

Solutions to Exam-style question (1)

(a) The sample correlation coefficient is given by:

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \times S_{YY}}},$$

where

$$\begin{aligned} S_{XY} &= \sum xy - n\bar{x}\bar{y} \\ S_{XX} &= \sum x^2 - n\bar{x}^2 \quad \text{and} \\ S_{YY} &= \sum y^2 - n\bar{y}^2 \end{aligned}$$

Since monthly figures were collected over a one-year period (2009), we have $n = 12$ and

$$\bar{x} = \frac{120}{12} = 10 \quad \text{and} \quad \bar{y} = \frac{1193}{12} = 99.42.$$

This gives:

$$\begin{aligned} S_{XY} &= 13086 - 12 \times 10 \times 99.42 = 1155.6, \\ S_{XX} &= 1450 - 12 \times 10^2 = 250 \quad \text{and} \\ S_{YY} &= 128997 - 12 \times 99.42^2 = 10384.9632 \end{aligned}$$

From these, we get:

$$\begin{aligned} r &= \frac{1155.6}{\sqrt{250 \times 10384.9632}} \\ &= 0.71 \end{aligned}$$

Comment: This indicates a moderate/strong positive linear relationship between temperature and sales.

(b) (i) For the simple linear regression model we have

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

We can compute $\hat{\beta}_1$ as

$$\begin{aligned} \hat{\beta}_1 &= \frac{S_{XY}}{S_{XX}} \\ &= \frac{1155.6}{250} \\ &= 4.6224. \end{aligned}$$

Also, $\hat{\beta}_0$ is given by:

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 99.42 - 4.6224 \times 10 \\ &= 53.196.\end{aligned}$$

Thus, the regression equation is

$$Y = 53.196 + 4.6224X + \epsilon.$$

These values should be entered in the **Regression equation** and in the **Coef** boxes in the **Minitab** output.

The test statistic **T** is equal to **Coef/SE Coef**; thus, for β_0 (the **Constant**), we have

$$3.40 = 53.196/\text{SE Coef}$$

So the **SE Coef** for β_0 is

$$\text{SE Coef} = 53.196/3.40 = 15.65.$$

Similarly, for β_1 we have

$$T = 4.6224/1.421 = 3.253.$$

Also, we can calculate **R-Sq** as the square of the estimated correlation coefficient, i.e.

$$\text{R-Sq} = 0.71 \times 0.71 = 0.5041.$$

This gives the following completed regression output:

Regression Analysis: Sales (£000) versus Average temp

The regression equation is

$$\text{Sales (000)} = 53.196 + 4.6224 \text{ Average temperature}$$

Predictor	Coef	SE Coef	T	P
Constant	53.196	15.65	3.40	0.007
Average temperature	4.6224	1.421	3.253	0.009

$$S = 22.466$$

$$\text{R-Sq} = 50.41\%$$

$$\text{R-Sq(adj)} = 46.6\%$$

(ii) The full estimated regression model is

$$Y = 53.196 + 4.6224X + \epsilon, \epsilon \sim N(0, 22.4668^2).$$

(iii) To test the significance of the slope term, we have

$$\begin{aligned} H_0 &: \beta_1 = 0 && \text{versus} \\ H_1 &: \beta_1 \neq 0. \end{aligned}$$

The test statistic is $t = 3.253$, which gives a p -value of 0.009, or 0.9%. Thus,

- * we have strong evidence against H_0 ;
- * we reject H_0 in favour of H_1 ;
- * the slope term *is* significantly different from zero, and so temperature *is* a good predictor of sales.

(iv) We have $R^2 = 50.41\%$, which tells us that just over half the variability in sales is explained by average temperature.

(v) The assumptions underlying this regression model are that the residuals are Normally distributed with zero mean and constant variance, i.e. $\epsilon \sim N(0, \sigma^2)$.

(c) If $X = 21$.

$$\begin{aligned} Y &= 53.196 + 4.6224 \times 21 \\ &= 150.2644, \end{aligned}$$

i.e. just over £150 million.

(d) If $Y = 81$, then

$$81 = 53.196 + 4.6224X$$

Rearranging, we get

$$\begin{aligned} X &= (81 - 53.196)/4.6224 \\ &= 6.025, \end{aligned}$$

i.e. about 6°C.