MAT 2377
Probability and Statistics for Engineers

Practice Set

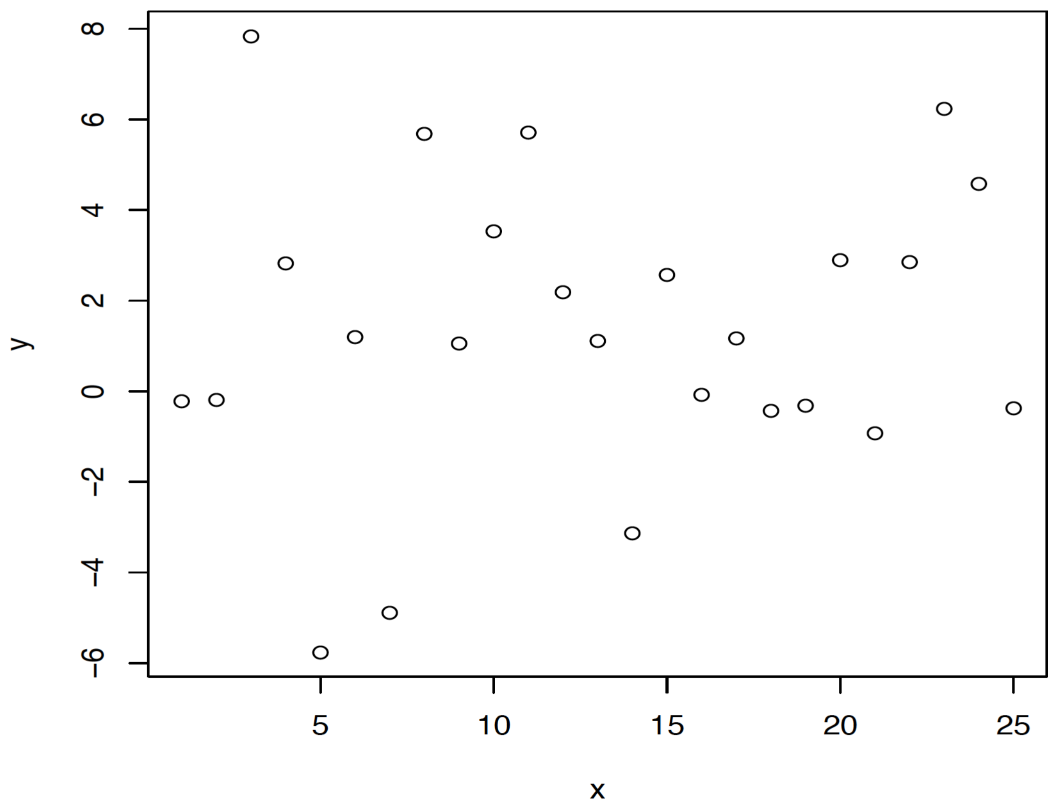
P. Boily (uOttawa)

Winter 2021

Based on course notes by Rafał Kulik

Q141. For the following data the correlation coefficient is most likely to be

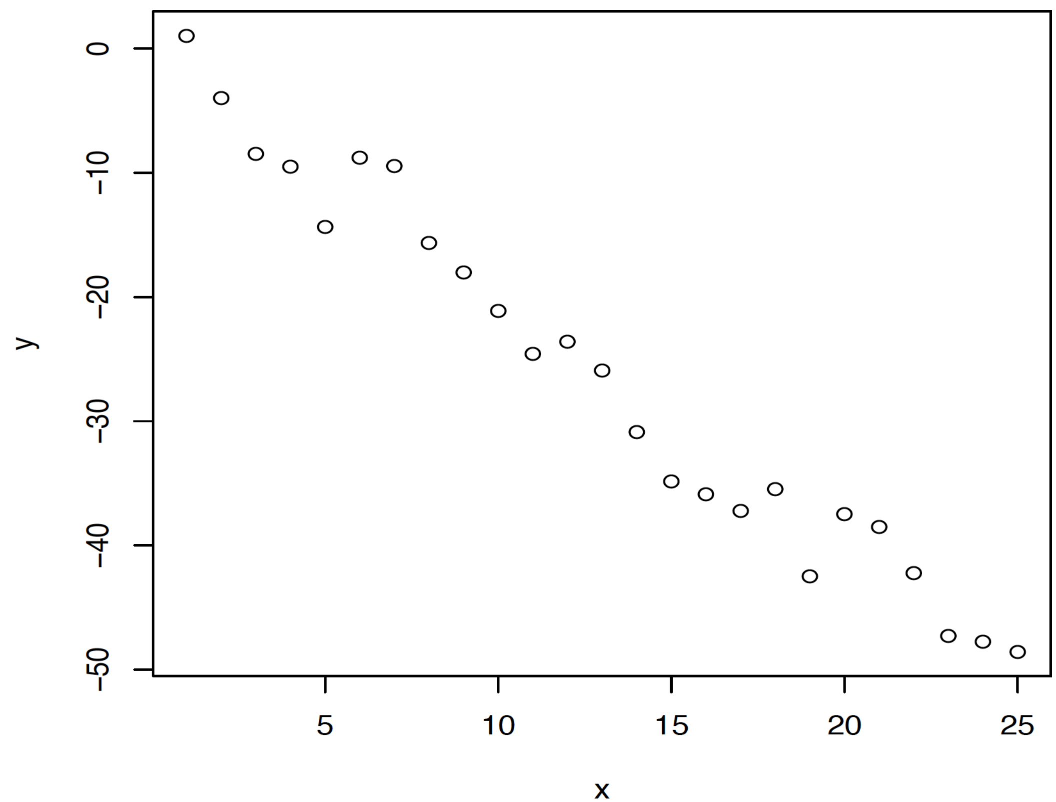
a) 0.01 b) 0.98 c) -0.5 d) -0.98



Solution: the scatterplot shows no real structure or relationship between x and y. The most likely answer si $\rho = 0.01$.

Q142. For the following data the correlation coefficient is most likely to be

a)0.01 b)0.98 c)-0.5 d)-0.98



Solution: the scatter plot shows a clear anti-correlated pattern between x and y – when x increases, y decreases and vice-versa. The most likely value is $\rho = -0.98$.

Q143. A company employs 10 part-time drivers for its fleet of trucks. Its manager wants to find a relationship between number of km driven (X) and number of working days (Y) in a typical week. The drivers are hired to drive half-day shifts, so that 3.5 stands for 7 half-day shifts.

The manager wants to use the linear regression model $Y = \beta_0 + \beta_1 x + \epsilon$ on the following data:

	1	2	3	4	5	6	7	8	9	10
x	825	215	1070	550	480	920	1350	325	670	1215
y	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

Note that $\sum x_i^2 = 7104300$, $\sum y_i^2 = 99.75$, and $\sum x_i y_i = 26370$. What is the fitted regression line?

Solution: we have

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 1297860$$

and

$$S_{xy} = 4653,$$

so that

$$b_1 = S_{xy}/S_{xx} = 0.0036,$$

and

$$b_0 = \sum_{i=1}^{n} y_i/n - b_1 \sum_{i=1}^{n} x_i = 0.1181;$$

hence the fitted line is $\hat{y} = 0.1181 + 0.0036x$.

Q144. Using the data from question **Q143**, what value is the correlation coefficient of x and y closest to?

a) 0.437 b) 0.949 c) 0.113 d) 1.123 e) none of the preceding

Solution: as in question **Q143**, we have $S_{xx} = 12978600$ and $S_{xy} = 4653$. Furthermore, we have

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 18.525,$$

so that the correlation coefficient is

$$\rho_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{4653}{\sqrt{18.525 \cdot 1297860}} \approx 0.949$$

Q145. We want to test significance of regression, i.e. $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. The value of the appropriate statistic and the decision for $\alpha = 0.05$ is:

a) 8.55; do not reject H_0

b) 2.31; reject H_0

c) 8.55; reject H_0

d) 2.31; do not reject H_0

e) none of the preceding

Solution: the estimated variance is

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{1.8434}{8} = 0.23.$$

Consequently, the test statistic is

$$t_0 = \frac{b_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \frac{0.0036}{\sqrt{0.23/1297860}} = 8.551701.$$

Since $t_{0.05/2}(n-2) = t_{0.025}(8) = 2.306$, we reject H_0.

Q146. Regression methods were used to analyze the data from a study investigating the relationship between roadway surface temperature in F (x) and pavement defection (y). Summary quantities were $n = 20$,

$$\sum y_i = 12.75, \sum y_i^2 = 8.86, \sum x_i = 1478 \sum x_i^2 = 143,215.8 \sum x_i y_i = 1083.67.$$

a) Calculate the least squares estimates of the slope and intercept. Estimate σ^2.

b) Use the equation of the fitted line to predict what pavement deflection would be observed when the surface temperature is 90F.

c) Give a point estimate of the mean pavement deflection when the surface is 85F.

d) What change in mean pavement deflection would be expected for a 1F change in surface temperature?

Solution:

a) We have

$$b_1 = \frac{S_{xy}}{S_{xx}}, \quad b_0 = \overline{y} - b_1\overline{x}, \quad \hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2},$$

where

$$\begin{aligned} S_{xy} &= \sum x_i y_i - \tfrac{1}{n}(\sum x_i)(\sum y_i) = 141.445 \\ S_{xx} &= \sum x_i^2 - \tfrac{1}{n}(\sum x_i)^2 = 33991.6 \\ S_{yy} &= \sum y_i^2 - \tfrac{1}{n}(\sum y_i)^2 = 0.731875, \end{aligned}$$

so that $b_1 = 0.00416$, $b_0 = 0.32999$, and $\hat{\sigma}^2 = 0.00797$

b) $\hat{y}(90) = b_0 + b_1 \cdot 90 = 0.70$

c) The question can be rephrased as "use the equation of the fitted line to predict what pavement deflection would be observed when the surface temperature is 85F", i.e. $\hat{y}(85) = b_0 + b_1 \cdot 85 = 0.68$.

d) That is simply the slope: $b_1 = 0.00416$

Q147. Consider the data from **Q146**.

a) Test for significance of regression using $\alpha = 0.05$. Find the p-value for this test. What conclusion can you draw?

b) Estimate the standard errors of the slope and intercept.

Solution:

a) We test for $H_0 : \beta_1 = 0$, against $H_1 : \beta_1 \neq 0$. The test statistic is $T_0 = \frac{b_1 - 0}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}} \sim t(n-2)$. Its observed value is $t_0 = \frac{b_1 - 0}{\sqrt{\hat{\sigma}^2/S_{xx}}} = 8.6$. The p-value (using $t(18)$ table) is $2P(t_{18} > 8.6) < 0.001$, and so we reject H_0 in favour of a linear relationship between x and y.

b) The standard errors are

$$\mathrm{se}(b_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \quad \mathrm{se}(b_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}.$$

So, $\mathrm{se}(b_1) = 0.00048$, $\mathrm{se}(b_0) = 0.04098$.