Name: Abdullah al Tamim

ID: 2020-1-60-127

Assignment-4

1) Peanson's connelation coefficient:-

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = \frac{865 - 812}{70.46}$$

$$= 0.75$$

So, there is a strong positive linean nelationship between x and y. That is, both inencase on deenease in the same dinection.

2] i] We know,

Estimated regression equation, $\hat{y} = a + bx$

Here,

sales in million (independent) $= x$

earning (dependent) $= y$

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{14084.9 - 11172}{}$$

$= \$0.06$ million per \$1 million sales.

If the sales increased by \$1 million, the expected earnings will be increased by \$60,000.

$$a = \frac{\Sigma y}{n} - b\frac{\Sigma x}{n} = \$1.71 \text{ million}$$

If there is no effect of x on y, then the average earning is \$1.71 million.

$\therefore \hat{y} = 1.71 + 0.06x$

ii) From (i): $\hat{y} = 1.71 + 0.06x$

If the sales of a small company, $x = \$50$ million

then the estimated earnings will be, $\hat{y} = 1.71 + 0.06 \times 50$

$= \$4.71$ million.

iii) We know,

Standard error of estimate, $S_y = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}} = \sqrt{\frac{15.97}{6}}$

$= \$1.63$ million

The difference between the actual and estimated earnings is $\$1.63$ million.

iv) We know,

Coefficient of determination, $R^2 = 1 - \frac{SSE}{SST}$.

$SSE = \Sigma(y-\hat{y})^2 = 15.97$

$SST = \Sigma(y-\bar{y})^2 = 37.96$

$\therefore R^2 = 1 - \frac{15.97}{37.96} = 0.58$

The variation in Y is 58% explained by the variation in X.

ii) From (i): $\hat{y} = 1.71 + 0.06x$

If the sales of a small company, $x = \$50$ million

then the estimated earnings will be, $\hat{y} = 1.71 + 0.06 \times 50$

$= \$4.71$ million.

iii) We know,

Standard error of estimate, $S_y = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}} = \sqrt{\frac{105.97}{6}}$

$= \$1.63$ million

The difference between the actual and estimated earnings is \$1.63 million.

iv) We know,

Coefficient of determination, $R^2 = 1 - \frac{SSE}{SST}$

$SSE = \Sigma(y-\hat{y})^2 = 15.97$

$SST = \Sigma(y-\bar{y})^2 = 37.96$

$\therefore R^2 = 1 - \frac{15.97}{37.96} = 0.58$

The variation in Y is 58% explained by the variation in X.

3) i) We know,

$$\text{Pearson's coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

Data in ascending order: 32, 32, 32, 37, 37, 40, 40, 40, 40, 40, 42, 46, 46, 46, 46, 46, 46, 52, 52, 52, 52, 52, 52.

Mean, $\bar{x} = 43.48$

Median, $= \frac{n+1}{2}$ th position $=$ 12th position $= 46$

Standard deviation, $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{1015.74}{22}}$

$= 6.79$

$\therefore$ Coefficient of skewness $= \frac{3(43.48-46)}{6.79}$

$= -1.11$

So the shape of the data is left skewed. That is the number of higher numbers of customers is higher.

ii) To draw a box plot we need the followings:-

Minimum $= 32$

$Q_1 = P_{25} = \frac{n+1}{100} \times 25 = 6$th pos $\therefore Q_1 = 40$

$Q_2 = \frac{n+1}{100} \times 50 = 12\text{th pos} \quad \therefore Q_2 = 46$

$Q_3 = P_{75} = \frac{n+1}{100} \times 75 = 18\text{th pos} \quad \therefore Q_3 = 52$

Maximum $= 52$

Upper limit $= Q_3 + 1.5 \times IQR = 70$

Lower limit $= Q_1 - 1.5 \times IQR = 22$

No value crosses the limit so there are no outliers.

Box plot:

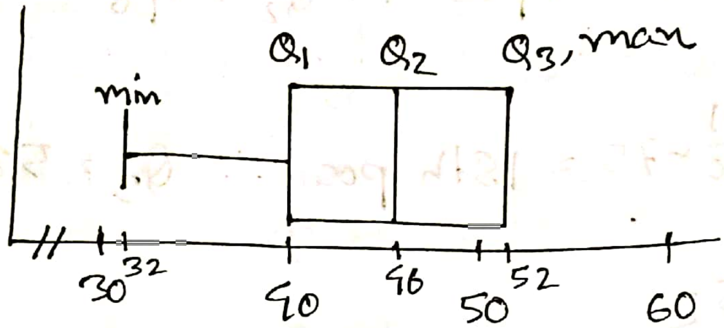

Fig: Box plot of number of visitors in Smith's True Value Hardware store:

4] i] The estimated median of the box plot is \$450. The tuition per credit hour of 50% public colleges is less than \$450 and the rest is above it.

ii] The estimated first quartile, Q_1 = \$300.

The tuition per credit hour of 25% public colleges is less than \$300 and the rest is above it.

The estimated third quartile, Q_3 = \$740.

The tuition per credit hour of 75% public colleges is less than \$740 and the rest is above it

iii) From (ii) : $Q_1 = \$300$, $Q_3 = \$740$

$\therefore IQR = Q_3 - Q_1 = 740 - 300 = \$440$

The interquartile range of tuition per credit hour of public colleges is \$440.

iv) A value will be considered as an outlier if that is less than \$0 or more than \$1340.

v) There is an outlier in the boxplot and the estimated value of that outlier ~~point~~ is \$1500.

vi) The distribution ~~is~~ is positively or right skewed. Because in the graph the difference between median and Q_3 is higher than median and Q_1.