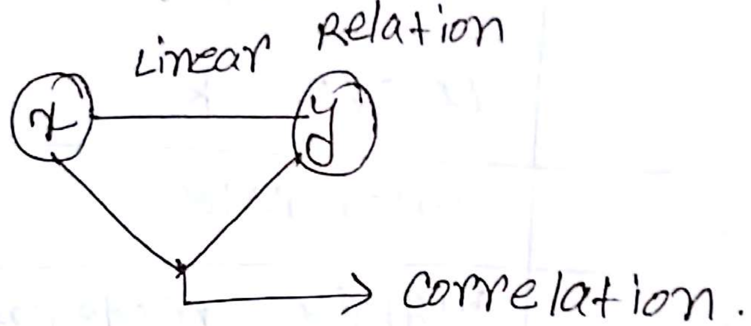
$$= \frac{9+9}{3}$$

$$= 6$$

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2} = \frac{3}{(6)^2} = \frac{3}{36} = 0.083$$

Flaty

Correlation



Kar Pearson Correlation Coefficient:

$$r = \frac{\sum_{i=1}^{n} \left[(x_i - \bar{x})(y_i - \bar{y})\right]}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

r

$r = 0$ No Linear relation.

$r = (-1 \text{ to } +1)$

Next

>0 to 0.2	very Low/very weak
0.21 to 0.4	Low/weak
0.41 to 0.6	moderate
0.61 to 0.8	Highly moderate
0.81 to 1	very High

Supply	Demand	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
5	2	0.7	-1.6	0.49	2.56
6	4	1.7	0.34	2.89	0.1156
2	5	-2.3	1.34	5.29	1.79
$\bar{x}=4.3$	$\bar{y}=3.66$	0.1	0.08	8.67	4.4656

8

$$r = \frac{\sum_{i=1}^{n}\left[(x_i-\bar{x})(y_i-\bar{y})\right]}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \frac{(.1 \times 0.08)}{(8.67 \times 4.46)}$$

$$= \frac{8 \times 10^{-3}}{38.71}$$

$$= 2.066 \times 10^{-4}$$

8

Coefficient of determination.

variation $r^2 = 90\%$

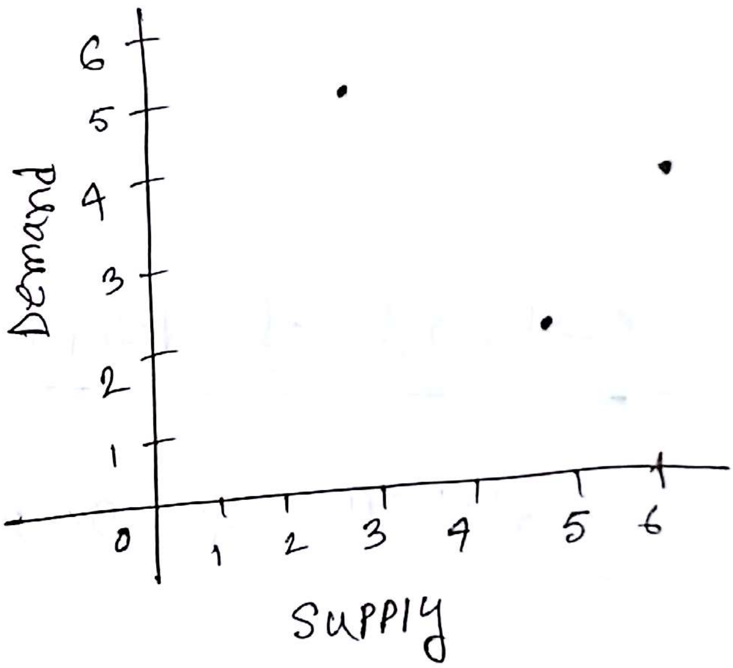
Correlation

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

10

Scatter plot

Supply	Demand
5	2
6	4
2	5



Scatter plot

Regression Analysis

x y 11

1) Simple Regression Analysis (two variables)
   i) independent var.
   2) Dependent var.

2) multiple Regression Analysis (more than two)
   i) more than 1 Independent var.
   2) 1 Dependent var.

Simple Regression Analysis:

Next

Estimated simple linear Regression equation /model/ Line:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$\hat{y}_i \rightarrow$ de.

$x_i \rightarrow$ ind

$\left.\begin{matrix}\beta_0 \\ \beta_1\end{matrix}\right] \rightarrow$ Regression Parameter.

x_i = independent var.

y_i = dependent var.

$$\hat{\beta}_1 = \text{Slope} = \frac{\sum_{i=1}^{n} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \text{intercept} = \boxed{\bar{y} - \hat{\beta}_1 \bar{x}}$$

12

x	y
1	4
2	5
3	6

de(y) Income	inde(x) Expenditure
20	18
7	6
15	15
12	11
$\bar{y} = 13.5$	$\bar{x} = 12.5$

$\beta_1 = n$

$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$
5.5	30.25	6.5	35.75
-6.5	42.25	-6.5	42.25
2.5	6.25	1.5	3.75
-1.5	2.25	-1.5	2.25
	81		84

13

$$\beta_1 = \frac{\sum_{i=0}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$$

$$= \frac{84}{81} = 1.037$$

$$\beta_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$= 13.5 - (1.037 \times 12.5)$$

$$= 0.53$$

Scanned by CamScanner

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \times x_i$$

$$= 0.53 + 1.037 x_i$$

(Ans:)

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

→ Regression Parameter.

$\hat{\beta}_0$ = Intercept

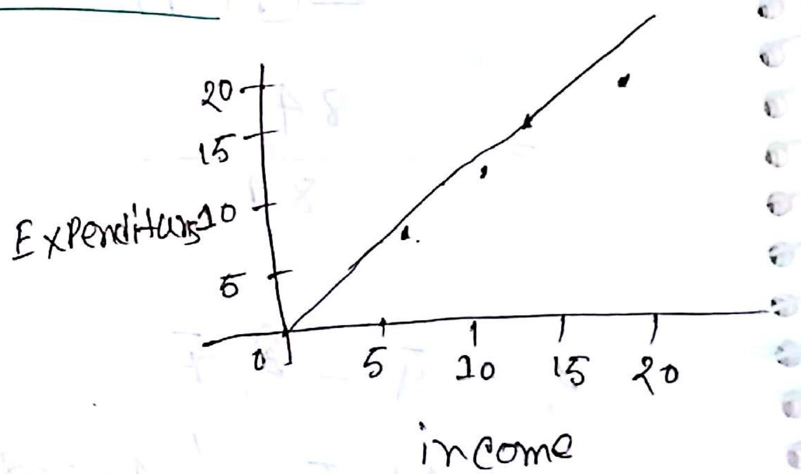
$\hat{\beta}_1$ = slope.

14

Standard Error

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Income	Expenditure
20	18
7	6
15	15
12	11



$$S_{x,y} = \sqrt{\frac{\Sigma (y_i - \hat{y}_i)^2}{n-2}}$$

$\hat{y}_i = -0.19 + 0.94 x_i$
$= -0.19 + 0.94 \times 20$ $= 18.61$
6.33
13.91
11.09

(15)

$$S_{x,y} = \sqrt{\frac{(18-18.61)^2 + (6-6.33)^2 + (15-13.91)^2 + (11-11.9)^2}{4-2}}$$

$$= \sqrt{\frac{0.372 + 1.089 + 1.18 + .81}{2}}$$

$= 1.11.$

Next

Q: calculating standed error?

$\hat{y}_i = -0.19 + 0.94 x_i$

Question:

Cost	Profit
10	8
5	7
12	11

Calculate Standed Error (SE)

16

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=0} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \bar{\beta}_1 \bar{x}$$

Cost	Profit	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
10	8	1	-.66	1	-0.66
5	7	-4	-1.66	16	6.64
12	11	3	2.34	9	7.02
				26	13

$\bar{x} = 9$

$\bar{y} = 8.66$

$$\hat{\beta}_1 = \frac{13}{26}$$

$$= \frac{1}{2} = 0.5$$

$$\beta_0 = \bar{y} - \bar{\beta}_1 \bar{x}$$

$$= 8.66 - .5 \times 9$$

$$= 4.16$$

$$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 \times 1$$

$$= 4.16 + .5 \times 10$$

$$= 9.16$$

$$\hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1 \times 2$$

$$= 4.16 + .5 \times 5$$

$$= 6.66$$

$$\hat{y}_3 = \hat{\beta}_0 + \hat{\beta}_1 \times 3$$

$$= 4.16 + .5 \times 12$$

$$= 10.16$$

$$S_{(x,y)} = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}}$$

$$= \sqrt{\frac{(10-9.16)^2 + (5-6.66)^2 + (12-10.17)^2}{3-2}}$$

$$= \sqrt{.7056 + 2.75 + 3.34}$$

$$= 2.60$$

(Ans)

17

Correlation

Correlation: A correlation is a Linear relationship between two variables. Correlation measures the Linear association between two variables.

~~EXP~~ **Example:**

* Is there any relationship between height and shoe size.
* Number of cigarattes smoked per day and lung cancer.
* Supply and demand.

Math:

Book	Page (x)	Price
History	500	84
Algebra	700	75
management	800	99
socilogo	600	72

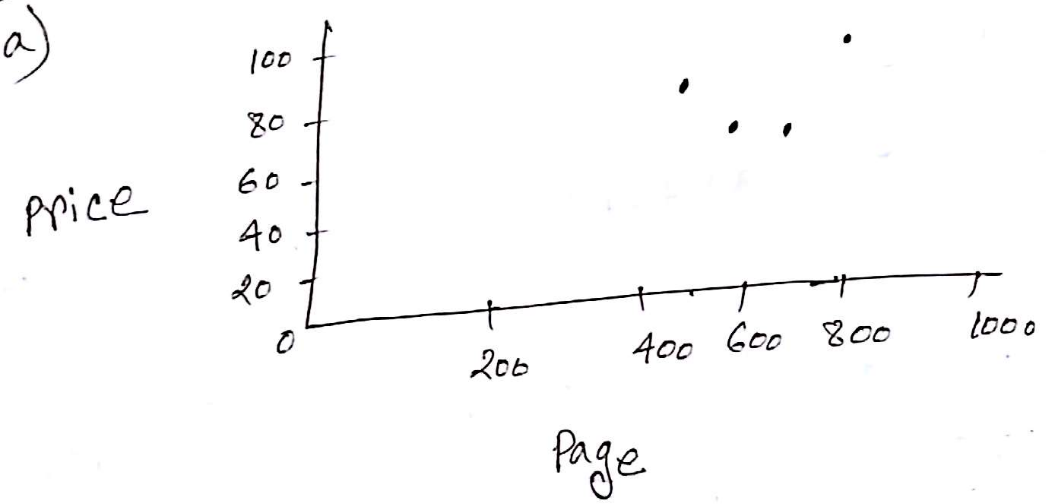
i) draw scatter diagram

ii) Determine the cofficient of correlation.

iii) Interpreted the result.

Solution:

a)



b) Here:

we know:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2600}{4} = 650.$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{330}{4} = 82.5.$$

x	y	$(x_i-\bar{x})$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
500	84	-150	1.5	22500	2.25	-225
700	75	50	-7.5	2500	56.25	-375
800	99	150	16.5	22500	272.25	2475
600	72	-50	-10.5	2500	110.25	525
				$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 50000$	$\sum_{i=1}^{n}(y_i-\bar{y})^2 = 441$	$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = 2400$

Now,

From the correlation coefficient.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \frac{2400}{\sqrt{50000 \times 441}}$$

$$= 0.511.$$

c. Interpretation: The correlation between the number of pages and the selling price of the book is 0.511. This indicates a moderate association between the variable.

Probobilith

1. Random Expremind
2. Event.
3. Malually Exclusive event
4. Sample space.

$$P(A) = \frac{m}{n}$$

$$= \frac{\text{Favorable no. of out come}}{\text{Total no of out come}}.$$

total 20.

Sak = 10

min = 8

mil = 2

$$P(\text{mil}) = \frac{2}{20}$$

A Inposiab event:

$P(A) = 0$

" ~~Digital Electronics~~ "

Adition Rule (two events)

→ specific addition rule:

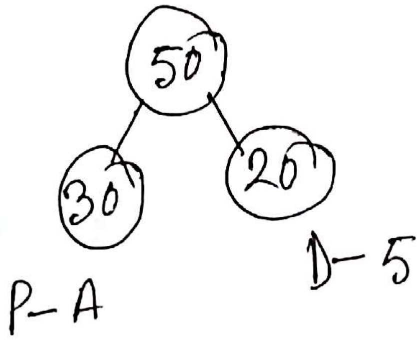
$P(A \text{ or } B) = P(A) + P(B)$

$\Rightarrow P(A \cup B) = P(A) + P(B)$.

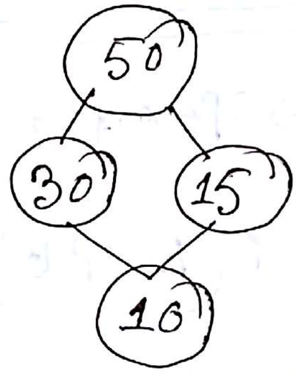
General addition rule:

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$$P(P \text{ or } D) = P(P) + P(D)$$
$$= 30/50 + 20/50$$



$$P(P \text{ or } D)$$
$$= P(P) + P(D) - P(P \cap D)$$
$$= 30/50 + 15/50 - P(P \cap D)$$
$$= 30/50 + 15/50 - 10/50$$

Laws / Propenties of Probability :

1. $0 \leq P(A) \leq 1$

2. $\sum_{i=1}^{n} P(A_i) = 1$

$\Rightarrow P(A_1) + P(A_2) + \cdots + P(A_n) = 1$

Scanned by CamScanner

$SUK = 5$ $P(SUK) = 5/20$

$min = 10$ $P(min) = 10/20$

$UH = 5$ $P(UH) = 5/20$.

Regressin Analysis

Regression Analysis: is concerned with the study of the dependence of one variable (the dependent variable) on one or more other variable (the independent variables). In regression analysis we use independen variables to estimating or predicting the average change of the dependent variables.

Examples:

The marking director of a company may wand to know how the demand for the company's product is related, to say advertising expenditure.

* suppose the sales manager of a company say x, wants to determine how the number of credit cards sell is related to the number of call.

Some terms related to regression analysis:

*A scatter plot: is a chart that portrays the relationship between two variables.

*Dependent variable: the dependent variable is the variable being predicted or estimated. It is denoted by y.

*Independent variable: The Independent variable provides the basis for estimation. It is the predictor variable. It is denoted by x

Linear Regression model: The equation the describes how y is related to x and an error terms is called the regression model.

the linear regression equation is:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

the estimated linear linear regression equatis is give as:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

Math:

Month	Advertising	Sales revenus
July	2	7
August	1	3
september	3	8
october	4	10

a) Draw Scatter Plot

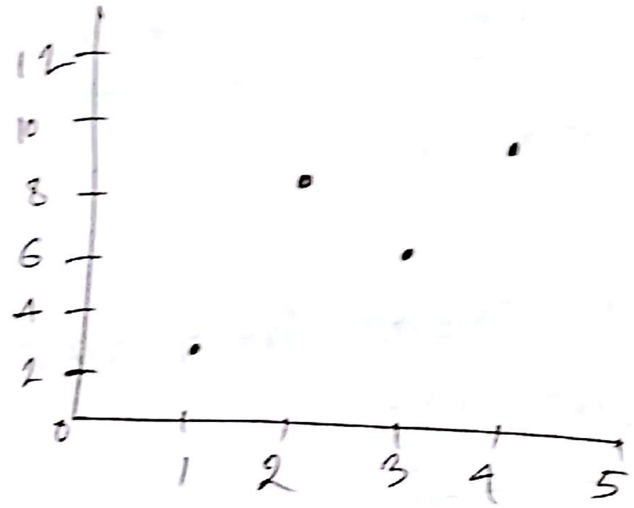
b) Determine the estimated regression equation.

c. Interpred the value β_0 and β_1

d. Estimate Sales when \$3 milion is spend on aderdising.

Solution:

(a)



(b) we know, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$\bar{x} = \frac{\sum_{i=1} x_i}{n} = \frac{10}{4} = 2.5 \text{ and}$$

$$\bar{y} = \frac{\sum_{i=1}^{4} y_i}{n} = \frac{28}{4} = 7.$$

Advertising expense(x)	Sales revenue(y)	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
2	7	-0.5	0	0.25	0
1	3	-1.5	-4	2.25	6
3	8	0.5	1	0.25	0.5
4	10	1.5	3	2.25	4.5
				$\sum_{i=1} (x_i-\bar{x})=5$	$\sum_{i=1}^{4}(x_i-\bar{x})(y_i-\bar{y})=11$

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{11}{5}$$

$$\beta_1 = 2.2$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} = 7 - 2.2 * 2.5$$

$$\beta_0 = 1.5$$

Now if we put the values of estimated regression equation is

$\hat{y}_i = 1.5 + 2.2x_i$

c. Interpretation:

Here ~~shreis~~ slape is $\hat{\beta}_1 = 2.2$ This mease that an increases of \$ 1 million in advertising cost, the sales revenue will increase \$2.2 million.

and $\hat{\beta}_0 = 1.5$ means that, is if there is no advertisement cost, then sales revenue would be \$1.5 million.

d) Now if $x = 3$ then

$\hat{y} = 1.5 + 2.2 * 3 = 8.1$

So when advertisement cost is \$ 3 million, the sales revenue would be 8.1 million.

Scanned by CamScanner

Standard error

The standard Error of Estimate measures the scatter, or dispersion, of the observed values arround the line of regression. The formula that is used to compute the standard error.

$$S_{(y,x)} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}}$$

Example: From the Previous example we calculate the value of $\hat{y}_i$ as we know.

Standard Error: $S_{yx} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}}$

Scanned by CamScanner

y_1	$\hat{y_i}$	$(y_1-\hat{y_1})$	$(y_i-\hat{y_1})^2$
7	5.9	1.1	1.21
3	3.7	-0.7	0.49
8	8.1	-0.1	0.01
10	10.3	-0.3	0.09

$$\sum_{i=1}^{4} (y_1-\bar{y}_i) = 1.8$$

Here $n = 4$

Now by putting value we get.

$$S_{(y,x)} = \sqrt{\frac{\sum_{i=1}^{n}(y_1-\bar{y}_i)^2}{4-2}}$$

$$= \sqrt{\frac{1.8}{2}}$$

$$= 0.95$$

So the standred error is 0.95.

Coefficien of Determination r^2:

the coefficient of determination tells the percent of the variation in the dependent.

Scanned by CamScanner

variable that is explained (determined) by the model and the explanatory variable.

Interpretation of r^2: Suppose $r^2 = 92.7\%$

Interpretation: Almost 93% of the variability of the dependend variable explained by the independend variables.

Relationship between regression and correlation:

$$\hat{r} = \hat{\beta}_1 \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

Regression versue correlation:

In correlation the primary objective is to measure the strength or degree of linear association between two variables.

* In regression analysis, we are not primary intersted in such a measure. Instend, we try to estimate or predicd the average value of one variable on the basis of the fixed values of other variables.

* In correlation analysis there is no distinction between the dependend and explanatory variables.

In regression anylysis there is an asymmetry in the dependent and explanatory variables are treated. The dependend variable is assumed to be statistical, random, or stochastic. The explanatory variables, on the other hand, are assumed to have fixed values.

Multiplication Rule

or → $\cup$ → (+)

and → $\cap$ → (×)

1. Specific Multiplication Rule:

$$P(A \cap B) = P(A) \times P(B)$$

$$\Rightarrow P(A \text{ and } B) = P(A) \times P(B)$$

[Independent events]

2. General Multiplication Rule:
(Dependent events)

$$P(A \cap B) = P(A) \cdot \underbrace{P(B/A)}_{\text{Conditional Proby.}}$$

$$\Rightarrow P(A \text{ and } B) = P(A) \cdot P(B/A)$$

Example: A company has two large computers the prob. that the newer one will breakdown on only particular month is 0.05, The proby. that

The older one will breakdown on any Particular month is 0.1

what is the ~~Proby~~ Proby. That they will both breakdown in a particular month?

$P(A) = 0.05$ $\qquad$ $P(A \cap B) = P(A) \times P(B)$

$P(B) = 0.1$

$= 0.05 \times .1$

$= 0.005$ (Ans:).

Complement Rule

$$P(\overline{A}) = 1 - P(A).$$

Example: Rahim can hit a target 5 out of 7 chance

Karim " " " " 6 " " 11 "

Find the chances/probability that the

target is hit onces they both they ?

$P(R) = \frac{5}{7}$

$P(K) = \frac{6}{11}$.

**

$P(R \text{ or } K) = P(R) + P(K) - P(R \text{ and } K)$

$= 5/7 + 6/11 - [P(R) . P(K)]$

$= 5/7 + \frac{6}{11} - (5/7 * 6/11)$

$\boxed{P(R \text{ and } K) = P(R) * P(K)}$

complementary

$P(R) = 5/7$

$P(K) = 6/11$

$P(\bar{R}) = 1 - 5/7 = \frac{2}{7}$

$P(\bar{K}) = (1 - 6/11) = 5/11$

$$P(\bar{R} \cap \bar{K}) = P(\bar{R}) \cdot P(\bar{K})$$

$$= 2/7 \times 5/11$$

Karim and Rahmim কেউ পারে না।

$$1 - P(\bar{R} \cap \bar{K}) = P(R \cup K)$$

$$\Rightarrow P(R \cup K) = 1 - \frac{10}{77}$$

$$= \frac{77 - 10}{77}$$

একজন পারবে

1. marginal Probality
2. Joint "
3. Conditional "

Condition Probality:

$$P(A/B) = \frac{P(A \text{ and } B)}{P(B)} \quad \Big| \quad P(B) \neq 0$$

or $P(B/A) = \frac{P(B \text{ and } A)}{P(A)}$; $P(A) \neq 0$

Marginal Probability:

$N = 130$

	Yes	No	total
Femal	50	10	60
male	40	30	70
total	90	40	130

$P(\text{yes}) = 90/130$

$P(\text{male}) = \frac{70}{130}$.

$P(\text{female}) = \frac{60}{130}$

Joint Probability

$P(\text{male and N}) = 30/130$

$P(\text{female and yes}) = 50/130$

Conditional Probability

$$P\left(\text{male}/\text{yes}\right) = \frac{P(\text{male and yes})}{P(\text{yes})}$$

$$= \frac{40/130}{90/130} = \frac{40}{90}$$

$$P\left(F/N\right) = \frac{10/130}{40/130} = 10/40$$

Examp: A humburger chain found that 75% of all customers use salad, 80%, use ketch up and 65% use both what are the probabilits that.

i) a ketchap used salad

ii) a salad " katchup.

$$P(S) = 0.75 = \frac{75}{100}$$

$$P(K) = 0.80$$

Scanned by CamScanner

$P(S \text{ and } K) = 0.65$

i) $P(S/K) = \dfrac{P(S \text{ and } K)}{P(K)} = .65/.80$ (Ans:)

ii) $P(K/S) = \dfrac{P(K \text{ and } S)}{P(S)} = \dfrac{.86}{.65}$ (Ans:)