Exercise 6 A worldwide association of geophysicians is searching for diamonds for 12 days in a row and suspects that amount of diamonds found per day is affected by the amount it rains.
Measuring the amount of rainfall and the amount of diamonds found, we get:

	1	2	3	4	5	6	7	8	9	10	11	12
Diamonds found p. d.	215	325	185	332	406	522	412	614	544	421	445	408
Rainfall	14.2	16.4	11.9	15.2	18.5	22.1	19.4	25.1	23.4	18.1	22.6	17.2

a) Carry out a linear regression by calculating the regression coefficients.

b) Assume all conditions for hypothesis testing are satisfied and check whether $H_0 : \beta_1 = 0$ can be rejected at a 5%-significance level.

c) State the confidence interval for the hypothesis test.

Solution for proposal

a) First we compute the sample means, sample standard deviations and the sample covariance of the random variable $Y :=$ Diamond and $X :=$ Rain.

$$\bar{y} = 402.42,\ \bar{x} = 18.68, cov(X,Y) = 484.09$$

For $r = r_{XY} = \frac{cov(X,Y)}{s_X \cdot s_Y}$ we have

$$\hat{\beta}_1 = r \cdot (\frac{s_Y}{s_X}) = \frac{\sum_{i=1}^{12}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{12}(x_i - \bar{x})^2}$$

and so we get

$$\hat{\beta}_1 = 30.09$$

Furthermore we have that

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 402.42 - 30.09 \cdot 18.68 = -159.66$$

b) Recall, that the the residual variance is definded as follows

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

and

$$\hat{\sigma}_{\hat{\beta}_1} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

and

$$\hat{\sigma}_{\hat{\beta}_0} = \hat{\sigma}\frac{\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

Note that

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated coefficients.
We calculate

$$\hat{\sigma} = \sqrt{\frac{1}{10}\sum_{i=1}^{12}(y_i - \hat{y}_i)^2} = 38.13$$

$$\hat{\sigma}_{\hat{\beta}_1} = 2.87$$

Our test problem is

$$H_0 : \beta_1 = 0 \;\; VS \;\; H_1 : \beta_1 \neq 0$$

Now we use our statistic:

$$t = \frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{30.09}{2.87} = 10.48$$

Since $|t| = 10.49 > 2.23 = t_{0.975}(10)$, we reject the null hypothesis.

c) Our confidence interval for $\hat{\beta}_1$ is given by

$$[\hat{\beta}_1 \pm \hat{\sigma}_{\hat{\beta}_1} \cdot t_{1-\frac{\alpha}{2}}(n-2)] = [30.09 \pm 2.87 \cdot 2.23] = [23.69, 36.49]$$