Simple linear Regression and correlation

Assignment No-01

The following results from an experiment in which different masses were placed on a spring and resulting length of the spring measured are shown bellow:

mass, x kg	20	40	60	80	100
length, y cm	48	55.1	56.3	61.2	68.

(i) Draw a scatter diagram.

(ii) Find the regression equation of Length on mass.

(iii) Find or predict the strength when the mass is 351 g.

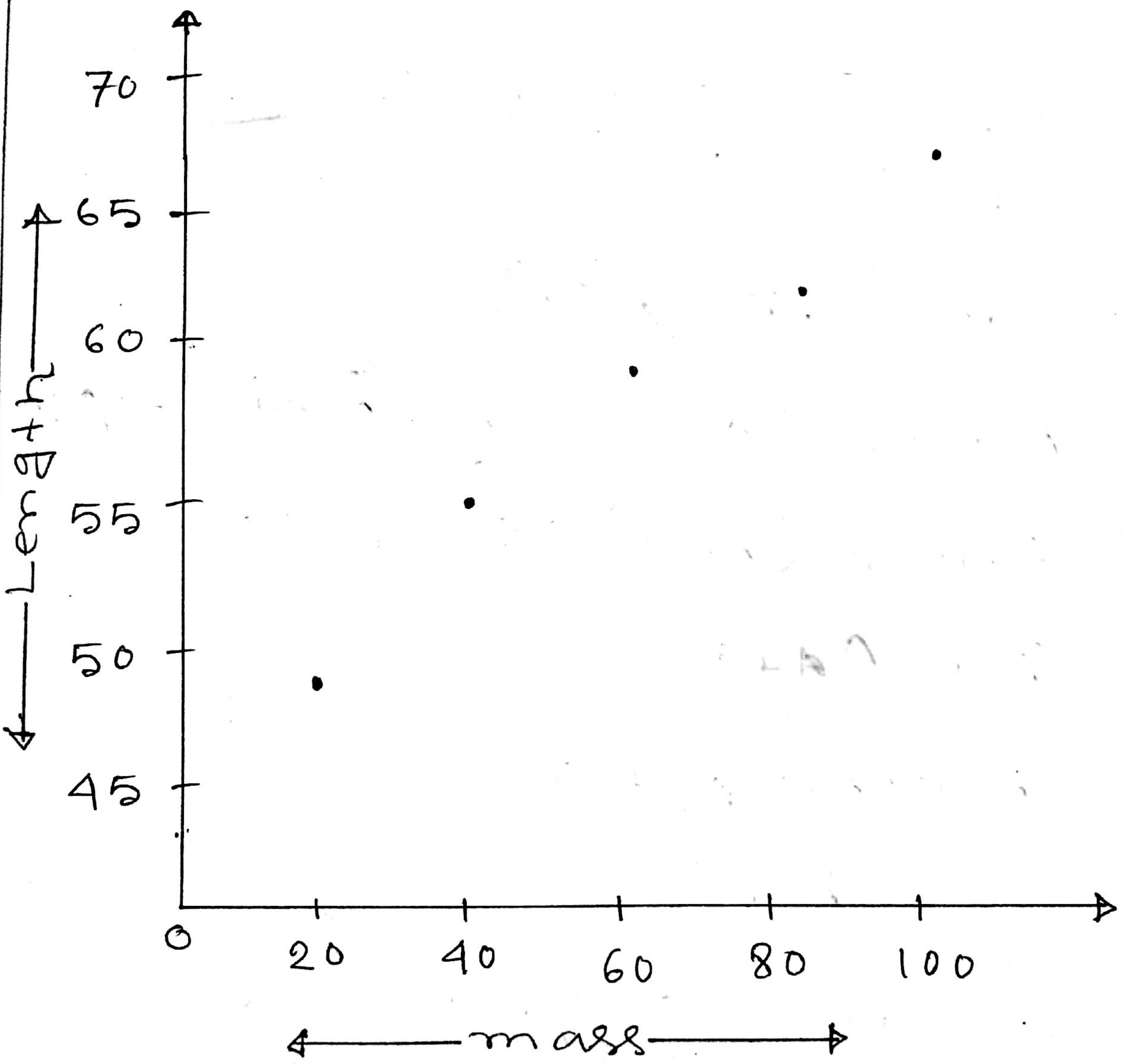
(iv) Explain the constants of these model.

(v) compute correlation coefficien

(vi) Determine the co-efficient of Determination and interpret

(i) **Solution:-** let, mass $= x_i$

length $= y_i$



scatter plot.

Comment: From scatter plot we can see that mass and length are positively associated.

(ii) **Solution:** Mass (x) is independent variable and length is dependent variable.

Let, the least square Regression Line is, $\hat{Y} = \hat{B}_0 + \hat{B}_1 x$

x_i	y_i	$x_i y_i$	x_i^2	y_i^2	
20	48	960	400	2304	
40	55.1	2204	1600	3036.01	$\bar{x} = \frac{300}{5} = 60$
60	56.3	3378	3600	3169.69	
80	61.2	4896	6400	3745.44	
100	68	6800	10000	4624	$\bar{y} = \frac{288.6}{5} = 57.72$

Here, $\Sigma x_i = 300$

$\Sigma y_i = 288.6$

$\Sigma x_i y_i = 18238$

$\Sigma x_i^2 = 22000$

$\Sigma y_i^2 = 16879.14$

where, $\hat{B}_1 = \dfrac{\sum x_i y_i - n\bar{x}\cdot\bar{y}}{\sum x_i^2 - n\bar{x}^2}$

$= \dfrac{18238 - 5\times 60\times 57.72}{22000 - 5\times(60)^2}$

$= \dfrac{922}{4000}$

$= 0.2305$

$\therefore \hat{B}_0 = \bar{y} - \hat{B}_1\bar{x}$

$= 57.72 - 0.231\times 60$

$= 43.86$

$\hat{y} = \hat{B}_0 + \hat{B}_1 x$

$= 43.86 + 0.2305x \longrightarrow (i)$

$\therefore b = 0.2305$

$a = 43.89$

(iii) when mass is 35 kg,

Length $= 43.89 + 0.2305 \times 35$

$= 51.96$ cm

(iii) $a = 43.89$ cm is the length of the spring when $x = 0$ i.e, when no mass is given on it.

$b = 0.2305$ is the amount by which the springs length increase for every extra 1 kg of mass.

(iv) Calculate the co-rrelation co-efficient :-

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

$$\text{or, } r = \frac{18238 - 5 \times 60 \times 57.72}{\sqrt{(22000 - 5 \times 60^2)(16879.14 - 5 \times 57.72^2)}}$$

$$\text{or, } r = \frac{922}{940.53} = 0.98$$

(v) $r^2 = 0.96$

That means masses can explain 96% variation of the 43.89 cm length of spring regression lines is 51.96 cm. Any prediction made by the regression line accepted/rejected.

Interpretation:- co-relation co-efficient $r = 0.98$ implies that there is hight positive co-relation between mass and length

Range: $-1 \leq r \leq 1$

$r^2 = 0.96$

that means 96% variation of length can be explained by mass.

Assignment-02

A researcher thinks that there is a link between a persons confidence and their height. He divises a test to measure the confidence 'c' of nine people and their height 'h' cm. The data are shown below:

h	179	169	187	166	162	193	161	177	168
c	569	561	579	561	540	598	542	565	573

(a) Draw a scatter diagram to represent these data & comment on the characteristies of co-rrelation

(b) calculate co-rrelation co-efficient and interpret.

(c) calculate the equation of the regression line of c and h.

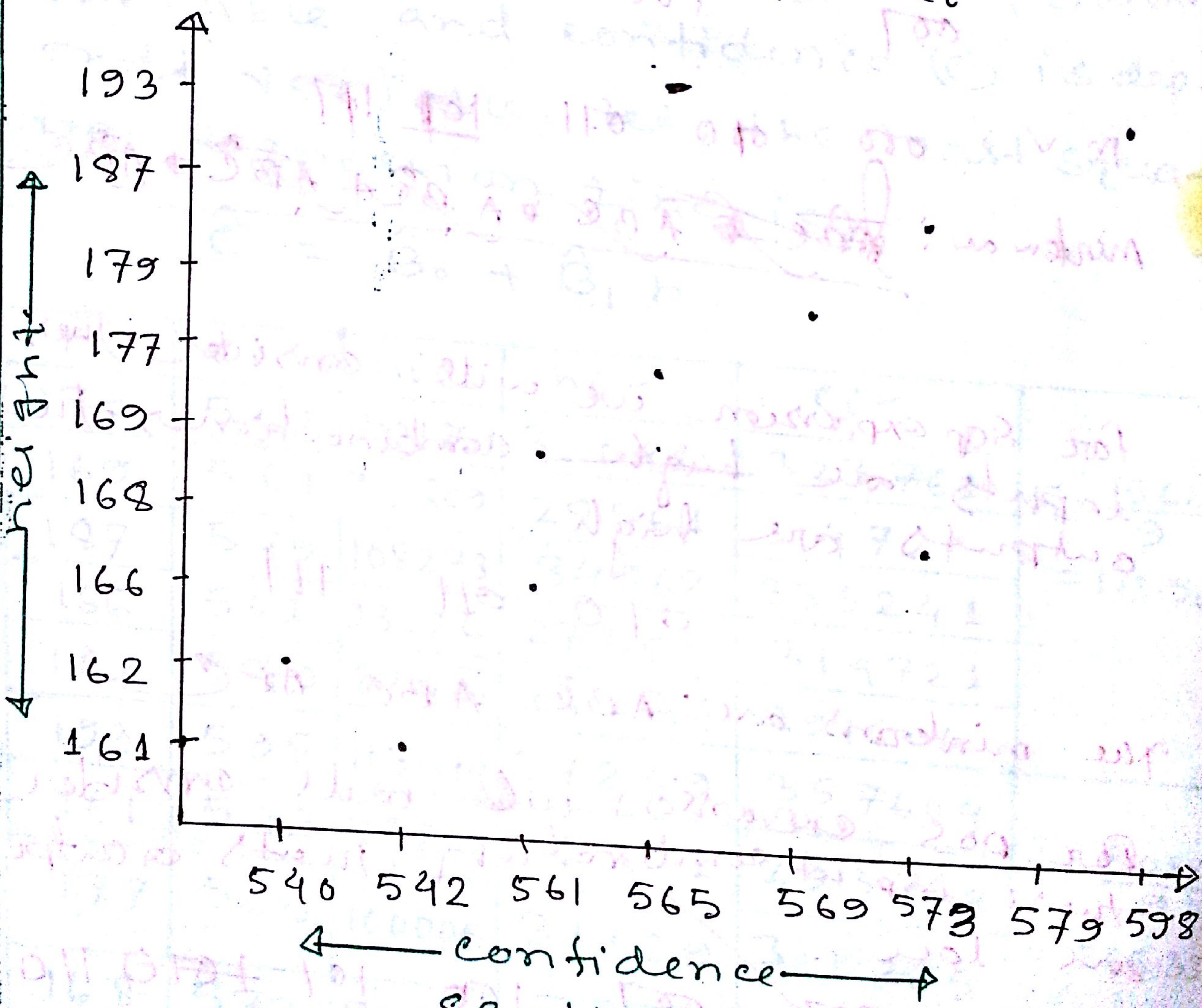
(d) draw this line on your scatter diagram.

(e) Interpret the gradient / slope of the regression line.

(f) Find the proposed confidence for the person who has a height of 172 cm.

(a) **Solution:** Let, height $= h_i$

confidence $= c_i$



scatter plot.

Comment: we can write from the data that the height and confidence are positively associated.

Solution: Height (h) is independent variable and confidence (c) is dependent variable. Let, the least square Regression Line is,

$$\hat{c} = \hat{\beta}_0 + \hat{\beta}_1 h.$$

h_i	c_i	$h_i c_i$	h_i^2	c_i^2
179	569	101851	32041	323761
169	561	94809	28561	314721
187	579	108273	34969	335241
166	561	93126	27556	314721
162	540	87480	26244	291600
193	598	115414	37249	357604
161	542	87262	25921	293764
177	565	100005	31329	319225
168	573	96264	28224	329329

$\bar{h} = \frac{1562}{9}$

$= 173.56$

$\bar{c} = \frac{5088}{9}$

$= 565.33$

Here, $\sum h_i = 1562$

$\sum c_i = 5088$

$\sum h_i c_i = 884484$

$\sum h_i^2 = 272094$

$\sum c_i^2 = 2878966$.

Where, $$\hat{B}_1 = \frac{\sum h_i c_i - n\bar{h}.\bar{c}}{\sum h_i^2 - n\bar{h}^2}$$

$$= \frac{884484 - 9 \times 173.56 \times 565.33}{272094 - 9 \times (173.56)^2}$$

$$= \frac{1415.93}{986.34}$$

$$= 1.436$$

$$\hat{B}_0 = \bar{c} - \hat{B}_1 \bar{h}$$

$$= 565.33 - 1.436 \times 173.56$$

$$= 565.33 - 249.23$$

$$= 316.1$$

$$\hat{c} = 316.1 + 1.436h \longrightarrow \text{(i)}$$

$\therefore b = 1.436$

$a = 316.1$

(iii) when height is 172 cm

$$\hat{c} = 316.1 + 1.436 \times 172$$
$$= 563.09$$

(1) 573.178 (4) 554.458 (7) 547.258
(2) 558.778 (5) 548.698 (8) 570.298
(3) 584.698 (6) 593.338 (9) 537.338

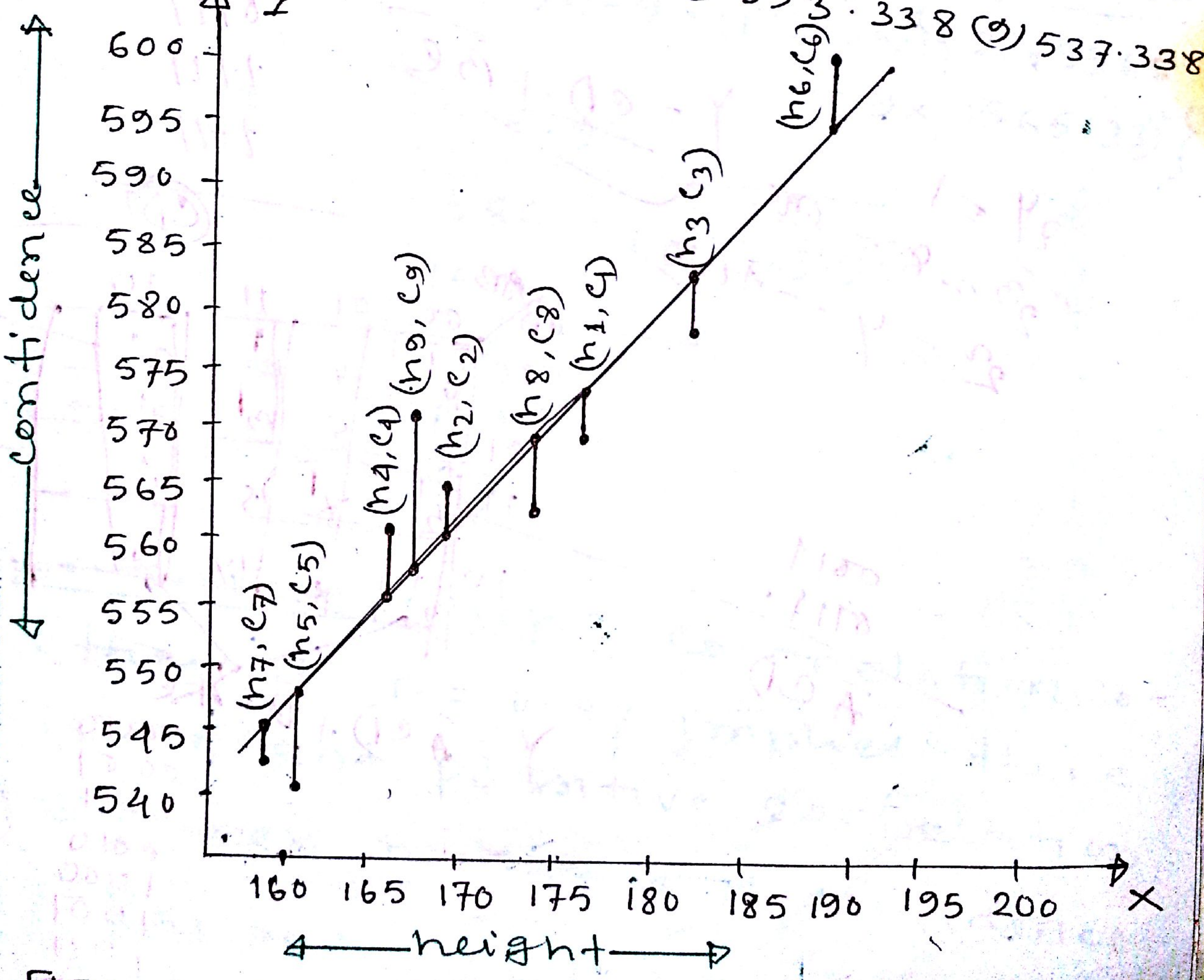


Fig: A scatter digram of the Regression Line

(b) calculate co-rrelation co-efficient and interpret,

$$r = \frac{\sum h_i c_i - n\bar{h}\bar{c}}{\sqrt{(\sum h_i^2 - n\bar{h}^2)(\sum c_i^2 - n\bar{c}^2)}}$$

$$= \frac{884484 - 9 \times 173.56 \times 565.33}{\sqrt{\{272094 - 9 \times (173.56)^2\}\{2878966 - 9 \times (565.33)^2\}}}$$

$$= \frac{1415.93}{\sqrt{986.34 \times 2583.92}}$$

$$= \frac{1415.93}{1596.44}$$

$$= 0.886$$

$\therefore r = 0.87.$

<u>Interpretation</u>:- co-rrelation co-efficient $r = 0.87$ implies there is height possitive correlation between height and confidence.

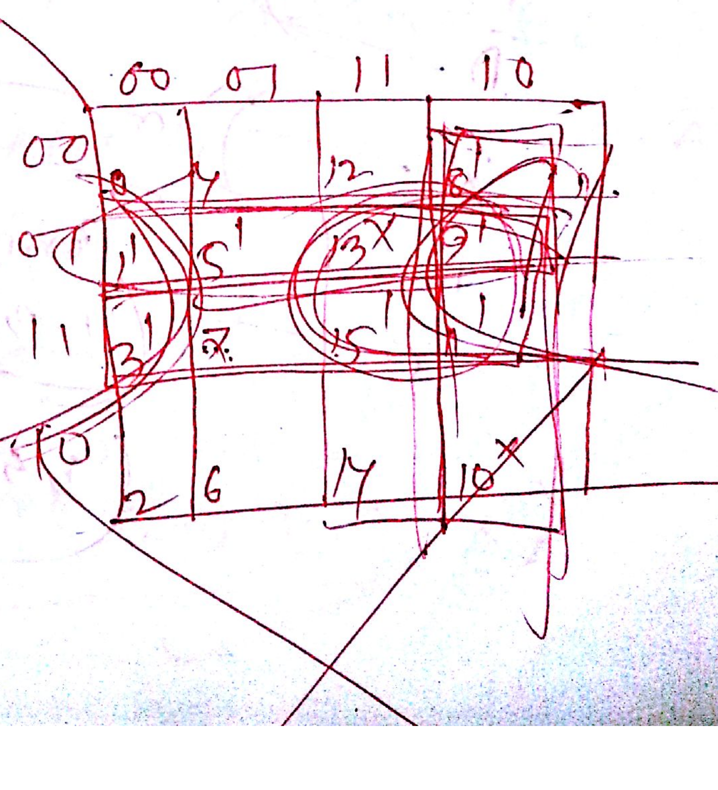
Range: $-1 \leq r \leq 1$

(e) For increasing each unit of height the amount of confidence increased 1.44 unit.

(v) $r^2 = 0.76$

height can explain 76% variation of the 361.1 confidence of spring regression lines fit 563.09. Any prediction made by the regression line accepted/rejected

Don't care condition



Assignment no-03

compute and interpret the co-relation co-efficient for the following grades of 6 students selected at random.

Math grade	70	92	80	74	65	83
English grade	74	84	63	87	78	90

Solution :- Let, Math = M_i

English = E_i

Step 1: $ABC + \bar{A}BC + A\bar{B}C + \bar{A}\bar{B}$ $\bar{A}\bar{B}(C+\bar{C}) = \bar{A}\bar{B}C$

Step 2: $ABC + \bar{A}BC + A\bar{B}C + \bar{A}\bar{B}C + \bar{A}\bar{B}\bar{C}$

111   011   101   001   000

val: 7   3   5   1   0

Step: function $= \sum m(0, 1, 3, 5, 7)$

M_i	E_i	M_iE_i	M_i^2	E_i^2
70	74	5180	4900	5476
92	84	7728	8464	7056
80	63	5040	6400	3969
74	87	6438	5476	7569
65	78	5070	4225	6084
83	90	7470	6889	8100
$\Sigma M_i = 464$	$\Sigma E_i = 476$	$\Sigma M_iE_i = 36926$	$\Sigma M_i^2 = 36354$	$\Sigma E_i^2 = 38254$

$\bar{M} = \frac{464}{6} = 77.33$

$\bar{E} = \frac{476}{6} = 79.33$

Calculate the correlation coefficient and interpretation:

$$r = \frac{\Sigma M_iE_i - n\bar{M}\bar{E}}{\sqrt{(\Sigma M_i^2 - n\bar{M}^2)(\Sigma E_i^2 - n\bar{E}^2)}}$$

$$= \frac{36926 - 6 \times 77.33 \times 79.33}{\sqrt{\{36354 - 6\times(77.33)^2\}\{38254 - 6\times(79.33)^2\}}}$$

$$= \frac{118.47}{\sqrt{474.43 \times 494.51}}$$

$$= \frac{118.47}{484.37}$$

$\therefore r = 0.244$

Interpretation: co-rrelation co-efficient $r = 0.244$ implies there is height possitive co-rrelation between MAth grade and English grade.

Range: $-1 \leq r \leq 1$.

$\rightarrow \overline{AB}$

Comment: The two variables are associated weakly.

Table-4

Terms with X	minterms of the [illegible] sop				
	0	1	3	5	7
$\overline{AB}$	X	X			
C		X	X	X	X

$$Y = \overline{AB} + C$$

Assignment-04

The following data were obtained in a study of the relationship between the weight and chest size of infants at birth.

Weight (kg)	Chest size (cm)
2.75	29.5
2.15	26.3
4.41	32.2
5.52	36.5
3.21	27.2
4.32	27.7
2.31	28.3
4.30	30.3
3.71	28.7

Solution:- Let, weight $= w_i$

chest size $= c_i$

w_i	e_i	w_ie_i	w_i^2	e_i^2	
2.75	29.75	81.125	7.56	870.25	
2.15	26.3	56.645	4.62	691.69	$\bar{w} = \frac{32.68}{9}$
4.41	32.2	142.002	19.45	1036.84	$= 3.63$
5.52	36.5	201.48	30.47	1332.25	
3.21	27.2	87.312	10.30	739.84	
4.32	27.7	119.664	18.66	767.29	$\bar{e} = \frac{266.7}{9}$
2.31	28.3	65.373	5.34	800.89	$= 29.63$
4.30	30.3	130.29	18.49	918.09	
3.71	28.7	106.477	13.76	823.69	
Σw_i 32.68	Σe_i 266.7	Σw_ie_i 990.37	Σw_i^2 128.65	Σe_i^2 7980.83	

(a) calculate, r

$$r = \frac{\Sigma w_ie_i - n\bar{w}\bar{e}}{\sqrt{(\Sigma w_i^2 - n\bar{w}^2)(\Sigma e_i^2 - n\bar{e}^2)}}$$

$$= \frac{990.37 - 9\times 3.63 \times 29.63}{\sqrt{\{128.65 - 9\times(3.63)^2\}\{7980.83 - 9\times(29.63)^2\}}}$$

$$= \frac{22.36}{\sqrt{10.66 \times 79.39}} = \frac{22.36}{28.26}$$

$\therefore r = 0.791.$

(b) $r^2 = 0.63$

that means 63% variation of in tant chest size is explained by weight.

Half adder: An Half adder is a logic circuit that adds two single bit numbers and produces a sum and a carry.

Truth table:

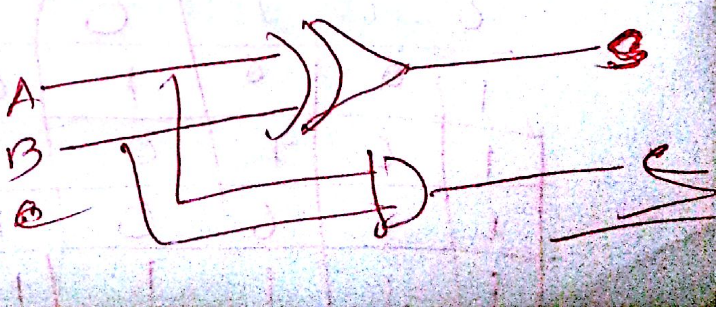
Inputs		Outputs	
A Augend	B Addend	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1		0	1

$\begin{array}{c} A \\ +B \\ \hline \end{array}$

Addition
A → Augend
B → Addend

$S = A\bar{B} + \bar{A}B = A \oplus B$

$C = AB$



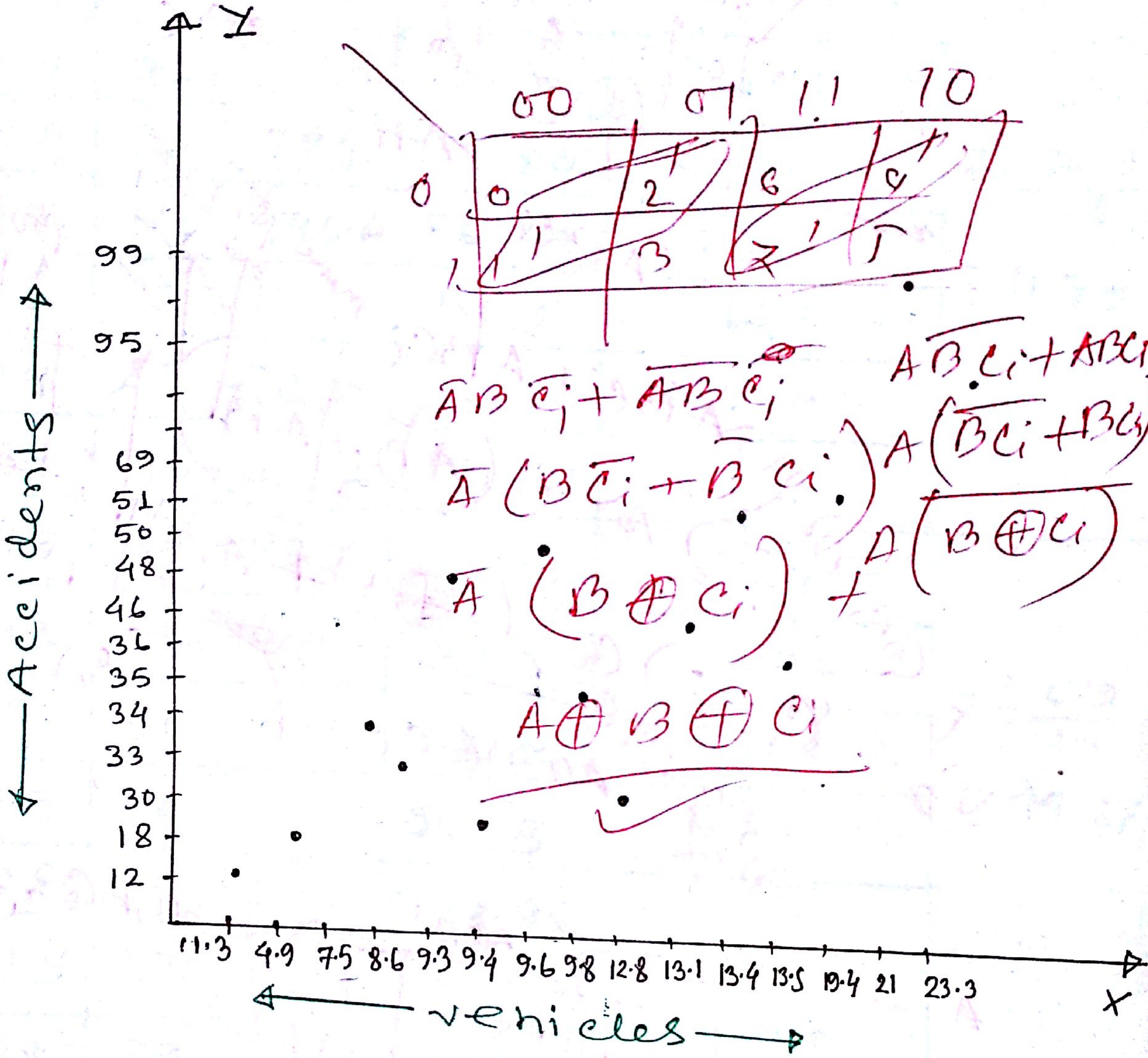
Assignment NO-05

The number of vehicles, x millions and the number of accidents Y thousands in 15 different country were:-

Country	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Vehicles, X millions	8.6	13.4	12.8	9.3	1.3	9.4	13.1	4.9	13.5	9.6	7.5	9.8	23.3	2.1	19.4
accidents, Y thousands	33	51	30	48	22	23	46	18	36	50	34	35	95	99	69

Solution: Let, Vehicles $= x_i$

Accidents $= Y_i$



$\bar{A}B\bar{C_i} + \bar{A}\bar{B}C_i$

$\bar{A}(B\bar{C_i} + \bar{B}C_i)$

$\bar{A}(B \oplus C_i)$ +

$A\bar{B}\bar{C_i} + ABC_i$

$A(\bar{B}\bar{C_i} + BC_i)$

$A(\overline{B \oplus C_i})$

$A \oplus B \oplus C_i$

<u>Comment:</u> The co-rrelation co-efficient between x and y is possitive.

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
8.6	33	283.8	73.96	1089
13.4	51	683.4	179.56	2601
12.8	30	384	163.84	900
9.3	48	446.4	86.49	2304
1.3	12	15.6	1.69	144
9.4	23	216.2	88.36	529
13.1	46	602.6	171.61	2116
4.9	18	88.2	24.01	324
13.5	36	486	182.25	1296
9.6	50	480	92.16	2500
7.5	34	255	56.25	1156
9.8	35	343	96.04	1225
23.3	95	2213.5	542.89	9025
21	99	2079	441	9801
19.4	69	1338.6	376.36	4761

$\bar{x} = \frac{176.9}{15}$

$= 11.79$

$\bar{y} = \frac{679}{15}$

$= 45.27$

Here, $\sum x_i = 176.9$

$\sum y_i = 679$

$\sum x_i y_i = 9915.3$

$\sum x_i^2 = 2576.47$

$\sum y_i^2 = 39771$

We know co-rrelation co-efficient,

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

$$= \frac{9915.3 - 15 \times 11.79 \times 45.27}{\sqrt{\{2576.47 - 15 \times (11.79)^2\}\{39771 - 15 \times (45.27)^2\}}}$$

$$= \frac{1909.30}{\sqrt{491.41 \times 9030.41}}$$

$$= \frac{1909.30}{2106.57}$$

$$= 0.906$$

The value of the co-rrelation co-efficient is 0.906. This is a strong possitive co-rrelation. That means the greater the number of vehicles the higher the number of accidents.