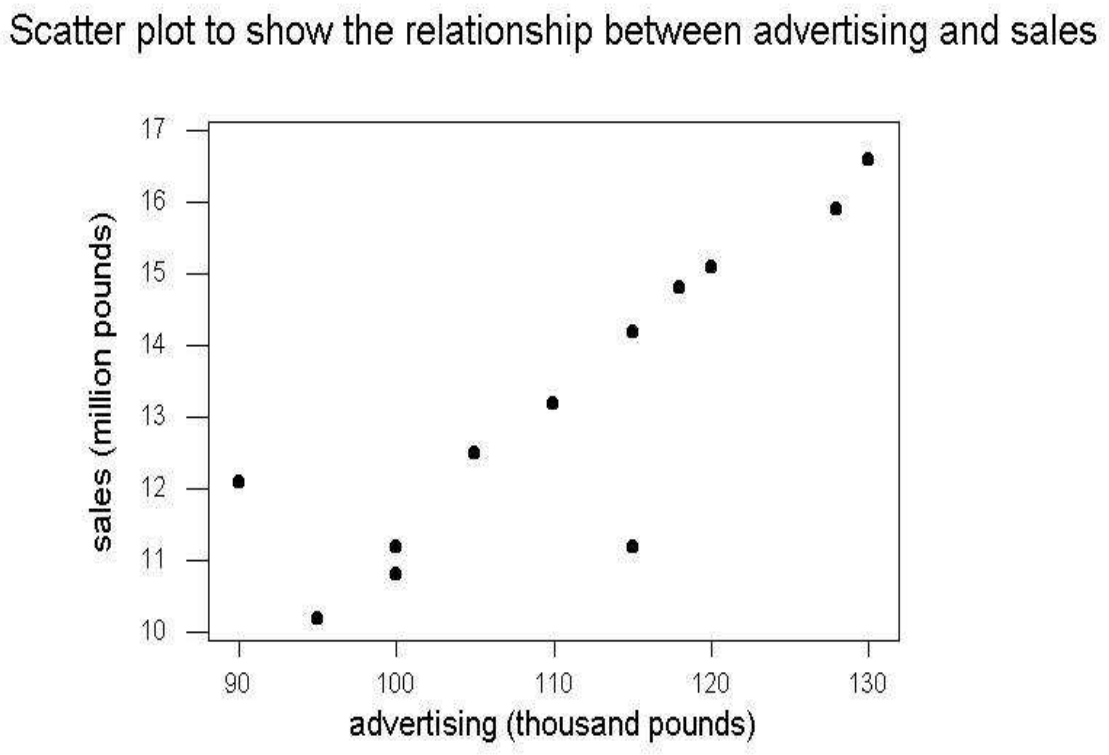
# Chapter 6: Model solutions

(a) The scatter plot for advertising and sales has been constructed in `Minitab`, and is shown below; yours (done by hand) should look similar to this! Notice how "sales" is on the $y$–axis, since this is the response variable – we're probably going to be interested in predicting sales based on the amount spent on advertising. From this plot, we can see that there is quite a strong, linear, positive relationship between the amount spent on advertising and the amount brought in through sales, i.e. the more the company spends on advertising, the more money they are likely to make!




(b) Recall that the sample correlation coefficient, $r$, is found as

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

where

$$\begin{aligned} S_{XY} &= \left(\sum xy\right) - n\bar{x}\bar{y}, \\ S_{XX} &= \left(\sum x^2\right) - n\bar{x}^2, \\ S_{YY} &= \left(\sum y^2\right) - n\bar{y}^2. \end{aligned}$$

We can draw up the following table to help us calculate $r$:

| $x$ | $y$ | $x^2$ | $y^2$ | $xy$ |
|---|---|---|---|---|
| 100 | 11.2 | 10000 | 125.44 | 1120 |
| 90 | 12.1 | 8100 | 146.41 | 1089 |
| 110 | 13.2 | 12100 | 174.24 | 1452 |
| 120 | 15.1 | 14400 | 228.01 | 1812 |
| 115 | 14.2 | 13225 | 201.64 | 1633 |
| 95 | 10.2 | 9025 | 104.04 | 969 |
| 105 | 12.5 | 11025 | 156.25 | 1312.5 |
| 130 | 16.6 | 16900 | 275.56 | 2158 |
| 118 | 14.8 | 13924 | 219.04 | 1746.4 |
| 100 | 10.8 | 10000 | 116.64 | 1080 |
| 115 | 11.2 | 13225 | 125.44 | 1288 |
| 128 | 15.9 | 16384 | 252.81 | 2035.2 |
| 1326 | 157.8 | 148308 | 2125.52 | 17695.1 |

Using the information in the above table,

$$\begin{aligned} \bar{x} &= \frac{1326}{12} \\ &= 110.5, \\ \bar{y} &= \frac{157.8}{12} \\ &= 13.15. \end{aligned}$$

Thus,

$$\begin{aligned} S_{XY} &= \left(\sum xy\right) - n \times \bar{x} \times \bar{y} \\ &= 17695.1 - 12 \times 110.5 \times 13.15 \\ &= 258.2, \\ S_{XX} &= \left(\sum x^2\right) - n \times \bar{x}^2 \\ &= 148308 - 12 \times 110.5 \times 110.5 \\ &= 1785 \qquad \text{and} \\ S_{YY} &= \left(\sum y^2\right) - n \times \bar{y}^2 \\ &= 2125.52 - 12 \times 13.15 \times 13.15 \\ &= 50.45. \end{aligned}$$

So the sample correlation coefficient is

$$\begin{aligned} r &= \frac{S_{XY}}{\sqrt{S_{XX} \times S_{YY}}} \\ &= \frac{258.2}{\sqrt{1785 \times 50.45}} \\ &= 0.8604 \text{ (to 4 decimal places).} \end{aligned}$$

Our sample correlation coefficient is $r = 0.8604$, which indicates quite a strong, positive, linear association between amount spent on advertising and sales. This *does* agree with the scatter plot in part (a); the plot shows sales increasing with amount spent on advertising, and indicates a strong, positive, linear relationship between the two!

(c) The simple linear regression model is given by

$$Y = \alpha + \beta X + \epsilon,$$

where we can estimate $\alpha$ and $\beta$ using

$$\begin{aligned} \hat{\beta} &= \frac{S_{XY}}{S_{XX}} \qquad \text{and} \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}. \end{aligned}$$

Thus,

$$\begin{aligned} \hat{\beta} &= \frac{258.2}{1785} \\ &= 0.1446, \qquad \text{and} \\ \hat{\alpha} &= 13.15 - 0.1446 \times 110.5 \\ &= -2.8283. \end{aligned}$$

Thus, the estimated linear regression equation is

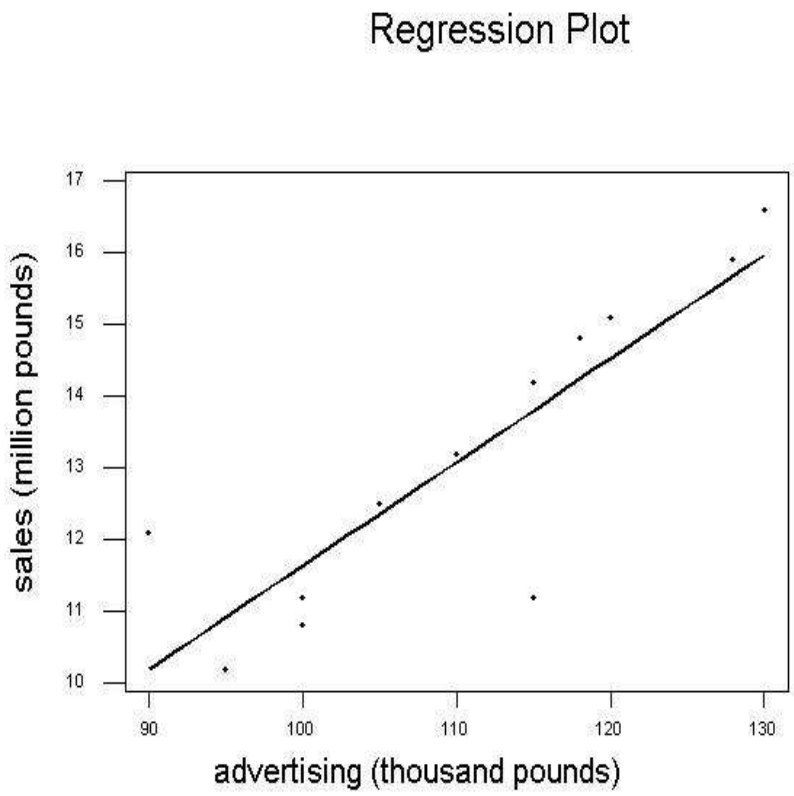
$$Y = -2.8283 + 0.1446X + \epsilon.$$

(d) The regression line has been superimposed on the original scatter plot, and is shown below. Remember, to do this, just pick two arbitrary $X$ values and use your estimated regression equation to find the corresponding $Y$ values. Then all you have to do is plot the two points and join up the line!

For example, $X = 100$ and $X = 110$:

- when $X = 100$, $Y = -2.8283 + 0.1446 \times 100 = 11.63$;
- when $X = 110$, $Y = -2.8283 + 0.1446 \times 110 = 13.08$.

So all you have to do now is plot the points $(100, 11.63)$ and $(110, 13.08)$, and join them up with a straight line!




(e) If the company were to spend £112,000, we can expect their sales to be

$$\begin{aligned} Y &= -2.8283 + 0.1446 \times 112 \\ &= 13.4, \end{aligned}$$

i.e. £13.4 million.