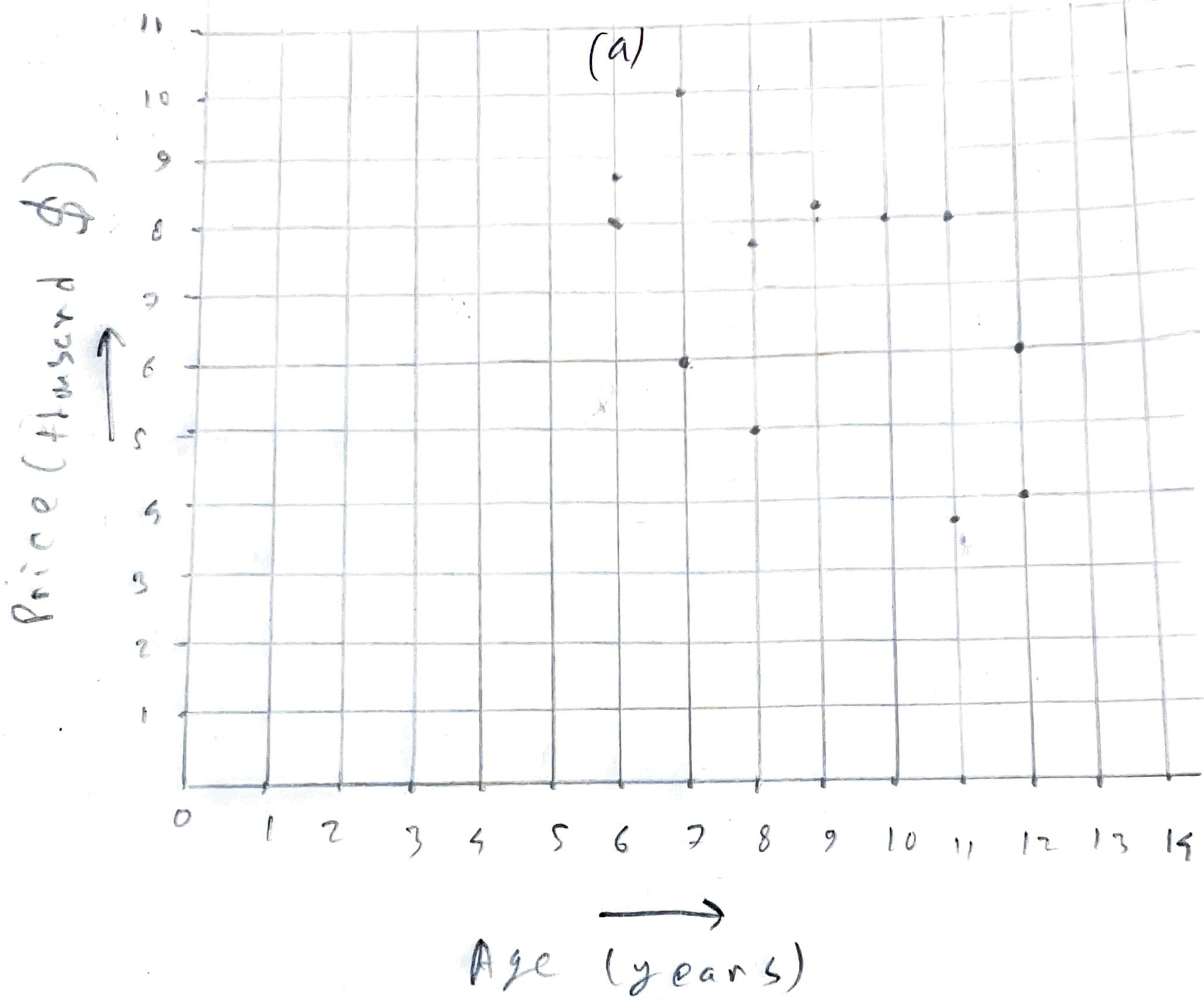
Name : Md Raihanul Islam Bhuiyan

Id : 20101239

Section: 10

Course : STA201

Ans to the Q: N: 7 (a)



There is a negative correlation between the age of a car and it's price.

(b)

Pearson correlation coefficient ~~cot~~ and coefficient of determination.

Age (years) x	Price (Thousands \$) y	x^2	y^2	xy
9	8.1	81	65.61	72.9
7	6	49	36	42
11	3.6	121	12.96	39.6
12	4	144	16	48
8	5	64	25	40
7	10	49	100	70
8	7.6	64	57.76	60.8
11	8	121	64	88
10	8	100	64	80
12	6	144	36	72
6	8.6	36	72.96	51.6
6	8	36	64	48
Σ(sum) 107	82.9	1009	615.29	712.9

$\sum x = 107.$

$\sum y = 82.9$

$\sum x^2 = 1009$

$\sum y^2 = 615.29$

$\sum xy = 712.9$

$$\therefore \bar{x} = \frac{\sum x}{n} = \frac{107}{12}$$

$$= 8.917$$

$$\therefore \bar{y} = \frac{\sum y}{n} = \frac{82.9}{12}$$

$$= 6.908$$

$$\therefore r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\left(\sum x^2 - n\bar{x}^2\right) \times \left(\sum y^2 - n\bar{y}^2\right)}}$$

$$= \frac{712.9 - 12 \times 8.917 \times 6.908}{\sqrt{(1009 - 12 \times 8.917^2)(615.29 - 12 \times 6.908^2)}}$$

$\therefore r = -0.5435$

$\therefore$ Pearson correlation coefficient, $r = -0.5435$

$\therefore$ The coefficient of determination, $r^2 = 0.2953$

$= 29.53\%$

$\therefore$ There is a negative correlation between the age and selling price of a car and 29.53% of the variation in the price can be explained by the age.

Ans to the Q: N: 2

(a)

Judge 1 (X)	Judge 2 (Y)	R_X	R_Y	d	d^2
650	900	8	4	4	16
760	720	2	9	-7	49
740	690	3	11.5	-8.5	72.25
700	850	5	6	-1	1
590	920	11	2.5	8.5	72.25
620	800	9	7	2	4
700	890	5	5	0	0
690	920	7	2.5	4.5	20.25
900	1000	1	1	0	0
500	690	12	11.5	0.5	0.25
610	700	10	10	0	0
710	760	5	8	-3	9

$\Sigma d^2 = 244$

Spearman rank Correlation,

$$r_s = 1 - \frac{6 \sum_{i=1}^{12} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 244}{12(144-1)}$$

$$= 0.14685$$ (Ans)

(b)

Association between the scores of the two judges is somewhat positive but very close to 0. So, it is a weak positive correlation.

Ans to the Q: N: 3

(a)

Number of Rooms (X)	Energy Consumption (thousand kWh) Y	X^2	Y^2	XY
12	9	144	81	108
9	7	81	49	63
14	10	196	100	140
6	5	36	25	30
10	8	100	64	80
8	6	64	36	48
10	8	100	64	80
10	10	100	100	100
5	4	25	16	20
7	7	49	49	49
Σ(Sum) $\Sigma x = 91$	$\Sigma Y = 74$	$\Sigma x^2 = 895$	$\Sigma Y^2 = 584$	$\Sigma XY = 718$

Regression equation for simple linear regression is,

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{n(\Sigma x_i y_i) - (\Sigma x_i)(\Sigma y_i)}{n(\Sigma x_i)^2 - (\Sigma x_i)^2}$$

$$= \frac{(10 \times 718) - (91 \times 74)}{10 \times 895 - (91)^2}$$

$$= 0.6667$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$= \frac{74}{10} - \left(0.6667 \times \frac{91}{10}\right)$$

$$= 1.333$$

$$\therefore \hat{y} = 1.333 + 0.6667x$$

(b)

Here, b_0 is the y-intercept of the regression line which measures the value of y when $x = 0$. B_1 is the slope which measures the average rate of change in dependent variable per unit change in independant variable.

$b_0 = 1.333$ indicates when number of rooms is 0, total energy consumption will be 1.333.

$b_1 = 0.6667$ indicates total energy consumption will increase by 0.6667 for increasing 1 room.

(c)

When room number is 0, $x = 0$

$\therefore \hat{Y} = 1.333 + (0.6667 \times 6)$

$= 5.3332$ (thousand kWh)

(d)

Goodness of the fit of the model is explained by r^2:

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

$\therefore SSE = \sum y_i^2 - b_0 \sum y_i - b_1 \sum x_i y_i$

$= 584 - (1.333 \times 74) - (0.6667 \times 718)$

$= 6.6674$

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

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

$$= 36.4$$

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

$$= 1 - \frac{6.6674}{36.4}$$

$$= 0.8168$$
$$= 81.68\%$$

$\therefore$ 81.68% of the variation in energy consumption can be explained by the number of rooms.

Ans to the Q: N: 4

(a)

Here, Estimated intercept, $b_0 = 356.12083$

Coefficient of x_1, $b_1 = -0.09875$

Coefficient of x_2, $b_2 = 122.86721$

$\therefore$ Regression equation,

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

$$\Rightarrow \hat{y} = 356.12083 - 0.09874 x_1 + 122.86721 x_2$$

(b)

Here, $b_0 = 356.12083$ indicates, when capacity (x_1) and comfort rating (x_2) are 0, the price of a backpack will be 356.12083.

$b_1 = -0.09874$ indicates when capacity (x_1) increases by 1 cubic inches, price will decrease by 0.09874 units when comfort rating(x_2) is constant.

$b_2 = 122.86721$ indicates when capacity (x_1) is fixed and comfort rating is increased by 1 point, the price is increased by 122.86721 units.

(c)

Capacity (x_1) = 4500 cubic inches

comfort rating (x_2) = 4

$\therefore \hat{Y} = 356.12083 - (0.09879 \times 4500) + (122.86721 \times 4)$

$= 403.25967$ (Ans)

(d)

Goodness of fit of the model is explained by Adjusted R-squared.

Adjusted R squared = 0.7838
= 78.38%

$\therefore$ 78.38% of the variation in the price of backpack can be explained by

~~Ale~~ the capacity and comfort rating.

Ans to the Q: N: 5

(a)

Here,

$b_0 = -471.441$

$b_1 = 6.394$

$b_2 = 1.347$

$$\therefore \hat{y} = \frac{e^{b_0 + b_1 x_1 + b_2 x_2}}{1 + e^{b_0 + b_1 x_1 + b_2 x_2}}$$

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

(Ans)

(b)

~~Q~~d

Here, ~~x~~ b_1 relates the probablity of a second heart attack within ~~o~~ the next 1 year with the age of the patient.

b_2 relates the second heart attack with the increase in the score of anxiety.

$b_1 = 6.394$

$\therefore$ odds ratio $= e^{b_1} = e^{6.394}$

$= 598.2448$

$\therefore$ odds of a second heart attack is increased by 598.2448 for every unit change of age in the patient.

$e^{b_2} = e^{1.347}$

$= 3.8459$

$\therefore$ Odds of a second heart attack within 1 year is increased by 3.8459 for every unit increase of the score in anxiety