Assignment - 02

STA201

Section :- 12

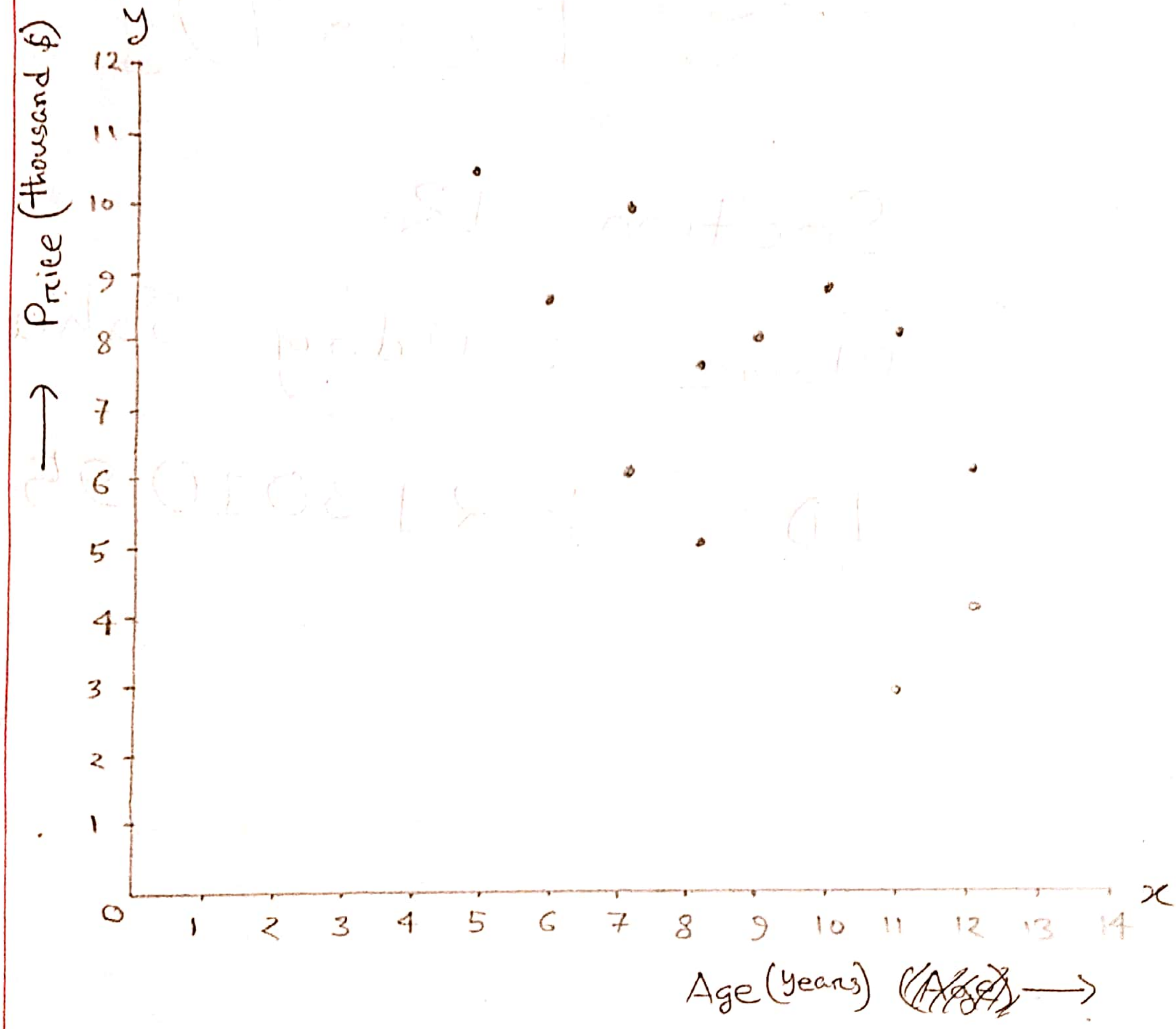
Name :- Udoy Saha

ID :- 21301095

Ans to the ques no:- 1

(a)

Age (Years)	9	7	11	12	8	7	8	11	10	12	6	5
Price (thousand $)	8.1	6	2.8	4	5	10	7.6	8	8.7	6	8.6	10.5



The diagram shows that, there is a low degree of negative correlation between age of the car and the selling price of that.

(b)

Age (years) (x)	Price (thousand \$) (y)	xy	x^2	y^2
9	8.1	72.9	81	65.61
7	6	42	49	36
11	2.8	30.8	121	7.84
12	4	48	144	16
8	5	40	64	25
7	10	70	49	100
8	7.6	60.8	64	57.76
11	8	88	121	64
10	8.7	87	100	75.69
12	6	72	144	36
6	8.6	51.6	36	73.96
5	10.5	52.5	25	110.25
$\Sigma x = 106$	$\Sigma y = 85.3$	$\Sigma xy = 715.6$	$\Sigma x^2 = 998$	$\Sigma y^2 = 668.11$

Here, $\bar{x} = \frac{\Sigma x}{n}$ and $\bar{y} = \frac{\Sigma y}{n}$

where, n = number of observations = 12

$\therefore \bar{x} = \frac{106}{12}$ $\therefore \bar{y} = \frac{85.3}{12}$

$= \frac{53}{6}$ $= \frac{853}{120}$

Now, Pearson Correlation Coefficient,

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

$$\therefore r = \frac{715.6 - 12 \times \frac{53}{6} \times \frac{853}{120}}{\sqrt{\left(998 - 12 \times \left(\frac{53}{6}\right)^2\right)\left(668.11 - 12 \times \left(\frac{853}{120}\right)^2\right)}}$$

$$= \frac{715.6 - \frac{45209}{60}}{\sqrt{\frac{185}{3} \times 61.76916667}}$$

$$= -0.61381$$

So, the Coefficien of Determination,

$$r^2 = (r)^2$$
$$= (-0.61381)^2$$
$$= 0.376768$$
$$= 37.6768\ \%$$

From 'r', we can see that there is a moderate degree of negative correlation between x and y.

From 'r^2', we can see that, 37.67681% data from y can be explained from x.

— o — x — o —

Ans to the ques no:- 2

(a)

Judge 1 (x)	Judge 2 (y)	R_x	R_y	$d = x - y$	d^2
650	920	5	10	−5	25
760	720	11	4	7	49
740	690	10	1.5	8.5	72.25
700	850	7.5	7	0.5	0.25
590	920	2	10	−8	64
620	800	4	6	−2	4
700	890	7.5	8	−0.5	0.25
690	920	6	10	−4	16
~~700~~ 950	1000	12	12	0	0
~~690~~ 500	690	1	1.5	−0.5	0.25
~~850~~ 610	700	3	3	0	0
~~5~~ 710	760	~~8~~ 9	5	4	16
					$\Sigma d^2 = 247$

$\therefore$ Spearman's rank correlation, $r_s = 1 - \dfrac{6 \Sigma d^2}{n(n^2-1)}$

Here, n = number of observations = 12

$$\therefore r_s = 1 - \frac{6 \times 247}{12(12^2-1)} = 0.1364$$

(b)

Here, $r = 0.1364$

and, $r^2 = 0.0185951$

$= 1.85951\%$

Here, the value of r shows that, there is a low degree of positive correlation between the scoring of the two judges.

And, r^2 shows that, ~~there~~ 1.85951% scorings of the Judge 2 can be explained by the scorings of Judge 1.

— o — x — o —

Ans to the ques no :- 3

(a)

Number of rooms (x)	Energy Consumption (thousand kWh) (y)	xy	x^2	y^2
13	9	117	169	81
9	7	~~[illegible]~~ 63	81	49
14	11	154	196	121
6	6	36	36	36
10	8	80	100	64
7	6	42	49	36
11	8	88	121	64
10	9	90	100	81
5	4	20	25	16
7	7	49	49	49
$\Sigma x = 92$	$\Sigma y = 75$	$\Sigma xy = 739$	$\Sigma x^2 = 926$	$\Sigma y^2 = 597$

The regression equation will be in the form,

$$\hat{y} = \beta_0 + \beta_1 x$$

Here, $\beta_1 = \dfrac{n \cdot \Sigma xy - (\Sigma x \cdot \Sigma y)}{n\, \Sigma x^2 - (\Sigma x)^2}$

Where, $n = 10$

$$\therefore \beta_1 = \frac{10 \times 739 - \cancel{926 \times 597}\ (92 \times 75)}{10 \times 926 - (92)^2}$$

$= 0.615578$

And, $\beta_0 = \bar{y} - \beta_1 \bar{x}$

Here, $\bar{x} = \frac{\Sigma x}{n} = \frac{92}{10} = 9.2$

$\bar{y} = \frac{\Sigma y}{n} = \frac{75}{10} = 7.5$

$\therefore \beta_0 = 7.5 - 0.615578 \times 9.2$

$= 1.836683$

$\therefore$ The regression equation becomes,

$\hat{y} = 1.836683 + 0.615578\,x$

$\boxed{(b)}$

From the equation, $\beta_0 = 1.836683$, which is the y-intercept means that, energy consumption (y) will be 1.836683 (thousand kWh) when number of rooms (x) is 0.

And, $\beta_1 = 0.615578$, which is the slope of the regression line denotes that, Energy consumption (y) will increase by 0.615578 (thousand kWh) with 1 unit increment of Number of rooms (x).

$\boxed{(C)}$

Here, number of rooms, $n = 7$

$$\therefore \hat{y} = 1.836683 + 0.615578 \times 7$$
$$= 6.145729$$

$\therefore$ For a 7 room house, the predicted energy consumption is 6.145729 (thousand) kWh.

(d)

We can know about the goodness of fit for the model through the coefficient of determination, r^2.

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

Where, $SSE = \Sigma y^2 - \beta_0 \Sigma y - \beta_1 \Sigma xy$

$= 597 - 1.836683 \times 75 - 0.615578 \times 739$

$= 4.336683417$

And, $SST = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$

$= 597 - \frac{75^2}{10}$

$= 34.5$

$\therefore r^2 = 1 - \frac{4.336683417}{34.5}$

$= 0.874299$

$= 87.4299\%$

$\therefore$ 87·4299 % of the variation can be explained through the model.

So, this is nearly a good fit or almost perfect fit.

— o —x— o —

Ans to the ques no:- 4

(a)

Let, capacity of the bag (cubic inches) = $x1$

comfort rating = $x2$

predicted price = $\hat{y}$

So, the regression equation will be,

$$\hat{y} = \beta_0 + \beta_1 x1 + \beta_2 \cdot x2$$

From the given data, plugging in the values for all parameters,

$$\hat{y} = 356·12083 - 0·09874·x1 + 122·86721·x2$$

(b)

From the progression model,

$\beta_0 = 356.12083$ means that, $\hat{y}$ will be 0, ie, predicted price will be 365.12083 when $x1$ and $x2$ both will be 0.

$\beta_1 = -0.09874$ means, if the comfort rating is constant, when the capacity of the bag increases by 1 unit, the predicted price decreases by 0.09874 unit.

$\beta_2 = 122.86721$ means, if capacity of the bag is constant, when the comfort rating increases by 1, the predicted price of the bag increases by 122.86721 unit.

(c)

Here, capacity = 5500 cubic inches

comfort rating = 4·5

$\therefore \hat{y} = 356.12083 - 0.09874 \times 5500 + 122.86721 \times 4.5$

$= 365.953275$

$\therefore$ Predicted price, $\hat{y} = 365.953275$ unit

(d)

Given that, R-squared $= r^2 = 0.8318$

$= 83.18\%$

There is a high percentage of variation which can be explained by our model.

So, the regression model is a good fit.

—o—x—o—

Ans to the ques no :- 5

$\boxed{(a)}$

Let,

Age = $x1$

Anxiety scale = $x2$

$\therefore$ the logistic equations general form,

$$E(y) = \frac{e^{\beta_0 + \beta_1 \cdot x1 + \beta_2 \cdot x2}}{1 + e^{\beta_0 + \beta_1 \cdot x1 + \beta_2 \cdot x2}}$$

Given, ~~~~ $\beta_0 = -471.441$

$\beta_1 = 6.394$

$\beta_2 = 1.347$

$$\therefore E(y) = \frac{e^{-471.441 + 6.394 \times x1 + 1.347 \times x2}}{1 + e^{-471.441 + 6.394 \times x1 + 1.347 \times x2}}$$

(b)

From the regression equation,

$\beta_0 = -471.441$ doesn't have so much significance in logistic regression.

$\beta_1 = 6.394$, so, odds ratio for age variable is, $e^{\beta_1} = e^{6.394}$

$= 598.2447792$

So, the odds of having a second heart ~~within 1 year~~ attack within 1 year increases by 598.2447792 with every unit increase in Age (keeping Anxiety level constant). As it is so larger than 1, so, age has a very heavy impact on heart attack.

Now, $\beta_2 = 1.347$, so, odds ratio for Anxiety scale is, $e^{\beta_2} = e^{1.347}$

$= 3.84587059$

Therefore, odds of having a second heart attack within 1 year Increases by 3·845870595 with every unit increase in Anxiety scale, while Age is constant. As it is larger than 1, so, Anxiety has a positive impact, ie, incremental impact in having heart attack.

—o—x—o—