# Chapter 7
# Simple linear regression and Correlation

**11.53** The following data represent the chemistry grades for a random sample of 12 freshmen at a certain college along with their scores on an intelligence test administered while they were still seniors in high school.

| Student | Test Score, $x$ | Chemistry Grade, $y$ |
|---|---|---|
| 1 | 65 | 85 |
| 2 | 50 | 74 |
| 3 | 55 | 76 |
| 4 | 65 | 90 |
| 5 | 55 | 85 |
| 6 | 70 | 87 |
| 7 | 65 | 94 |
| 8 | 70 | 98 |
| 9 | 55 | 81 |
| 10 | 70 | 91 |
| 11 | 50 | 76 |
| 12 | 55 | 74 |

(a) Compute and interpret the sample correlation coefficient.

a) $$r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{\sum x_i y_i - n\,\bar{X}\bar{Y}}{\sqrt{(\sum x_i^2 - n\,\bar{X}^2)(\sum y_i^2 - n\,\bar{Y}^2)}}$$

$$\bar{X} = 60.4167, \bar{Y} = 84.25 \;\; ; \sum x_i^2 = 44475; \sum y_i^2 = 85905; \sum x_i y_i = 61685$$

Thus, $$r = \frac{61685 - (12)(60.4167)(84.25)}{\sqrt{(44475 - (12)(60.4167)^2)(85905 - (12)(84.25)^2)}} = 0.862$$

**Strong positive correlation**

**We can use the calculator to find the Correlation**

MODE → 3: Stat → 2: A+BX → | x | y | → AC

Shift → STAT 1 → 5: Reg → 3: r → =

**11.9** A study was made by a retail merchant to determine the relation between weekly advertising expenditures and sales.

| Advertising Costs ($) | Sales ($) |
|---|---|
| 40 | 385 |
| 20 | 400 |
| 25 | 395 |
| 20 | 365 |
| 30 | 475 |
| 50 | 440 |
| 40 | 490 |
| 20 | 420 |
| 50 | 560 |
| 40 | 525 |
| 25 | 480 |
| 50 | 510 |

(a) Find the equation of the regression line to predict weekly sales from advertising expenditures.
(b) Estimate the weekly sales when advertising costs are $35.

a)

$$\hat{Y}_i = b_0 + b_1 x, \qquad b_1 = \frac{\sum x_i y_i - n\,\bar{X}\bar{Y}}{\sum x_i^2 - n\bar{X}^2}, \qquad b_0 = \bar{Y} - b_1\bar{X}$$

$\bar{X} = \frac{\sum x}{n} = 34.1667$ , $\bar{Y} = \frac{\sum y}{n} = 453.75.$

$\sum x_i y_i = 191325$ ; $\sum x_i^2 = 15650$

Thus, $b_1 = \frac{191325-(12)(34.1667)(453.75)}{15650-(12)(34.1667)^2} = 3.22$

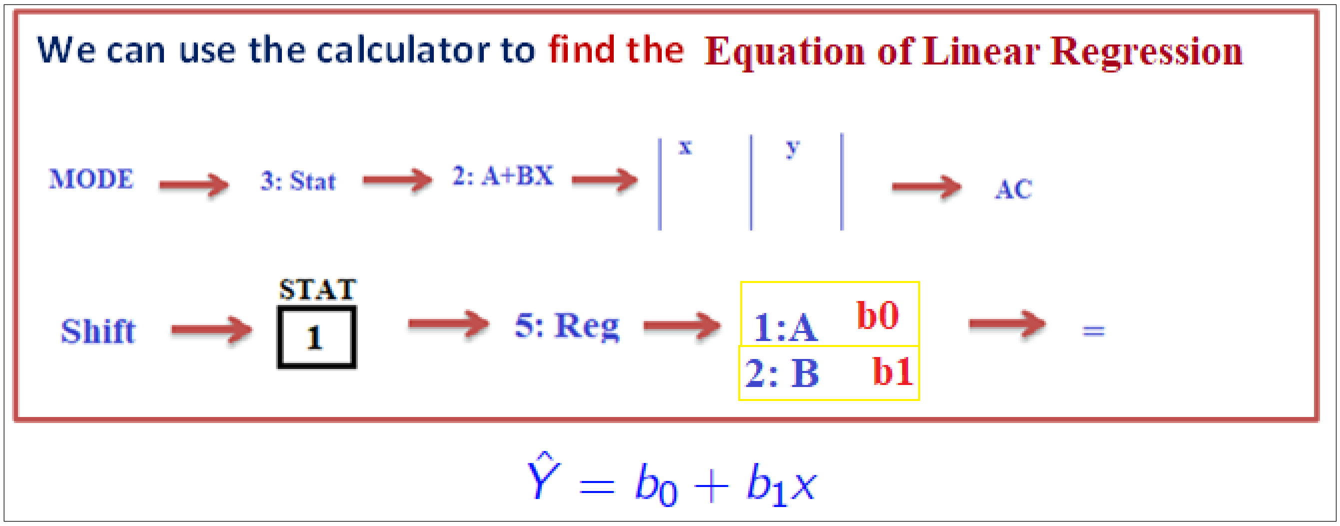
$$b_0 = 453.75 - (3.22)(34.1667) = 343.7$$

**The equation of regression line is:**

$$\hat{Y}_i = 343.7 + 3.22\, x_i$$

b) **Point Estimation of weekly sales when advertising costs are $35**

$$x_k = 35,\ \hat{Y}_k = 343.7 + 3.22\,(35) = 456.434.$$




**11.21** Test the hypothesis that β1 = 6 in (**Exercise 11.9 on page** 399) against the alternative that β1 < 6. Use a 0.025 level of significance.

**The hypotheses :**

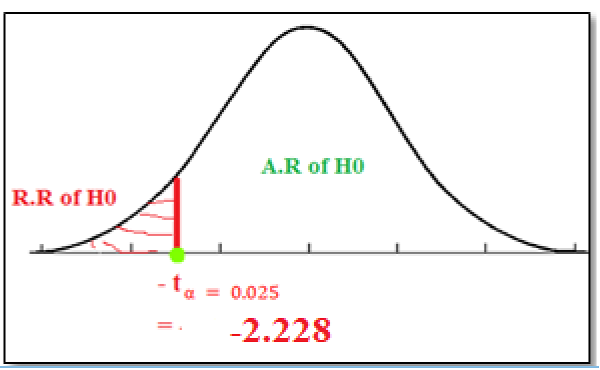
$$H_0: \beta_1 = 6 \qquad vs \qquad H_1: \beta_1 < 6$$

**Test statistics:**

$$t = \frac{b_1 - \beta_1}{\hat{\sigma}/\sqrt{S_{xx}}} = \frac{3.22 - 6}{50.266/\sqrt{1641.64}} = -2.24$$

*degrees of freedom= n -2 = 12- 2 =10*

**R.R&A.R of $H_0$:**



**Decision**:

$$If\ t < -t_\alpha\ , we\ reject\ H_0$$

$\dashrightarrow t = -2.24 < -2.228$ , $we\ Reject\ H0\ and\ conclude\ \beta_1 < 6$

**We can also use the p-value to get the decision**

$p - value\ = 0.02\ < \ \alpha\ =\ 0.025$ ,, $Reject\ H_0\ and\ conclude\ \beta_1 < 6$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 = 1641.64$$

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2} = \frac{25226.22}{12-2} = 2522.66$$

| y | x | $\hat{y} = 343.7 + 3.33x$ | $(y_i - \hat{y}_i)^2$ | |
|---|---|---|---|---|
| 385 | 40 | 472.5 | 7656.25 | |
| 400 | 20 | 408.1 | 65.61 | |
| 395 | 25 | 424.2 | 852.64 | |
| 365 | 20 | 408.1 | 1857.61 | |
| 475 | 30 | 440.3 | 1204.09 | |
| 440 | 50 | 504.7 | 4186.09 | |
| 490 | 40 | 472.5 | 306.25 | |
| 420 | 20 | 408.1 | 141.61 | |
| 560 | 50 | 504.7 | 3058.09 | |
| 525 | 40 | 472.5 | 2756.25 | |
| 480 | 25 | 424.2 | 3113.64 | |
| 510 | 50 | 504.7 | 28.09 | |
| | | | 25226.22 | Total |

| x | $(x_i - \bar{x})^2$ | |
|---|---|---|
| 40 | 34.024 | |
| 20 | 200.7 | |
| 25 | 84.034 | |
| 20 | 200.7 | |
| 30 | 17.364 | |
| 50 | 250.68 | |
| 40 | 34.024 | |
| 20 | 200.7 | |
| 50 | 250.68 | |
| 40 | 34.024 | |
| 25 | 84.034 | |
| 50 | 250.68 | |
| | 1641.7 | Total |

**11.21 With reference to Exercise 11.9**

a) Construct a 99% confidence interval for $\beta_1$.

b) Find and interpret the Coefficient of Determination $R^2$ .

**a) a 99% C.I of $\beta_1$:**

$$b_1 \pm t_{\frac{\alpha}{2},n-2} \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

$$3.22 \pm (3.169)\left(\frac{50.266}{\sqrt{1641.64}}\right)$$

$$\beta_1 \in (\,-0.71\,,\quad 7.15\,)$$

$\alpha = 0.01\,, df = n - 1 = 10$

$t_{\frac{\alpha}{2}} = t_{0.005} = 3.169$

**b)** $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{25226.22}{42256.25} = 0.4030$

This means that the 40.30% of change in the mean sales for retail merchant is by advertising expenditures.

| x | y | $\hat{y} = 343.7 + 3.22x$ | $(y-\hat{y})^2$ | $(y-\bar{y})^2$ | $(x-\bar{x})^2$ | |
|---|---|---|---|---|---|---|
| 40 | 385 | 472.5 | 7656.25 | 4726.5625 | 34.02778 | |
| 20 | 400 | 408.1 | 65.61 | 2889.0625 | 200.6944 | |
| 25 | 395 | 424.2 | 852.64 | 3451.5625 | 84.02778 | |
| 20 | 365 | 408.1 | 1857.61 | 7876.5625 | 200.6944 | |
| 30 | 475 | 440.3 | 1204.09 | 451.5625 | 17.36111 | |
| 50 | 440 | 504.7 | 4186.09 | 189.0625 | 250.6944 | |
| 40 | 490 | 472.5 | 306.25 | 1314.0625 | 34.02778 | |
| 20 | 420 | 408.1 | 141.61 | 1139.0625 | 200.6944 | |
| 50 | 560 | 504.7 | 3058.09 | 11289.0625 | 250.6944 | |
| 40 | 525 | 472.5 | 2756.25 | 5076.5625 | 34.02778 | |
| 25 | 480 | 424.2 | 3113.64 | 689.0625 | 84.02778 | |
| 50 | 510 | 504.7 | 28.09 | 3164.0625 | 250.6944 | |
| | | | 25226.22 | 42256.25 | 1641.667 | Total |
| | | | SSE | SST $S_{yy}$ | $S_{xx}$ | |

| | |
|---|---|
| $\bar{y} =$ | 453.75 |
| $\bar{x} =$ | 34.16666667 |