# MA 542 : Regression Analysis HW-I - Solutions.

Q1] a) $\sum Y_i = \sum \hat{Y}_i$

Proof:

consider, $\sum Y_i - \sum \hat{Y}_i$

$$= \sum (Y_i - (b_0 + b_1 X_i))$$

$$= \sum Y_i - nb_0 + b_1 \sum X_i$$

$$= 0 \quad (\because \text{Normal equation } 1).$$

$$\therefore \sum Y_i = \sum \hat{Y}_i.$$

b) $\sum X_i e_i = 0$

Proof:

$$\sum X_i e_i = \sum X_i (Y_i - b_0 - b_1 X_i)$$

$$= \sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2$$

$$= 0 \quad (\because \text{Normal Equation } 2).$$

c) $\sum \hat{Y}_i e_i = 0$

Proof

$$\sum \hat{Y}_i e_i = \sum (b_0 + b_1 X_i) e_i$$

$$= b_0 \underbrace{\sum e_i}_{=0} + b_1 \underbrace{\sum X_i e_i}_{=0}$$

$$= 0 \quad (\because \sum e_i = 0 \text{ and } \sum X_i e_i = 0).$$

**Q2**

a)

| $X_i$ | $Y_i$ | $X_i - \bar{X}$ | $Y_i - \bar{Y}$ | $(X_i-\bar{X})(Y_i-\bar{Y})$ | $(X_i-\bar{X})^2$ |
|---|---|---|---|---|---|
| 1 | 16 | 0 | 1.8 | 0 | 0 |
| 0 | 9 | −1 | −5.2 | 5.2 | 1 |
| 2 | 17 | 1 | 2.8 | 2.8 | 1 |
| 0 | 12 | −1 | −2.2 | 2.2 | 1 |
| 3 | 22 | 2 | 7.8 | 15.6 | 4 |
| 1 | 13 | 0 | −1.2 | 0 | 0 |
| 0 | 8 | −1 | −6.2 | 6.2 | 1 |
| 1 | 15 | 0 | 0.8 | 0 | 0 |
| 2 | 19 | 1 | 4.8 | 4.8 | 1 |
| 0 | 11 | −1 | −3.2 | 3.2 | 1 |
| $\Sigma X_i = 10$ | $\Sigma Y_i = 142$ | | | $\Sigma (X_i-\bar{X})(Y_i-\bar{Y}) = 40$ | $\Sigma (X_i-\bar{X})^2 = 10$ |

$\Rightarrow \bar{X} = 1 \qquad \rightarrow \bar{Y} = 14.2$

$$b_1 = \frac{\Sigma (X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma (X_i-\bar{X})^2} = \frac{40}{10} = 4$$

$$b_0 = \bar{Y} - b_1\bar{X} = 14.2 - 4 \times 1 = 10.2$$

∴ Estimated Regression Function is $\hat{Y} = 10.2 + 4X$.

b) $E(Y) = 10.2 + 4(1)$

$= 14.2$.

c) $E(Y)_{X=2} - E(Y)_{X=1} = b_1 = 4$.

d) $Y = 10.2 + 4(\bar{X}) = 10.2 + 4(1) = 14.2 = \bar{Y}$

∴ Regression line $Y = \overset{10.2}{\cancel{b_0 + b_1}} X$ goes throug $(\bar{X}, \bar{Y}) = (1, 14.2)$.

Q3 

| $X_i$ | $Y_i$ | $X_i-\bar{X}$ | $Y_i-\bar{Y}$ | $(X_i-\bar{X})(Y_i-\bar{Y})$ | $(X_i-\bar{X})^2$ |
|---|---|---|---|---|---|
| 16 | 199 | -12 | -26.5625 | 318.75 | 144 |
| 16 | 205 | -12 | -20.5625 | 246.75 | 144 |
| 16 | 196 | -12 | -29.5625 | 354.75 | 144 |
| 16 | 200 | -12 | -25.5625 | 306.75 | 144 |
| 24 | 218 | -4 | -7.5625 | 30.25 | 16 |
| 24 | 220 | -4 | -5.5625 | 22.25 | 16 |
| 24 | 215 | -4 | -10.5625 | 42.25 | 16 |
| 24 | 223 | -4 | -2.5625 | 10.25 | 16 |
| 32 | 237 | 4 | 11.4375 | 45.25 | 16 |
| 32 | 234 | 4 | 8.4375 | 33.75 | 16 |
| 32 | 235 | 4 | 9.4375 | 37.75 | 16 |
| 32 | 230 | 4 | 4.4375 | 17.75 | 16 |
| 40 | 250 | 12 | 24.4375 | 293.25 | 144 |
| 40 | 248 | 12 | 22.4375 | 269.25 | 144 |
| 40 | 253 | 12 | 27.4375 | 329.25 | 144 |
| 40 | 246 | 12 | 20.4375 | 245.25 | 144 |

$\Sigma X_i = 448$ $\quad \Sigma Y_i = 3609$ $\quad \Sigma(X_i-\bar{X})(Y_i-\bar{Y}) = 2604$ $\quad \Sigma(X_i-\bar{X})^2 = 1280$

$\Rightarrow \bar{X} = 28$ $\quad \Rightarrow \bar{Y} = 225.5625$

a) $b_1 = \dfrac{2604}{1280} = 2.034$ , $\quad b_0 = \bar{Y} - b_1\bar{X} = 225.5625 - 2.034 \times 28$

$= 168.61$

$\therefore$ Estimated Regression line : $\hat{Y} = 168.61 + 2.034X$.

b) when $X = 40$,

$E(Y) = 168.61 + 2.034(40) = 249.97$.

c) change in mean hardness when $X$ is increased by 1 hour is $\hat{\beta}_1 = b_1 = 2.034$.

Q4

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 199 | 205 | 196 | 200 | 218 | 220 | 215 | 223 | 237 | 234 |
| $\hat{y}_i$ | 201.15 | 201.15 | 201.15 | 201.15 | 217.43 | 217.43 | 217.43 | 217.43 | 233.70 | 233.70 |
| $e_i$ | -2.15 | 3.85 | -5.15 | -1.15 | 0.57 | 2.57 | -2.43 | 5.57 | 3.30 | 0.30 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $y_i$ | 235 | 230 | 250 | 248 | 253 | 246 |
| $\hat{y}_i$ | 233.70 | 233.70 | 249.98 | 249.98 | 249.98 | 249.98 |
| $e_i$ | 1.30 | -3.70 | 0.02 | -1.98 | 3.02 | -3.98 |

$$\sum e_i = -2.15 + 3.85 - 5.15 - 1.15 + 0.57 + 2.57 - 2.43 + 5.57 + 3.30 + 0.30 + 1.30 - 3.70 + 0.02 - 1.98 + 3.02 - 3.98 = \underline{0.00}$$

Q5 If $\beta_1 = 0$, then there is no ~~a~~ linear relationship between the response variable and the predictor variable.

Regression function with $\beta_1 = 0$ is a horizontal line,

Q6
$$\sum y_i = nb_0 + b_1 \sum x_i \longrightarrow (1)$$
$$\sum x_i y_i = b_0 \sum x_i + b_1 \sum x_i^2 \longrightarrow (2)$$

(2) $\times n$ − (1) $\times \sum x_i$ :

$$n\sum x_i y_i - \sum x_i \sum y_i = nb_1 \sum x_i^2 - b_1 (\sum x_i)^2$$

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

$$= \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}$$

$$b_1 = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i - \frac{1}{n}\sum x_i \sum y_i + \frac{1}{n}\sum x_i y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2 - \frac{1}{n}(\sum x_i)^2 + \frac{1}{n}(\sum x_i)^2}$$

$$= \frac{\sum x_i y_i - \bar{x}\sum \cancel{x}y_i - \bar{y}\sum x_i + n\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}$$

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

Q7) $y_i = \beta_0 + \varepsilon_i$

consider,

$$Q = \sum_{i=1}^{n} (y_i - \beta_0)^2$$

d.w.r.t. $\beta_0$,

$$\frac{dQ}{d\beta_0} = 2\sum (y_i - \beta_0)(-1) \overset{\text{set}}{=} 0$$

$$\Rightarrow \sum y_i - n\beta_0 = 0$$

$$\Rightarrow \hat{\beta}_0 = b_0 = \frac{\sum y_i}{n} = \bar{y} .$$

$$\frac{d^2 Q}{d\beta_0^2} = -2n < 0$$

$\therefore$ Least square estimator of $\beta_0 = \hat{\beta}_0 = b_0 = \bar{y}$.

Q8

$$E(b_0) = E(\bar{Y}) = E\left(\frac{\sum Y_i}{n}\right)$$
$$= \frac{1}{n}\sum E(Y_i) = \frac{1}{n}\sum E(\beta_0 + \varepsilon_i) \quad (\because Y = \beta_0 + \varepsilon_i)$$
$$= \frac{1}{n}\sum [\beta_0 + E(\varepsilon_i)]$$
$$= \frac{1}{n}[n\beta_0 + 0] = \beta_0$$

i.e. $E(b_0) = \beta_0$

$\therefore$ $b_0$ is an unbiased estimator.