Assignment 3

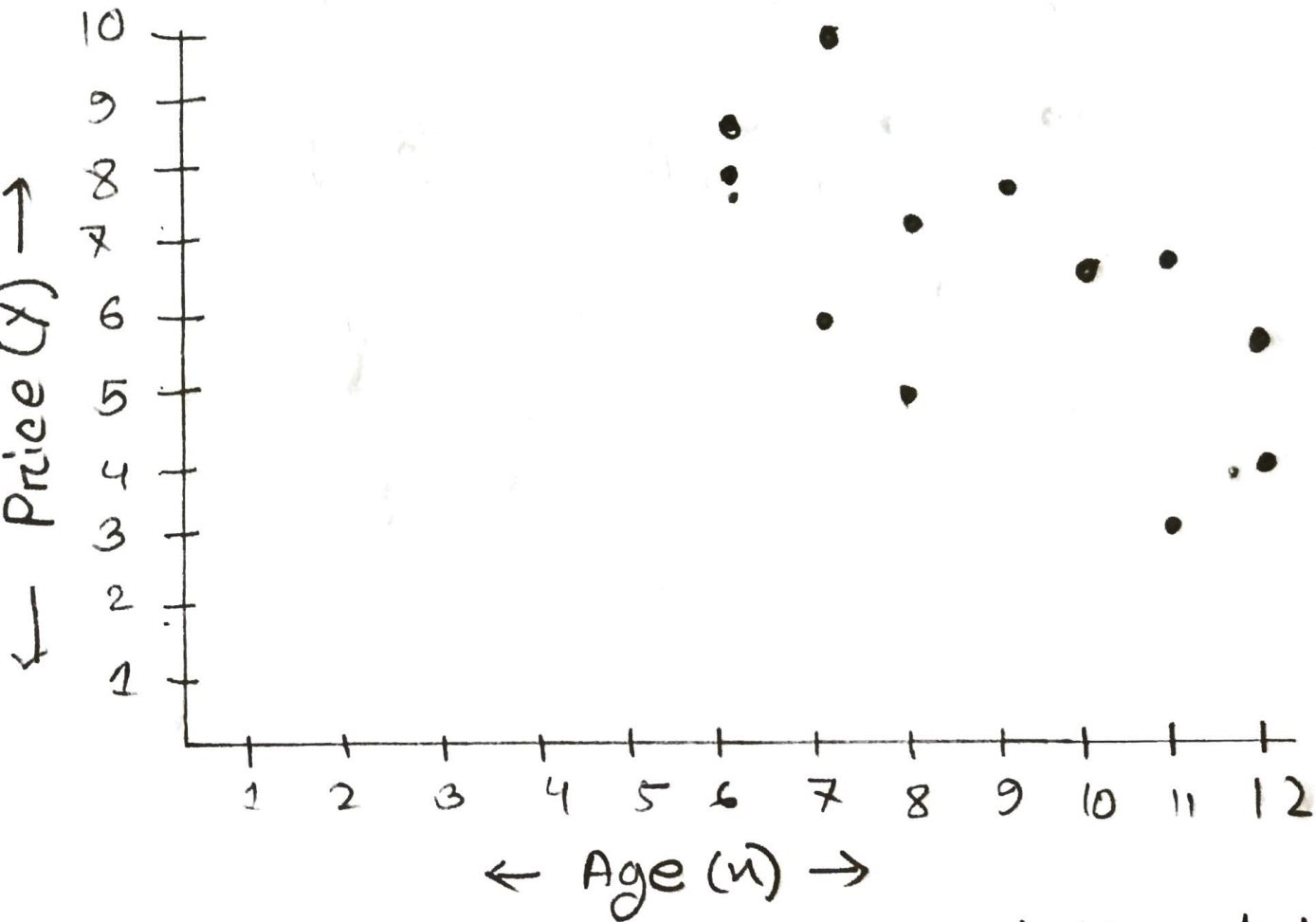
Akib Zabed

ID : 20101113

Sec : 05

Problem 1

a



There is moderate negative (-) correlation between Age(X) and Price (Y)

b

Age (x)	Price (Y)	XY	X^2	Y^2
9	8.1	72.9	81	65.61
7	6	42	49	36
11	3.6	39.6	121	12.96
12	4	48	144	16
8	5	40	64	25
7	10	70	49	100
8	7.86	60.8	64	57.76
11	8	88	121	64
10	8	80	100	64
12	6	72	144	36
6	8.6	51.6	36	73.96
6	8	48	36	64
Total: 107	82.9	712.9	1009	615.29

$$\bar{x} = \frac{107}{12} = 8.91$$

$$\bar{y} = \frac{82.9}{12} = 6.90$$

∴ pearson correlation coefficient

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

$$= \frac{712.9 - 12 \times 8.91 \times 6.90}{\sqrt{(1009 - 12 \times 8.91^2)(615.29 - 12 \times 6.90^2)}}$$

$$= \frac{712.9 - 12 \times 8.91 \times 6.90}{\sqrt{(1009 \times 12 \times 79.50)(615.29 \times 102 \times 97.72)}}$$

$r = -0.543$

$\therefore r^2 = 0.2955$

$\because r = -0.543$ means there exists moderate degree of negative correlation between Age(x) and Price (Y)

and $r^2 = 0.2955$

$\Rightarrow$ 29.55% of total variation in price (y) can be explained by linear relation between Age and Price. The other 70.45% of total variation remains unexplained.

Problem 2

a

Judge 1 (x)	Judge 2 (y)	R_x	R_y	$d = R_x - R_y$	d^2
650	900	5	9	−4	16
760	720	11	4	7	49
740	690	10	1.5	8.5	72.25
700	850	8	7	1	1
590	920	2	10.5	−8.5	72.25
620	800	4	6	−2	4
700	890	8	8	0	0
690	920	6	10.5	−4.5	20.25
900	1000	12	12	0	0
500	690	1	1.5	−0.5	0.25
610	700	3	3	0	0
700	760	8	5	3	9
				Total:	244

∴ Spearman Rank correlation:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$n = 12$

$$= 1 - \frac{6 \times 244}{12(12^2-1)}$$

$$= 0.146$$

b

from a. $r = 0.146$

$\therefore r^2 = 0.0216 = 2.16\%$.

$\therefore$ the value of r is between $0 - 0.25$, which exists in low degree positive correlative between judge 1 (x) and judge 2 (y).

and $r^2 = 2.16\%$ of variation of judge 2 (y) can be explained by judge 1 (x)

Problem 3

No of rooms (x)	Energy (y)	xy	x^2	y^2
12	9	108	144	81
9	7	63	81	49
14	10	140	196	100
6	5	30	36	25
10	8	80	100	64
8	6	48	64	36
10	8	80	100	64
10	10	100	100	100
5	4	20	25	16
7	7	49	49	49
Total: 91	= 74	= 718	= 895	= 584

a) $$b_1 = \frac{\Sigma x_i y_i - \left(\frac{\Sigma x \Sigma y}{n}\right)}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{718 - \frac{9 \times 74}{10}}{895 - \frac{(91)^2}{10}}$$

$$= 0.667$$

$$b_0 = \bar{y} - b_1 \bar{x} \Rightarrow \frac{\Sigma y}{n} - b_1 \frac{\Sigma x}{n}$$

$$\Rightarrow \frac{74}{10} - \frac{2}{3} \times \frac{91}{10}$$

$$= 1.333$$

$\therefore \hat{y} = 1.333 + 0.667x$ (Ans)

b from a,

$b_0 = 1.333$ which means the energy consumption will be 1.333 when the number of room is zero.

again, $b_1 = 0.667$ means that the averge energy consumption will increase by 0.667 (thousand kwh) when the number of room will increase by 1.

c the energy consumption, in thousand kwh for a six-room house will be,

$y = 1.333 + (0.667 \times 6)$

$= 5.33$ (Ans)

d $r^2 = 1 - \frac{SSE}{SST}$

here, $SSE = \Sigma y_i^2 - b_0 \Sigma y_i - b_1 \Sigma x_i y_i$

$= 584 - (1.33 \times 74) - (0.667 \times 718)$

$= 6.67$

$SST = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$

$= 584 - \frac{(74)^2}{10} = 36.4$

$\therefore r^2 = 1 - \frac{6.67}{36.4}$

$= 0.8167$

$\therefore (0.8167 \times 100)\%$ or 81.67% variation in energy consumption can be explained by the number of rooms.

(Ans)

Problem 4

a

$b_0 = 356.12083$

$b_1 = -0.09874$ [from the code]

$b_2 = 122.86721$

So the estimated regression equation is,

$$\hat{y} = 356.121 - 0.09874x + 122.8672x_2$$

(Ans)

b

from a,

$b_0 = 356.12083$

which means the price of backpack will be 356.12089, when the capacity and comfort rating are zero.

$b_1 = -0.098$ means the average price of backpack will increase by -0.098 when the capacity will increase by 1 unit and keeping comfort rating fixed.

and finally $b_2 = 122.867$ means that the average price of backpack will increase by 122.867 when the comfort rating will increase by 1 unit and keeping the capacity fixed.

c

with comfort rating of 4 :

$$\hat{y} = 856.121 - (0.09874 \times 4500) + (122.86721 \times 4)$$

$$= 403.2868$$

∴ the predicted price of the backpack is approximately 403.268 (Ans)

d

comment on goodness of fit of the model:

Given, $R^2 = 0.8318$

83.18 %. variation in total price can be explained by the capacity and the comfort ratings.

Adjusted $R^2 = 0.7838$

78.38 %. variation in total price for backpack can be explained by the capacity and comfort ratings.

Problem 5

a

Given,

$b_0 = -471.441$

$b_1 = 6.394$

$b_2 = 1.347$

∴ Estimated logistic regression equation:

$$\hat{y} = \frac{e^{-471.441 + 6.394x_1 + 1.347x_2}}{1 + e^{-471.441 + 6.394x_1 + 1.347x_2}}$$

b

from a, $b_1 = 6.394$

∴ odds radio $= e^{6.394} = 598.2448$

Again,

$b_2 = 1.347$

odds radio $= e^{1.347} = 3.8459$

b

The odds of having a second heart attack within one year increase by 598.24 for year of increase is the age of that patient and the odds of having a 2nd heart attack within one year = 3.846 for every unit anxiety increase.