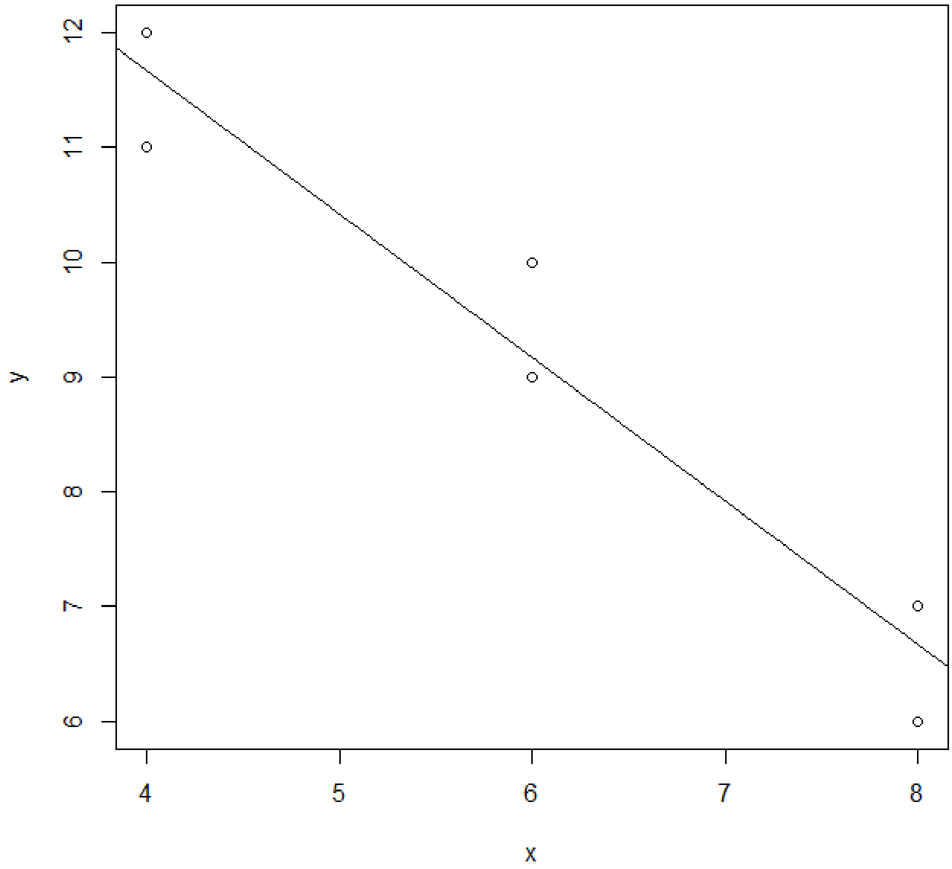
2. Consider the following data:

x	4	4	6	6	8	8
y	12	11	10	9	7	6

(a) Draw a scatter diagram. Does the model $y = \alpha + \beta x$ seem appropriate?
(b) Estimate the linear regression line. Explain the meaning of the parameter(s) in the model.
(c) Would we conclude that there is a negative linear relationship between between the x and y values of the population? Use $\alpha = 0.01$.
(d) Find a 99% confidence interval for the slope parameter β.
(e) Find the correlation coefficient.

Soln

(a)



Yes the model seems appropriate since there's a linear trend between x and y.

(b)

$$\sum_{i=1}^{n} x_i = 36 \qquad \sum_{i=1}^{n} x_i^2 = 232 \qquad n = 6$$
$$\sum_{i=1}^{n} y_i = 55 \qquad \sum_{i=1}^{n} y_i^2 = 531 \qquad \sum_{i=1}^{n} x_i y_i = 310$$

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$
$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\begin{aligned} S_{xy} &= \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n} \\ &= 310 - \frac{36(55)}{6} \\ &= -20 \end{aligned}$$

$$\begin{aligned} S_{xx} &= \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \\ &= 232 - \frac{(36)^2}{6} \\ &= 16 \end{aligned}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{-20}{16} = -1.25$$

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\ &= \frac{\sum_{i=1}^{n} y_i}{n} - \hat{\beta}\,\frac{(\sum_{i=1}^{n} x_i)}{n} \\ &= \frac{55}{6} - (-1.25)\left(\frac{36}{6}\right) \\ &= 16.6667 \end{aligned}$$

$$\therefore \ \hat{y} = 16.67 - 1.25x$$

On average y reduces by 1.25 when x increases by one unit.

Note: the intercept has no specific meaning in this case.

(c)

1. Testing problem

$$H_0: \beta = 0$$
$$H_1: \beta < 0$$

2. Test statistic

$$T = \frac{\hat{\beta} - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$$

3. Critical region

Reject H_0 if

$$t^* < -t_{\alpha,n-2} = -t_{0.01,4} = -3.747$$

4. Calculation

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-2}$$

$$SSE = S_{yy} - \hat{\beta}S_{xy}$$

$$\begin{aligned} S_{yy} &= \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} \\ &= 531 - \frac{55^2}{6} \\ &= 26.833333 \end{aligned}$$

$$SSE = 26.8333333 - (-1.25)(-20) = 1.8333333$$

$$\hat{\sigma}^2 = \frac{1.833333}{4} = 0.4583333$$

$$t^* = \frac{\hat{\beta} - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} = \frac{-1.25 - 0}{\sqrt{\frac{0.4583333}{16}}} = -7.385$$

i.e. $t^* < -3.747$

5. Conclusion

Reject H_0 at $\alpha = 0.01$ level of significance. Therefore we conclude that there is a significant linear relationship between x and y.

(d)

$$\hat{\beta} \pm t_{\frac{\alpha}{2},n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

$$-1.25 \pm t_{0.005,4}\sqrt{\frac{0.45833333}{16}}$$

$$-1.25 \pm 4.604\sqrt{\frac{0.45833333}{16}}$$

$$(-2.029, -0.471)$$

(e)

$$\begin{aligned} r &= \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \\ &= \frac{-20}{\sqrt{16(26.833333)}} \\ &= -0.965 \end{aligned}$$

OR

$$\begin{aligned} r &= sign\sqrt{R^2} \\ &= -\sqrt{1 - \frac{SSE}{SST}} \\ &= -\sqrt{1 - \frac{SSE}{S_{yy}}} \\ &= -\sqrt{1 - \frac{1.8333333}{26.833333}} \\ &= -0.965 \end{aligned}$$

Note:
The sign of the correlation in this example is negative since there is a negative linear relationship between x and y.