

**Figure 13.11. SIMPLE LINEAR REGRESSION: Case Study 3**

**Example 13.11.1:** On Saturday, March 22, 1997, *The Globe and Mail* published two articles (on pages B1 and B4) about negative customer experiences when buying a new car in Canada and in the United States; the articles also described how the process of car buying is likely to change under the influence of the Internet. The second article gave the following data on the number of new car dealers in Canada ( $x$ ) and the sales of light vehicles (presumably cars and vans) per dealer ( $y$ ) over a 16-year period:

| Year | 1981  | 1982  | 1983  | 1984  | 1985  | 1986  | 1987  | 1988  | 1989  | 1990  | 1991  | 1992  | 1993  | 1994  | 1995  | 1996  |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $x$  | 4,125 | 3,707 | 3,744 | 3,869 | 3,993 | 4,023 | 4,052 | 4,049 | 4,064 | 4,014 | 3,967 | 3,930 | 3,872 | 3,856 | 3,819 | 3,714 |
| $y$  | 277   | 242   | 283   | 323   | 376   | 367   | 368   | 375   | 355   | 321   | 319   | 307   | 301   | 320   | 296   | 316   |

- Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of  $\mathbf{Y}$  on  $\mathbf{X}$ .
- Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of  $\sigma$  for these data.
- Assess how well the regression model fits the data; give your reasons completely but concisely.
- Find a 95% confidence interval for the slope of the study population regression of  $\mathbf{Y}$  on  $\mathbf{X}$ .
- Find a 95% *confidence* interval for the *average* sales per dealer when there are 3,800 dealers in Canada.
- Find a 95% *prediction* interval for the sales of a dealer when there are 3,800 dealers in Canada.
- Investigate the effect on the estimated regression of omitting one or both points with estimated residuals of largest magnitude.

**Solution:** (a) From the  $n = 16$  observations ( $x_i, y_i$ ) given in the statement of the question, we find the following:

$$\sum x_i = 62,798,$$

$$\sum x_i^2 = 246,735,812;$$

$$\sum y_i = 5,146,$$

$$\sum y_i^2 = 1,677,374;$$

$$\sum x_i y_i = 20,242,915;$$

$$\bar{x} = 3,924.875,$$

$$\bar{y} = 321.625;$$

$$SS_{xy} = 45,508.25,$$

$$SS_x = 261,511.75,$$

$$SS_y = 22,291.75;$$

hence, the *estimates* of  $\beta_1$  (the slope) and  $\beta_0$  (the intercept) of the regression of  $\mathbf{Y}$  on  $\mathbf{X}$  are:

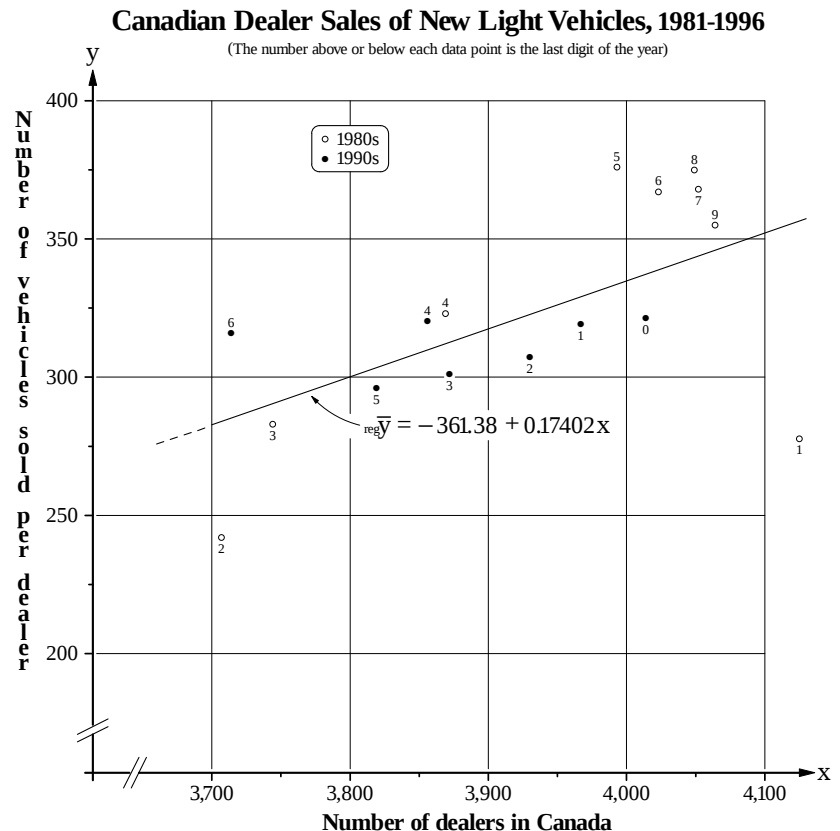
$$b_1 = 0.174\ 019\ 905,$$

$$b_0 = -361.381\ 375,$$

so the equation of the straight-line model is:

$$\begin{aligned} \text{reg } \bar{y} &= -361.38 + 0.17402x \\ &= 321.625 + 0.17402(x - 3924.875). \end{aligned}$$

The scatter diagram of the data for Canadian dealer sales of new light vehicles, with the estimated regression of  $\mathbf{Y}$  on  $\mathbf{X}$  superimposed on it, is shown above at the right.



- (b) The ANOVA table for these data (16 observations) is:

| SOURCE             | SUM of SQUARES                | Df | MEAN SQUARE                 | F-RATIO                                                     |
|--------------------|-------------------------------|----|-----------------------------|-------------------------------------------------------------|
| Model              | $b_1 SS_{xy} = 7,919.341\ 36$ | 1  | $MSM = 7,919.341\ 361$      | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 7.71$ |
| Estimated residual | $SSE = 14,372.408\ 64$        | 14 | $MSE = 1,026.600\ 617$      |                                                             |
| Total              | $SS_y = 22,291.75$            | 15 | $(s_y^2 = 1,486.11\dot{6})$ | -----                                                       |

**Example 13.11.1:** Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:  
(continued)

$$r^2 = 0.355\,258\,845, \quad r = 0.596\,035\,943, \quad \hat{\sigma} = \sqrt{\text{MSE}} = 32.040\,608\,8753.$$

(c) Three matters provide an assessment of how well the regression model fits the data:

- Visual inspection of the scatter diagram with the estimated regression of  $\mathbf{Y}$  on  $\mathbf{X}$  superimposed on it shows that 13 of the 16 points lie reasonably close to the line – those for 1981, 1982 and 1985 have appreciably larger estimated residuals, especially 1981 (*below* the fitted line);
- the points appear to be scattered without obvious pattern on both sides of the fitted line;
- there does not appear to be any systematic change in the magnitude of the estimated residuals with increasing values of  $x$ , which is consistent with the assumption of constant  $\sigma$ .
- The  $F$ -ratio (771) is reasonably high and provides statistically significant evidence ( $P \approx 0.0141$ ) against the hypothesis  $\beta_1 = 0$ , indicating a meaningful regression.
- The value of the coefficient of determination shows that around 35% of the variation of the  $y$ 's about their average has been accounted for by the estimated regression of  $\mathbf{Y}$  on  $\mathbf{X}$ .

These matters show the straight-line model is only a *moderate* fit to the data.

(d) We want a 95% confidence interval (CI) for  $\beta_1$ , representing the *slope* of the regression of  $\mathbf{Y}$  on  $\mathbf{X}$ .

$$\text{We have: } \hat{\sigma} = 32.040\,608\,8753, \quad b_1 = 0.174\,019\,905, \quad \Pr[-2.14479 \leq t_{14} \leq 2.14479] = 0.95,$$

$$\text{so that: } \hat{\sigma} \sqrt{\frac{1}{SS_x}} = \hat{\sigma} \sqrt{\frac{1}{261,511.75}} \approx 0.062\,654\,9167;$$

$$\text{hence, a 95\% CI for } \beta_1 \text{ is: } 0.174\,019\,905 \pm 0.134\,381\,639 \Rightarrow (0.039\,638\,266, 0.308\,401\,544) \\ \text{or about } (0.040, 0.308) \text{ vehicles/dealer.}$$

(e) We want a 95% *confidence* interval (CI) for the *mean*  $\mu_Y(x_j = 3,800)$ , representing the *average* of the response  $\mathbf{Y}$  when the study population size, represented by the explanatory variate  $\mathbf{X}$ , is 3,800 dealers.

$$\text{We have: } \hat{\sigma} = 32.040\,608\,8753, \quad \hat{\mu}_Y(x_j = 3,800) = 299.894\,265, \quad \Pr[-2.14479 \leq t_{14} \leq 2.14479] = 0.95,$$

$$\text{so that: } \hat{\sigma} \sqrt{\frac{(x_j - \bar{x})^2}{SS_x} + \frac{1}{n}} = \hat{\sigma} \sqrt{\frac{(3,800 - 3,924.875)^2}{261,511.75} + \frac{1}{16}} \approx 11.197\,232\,9905;$$

$$\text{hence, a 95\% CI for } \mu_Y(x_j = 3,800) \text{ is: } 299.894\,264 \pm 24.015\,713\,3 \Rightarrow (275.878\,551, 323.909\,977) \\ \text{or about } (276, 324) \text{ vehicles.}$$

(f) We want a 95% *prediction* interval (PI) for the *random variable*  $Y(x_j = 3,800)$ , representing the response  $\mathbf{Y}$  for an *individual* randomly-selected dealer when the study population size, represented by the explanatory variate  $\mathbf{X}$ , is 3,800 dealers.

$$\text{We have: } \hat{\sigma} = 32.040\,608\,8753, \quad \hat{\mu}_Y(x_j = 3,800) = 299.894\,265, \quad \Pr[-2.14479 \leq t_{14} \leq 2.14479] = 0.95,$$

$$\text{so that: } \hat{\sigma} \sqrt{\frac{(x_j - \bar{x})^2}{SS_x} + \frac{1}{n} + 1} = \hat{\sigma} \sqrt{\frac{(3,800 - 3,924.875)^2}{261,511.75} + \frac{1}{16} + 1} \approx 33.940\,810\,888;$$

$$\text{hence, a 95\% PI for } Y(x_j = 3,800) \text{ is: } 299.894\,264 \pm 72.795\,9118 \Rightarrow (227.098\,352, 372.690\,176) \\ \text{or about } (227, 373) \text{ vehicles.}$$

(g) Looking at the scatter diagram on the first side of the Figure, we see the point for 1981 has the estimated residual of *largest* magnitude; omitting the data for this year from the least squares calculations for fitting the line, the remaining 15 observations yield:

$$b_1 = 0.251\,536\,926, \quad b_0 = -659.295\,070,$$

$$\text{so the equation of the straight-line model is now: } \text{reg } \bar{y} = -659.30 + 0.25154x = 324.6 + 0.25154(x - 3911.53).$$

The ANOVA table for the 15 observations is:

| SOURCE             | SUM of SQUARES                  | Df | MEAN SQUARE            | F-RATIO                                                      |
|--------------------|---------------------------------|----|------------------------|--------------------------------------------------------------|
| Model              | $b_1 SS_{xy} = 13,843.133\,484$ | 1  | MSM = 13,843.133 4836  | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 28.45$ |
| Estimated residual | $SSE = 6,324.466\,516$          | 13 | MSE = 486.497 4243     |                                                              |
| Total              | $SS_y = 20,167.6$               | 14 | $(s_y^2 = 1,440.5429)$ | -----                                                        |

Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:

$$r^2 = 0.686\,404\,604, \quad r = 0.828\,495\,385, \quad \hat{\sigma} = \sqrt{\text{MSE}} = 22.056\,686\,6129.$$

**Figure 13.11. SIMPLE LINEAR REGRESSION: Case Study 3 (continued 1)**

**Example 13.11.1:** Proceeding in the same way but *also* omitting the data for 1985, the point on the scatter diagram with the estimated residual of the *second*-largest magnitude (although that for 1982 is comparable), the remaining 14 observations yield:

$$b_1 = 0.238\,792\,090, \quad b_0 = -611\,725\,103,$$

so the equation of the straight-line model is now:  $\text{reg}\bar{y} = -611.73 + 0.23879x = 320.9 + 0.23879(x - 3905.71)$ .

The ANOVA table for the 14 observations is:

| SOURCE             | SUM of SQUARES                 | Df | MEAN SQUARE                | F-RATIO                                                      |
|--------------------|--------------------------------|----|----------------------------|--------------------------------------------------------------|
| Model              | $b_1SS_{xy} = 12,070.394\,338$ | 1  | MSM = 12,070.394 338       | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 27.50$ |
| Estimated residual | SSE = 5,266.534 233            | 12 | MSE = 438.877 8528         |                                                              |
| Total              | $SS_y = 17,336.928\,571$       | 13 | $(s_y^2 = 1,333.609\,892)$ | -----                                                        |

Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:

$$r^2 = 0.696\,224\,495, \quad r = 0.834\,400\,680, \quad \hat{\sigma} = \sqrt{MSE} = 20.949\,411\,7529.$$

We also examine the effect of omitting the data for 1982 *instead* of 1985; the relevant 14 observations yield:

$$b_1 = 0.212\,295\,565, \quad b_0 = -503.002\,717,$$

so the equation of the straight-line model is now:  $\text{reg}\bar{y} = -503.00 + 0.21230x = 330.5 + 0.21230(x - 3926.14)$ .

The ANOVA table for these 14 observations is:

| SOURCE             | SUM of SQUARES                 | Df | MEAN SQUARE               | F-RATIO                                                      |
|--------------------|--------------------------------|----|---------------------------|--------------------------------------------------------------|
| Model              | $b_1SS_{xy} = 7,840.712\,3596$ | 1  | MSM = 7,840.712 3596      | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 18.75$ |
| Estimated residual | SSE = 5,016.787 6404           | 12 | MSE = 418.065 6367        |                                                              |
| Total              | $SS_y = 12,857.5$              | 13 | $(s_y^2 = 989.038\,4615)$ | -----                                                        |

Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:

$$r^2 = 0.609\,816\,244, \quad r = 0.780\,907\,321, \quad \hat{\sigma} = \sqrt{MSE} = 20.446\,653\,4352.$$

We compare the results from fitting a straight-line model to the original data and to subsets of them in Table 13.11.1 below:

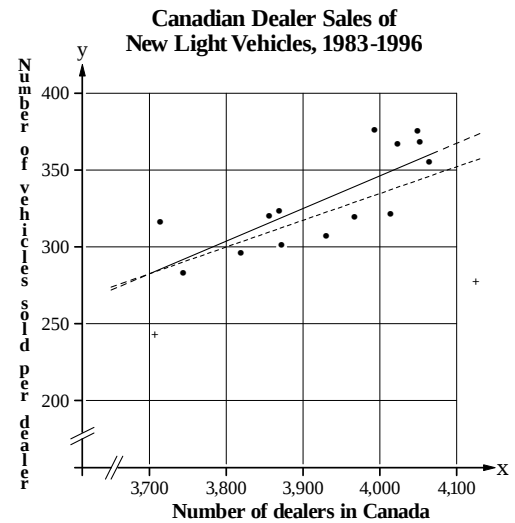
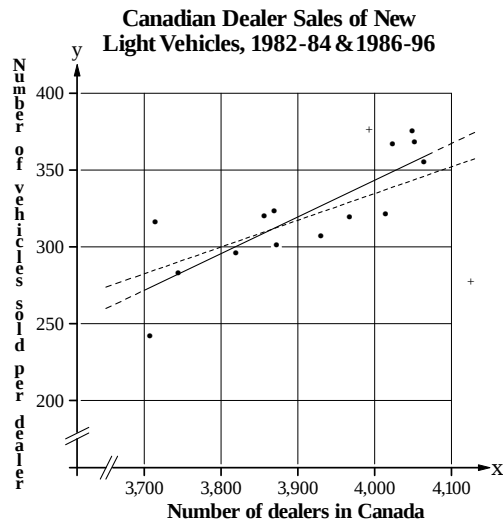
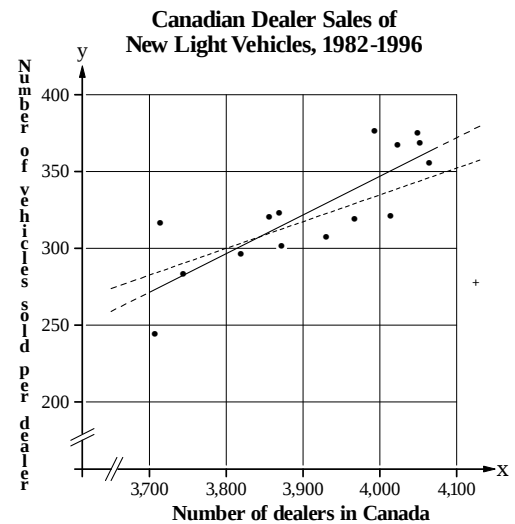
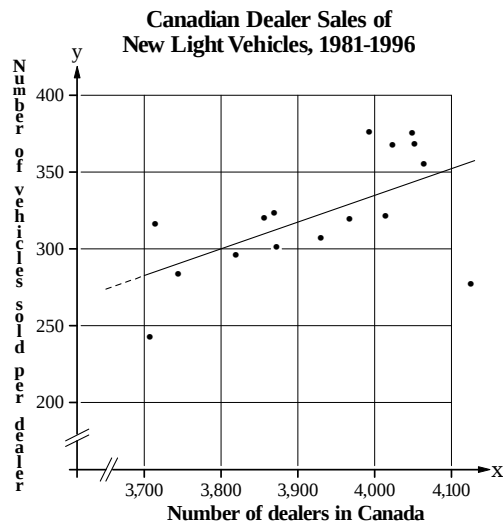
| Table 13.11.1: Data                             | $b_1$  | $b_0$  | $\bar{x}$ | $\bar{y}$ | $r^2$  | $r$    |
|-------------------------------------------------|--------|--------|-----------|-----------|--------|--------|
| 16 observations for <i>all</i> the years        | 0.1740 | -361.4 | 3,925     | 321.6     | 0.3553 | 0.5960 |
| 15 observations omitting the 1981 data          | 0.2515 | -659.3 | 3,912     | 324.6     | 0.6864 | 0.8285 |
| 14 observations omitting the 1981 and 1985 data | 0.2388 | -611.7 | 3,906     | 320.9     | 0.6962 | 0.8344 |
| 14 observations omitting the 1981 and 1982 data | 0.2123 | -503.0 | 3,926     | 330.5     | 0.6098 | 0.7809 |

A *graphical* comparison of the four cases is given overleaf on page 13.56 – in the three diagrams for *subsets* of the data, the point(s) omitted are shown as a ‘+’ sign and the straight-line model for all 16 observations is included as a *dashed* line. We note the following matters:

- there is a moderate overall positive association between number of new vehicles sold per dealer and the number of dealers in Canada, but with noticeable scatter of the data;
- the strength of the association is appreciably enhanced when the data for the *first* of the sixteen years (1981, the point with the estimated residual of *largest* magnitude) is omitted, suggesting that this year is somewhat *different* from the other 15 years;
  - if a follow-up study of the data were to be undertaken, it might be of interest to investigate in what way(s) 1981 was different from the other 15 years in the period;
- although 1982 and 1985 also have substantial estimated residuals, their omission does *not* greatly influence the fit of the straight-line model, suggesting that, despite their appreciable estimated residuals, these years are not *essentially* different from the set of 15 years which *includes* them but *excludes* 1981.

**Example 13.11.2:** *The Globe and Mail* article cited in Example 13.11.1 on the first side of Figure (page 13.53) also gave the 16 years of corresponding U.S. data on number of new car dealers (x) and the sales per dealer (y) as follows:

| Year | 1981   | 1982   | 1983   | 1984   | 1985   | 1986   | 1987   | 1988   | 1989   | 1990   | 1991   | 1992   | 1993   | 1994   | 1995   | 1996   |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| x    | 27,891 | 27,095 | 26,281 | 25,490 | 24,724 | 24,823 | 25,156 | 24,938 | 24,913 | 24,724 | 24,272 | 23,361 | 22,845 | 22,451 | 22,426 | 22,288 |
| y    | 365    | 379    | 458    | 560    | 628    | 650    | 593    | 622    | 584    | 562    | 511    | 553    | 610    | 673    | 657    | 677    |

**Example 13.11.1:**  
(continued)**Example 13.11.2:** Compare and contrast straight-line modelling of these data with those for Canada in Example 13.11.1.  
(continued)

**Solution:** From the  $n = 16$  observations  $(x_i, y_i)$  given in the statement of the question, we find the following:

$$\begin{aligned}\Sigma x_i &= 393,678, & \Sigma y_i &= 9,082, \\ \Sigma x_i^2 &= 9,727,085,548; & \Sigma y_i^2 &= 5,294,484; \\ \Sigma x_i y_i &= 221,484,420; & SS_{xy} &= -1,977,054.75, \\ \bar{x} &= 24,604.875, & SS_x &= 40,687,567.75, \\ \bar{y} &= 567.625; & SS_y &= 139,313.75;\end{aligned}$$

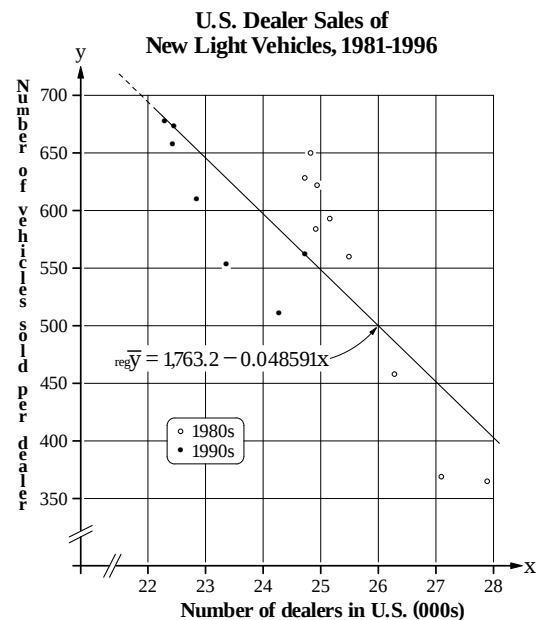
hence, the estimates of  $\beta_1$  (the slope) and  $\beta_0$  (the intercept) of the regression of  $\bar{Y}$  on  $\bar{X}$  are:

$$b_1 = -0.04859126,$$

$$b_0 = 1,763.20359,$$

so the equation of the straight-line model is:

$$\begin{aligned}\text{reg } \bar{Y} &= 1,763.20 - 0.048591x \\ &= 567.62 - 0.048591(x - 24,605).\end{aligned}$$



**Figure 13.11. SIMPLE LINEAR REGRESSION: Case Study 3 (continued 2)**

**Example 13.11.2: (continued)** The scatter diagram of the data for the U.S. dealer sales of new light vehicles, with the estimated regression of  $\mathbf{Y}$  on  $\mathbf{X}$  superimposed on it, is shown at the bottom right of the facing page. We note the following matters of interest from comparing this diagram with the one for the Canadian data on the first side (page 13.53) of the Figure:

- the scatter of the points about the line is clearly *less* for the 16 U.S. observations than for the full set of Canadian data, meaning a straight-line model is a better fit to the U.S. than to the Canadian data – this is quantified by the higher values of the  $F$ -ratio (31.10 vs. 7.71) and coefficient of determination (0.6896 vs. 0.3553) given below.
- the *trend* over time is dramatically different – sales per dealer in Canada tend to *increase* with the number of dealers in Canada whereas there is a clear *decrease* in sales per dealer with the number of dealers in the U.S.;
- sales per dealer tend to be *higher* in the U.S. than in Canada – the respective averages are about 568 and 322 vehicles over the 16-year period; at the same time, the ratio of the number of dealers is *smaller* than the approximately 10-to-1 ratio of the two country's populations – the respective average numbers of dealers over the 16 years were about 24,605 and 3,925 (around 6-to-1).

The ANOVA table for these data is:

| SOURCE             | SUM of SQUARES                 | Df | MEAN SQUARE                 | F-RATIO                                                      |
|--------------------|--------------------------------|----|-----------------------------|--------------------------------------------------------------|
| Model              | $b_1 SS_{xy} = 96,067.317\ 38$ | 1  | $MSM = 96,067.317\ 381$     | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 31.10$ |
| Estimated residual | $SSE = 43,246.432\ 62$         | 14 | $MSE = 3,089.030\ 901$      |                                                              |
| Total              | $SS_y = 139,313.75$            | 15 | $(s_y^2 = 9,257.58\dot{3})$ | -----                                                        |

Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:

$$r^2 = 0.689\ 575\ 274, \quad r = -0.830\ 406\ 692, \quad \hat{\sigma} = \sqrt{MSE} = 55.579\ 050\ 9217.$$

Looking at the scatter diagram for the data at the bottom right of the facing page 13.56, it seems that the data divide into two groups – those for 1981-90 and those for 1991-1996; the scatter diagram at the right shows the two straight-line models for the groups of 10 and 6 observations – the original model for all 16 observations is shown dashed.

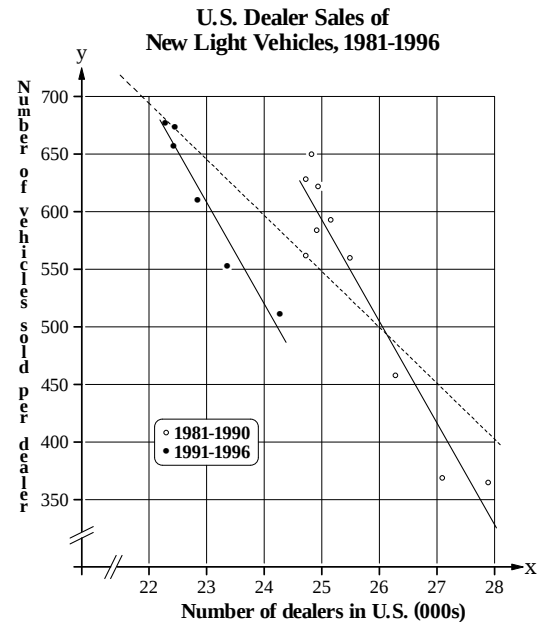
**NOTE:** Without subject-matter knowledge of the data, this separation into two groups is largely arbitrary, although the consecutive years within the groups provides some justification. If further investigating of the U.S. data were to be undertaken, it would be of interest to see if the separation suggested by the scatter diagram has an identifiable basis in fact.

The calculations for the straight-line models for the two groups within the U.S. data are:

for the  $n = 10$  observations  $(x_i, y_i)$  for 1981-1990:

$$\begin{aligned} \sum x_i &= 256,035, & \sum y_i &= 5,401, \\ \sum x_i^2 &= 6,566,597,397; & \sum y_i^2 &= 3,012,547; \\ \sum x_i y_i &= 137,294,964; & SS_{xy} &= -989,539.5, \\ \bar{x} &= 25,603.5, & SS_x &= 11,205,274.5, \\ \bar{y} &= 540.1; & SS_y &= 95,466.9; \end{aligned}$$

hence, the *estimates* of  $\beta_1$  (the slope) and  $\beta_0$  (the intercept) of the regression of  $\mathbf{Y}$  on  $\mathbf{X}$  are:



for the  $n = 6$  observations  $(x_i, y_i)$  for 1991-1996:

$$\begin{aligned} \sum x_i &= 137,643, & \sum y_i &= 3,681, \\ \sum x_i^2 &= 3,160,488,151; & \sum y_i^2 &= 2,281,937; \\ \sum x_i y_i &= 84,189,456; & SS_{xy} &= -254,524.5, \\ \bar{x} &= 22,940.5, & SS_x &= 2,888,909.5, \\ \bar{y} &= 613.5; & SS_y &= 23,643.5; \end{aligned}$$

hence, the *estimates* of  $\beta_1$  (the slope) and  $\beta_0$  (the intercept) of the regression of  $\mathbf{Y}$  on  $\mathbf{X}$  are:

(continued overleaf)

**Example 13.11.2:**  $b_1 = -0.088\ 310\ 152$ ,  
(continued)  $b_0 = 2,801.148\ 995$ ,

so the equation of the straight-line model is:

$$\begin{aligned}\text{reg}\bar{y} &= 2,801.15 - 0.088310x \\ &= 540.1 - 0.088310(x - 25,604).\end{aligned}$$

$$b_1 = -0.088\ 104\ 006,$$

$$b_0 = 2,634.649\ 96,$$

so the equation of the straight-line model is:

$$\begin{aligned}\text{reg}\bar{y} &= 2,634.65 - 0.088104x \\ &= 613.5 - 0.088104(x - 22,940).\end{aligned}$$

The ANOVA table for the 1981-1990 data (10 observations) is:

| SOURCE             | SUM of SQUARES                 | Df | MEAN SQUARE                 | F-RATIO                                                      |
|--------------------|--------------------------------|----|-----------------------------|--------------------------------------------------------------|
| Model              | $b_1SS_{xy} = 87,386.384\ 159$ | 1  | MSM = 87,386.384 159        | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 86.52$ |
| Estimated residual | SSE = 8,080.515 841            | 8  | MSE = 1,010.064 480         |                                                              |
| Total              | $SS_y = 95,466.9$              | 9  | $(s_y^2 = 10,607.4\dot{3})$ | -----                                                        |

Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:

$$r^2 = 0.915\ 357\ 932, \quad r = -0.956\ 743\ 399, \quad \hat{\sigma} = \sqrt{MSE} = 31.781\ 511\ 6094.$$

The ANOVA table for the 1991-1996 data (6 observations) is:

| SOURCE             | SUM of SQUARES                 | Df | MEAN SQUARE          | F-RATIO                                                      |
|--------------------|--------------------------------|----|----------------------|--------------------------------------------------------------|
| Model              | $b_1SS_{xy} = 22,424.628\ 082$ | 1  | MSM = 22,424.628 082 | $\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \approx 73.59$ |
| Estimated residual | SSE = 1,218.871 918            | 4  | MSE = 304.717 9795   |                                                              |
| Total              | $SS_y = 23,643.5$              | 5  | $(s_y^2 = 4,728.7)$  | -----                                                        |

Also, the coefficient of determination, the correlation coefficient and the estimate of  $\sigma$  are:

$$r^2 = 0.948\ 447\ 907, \quad r = -0.973\ 882\ 902, \quad \hat{\sigma} = \sqrt{MSE} = 17.456\ 173\ 1058.$$

Looking at the scatter diagram overleaf on page 13.57 with the superimposed estimated regression lines for the two groups within the U.S. data, or at the two sets of calculations which begin below the scatter diagram and continue above, we see that:

- there is striking *similarity* in the *slopes* of the two fitted lines (*viz.*  $-0.08831$  vs.  $-0.08810$ ; ratio 1.0023) although this *may* be only coincidence;
  - the *mathematical* interpretation of this parallelism is that the *rate* of decrease of sales per dealer is the *same* over the two time periods despite the *difference* in the magnitudes of the numbers of dealers in the U.S. in these two time periods;
- the fit of the model within *each* group is *better* than for all 16 observations taken together – the points generally lie *closer* to the relevant line,  $r^2$  is over 90% for both lines and both *F*-ratios provide appreciably more highly statistically significant evidence against the hypotheses  $\beta_1 = 0$ ;
  - the point with the estimated residual of largest magnitude is for 1990, the year of subdivision of the 16 observations into two groups.