4- Forming confidence limits for the mean response at $X_0 = 40$.

By solving the above example, the required values for the confidence limits for the mean response are as follows:

First, we substitute the value of $X_0 = 40$ into the estimated regression line equation to obtain $\hat{y}_0$ as follows:

$$\hat{y}_{X=40} = 85.043 + 1.1396(40)$$
$$\hat{y}_{X=40} = 130.627$$

The average response variance will be:

$$V(\hat{y}_0) = \frac{\hat{\sigma}^2}{n} + (X_0 - \bar{X})^2 \frac{\hat{\sigma}^2}{S_{XX}}$$
$$V(\hat{y}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{XX}} \right)$$
$$V(\hat{y}_0) = (77.9)\left(\frac{1}{10} + \frac{(40 - 49.1)^2}{2048.9} \right)$$
$$V(\hat{y}_0) = 10.938$$

The standard deviation of the variance of the average response will be:

$$\sqrt{V(\hat{y}_0)} = \sqrt{10.938} = 3.31$$

Thus, the confidence limits for the mean response at point $X_0 = 40$ are as follows:

$$\hat{y}_0 - t(\frac{\alpha}{2}, n-2) S(\hat{y}_0) \le E(y / X_0) \le \hat{y}_0 + t(\frac{\alpha}{2}, n-2) S(\hat{y}_0)$$
$$130.6 - (2.306)(3.31) \le E(y / X_0) \le 130.6 + (2.306)(3.31)$$
$$123.0 \le E(y / X_0) \le 138.2$$

Regression through the origin

In some cases of regression, the regression line may pass through the origin. In this case, the linear model is without the parameter β_0 (No intercept).

$$y_i = \beta_1 x_{i1} + e_i \quad , \ i = 1, 2, ..., n$$

That is, the equation of the regression line is as shown in the following figure:

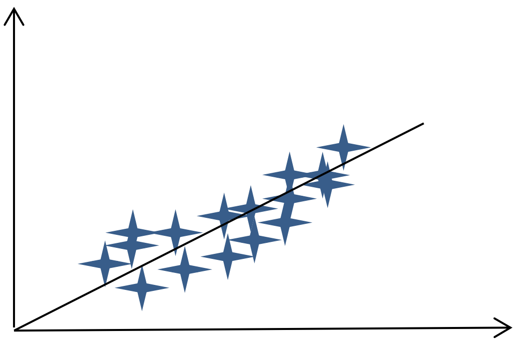


Figure (5) shows the relationship between x and y represented by a straight line.

To find the estimate of parameter β_1 using the least squares method, we follow the following

From the above equation we find that:

$$e_i = y_i - \beta_1 x_i$$

By squaring both sides and taking their sum, we get:

$$Q = \sum e_i^2 = \sum (y_i - \beta_1 x_i)^2$$
$$= \sum (y_i^2 - 2\beta_1 x_i y_i + \beta_1^2 X_i^2)$$
$$= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1^2 \sum X_i^2$$

We differentiate with respect to β_1 to get:

$$\frac{\partial Q}{\partial \beta_1} = 0 - 2\sum x_i y + 2\beta_1 \sum X_i^2$$

Thus:

$$2\beta_1 \sum X_i^2 = 2\sum x_i y$$

Thus, the estimated parameter $\hat{\beta}_1$ will be:

$$\hat{\beta}_1 = \frac{\sum x_i y}{\sum X_i^2}$$

Thus:

$$\hat{y}_i = \hat{\beta}_1 x_i$$

The estimated value of the variance of the population σ^2 can be obtained as follows:

$$SSe = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$$
$$= \sum (y_i - \hat{\beta}_1 x_i)^2$$
$$= \sum (y_i^2 - 2\beta_1 x_i y_i + \beta_1^2 X_i^2)$$
$$= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1^2 \sum X_i^2$$
$$= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1 \frac{\sum x_i y}{\sum X_i^2} \sum X_i^2$$
$$= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1 \sum x_i y$$
$$\therefore SSe = \sum y_i^2 - \beta_1 \sum x_i y$$

So MSe in this case will be:

$$\therefore MSe = \frac{SSe}{n-1} = \frac{\sum y_i^2 - \beta_1 \sum x_i y}{n-1}$$

The variance of parameter β_1 will be as follows:

$$S_{\hat{\beta}_1}^2 = \frac{MSe}{\sum X_i^2}$$

The question now is which model do we use, with β_0 or without it?
In this case we suggest the following:

1- Draw the scatter plot which gives some indication about the regression line in terms of whether it may pass through the origin or not.
2- Use the model that contains β_0 and then test whether $H_0: \beta_0 = 0$ or not, if the null hypothesis is accepted then use the model that does not contain β_0 as it may be more suitable for the data.
3- Use both models and compare the mean squares of the residuals MSe, the model that has a lower MSe is the best fit for the data.

Example:

The following data represents the number of units of production manufactured X and their different total labor costs y during a certain period and taken from 12 similar factories.

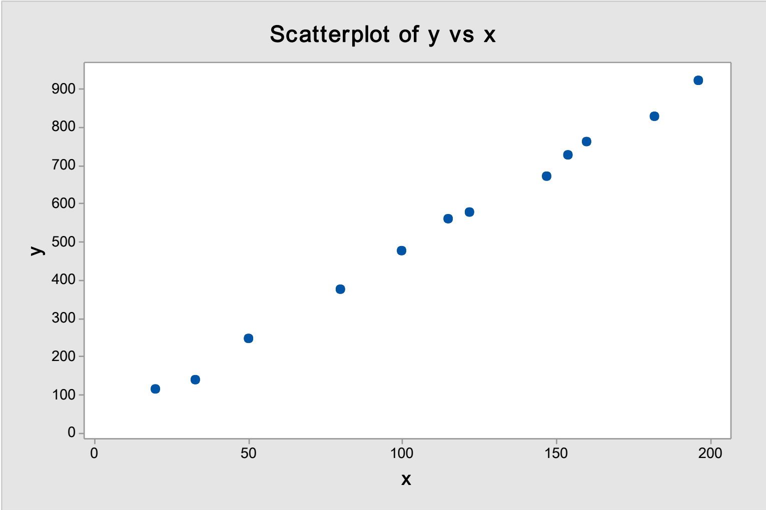
y_i^2	X_i^2	$X_i y_i$	X_i	y_i	ت
12996	400	2280	20	114	1
848241	38416	180516	196	921	2
313600	13225	64400	115	560	3
60025	2500	12250	50	245	4
330625	14884	70150	122	575	5
225625	10000	47500	100	475	6
19044	1089	4554	33	138	7
528529	23716	111958	154	727	8
140625	6400	30000	80	375	9
448900	21609	98490	147	670	10
685584	33124	150696	182	828	11
580644	25600	121920	160	762	12
4194438	**190963**	**894714**	**1359**	**6390**	**المجموع**

Note here that when there are no units produced, the cost is zero, i.e. when X=0, Y=0, so the model without the β_0 parameter may be more suitable for the data. In addition, if we notice the scatter plot, it indicates that the regression line may pass through the origin (0,0).

Sol:

1-We draw the scatter diagram between X and Y as follows:

From the figure, it is noted that most of the pattern of points indicates that they pass through the origin. Thus, we will estimate the model without the parameter.



2-We find the columns for each of $X_i y_i$, X_i^2 and y_i^2.

Thus, the estimated parameter $\hat{\beta}_1$ is as follows:

$$\hat{\beta}_1 = \frac{\sum x_i y}{\sum X_i^2} = \frac{894714}{190963} = 4.685$$

Thus, the equation of the regression line is as follows:

$$\hat{y}_i = 4.685 x_i$$

The estimate of the variance of the population σ^2 will be as follows:

$$\therefore MSe = \frac{SSe}{n-1} = \frac{\sum y_i^2 - \beta_1 \sum x_i y}{n-1}$$
$$= \frac{4194438 - (4.685)(894714)}{12-1}$$
$$= 223.399$$

The variance of parameter $\hat{\beta}_1$ is as follows:

$$S_{\hat{\beta}_1}^2 = \frac{MSe}{\sum X_i^2}$$
$$S_{\hat{\beta}_1}^2 = \frac{223.399}{190963}$$
$$S_{\hat{\beta}_1}^2 = 0.00117$$

The confidence limits will be as follows:

$$\hat{\beta}_1 - t(\frac{\alpha}{2}, n-2) S(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t(\frac{\alpha}{2}, n-2) S(\hat{\beta}_1)$$
$$4.685 - (2.201)(0.0342) \le \beta_1 \le 4.685 + (2.201)(0.0342)$$
$$4.61 \le \beta_1 \le 4.76$$

If we use the linear model which contains the parameter β_0, we find that:

$$S_{XX} = 37056.25$$
$$S_{yy} = 791763$$
$$\hat{\beta}_0 = 9.75$$
$$\hat{\beta}_1 = 4.615$$
$$\hat{y}_i = 9.75 + 4.62 X_i$$
$$SSe = 2236.13$$
$$MSe = \frac{2236.13}{10} = 223.612$$
$$S_{\hat{\beta}_0} = 9.799$$

If we conduct a t-test on parameter β_0 according to the following hypothesis:

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \ne 0$$

This will result in the following: