MAT 2377
Probability and Statistics for Engineers

Practice Set

P. Boily (uOttawa)

Winter 2021

Based on course notes by Rafał Kulik

Q131. For a set of 12 pairs of observations on (x_i, y_i) from an experiment, the following summary for x and y is obtained:

$$\sum_{i=1}^{12} x_i = 25,\ \sum_{i=1}^{12} y_i = 432,\ \sum_{i=1}^{12} x_i^2 = 59,\ \sum_{i=1}^{12} x_i y_i = 880.5,\ \sum_{i=1}^{12} y_i^2 = 15648.$$

The estimated value of y at $x = 5$ from the least squares regression line is:

a)27.78 b)47.77 c)41.87 d)55.97 e)none of the preceding

Solution: assuming the linear regression model is warranted, the estimated value at $x = 5$ is given by

$$\hat{y}(5) = b_0 + b_1(5).$$

We have

$$\overline{x} = \frac{1}{12}\sum_{i=1}^{12} x_i = \frac{25}{12}, \quad \overline{y} = \frac{1}{12}\sum_{i=1}^{12} y_i = 36$$

$$S_{xx} = \sum_{i=1}^{12} x_i^2 - 12\overline{x}^2 = \frac{83}{12}, \quad S_{xy} = \sum_{i=1}^{12} x_i y_i - 12\overline{xy} = -19.5,$$

$$b_1 = -\frac{19.5}{83/12} = -2.82, \quad b_0 = 36 - (-2.82)(25/12) = 41.87,$$

$$\hat{y}(5) = 41.87 - 2.82(5) = 27.78.$$

Q132. Assuming that the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ is appropriate for $n = 14$ observations, the estimated regression line is computed to be

$$\hat{y} = 0.66490 + 0.83075x.$$

Given that $S_{yy} = 4.1289$ and $S_{xy} = 4.49094$, compute the estimated standard error for the slope.

a)0.3176 b)0.0783 c)0.0855 d)0.0073 e)none of the preceding

Solution: the estimated standard error for the slope is

$$\mathrm{se}(b_1) = \sqrt{\hat{\sigma}^2/S_{xx}}.$$

But

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} \quad \text{and} \quad S_{xx} = \frac{S_{xy}}{b_1}$$

so that

$$\mathrm{se}(b_1) = \sqrt{\frac{b_1(S_{yy} - b_1 S_{xy})}{(n-2)S_{xy}}} = \sqrt{\frac{0.83075(4.1289 - 0.83075 \cdot 4.49094)}{(14-2)4.49094}}$$

$$= 0.07833.$$

Q137. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. The point estimate for the slope of the regression line is

a) 1.99 b) -1.99 c) 0.49 d) 0.59 e) none of the preceding

Solution: we have $b_1 = \frac{S_{xy}}{S_{xx}}$. Note that

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 13,\ \overline{y} = \frac{1}{25}\sum_{i=1}^{25} y_i = 26.359,\ S_{xx} = \sum_{i=1}^{25} x_i^2 - 25\overline{x}^2 = 1300$$

$$S_{xy} = \sum_{i=1}^{25} x_i y_i - 25\overline{xy} = 2586.952,\ S_{yy} = \sum_{i=1}^{25} y_i^2 - 25\overline{y}^2 = 5261.613,$$

$$b_1 = \frac{2586.952}{1300} = 1.99, \quad b_0 = 26.359 - (1.990)(13) = 0.49.$$

$\hat{y}(30) = 0.49 + 1.99(30) = 60.19.$

Q138. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000,\ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000,\ \sum_{i=1}^{n} x_i y_i = 11153.588,\ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. The point estimate for the intercept of the regression line is

a)1.99 b)-1.99 c)0.49 d)0.59 e)none of the preceding

Solution: see answer to **Q137**.

Q139. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. What is the prediction of y for $x = 30$?

a)60.19 b)16.67 c)30 d)30.54 e)none of the preceding

Solution: see answer to **Q137**.

Q140. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. Is the linear regression significant?

Solution: we are testing for $H_0 : \beta_1 = 0$ against $H_0 : \beta_1 \neq 0$. Let's use $\alpha = 0.05$. If we reject H_0 in favour of H_1, then the evidence suggests that there is a linear relationship between X and Y.

Under H_0, the test statistic $T_0 = \frac{b_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \sim t_{0.05/2}(23)$. Using the results from the previous questions, we have

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{5261.613 - 1.99 \cdot 2586.952}{23} = 4.94;$$

the observed statistic is thus

$$t_0 = \frac{1.99}{\sqrt{4.94/1300}} = 32.27 < t_{0.05/2}(23) = 2.069.$$

Thus we reject H_0 in favour of significance of regression.