CHAPTER

A simple linear regression model

13

This is a rather extensive chapter on an important subject matter with an abundance of diverse applications. The basic idea involved may be described as follows. There is a stimulus, denoted by x, and a response to it, denoted by y. At different levels of x, one observes the respective responses. How are the resulting (x, y) pairs related, if they are related at all? There are all kind of possibilities, and the one discussed in this chapter is the simplest such possibility, namely, the pairs are linearly related.

In reality, what one, actually, observes at x, due to errors, is a value of a r.v. Y, and then the question arises as to how we would draw a straight line, which would lie "close" to most of the (x, y) pairs. This leads to the Principle of Least Squares. On the basis of this principle, one is able to draw the so-called fitted linear regression line by computing the Least Squares Estimates (LSE's) of parameters involved. Also, some properties of these estimates are established. These things are done in the first two sections of the chapter.

Up to this point, the errors are not required to have any specific distribution, other than having zero mean and finite variance. However, in order to proceed with further statistical inference about the parameters involved, such as constructing confidence intervals and testing hypotheses, one has to stipulate a distribution for the errors; this distribution, reasonably enough, is assumed to be Normal. As a consequence of it, one is in a position to specify the distribution of all estimates involved and proceed with the inference problems referred to above. These issues are discussed in Sections 13.3 and 13.4.

In the following section, Section 13.5, the problem of predicting the expected value of the observation Y_0 at a given point x_0 and the problem of predicting a single value of Y_0 are discussed. Suitable predictors are provided, and also confidence intervals for them are constructed.

The chapter is concluded with Section 3.7 indicating extensions of the model discussed in this chapter to more general situations covering a much wider class of applications.

13.1 SETTING UP THE MODEL—THE PRINCIPLE OF LEAST SQUARES

As has already been mentioned elsewhere, Examples 22 and 23 in Chapter 1 provide motivation for the statistical model to be adopted and studied in this chapter.

An Introduction to Probability and Statistical Inference

Example 22, in particular, will serve throughout the chapter to illustrate the underlying general results. For convenience, the data related to this example are reproduced here in Table 13.1.

Table 13.1 The Data x = Undergraduate GPA and y = Score in the Graduate Management Aptitude Test (GMAT); There Are 34 $(x,\ y)$ Pairs Altogether

Data of Undergraduate GPA (*x*) and GMAT Score (*y*)					
x	***y***	***x***	***y***	***x***	***y***
3.63	447	2.36	399	2.80	444
3.59	588	2.36	482	3.13	416
3.30	563	2.66	420	3.01	471
3.40	553	2.68	414	2.79	490
3.50	572	2.48	533	2.89	431
3.78	591	2.46	509	2.91	446
3.44	692	2.63	504	2.75	546
3.48	528	2.44	336	2.73	467
3.47	552	2.13	408	3.12	463
3.35	520	2.41	469	3.08	440
3.39	543	2.55	538	3.03	419
				3.00	509

The first question which arises is whether the pairs (x, y) are related at all and, if they are, how? An indication that those pairs are, indeed, related is borne out by the scatter plot depicted in Figure 13.1. Indeed, taking into consideration that we are operating in a random environment, one sees a conspicuous, albeit somewhat loose, linear relationship between the pairs (x, y).

So, we are not too far off the target by assuming that there is a straight line in the xy-plane which is "close" to most of the pairs (x, y). The question now is how to quantify the term "close." The first step toward this end is the adoption of the model described in relation (11) of Chapter 8. Namely, we assume that, for each $i = 1, \ldots, 34$, the respective y_i is the observed value of a r.v. Y_i associated with x_i, and if it were not for the random errors involved, the pairs (x_i, y_i), $i = 1, \ldots, 34$ would lie on a straight line $y = \beta_1 + \beta_2 x$; i.e., we would have $y_i = \beta_1 + \beta_2 x_i$, $i = 1, \ldots, 34$. Thus, the r.v. Y_i itself, whose y_i are simply observed values, would be equal to $\beta_1 + \beta_2 x_i$ except for fluctuations due to a random error e_i. In other words, $Y_i = \beta_1 + \beta_2 x_i + e_i$. Next, arguing as in Section 8.4, it is reasonable to assume that the e_i's are independent r.v.'s with $Ee_i = 0$ and $Var(e_i) = \sigma^2$ for all i's, so that one arrives at the model described in relation (11) of Chapter 8; namely, $Y_1, \ldots, Y_{34}$ are independent r.v.'s having the structure:

$$Y_i = \beta_1 + \beta_2 x_i + e_i, \quad \text{with } Ee_i = 0 \quad \text{and} \quad Var(e_i) = \sigma^2 (< \infty), \quad i = 1, \ldots, 34. \tag{1}$$

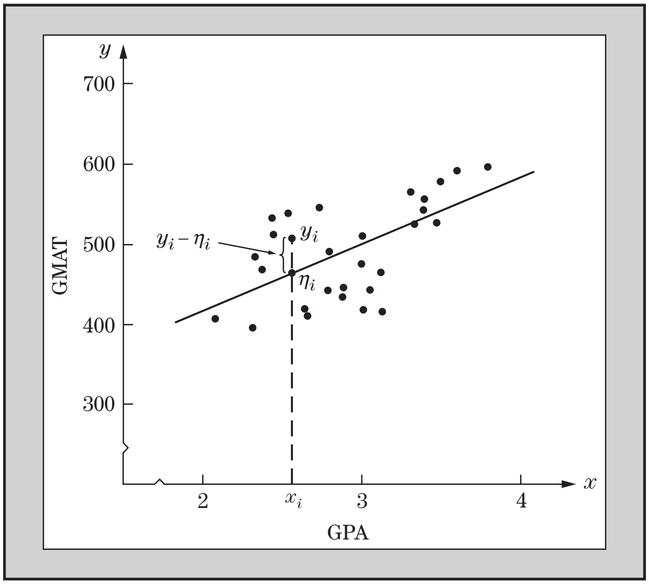

FIGURE 13.1

Scatter diagram for Table 13.1.

Set $EY_i = \eta_i$. Then, because of the errors involved, it is, actually, the pairs $(x_i, \eta_i), i = 1, \ldots, 34$ which lie on a straight line $y = \beta_1 + \beta_2 x$; i.e., $\eta_i = \beta_1 + \beta_2 x_i, i = 1, \ldots, 34$. It is in the determination of a particular straight line where the Principle of Least Squares enters the picture. According to this principle, one argues as follows: On the basis of the model described in (1), what we would *expect* to have observed at x_i would be η_i, whereas what is, *actually*, observed is y_i. Thus, there is a deviation measured by $y_i - \eta_i, i = 1, \ldots, 34$ (see Figure 13.1). Some of these deviations are positive, some are negative, and, perhaps, some are zero. In order to deal with non-negative numbers, look at $|y_i - \eta_i|$, which is, actually, the distance between the points (x_i, y_i) and (x_i, η_i). Then, draw the line $y = \beta_1 + \beta_2 x$, so that these distances are simultaneously minimized. More formally, first look at the squares of these distances $(y_i - \eta_i)^2$, as it is much easier to work with squares as opposed to absolute values, and in order to account for the simultaneous minimization mentioned earlier, consider the sum $\sum_{i=1}^{34} (y_i - \eta_i)^2$ and seek its minimization. At this point, replace the observed value y_i by the r.v. Y_i itself and set

$$\mathcal{S}(\boldsymbol{Y}, \boldsymbol{\beta}) = \sum_{i=1}^{34} (Y_i - \eta_i)^2 = \sum_{i=1}^{34} [Y_i - (\beta_1 + \beta_2 x_i)]^2 \left(= \sum_{i=1}^{34} e_i^2 \right), \tag{2}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_{34})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)$.

Then the *Principle of Least Squares* calls for the determination of β_1 and β_2 which *minimize the sum of squares of errors*; i.e., the quantity $\mathcal{S}(\boldsymbol{Y}, \boldsymbol{\beta})$ in (2). The actual minimization is a calculus problem. If there is a unique straight line so determined, then, clearly, this would be the line which lies "close" to most pairs $(x_i, Y_i), i = 1, \ldots, 34$, in the Least Squares sense. It will be seen below that this is, indeed, the case.

In a more general setting, consider the model below:

$$\begin{aligned}
&Y_i = \beta_1 + \beta_2 x_i + e_i, \text{ where the random errors}\\
&e_i, i = 1, \ldots, n \text{ are i.i.d. r.v.'s. with}\\
&Ee_i = 0, \text{and } Var(e_i) = \sigma^2 (< \infty), \text{which imply that the r.v.'s}\\
&Y_i, i = 1, \ldots, n \text{ are independent, but not identically distributed, with}\\
&EY_i = \eta_i = \beta_1 + \beta_2 x_i \text{ and } Var(y_i) = \sigma^2.
\end{aligned} \tag{3}$$

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the unique values of β_1 and β_2, respectively, which minimize the sum of squares of errors $\mathcal{S}(\boldsymbol{Y}, \boldsymbol{\beta}) = \sum_{i=1}^{n}[Y_i - (\beta_1 + \beta_2 x_i)]^2 (= \sum_{i=1}^{n} e_i^2)$. These values, which are functions of the Y_i's as well as the x_i's, are the LSE's of β_1 and β_2. Any line $y = \beta_1 + \beta_2 x$ is referred to as a *regression line* and, in particular, the line $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ is known as the *fitted regression line*. For this line, the $\hat{y}_i$'s corresponding to the x_i's are $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i,\ i = 1, \ldots, n$.

13.2 THE LEAST SQUARES ESTIMATES OF β_1 AND β_2 AND SOME OF THEIR PROPERTIES

In this section, the LSE's of β_1 and β_2 are derived and some of their properties are obtained. Also, the (unknown) variance σ^2 is estimated.

Theorem 1. *In reference to the model described in (3), the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ of β_1 and β_2, respectively, are given by the following expressions (which are also appropriate for computational purposes):*

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} Y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i Y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{4}$$

and

$$\hat{\beta}_2 = \frac{n\sum_{i=1}^{n} x_i Y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \tag{5}$$

Proof. Consider the partial derivatives:

$$\begin{aligned}
\frac{\partial}{\partial \beta_1}\mathcal{S}(\boldsymbol{Y}, \boldsymbol{\beta}) &= 2\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 x_i)(-1) = -2\left(\sum_{i=1}^{n} Y_i - n\beta_1 - \beta_2 \sum_{i=1}^{n} x_i\right),\\
\frac{\partial}{\partial \beta_2}\mathcal{S}(\boldsymbol{Y}, \boldsymbol{\beta}) &= 2\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 x_i)(-x_i)
\end{aligned} \tag{6}$$

$$= -2\left(\sum_{i=1}^n x_iY_i - \beta_1\sum_{i=1}^n x_i - \beta_2\sum_{i=1}^n x_i^2\right), \tag{7}$$

and solve the so-called *Normal equations:* $\frac{\partial}{\partial\beta_1}\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta}) = 0$ and $\frac{\partial}{\partial\beta_2}\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta}) = 0$, or $n\beta_1 + (\sum_{i=1}^n x_i)\beta_2 = \sum_{i=1}^n Y_i$ and $(\sum_{i=1}^n x_i)\beta_1 + (\sum_{i=1}^n x_i^2)\beta_2 = \sum_{i=1}^n x_iY_i$ to find:

$$\hat{\beta}_1 = \frac{\begin{vmatrix} \sum_{i=1}^n Y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_iY_i & \sum_{i=1}^n x_i^2 \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix}} = \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n Y_i\right) - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n x_iY_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2},$$

and

$$\hat{\beta}_2 = \frac{\begin{vmatrix} n & \sum_{i=1}^n Y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_iY_i \end{vmatrix}}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{n\sum_{i=1}^n x_iY_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n Y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}.$$

It remains to show that $\hat{\beta}_1$ and $\hat{\beta}_2$, actually, minimize $\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta})$. From (6) and (7), we get:

$$\frac{\partial^2}{\partial\beta_1^2}\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta}) = 2n, \quad \frac{\partial^2}{\partial\beta_1\,\partial\beta_2}\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta}) = \frac{\partial^2}{\partial\beta_2\,\partial\beta_1}\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta}) = 2\sum_{i=1}^n x_i,$$
$$\frac{\partial^2}{\partial\beta_2^2}\mathcal{S}(\boldsymbol{Y},\boldsymbol{\beta}) = 2\sum_{i=1}^n x_i^2,$$

and the 2×2 matrix below is positive semidefinite for all β_1, β_2, since, for all λ_1, λ_2 reals not both 0,

$$\begin{aligned}
&(\lambda_1,\lambda_2)\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix} \\
&= \left(\lambda_1 n + \lambda_2\sum_{i=1}^n x_i \quad \lambda_1\sum_{i=1}^n x_i + \lambda_2\sum_{i=1}^n x_i^2\right)\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix} \\
&= \lambda_1^2 n + 2\lambda_1\lambda_2\sum_{i=1}^n x_i + \lambda_2^2\sum_{i=1}^n x_i^2 \\
&= \lambda_1^2 n + 2n\lambda_1\lambda_2\bar{x} + \lambda_2^2\sum_{i=1}^n x_i^2 \quad \left(\text{where } \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i\right) \\
&= \lambda_1^2 n + 2n\lambda_1\lambda_2\bar{x} + \lambda_2^2\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) + \lambda_2^2 n\bar{x}^2 \\
&= n\left(\lambda_1^2 + 2\lambda_1\lambda_2\bar{x} + \lambda_2^2\bar{x}^2\right) + \lambda_2^2\sum_{i=1}^n (x_i - \bar{x})^2 \\
&= n(\lambda_1 + \lambda_2\bar{x})^2 + \lambda_2^2\sum_{i=1}^n (x_i - \bar{x})^2 \ge 0.
\end{aligned}$$

This completes the proof of the theorem. ■

Corollary. *With $\bar{x} = (x_1 + \cdots + x_n)/n$ and $\bar{Y} = (Y_1 + \cdots + Y_n)/n$, the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ may also be written as follows (useful expressions for noncomputational purposes):*

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{x}, \quad \hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x})Y_i. \tag{8}$$

Proof. In the first place,

$$\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) = \sum_{i=1}^n (x_i - \bar{x})Y_i - \bar{Y}\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})Y_i,$$

which shows that $\hat{\beta}_2$ is as claimed.

Next, replacing $\hat{\beta}_2$ by its expression in (5), we get

$$\begin{aligned}
\bar{Y} - \hat{\beta}_2\bar{x} &= \bar{Y} - \frac{n\sum_{i=1}^n x_iY_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n Y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\bar{x} \\
&= \bar{Y} - \frac{n\sum_{i=1}^n x_iY_i - n^2\bar{x}\bar{Y}}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2}\bar{x} \\
&= \frac{n\bar{Y}\sum_{i=1}^n x_i^2 - n^2\bar{x}^2\bar{Y} - n\bar{x}\sum_{i=1}^n x_iY_i + n^2\bar{x}^2\bar{Y}}{n\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n Y_i\right) - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n x_iY_i\right)}{n\sum_{i=1}^n (x_i - \bar{x})^2} = \hat{\beta}_1 \text{ by (4).}
\end{aligned}$$

■

The following notation is suggested (at least in part) by the expressions in the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$, and it will be used extensively and conveniently throughout the rest of this chapter.

Set

$$SS_x = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2,$$

and likewise,

$$SS_y = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = \sum_{i=1}^n Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^n Y_i\right)^2, \tag{9}$$

and

$$\begin{aligned}
SS_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) = \sum_{i=1}^n (x_i - \bar{x})Y_i \\
&= \sum_{i=1}^n x_iY_i - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n Y_i\right).
\end{aligned}$$

Then the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ may be rewritten as follows:

$$\hat{\beta}_1 = \frac{1}{n}\sum_{i=1}^n Y_i - \hat{\beta}_2\left(\frac{1}{n}\sum_{i=1}^n x_i\right), \quad \hat{\beta}_2 = \frac{SS_{xy}}{SS_x}. \tag{10}$$

Table 13.2 GPA's and the corresponding GMAT scores

	x	y	x^2	xy	x	y	x^2	xy
	3.63	447	13.1769	1622.61	3.48	528	12.1104	1837.44
	3.59	588	12.8881	2110.92	3.47	552	12.0409	1915.44
	3.30	563	10.8900	1857.90	3.35	520	11.2225	1742.00
	3.40	553	11.5600	1880.20	3.39	543	11.4921	1840.77
	3.50	572	12.2500	2002.00	2.36	399	5.5696	941.64
	3.78	591	14.2884	2233.98	2.36	482	5.5696	1137.52
	3.44	692	11.8336	2380.48	2.66	420	7.0756	1117.20
	2.68	414	7.1824	1109.52	3.01	471	9.0601	1417.71
	2.48	533	6.1504	1321.84	2.79	490	7.7841	1367.10
	2.46	509	6.0516	1252.14	2.89	431	8.3521	1245.59
	2.63	504	6.9169	1325.52	2.91	446	8.4681	1297.86
	2.44	336	5.9536	819.84	2.75	546	7.5625	1501.50
	2.13	408	4.5369	869.04	2.73	467	7.4529	1274.91
	2.41	469	5.8081	1130.29	3.12	463	9.7344	1444.56
	2.55	538	6.5025	1371.90	3.08	440	9.4864	1355.20
	2.80	444	7.8400	1243.20	3.03	419	9.1809	1269.57
	3.13	416	9.7969	1302.08	3.00	509	9.0000	1527.00
Total	50.35	8577	153.6263	25,833.46	50.38	8126	151.1622	24,233.01

Also, recall that the fitted regression line is given by:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x \quad \text{and that} \quad \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i, \qquad i = 1, \ldots, n. \tag{11}$$

Before we go any further, let us discuss the example below.

Example 1. In reference to Table 13.1, compute the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ and draw the fitted regression line $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$.

Discussion. The application of formula (10) calls for the calculation of SS_x and SS_{xy} given in (9). Table 13.2 facilitates the calculations.

$$\sum_i x_i = 100.73, \quad \sum_i y_i = 16{,}703, \quad \sum_i x_i^2 = 304.7885, \quad \sum_i x_i y_i = 50{,}066.47,$$

and then

$$\begin{aligned} SS_x &= 304.7885 - \frac{(100.73)^2}{34} \simeq 304.7885 - 298.4274 \simeq 6.361, \\ SS_{xy} &= 50{,}066.47 - \frac{(100.73) \times (16{,}703)}{34} \simeq 50{,}066.47 - 49{,}485.094 \\ &= 581.376. \end{aligned}$$

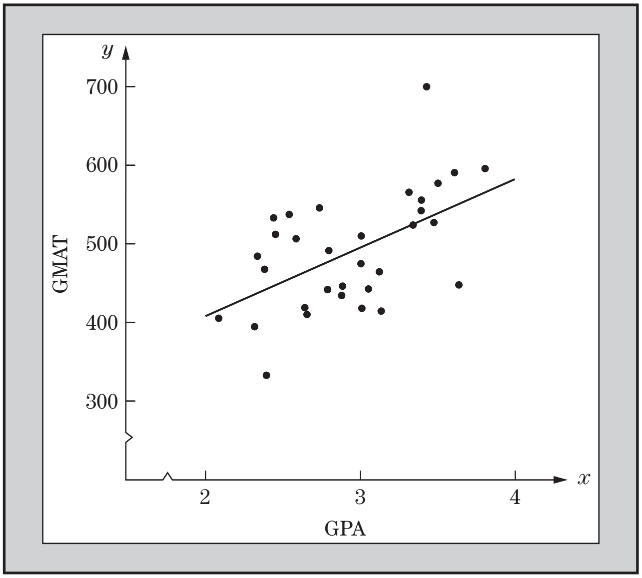

FIGURE 13.2

The fitted regression line $\hat{y} = 220.456 + 91.397x$.

Thus,

$$\hat{\beta}_2 = \frac{581.376}{6.361} \simeq 91.397 \quad \text{and} \quad \hat{\beta}_1 = \frac{16{,}703}{34} - (91.397) \times \frac{100.73}{34}$$
$$\simeq 491.265 - 270.809 = 220.456,$$

and the fitted regression line $\hat{y} = 220.456 + 91.397x$ is depicted in the Figure 13.2.

The LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ have the desirable property of being unbiased, as shown in the following theorem.

Theorem 2. *The LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased; i.e., $E\hat{\beta}_1 = \beta_1$ and $E\hat{\beta}_2 = \beta_2$. Furthermore,*

$$Var(\hat{\beta}_1) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SS_x}\right) \quad \text{and} \quad Var(\hat{\beta}_2) = \frac{\sigma^2}{SS_x},$$

where SS_x is given in (9).

Proof. In this proof and also elsewhere, the range of the summation is not explicitly indicated, since it is always from 1 to n. Consider $\hat{\beta}_2$ as given in (8). Then: $SS_x\hat{\beta}_2 = \sum_i (x_i - \bar{x})Y_i$, so that, by taking expectations:

$$SS_x E\hat{\beta}_2 = \sum_i (x_i - \bar{x})EY_i = \sum_i (x_i - \bar{x})(\beta_1 + \beta_2 x_i)$$

$$
\begin{aligned}
&= \beta_1 \sum_i (x_i - \bar{x}) + \beta_2 \sum_i x_i(x_i - \bar{x}) = \beta_2 \sum_i x_i(x_i - \bar{x})\\
&= \beta_2 \left(\sum_i x_i^2 - n\bar{x}^2 \right) = \beta_2 \sum_i (x_i - \bar{x})^2 = SS_x \beta_2.
\end{aligned}
$$

Therefore, dividing through by SS_x, we get $E\hat{\beta}_2 = \beta_2$. Next, also from (8),

$$
\begin{aligned}
E\hat{\beta}_1 &= E(\bar{Y} - \hat{\beta}_2 \bar{x}) = E\bar{Y} - \bar{x}E\hat{\beta}_2 = \frac{1}{n}\sum_i (\beta_1 + \beta_2 x_i) - \bar{x}\beta_2\\
&= \beta_1 + \beta_2\bar{x} - \beta_2\bar{x} = \beta_1.
\end{aligned}
$$

Regarding the variances, we have from (8): $SS_x \hat{\beta}_2 = \sum_i (x_i - \bar{x})Y_i$, so that:

$$
\begin{aligned}
SS_x^2\, Var(\hat{\beta}_2) &= Var\left(\sum_i (x_i - \bar{x})Y_i \right) = \sum_i (x_i - \bar{x})^2 Var(Y_i)\\
&= \sigma^2 \sum_i (x_i - \bar{x})^2 = \sigma^2 SS_x,
\end{aligned}
$$

so that $Var(\hat{\beta}_2) = \sigma^2 / SS_x$. Finally, from (8),

$$
\hat{\beta}_1 = \bar{Y} - \bar{x}\hat{\beta}_2 = \frac{1}{n}\sum_i Y_i - \frac{\bar{x}}{SS_x}\sum_i (x_i - \bar{x})Y_i = \sum_i \left[\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{SS_x} \right] Y_i, \qquad (12)
$$

so that

$$
Var(\hat{\beta}_1) = \sigma^2 \sum_i \left[\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{SS_x} \right]^2 = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SS_x^2} SS_x \right) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SS_x} \right). \quad \blacksquare
$$

Example 2. In reference to Example 1, the variances of the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ are given by:

$$
Var(\hat{\beta}_1) \simeq \sigma^2 \left(\frac{1}{34} + \frac{8.777}{6.361} \right) \simeq \sigma^2 (0.029 + 1.380) = 1.409\sigma^2,
$$

and

$$
Var(\hat{\beta}_2) = \frac{\sigma^2}{6.361} \simeq 0.157\sigma^2.
$$

In fitting a regression line, there are various deviations which occur. At this point, these deviations will be suitably attributed to several sources, certain pieces of terminology will be introduced, and also some formal relations will be established. To this end, look at the observable Y_i and split it as follows: $Y_i = \hat{y}_i + (Y_i - \hat{y}_i)$. The component $\hat{y}_i$ is associated with the point $(x_i, \hat{y}_i)$ which lies on the fitted regression line $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$, and the difference $Y_i - \hat{y}_i$ is the deviation of Y_i from $\hat{y}_i$. We may refer to the component $\hat{y}_i$ as that part of Y_i which is *due to the linear regression*, or it *is explained by the linear regression*, and the component $Y_i - \hat{y}_i$ of Y_i as the *residual*, or the *deviation from the linear regression*, or *variability unexplained by the linear* *regression*. We can go through the same arguments with reference to the sample mean $\bar{Y}$ of the Y_i's. That is, we consider:

$$Y_i - \bar{Y} = (\hat{y}_i - \bar{Y}) + (Y_i - \hat{y}_i).$$

The interpretation of this decomposition is the same as the one given above, but with reference to $\bar{Y}$. Next, look at the squares of these quantities:

$$(Y_i - \bar{Y})^2, \quad (\hat{y}_i - \bar{Y})^2, \quad (Y_i - \hat{y}_i)^2,$$

and, finally, at their sums:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2, \quad \sum_{i=1}^{n}(Y_i - \hat{y}_i)^2.$$

At this point, assume for a moment that:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{y}_i)^2. \tag{13}$$

Then this relation would state that the *total variability* (of the Y_i's in reference to their mean $\bar{Y}$), $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, is the sum of the variability $\sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2$ *due to the linear regression*, or *explained by the linear regression*, and the *sum of residual variability*, $\sum_{i=1}^{n}(Y_i - \hat{y}_i)^2$, or *variability unexplained by the linear regression.*

We proceed in proving relation (13).

Theorem 3. *Let SS_T(=SS_y, see (9)), SS_R, and SS_E, respectively, be the total variability, the variability due to the linear regression (or explained by the linear regression), and the residual variability (or variability not explained by the linear regression); i.e.,*

$$SS_T(= SS_y) = \sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2, \quad SS_E = \sum_{i=1}^{n}(Y_i - \hat{y}_i)^2, \tag{14}$$

where $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i, i = 1, \ldots, n$, the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ are given by (10) (or (8)), and $\bar{Y}$ is the mean of the Y_i's. Then:

(i)

$$SS_T = SS_R + SS_E. \tag{15}$$

Furthermore,

(ii)

$$SS_T = SS_y, \quad SS_R = \frac{SS_{xy}^2}{SS_x}, \quad \text{and hence} \quad SS_E = SS_y - \frac{SS_{xy}^2}{SS_x}, \tag{16}$$

where SS_x, SS_y, and SS_{xy} are given in (9).

Proof.

(i) We have:

$$SS_T = \sum_i (Y_i - \bar{Y})^2 = \sum_i [(\hat{y}_i - \bar{Y}) + (Y_i - \hat{y}_i)]^2$$

$$
\begin{aligned}
&= \sum_i (\hat{y}_i - \bar{Y})^2 + \sum_i (Y_i - \hat{y}_i)^2 + 2\sum_i (\hat{y}_i - \bar{Y})(Y_i - \hat{y}_i) \\
&= SS_R + SS_E + 2\sum_i (\hat{y}_i - \bar{Y})(Y_i - \hat{y}_i).
\end{aligned}
$$

So, we have to show that the last term on the right-hand side above is equal to 0. To this end, observe that $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ and $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x}$ (by (8)), so that

$$\hat{y}_i - \bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 x_i - \bar{Y} = \bar{Y} - \hat{\beta}_2 \bar{x} + \hat{\beta}_2 x_i - \bar{Y} = \hat{\beta}_2 (x_i - \bar{x}),$$

and

$$Y_i - \hat{y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i = Y_i - \bar{Y} + \hat{\beta}_2 \bar{x} - \hat{\beta}_2 x_i = (Y_i - \bar{Y}) - \hat{\beta}_2 (x_i - \bar{x}),$$

so that

$$
\begin{aligned}
(\hat{y}_i - \bar{Y})(Y_i - \hat{y}_i) &= \hat{\beta}_2 (x_i - \bar{x})[(Y_i - \bar{Y}) - \hat{\beta}_2 (x_i - \bar{x})] \\
&= \hat{\beta}_2 (x_i - \bar{x})(Y_i - \bar{Y}) - \hat{\beta}_2^2 (x_i - \bar{x})^2.
\end{aligned}
$$

Therefore, by (9) and (10):

$$\sum_i (\hat{y}_i - \bar{Y})(Y_i - \hat{y}_i) = \frac{SS_{xy}}{SS_x} \times SS_{xy} - \frac{SS_{xy}^2}{SS_x^2} \times SS_x = \frac{SS_{xy}^2}{SS_x} - \frac{SS_{xy}^2}{SS_x} = 0. \tag{17}$$

Thus, $SS_T = SS_R + SS_E$.

(ii) That $SS_T = SS_y$ is immediate from relations (9) and (14). Next, $\hat{y}_i - \bar{Y} = \hat{\beta}_2 (x_i - \bar{x})$ as was seen in the proof of part (i), so that, by (10),

$$SS_R = \sum_i (\hat{y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum_i (x_i - \bar{x})^2 = \frac{SS_{xy}^2}{SS_x^2} \times SS_x = \frac{SS_{xy}^2}{SS_x},$$

as was to be seen. Finally, by part (i), $SS_E = SS_T - SS_R = SS_y - \frac{SS_{xy}^2}{SS_x}$, as was to be seen. ∎

This section is closed with some remarks.

Remark 1.

(i) The quantities SS_T, SS_R, and SS_E, given in (14), are computed by way of SS_x, SS_y, and SS_{xy} given in (9). This is so because of (16).

(ii) In the next section, an estimate of the (unknown) variance σ^2 will also be given, based on the residual variability SS_E. That this should be the case is intuitively clear by the nature of SS_E, and it will be formally justified in the following section.

(iii) From the relation $SS_T = SS_R + SS_E$ given in (15) and the definition of the variability due to regression, SS_R, given in (14), it follows that the better the regression fit is, the smaller the value of SS_R is. Then, its ratio to the total variability, SS_T, $r = SS_R/SS_T$, can be used as an index of how good the linear regression fit is; the larger $r(\leq 1)$ the better the fit.

13.3 NORMALLY DISTRIBUTED ERRORS: MLE'S OF β_1, β_2, AND σ^2, SOME DISTRIBUTIONAL RESULTS

It is to be noticed that in the linear regression model as defined in relation (3), no distribution assumption about the errors e_i, and therefore the r.v.'s Y_i, was made. Such an assumption was not necessary, neither for the construction of the LSE's of β_1, β_2, nor in proving their unbiasedness and in calculating their variances. However, in order to be able to construct confidence intervals for β_1 and β_2 and test hypotheses about them, among other things, we have to assume a distribution for the e_i's. The e_i's being errors, it is not unreasonable to assume that they are Normally distributed, and we shall do so. Then the model (3) is supplemented as follows:

$$\begin{aligned} &Y_i = \beta_1 + \beta_2 x_i + e_i, \quad \text{where } e_i, i = 1, \ldots, n \text{ are independent r.v.'s} \\ &\quad \sim N(0, \sigma^2), \quad \text{which implies that } Y_i, i = 1, \ldots, n \text{ are} \\ &\text{independent r.v.'s and } Y_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2). \end{aligned} \tag{18}$$

We now proceed with the following theorem.

Theorem 4. *Under model (18):*

(i) *The LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$ of β_1 and β_2, respectively, are also MLE's.*
(ii) *The MLE $\hat{\sigma}^2$ of σ^2 is given by: $\hat{\sigma}^2 = SS_E/n$.*
(iii) *The estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are Normally distributed as follows:*

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SS_x}\right)\right), \quad \hat{\beta}_2 \sim N\left(\beta_2, \frac{\sigma^2}{SS_x}\right),$$

where SS_x is given in (9).

Proof.

(i) The likelihood function of the Y_i's is given by:

$$L(y_1, \ldots, y_n; \beta_1, \beta_2, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_i (y_i - \beta_1 - \beta_2 x_i)^2\right].$$

For each fixed σ^2, maximization of the likelihood function with respect to β_1 and β_2, is, clearly, equivalent to minimization of $\sum_i (y_i - \beta_1 - \beta_2 x_i)^2$ with respect to β_1 and β_2, which minimization has produced the LSE's $\hat{\beta}_1$ and $\hat{\beta}_2$.

(ii) The MLE of σ^2 is to be found by minimizing, with respect to σ^2, the expression:

$$\log L(y_i, \ldots, y_n; \hat{\beta}_1, \hat{\beta}_2, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}SS_E,$$

since, by (14) and (11), and by using the same notation SS_E both for $\sum_{i=1}^n (Y_i - \hat{y}_i)^2$ and $\sum_{i=1}^n (y_i - \hat{y}_i)^2$, we have $\sum_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 =$

$\sum_i (y_i - \hat{y}_i)^2 = SS_E$. From this expression, we get:

$$\frac{d}{d\sigma^2} \log L(y_1, \ldots, y_n; \hat{\beta}_1, \hat{\beta}_2, \sigma^2) = -\frac{n}{2} \times \frac{1}{\sigma^2} + \frac{SS_E}{2(\sigma^2)^2} = 0,$$

so that $\sigma^2 = SS_E/n$. Since

$$\frac{d^2}{d(\sigma^2)^2} \log L\big(y_1, \ldots, y_n; \hat{\beta}_1, \hat{\beta}_2, \sigma^2\big)\big|_{\sigma^2 = SS_E/n} = -\frac{n^3}{2SS_E^2} < 0,$$

it follows that $\hat{\sigma}^2 = SS_E/n$ is, indeed, the MLE of σ^2.

(iii) From (12), we have: $\hat{\beta}_1 = \sum_i [\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{SS_x}] Y_i$, and we have also seen in Theorem 2 that:

$$E\hat{\beta}_1 = \beta_1, \quad Var(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SS_x}\right).$$

Thus, $\hat{\beta}_1$ is Normally distributed as a linear combination of independent Normally distributed r.v.'s, and its mean and variance must be as stated above. Next, from (8), we have that: $\hat{\beta}_2 = \sum_i (\frac{x_i - \bar{x}}{SS_x}) Y_i$, so that, as above, $\hat{\beta}_2$ is Normally distributed. Its mean and variance have been computed in Theorem 2 and they are β_2 and σ^2/SS_x, respectively. ■

Before proceeding further, we return to Example 1 and compute an estimate for σ^2. Also, discuss Example 23 in Chapter 1 and, perhaps, an additional example to be introduced here.

Example 3. In reference to Example 1, determine the MLE of σ^2.

Discussion. By Theorem 4(ii), this estimate is: $\hat{\sigma}^2 = \frac{SS_E}{n}$. For the computation of SS_E by (16), we have to have the quantity $\sum y_i^2$ from Table 13.2, which is calculated to be:

$$\sum_i y_i^2 = 8{,}373{,}295. \tag{19}$$

Then, by (9),

$$SS_y = 8{,}373{,}295 - \frac{(16{,}703)^2}{34} \simeq 8{,}373{,}295 - 8{,}205{,}594.382 = 167{,}700.618,$$

and therefore

$$SS_E = 167{,}700.618 - \frac{(581.376)^2}{6.361} \simeq 167{,}700.618 - 53{,}135.993 = 114{,}564.625;$$

i.e.,

$$SS_E = 114{,}564.625 \quad \text{and then} \quad \hat{\sigma}^2 = \frac{114{,}564.625}{34} \simeq 3369.548.$$

Since $SS_T = SS_y = 167{,}700.618$ and $SS_R = 53{,}135.993$, it follows that only $\frac{53{,}135.993}{167{,}700.618} \simeq 31.685\%$ of the variability is explained by linear regression and $\frac{114{,}564.625}{167{,}700.618} \simeq 68.315\%$ is not explained by linear regression. The obvious outlier (3.44, 692) may be mainly responsible for it.

Example 4. In reference to Example 23 in Chapter 1, assume a linear relationship between the dose of a compost fertilizer x and the yield of a crop y. On the basis of the following summary data recorded:

$$n = 15, \quad \bar{x} = 10.8, \quad \bar{y} = 122.7, \quad SS_x = 70.6, \quad SS_y = 98.5, \quad SS_{xy} = 68.3:$$

(i) Determine the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, and draw the fitted regression line.
(ii) Give the MLE $\hat{\sigma}^2$ of σ^2.
(iii) Over the range of x values covered in the study, what would your conjecture be regarding the average increase in yield per unit increase in the compost dose?

Discussion.

(i) By (10),

$$\hat{\beta}_2 = \frac{68.3}{70.6} \simeq 0.967, \quad \hat{\beta}_1 = 122.7 - 0.967 \times 10.8 \simeq 112.256,$$
$$\text{and} \quad \hat{y} = 112.256 + 0.967x.$$

(ii) We have: $\hat{\sigma}^2 = \frac{SS_E}{15}$, where $SS_E = 98.5 - \frac{(68.3)^2}{70.6} \simeq 32.425$, so that $\hat{\sigma}^2 = \frac{32.425}{15} = 2.162$.
(iii) The conjecture would be a number close to the slope of the fitted regression line, which is 0.967 (Figure 13.3).

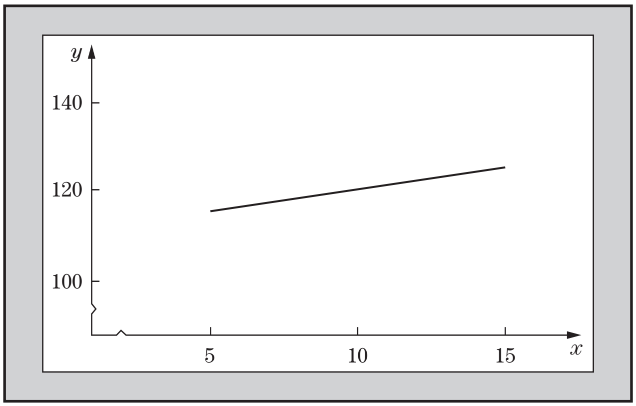

FIGURE 13.3

The fitted regression line $\hat{y} = 112.256 + 0.967x$.

Table 13.3 Dosage (x) (in milligrams) and the number of days of relief (y) from allergy for 10 patients

	x	y	x^2	y^2	xy
	3	9	9	81	27
	3	5	9	25	15
	4	12	16	144	48
	5	9	25	81	45
	6	14	36	196	84
	6	16	36	256	96
	7	22	49	484	154
	8	18	64	324	144
	8	24	64	576	192
	9	22	81	484	198
Total	59	151	389	2651	1003

Example 5. In one stage of the development of a new medication for an allergy, an experiment is conducted to study how different dosages of the medication affect the duration of relief from the allergic symptoms. Ten patients are included in the experiment. Each patient receives a specific dosage of the medication and is asked to report back as soon as the protection of the medication seems to wear off. The observations are recorded in Table 13.3, which shows the dosage (x) and respective duration of relief (y) for the 10 patients.

(i) Draw the scatter diagram of the data in Table 13.3 (which indicate tendency toward linear dependence).
(ii) Compute the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, and draw the fitted regression line.
(iii) What percentage of the total variability is explained by the linear regression and what percentage remains unexplained?
(iv) Compute the MLE $\hat{\sigma}^2$ of σ^2.

Discussion.

(i),(ii) First, $SS_x = 389 - \frac{59^2}{10} = 40.9$ and $SS_{xy} = 1003 - \frac{59\times151}{10} = 112.1$, and hence:

$$\hat{\beta}_2 = \frac{112.1}{40.9} \simeq 2.741 \quad \text{and} \quad \hat{\beta}_1 = \frac{151}{10} - 2.741 \times \frac{59}{10} \simeq -1.072.$$

Then the fitted regression line is $\hat{y} = -1.072 + 2.741x$ (Figure 13.4).

(iii) Since $SS_T = SS_y = 2651 - \frac{151^2}{10} = 370.9$ and $SS_R = \frac{(112.1)^2}{40.9} \simeq 307.247$, it follows that $SS_E = 370.9 - 307.247 = 63.653$. Therefore $\frac{307.247}{370.9} \simeq 82.838\%$

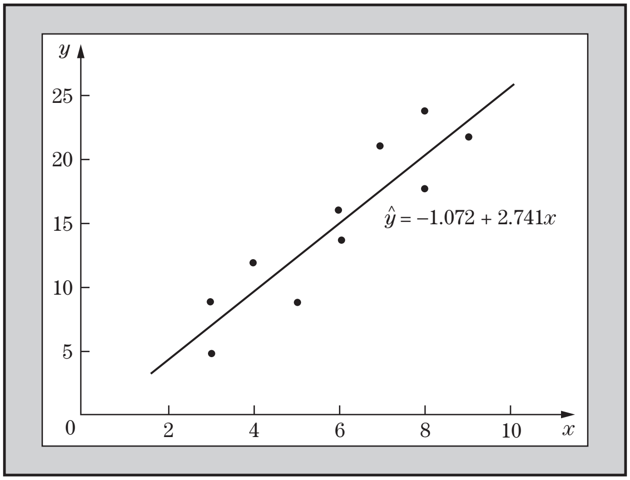

FIGURE 13.4

Scatter diagram and the fitted regression line $\hat{y} = -1.072 + 2.741x$.

of the variability is explained by the linear regression and $\frac{63.653}{370.9} \simeq 17.162\%$ remains unexplained.

(iv) We have: $\hat{\sigma}^2 = \frac{63.653}{10} = 6.3653 \simeq 6.365$.

For the purpose of constructing confidence intervals for the parameters of the model, and also testing hypotheses about them, we have to know the distribution of SS_E and also establish independence of the statistics $\hat{\beta}_1$ and SS_E, as well as independence of the statistics $\hat{\beta}_2$ and SS_E. The relevant results are stated in the following theorem, whose proof is deferred to Section 13.6.

Theorem 5. *Under model (18):*

(i) *The distribution of SS_E/σ^2 is χ^2_{n-2}.*
(ii) *The following statistics are independent:*

(a) SS_E and $\hat{\beta}_2$; (b) $\bar{Y}$ and $\hat{\beta}_2$; (c) $SS_E, \bar{Y}$ and $\hat{\beta}_2$; (d) SS_E and $\hat{\beta}_1$.

Proof. Deferred to Section 13.6. ■

To this theorem, in conjunction with Theorem 4, there is the following corollary.

Corollary. *Under model (18):*

(i) *The MLE $\hat{\sigma}^2$ of σ^2 is a biased estimate of σ^2, but $\frac{n}{n-2}\hat{\sigma}^2 = \frac{SS_E}{n-2}$, call it S^2, is an unbiased estimate of σ^2.*

(ii)

$$\frac{\hat{\beta}_1 - \beta_1}{S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}}} \sim t_{n-2}, \quad (iii) \quad \frac{\hat{\beta}_2 - \beta_2}{S/\sqrt{SS_x}} \sim t_{n-2}, \tag{20}$$

where

$$S^2 = SS_E/(n-2). \tag{21}$$

Proof.

(i) It has been seen in Theorem 4(ii) that $\hat{\sigma}^2 = \frac{SS_E}{n} = \frac{n-2}{n} \times \frac{SS_E}{n-2}$. Since $\frac{SS_E}{\sigma^2} \sim \chi^2_{n-2}$, it follows that $E(\frac{SS_E}{\sigma^2}) = n-2$, or $E(\frac{SS_E}{n-2}) = \sigma^2$, so that $\frac{SS_E}{n-2}$ is an unbiased estimate of σ^2. Also, $E\hat{\sigma}^2 = \frac{n-2}{n}E(\frac{SS_E}{n-2}) = \frac{n-2}{n}\sigma^2$, so that $\hat{\sigma}^2$ is biased.

(ii) By Theorem 4(iii),

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}}} \sim N(0, 1),$$

and $\frac{SS_E}{\sigma^2} = \frac{(n-2)SS_E}{(n-2)\sigma^2} = \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$. Furthermore, $\frac{\hat{\beta}_1 - \beta_1}{\sigma(\hat{\beta}_1)}$ and $\frac{SS_E}{\sigma^2}$ are independent, since $\hat{\beta}_1$ and SS_E are so. It follows that:

$$\frac{(\hat{\beta}_1 - \beta_1)/\sigma(\hat{\beta}_1)}{\sqrt{\frac{SS_E}{\sigma^2}/(n-2)}} \sim t_{n-2}, \quad \text{or} \quad \frac{(\hat{\beta}_1 - \beta_1)/\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}}}{\sqrt{S^2/\sigma^2}} \sim t_{n-2},$$

or, finally,

$$\frac{\hat{\beta}_1 - \beta_1}{S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}}} \sim t_{n-2}.$$

(iii) Again by Theorem 4(iii),

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - \beta_2}{\sigma/\sqrt{SS_x}} \sim N(0, 1),$$

and $\frac{\hat{\beta}_2 - \beta_2}{\sigma(\hat{\beta}_2)}$ and $\frac{SS_E}{\sigma^2}$ are independent, since $\hat{\beta}_2$ and SS_E are so. Then:

$$\frac{(\hat{\beta}_2 - \beta_2)/\sigma(\hat{\beta}_2)}{\sqrt{\frac{SS_E}{\sigma^2}/(n-2)}} \sim t_{n-2}, \quad \text{or} \quad \frac{(\hat{\beta}_2 - \beta_2)/\frac{\sigma}{\sqrt{SS_x}}}{\sqrt{S^2/\sigma^2}} \sim t_{n-2},$$

or, finally, $\frac{\hat{\beta}_2 - \beta_2}{S/\sqrt{SS_x}} \sim t_{n-2}$. ■

EXERCISES

3.1 Verify the result $\frac{d^2}{dt^2}\log(y_1,\ldots,y_n;\hat\beta_1,\hat\beta_2,t)\big|_{t=SS_E/n} = -\frac{n^3}{2SS_E^2}$ as claimed in the proof of Theorem 4(ii), where $t=\sigma^2$.

3.2 Consider Table 13.1 and leave out the "outlier" pairs (3.63, 447), (3.44, 692), and (2.44, 336). Then recalculate all quantities below:

$$\sum_i x_i,\quad \sum_i y_i,\quad \sum_i x_i^2,\quad \sum_i x_i y_i,\quad \sum_i y_i^2,\quad SS_x,\quad SS_y,\quad SS_{xy}.$$

3.3 Use the calculations in Exercise 3.2 to compute the estimates $\hat\beta_1,\hat\beta_2$ and the fitted regression line.

3.4 Refer to Exercise 3.2, and compute the variances $Var(\hat\beta_1)$, $Var(\hat\beta_2)$, and the MLE of σ^2.

3.5 By Theorem 5, the r.v. SS_E/σ^2 is distributed as χ^2_{n-2}. Therefore, in the usual manner,

$$\left[\frac{SS_E}{\chi^2_{n-2\,;\frac{\alpha}{2}}},\quad \frac{SS_E}{\chi^2_{n-2\,;1-\frac{\alpha}{2}}}\right]$$

is a confidence interval for σ^2 with confidence coefficient $1-\alpha$, where SS_E is given in (16) and (9). That is, $SS_E = SS_y - \frac{SS_{xy}^2}{SS_x}$, where SS_x, SS_y, and SS_{xy} are given in (9).

- **(i)** Refer to Example 1 (see also Example 3), and construct a 95% confidence interval for σ^2.
- **(ii)** Refer to Example 4 and do the same as in part (i).
- **(iii)** Refer to Example 5 and do the same as in part (i).
- **(iv)** Refer to Exercise 3.2 and do the same as in part (i).

3.6 Consider the linear regression model given in relation (18), and let x_0 be an unknown point at which observations $Y_{0i}, i=1,\ldots,m$ are taken. It is assumed that the Y_{0i}'s and the Y_j's are independent, and set $Y_0 = \frac{1}{m}\sum_{i=1}^m Y_{0i}$. Set $\boldsymbol{y}=(y_1,\ldots,y_n)$, $\boldsymbol{y}_0=(y_{01},\ldots,y_{0m})$ for the observed values of the Y_j's and Y_{0i}'s, respectively, and form their joint log-likelihood function:

$$\begin{aligned}\Lambda &= \Lambda(\beta_1,\beta_2,\sigma^2,x_0) = \log L(\beta_1,\beta_2,\sigma^2,x_0\,|\boldsymbol{y},\boldsymbol{y}_0)\\ &= -\frac{m+n}{2}\log(2\pi) - \frac{m+n}{2}\log\sigma^2\\ &\quad -\frac{1}{2\sigma^2}\left[\sum_{j=1}^n (y_j-\beta_1-\beta_2x_j)^2 + \sum_{i=1}^m (y_{0i}-\beta_1-\beta_2x_0)^2\right].\end{aligned}$$

- **(i)** Show that the log-likelihood equations $\frac{\partial\Lambda}{\partial\beta_1}=0$, $\frac{\partial\Lambda}{\partial\beta_2}=0$, and $\frac{\partial\Lambda}{\partial x_0}=0$ produce the equations:

$$(m+n)\beta_1 + (mx_0+n\bar x)\beta_2 = my_0 + n\bar y \tag{a}$$

$$(mx_0 + n\bar{x})\beta_1 + \left(mx_0^2 + \sum_j x_j^2\right)\beta_2 = mx_0 y_0 + \sum_j x_j y_j \tag{b}$$

$$\beta_1 + x_0\beta_2 = y_0. \tag{c}$$

(ii) In (c), solve for β_1, $\beta_1 = y_0 - x_0\beta_2$, replace it in (a) and (b), and solve for β_2 to obtain, by assuming here and in the sequel that all divisions and cancellations are legitimate,

$$\beta_2 = \frac{\bar{y} - y_0}{\bar{x} - x_0}, \quad \beta_2 = \frac{\sum_j x_j y_j - n\bar{x}y_0}{\sum_j x_j^2 - n x_0 \bar{x}}. \tag{d}$$

(iii) Equate the β_2's in (ii), and solve for x_0 to obtain:

$$\left.\begin{aligned} x_0 &= \left[y_0 \sum_j (x_j - \bar{x})^2 + \bar{x}\sum_j x_j y_j - \bar{y}\sum_j x_j^2\right] \\ &\qquad \Big/ \left(\sum_j x_j y_j - n\bar{x}\bar{y}\right) \\ &= \left[n y_0 \sum_j (x_j - \bar{x})^2 + n\bar{x}\textstyle\sum_j x_j y_j - n\bar{y}\displaystyle\sum_j x_j^2\right] \\ &\qquad \Big/ \left[n\sum_j x_j y_j - \left(\sum_j x_j\right)\left(\sum_j y_j\right)\right]. \end{aligned}\right\} \tag{e}$$

(iv) Replace x_0 in the first expression for β_2 in (d) in order to get, after some simplifications:

$$\beta_2 = \frac{n\sum_j x_j y_j - \left(\sum_j x_j\right)\left(\sum_j y_j\right)}{n\sum_j x_j^2 - \left(\sum_j x_j\right)^2}, \tag{f}$$

and observe that this expression is the MLE (LSE) of β_2, calculated on the basis of y_j and $x_j, j = 1, \ldots, n$ only (see relation (5) and Theorem 4(i)).

(v) Replace x_0 and β_2 in the expression $\beta_1 = y_0 - x_0\beta_2$ in order to arrive at the expression:

$$\beta_1 = \frac{\left(\sum_j x_j^2\right)\left(\sum_j y_j\right) - \left(\sum_j x_j\right)\left(\sum_j x_j y_j\right)}{n\sum_j x_j^2 - \left(\sum_j x_j\right)^2}, \tag{g}$$

after some calculations, and observe that this is the MLE (LSE) of β_1, calculated on the basis of y_j and $x_j, j = 1, \ldots, n$ only (see relation (4) and Theorem 4(i)).

(vi) It follows that the MLE's of β_1, β_2, and x_0, to be denoted by $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{x}_0$, respectively, are given by the expressions:

$$\hat{\beta}_1 = \frac{\left(\sum_j x_j^2\right)\left(\sum_j y_j\right) - \left(\sum_j x_j\right)\left(\sum_j x_j y_j\right)}{n\sum_j x_j^2 - \left(\sum_j x_j\right)^2},$$

$$\hat{\beta}_2 = \frac{n\sum_j x_j y_j - \left(\sum_j x_j\right)\left(\sum_j y_j\right)}{n\sum_j x_j^2 - \left(\sum_j x_j\right)^2},$$

and

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_1}{\hat{\beta}_2}.$$

(vii) Differentiate the log-likelihood function with respect to σ^2, equate the derivative to zero, and replace β_1, β_2, and x_0 by their MLE's in order to obtain the MLE $\hat{\sigma}^2$ of σ^2, which is given by the expression:

$$\hat{\sigma}^2 = \frac{1}{m+n}(SS_E + SS_{0E}),$$

where

$$SS_E = \sum_{j=1}^{n}(y_j - \hat{y}_j)^2 = \sum_{j=1}^{n}(y_j - \hat{\beta}_1 - \hat{\beta}_2 x_j)^2,$$

and

$$SS_{0E} = \sum_{i=1}^{m}(y_{0i} - \hat{\beta}_1 - \hat{\beta}_2\hat{x}_0)^2 = \sum_{i=1}^{m}(y_{0i} - y_0)^2.$$

Also, by means of (14) and (16),

$$SS_E = SS_y - \frac{SS_{xy}^2}{SS_x}, \quad \text{where } SS_x = \sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2,$$

$$SS_y = \sum_{j=1}^{n} y_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} y_j\right)^2, \quad SS_{xy} = \sum_{j=1}^{n} x_j y_j - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} y_j\right),$$

and

$$SS_{0E} = \sum_{i=1}^{m} y_{0i}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} y_{0i}\right)^2.$$

(viii) Observe that, by Theorem 5(i), $\frac{SS_E}{\sigma^2} \sim \chi^2_{n-2}$, whereas $\frac{SS_{0E}}{\sigma^2} \sim \chi^2_{m-1}$. Then, by independence of the Y_j's and the Y_{0i}'s, it follows that $\frac{1}{\sigma^2}(SS_E + SS_{0E}) \sim \chi^2_{m+n-3}$.

(ix) Observe that, by Theorem 8(i), $\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0 \sim N(\beta_1 + \beta_2 x_0, \sigma^2(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{SSx}))$, whereas $Y_0 \sim N(\beta_1 + \beta_2 x_0, \frac{\sigma^2}{m})$, and $\hat{y}_0$ and Y_0 are independent, so that the r.v. $V = Y_0 - \hat{y}_0 \sim N(0, \sigma_V^2)$, where $\sigma_V^2 = \sigma^2(\frac{1}{m} + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{SSx})$, and $\frac{V}{\sigma_V} \sim N(0,1)$.

(x) Observe that, by Theorem 5, the r.v.'s V/σ_V and $(SS_E + SS_{0E})/\sigma^2$ are independent, so that

$$\frac{V/\sigma_V}{\frac{1}{\sigma}\sqrt{\frac{SS_E+SS_{0E}}{m+n-3}}} = \frac{V/\sigma\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{SSx}}}{\frac{1}{\sigma}\sqrt{\frac{SS_E+SS_{0E}}{m+n-3}}}$$

$$= \frac{\sqrt{m+n-3}\,V}{\sqrt{\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{SSx}\right](SS_E+SS_{0E})}} \sim t_{m+n-3}.$$

13.4 CONFIDENCE INTERVALS AND HYPOTHESES TESTING PROBLEMS

The results obtained in the corollary to Theorem 5 allow the construction of confidence intervals for the parameters of the model, as well as testing hypotheses about them.

Theorem 6. *Under model (18),* $100(1-\alpha)\%$ *confidence intervals for* β_1 *and* β_2 *are given, respectively, by:*

$$\left[\hat{\beta}_1 - t_{n-2\,;\alpha/2}S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SS_x}},\quad \hat{\beta}_1 + t_{n-2\,;\alpha/2}S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SS_x}}\right], \tag{22}$$

and

$$\left[\hat{\beta}_2 - t_{n-2\,;\alpha/2}\frac{S}{\sqrt{SS_x}},\quad \hat{\beta}_2 + t_{n-2\,;\alpha/2}\frac{S}{\sqrt{SS_x}}\right], \tag{23}$$

where $S = \sqrt{SS_E/(n-2)}$, *and* SS_E, SS_x, *and* $\hat{\beta}_1, \hat{\beta}_2$ *are given by (16), (9), and (10).*

Proof. The confidence intervals in (22) and (23) follow immediately from results (ii) and (iii), respectively, in the corollary to Theorem 5, and the familiar procedure of constructing confidence intervals. ■

Remark 2. A confidence interval can also be constructed for σ^2 on the basis of the statistic SS_E and the fact that SS_E/σ^2 is distributed as χ^2_{n-2}.

Procedures for testing some hypotheses are summarized below in the form of a theorem. The tests proposed here have an obvious intuitive interpretation. However, their justification rests on that they are likelihood ratio tests. For the case of simple hypotheses, this fact can be established directly. For composite hypotheses, it follows as a special case of more general results of testing hypotheses regarding the entire mean $\eta = \beta_1 + \beta_2 x$. See, e.g., Chapter 16 and, in particular, Examples 2 and 3 in the book *A Course in Mathematical Statistics*, 2nd edition (1997), Academic Press, by G. G. Roussas.

Theorem 7. *Under model (18):*

(i) *For testing the hypothesis H_0: $\beta_1 = \beta_{10}$ against the alternative H_A: $\beta_1 \neq \beta_{10}$ at level of significance α, the null hypothesis H_0 is rejected whenever*

$$|t| > t_{n-2;\alpha/2}, \quad \text{where } t = (\hat{\beta}_1 - \beta_{10}) \Big/ S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}}. \tag{24}$$

(ii) *For testing the hypothesis H_0: $\beta_2 = \beta_{20}$ against the alternative H_A: $\beta_2 \neq \beta_{20}$ at level of significance α, the null hypothesis H_0 is rejected whenever*

$$|t| > t_{n-2;\alpha/2}, \quad \text{where } t = (\hat{\beta}_2 - \beta_{20}) \Big/ \frac{S}{\sqrt{SS_x}}. \tag{25}$$

When the alternative is of the form H_A: $\beta_2 > \beta_{20}$, the null hypothesis is rejected whenever $t > t_{n-2\,;\alpha}$, and it is rejected whenever $t < -t_{n-2\,;\alpha}$ if the alternative is of the form H_A: $\beta_2 < \beta_{20}$.

Remark 3.

(i) In the reference cited above, the test statistic used, actually, has the F-distribution under the null hypothesis. It should be recalled, however, that if t has the t-distribution with r d.f.; i.e., $t = Z/\sqrt{\chi_r^2/r}$, where χ_r^2 has the χ^2-distribution with r d.f. $Z \sim N(0,1)$ and Z and χ_r^2 are independent then, $t^2 = \frac{Z^2}{\chi_r^2/r}$ has the F-distribution with 1 and r d.f.

(ii) Hypotheses can also be tested about σ^2 on the basis of the fact that $\frac{SS_E}{\sigma^2} \sim \chi_{n-2}^2$.

Example 6. In reference to Example 1:

(i) Construct 95% confidence intervals for β_1 and β_2.

(ii) Test the hypothesis that the GMAT scores increase with increasing GPA scores.

Discussion. (i) The required confidence intervals are given by (22) and (23). In the discussion of Example 1, we have found that: $\bar{x} \simeq 2.963, SS_x \simeq 6.361, \hat{\beta}_1 \simeq 220.456$, and $\hat{\beta}_2 \simeq 91.397$. Also, in the discussion of Example 3, we saw that $SS_E \simeq 114{,}564.625$, so that, by (21), $S = (\frac{114{,}564.625}{32})^{1/2} \simeq 59.834$. Finally, $t_{32\,;0.025} = 2.0369$. Then

$$\hat{\beta}_1 - t_{n-2\,;\alpha/2} S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}} = 220.456 - 2.0369 \times 59.834 \times \sqrt{\frac{1}{34} + \frac{(2.963)^2}{6.361}}$$
$$\simeq 220.456 - 121.876 \times 1.187 \simeq 220.456 - 144.667 = 75.789,$$

and

$$\hat{\beta}_1 + t_{n-2\,;\alpha/2} S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}} \simeq 220.456 + 144.667 = 365.123.$$

So the observed conference interval is [75.789, 365.123].

Likewise, $t_{n-2\,;\alpha/2} \frac{S}{\sqrt{SS_x}} \simeq 1.6939 \times \frac{59.834}{\sqrt{6.361}} = 1.6939 \times 23.725 \simeq 40.188$, and therefore the observed confidence interval for β_2 is: [51.209, 131.585].

(ii) Here we are to test H_0: $\beta_2 = 0$ against the alternative H_A: $\beta_2 > 0$. Let us take $\alpha = 0.05$, so that $t_{32;0.05} = 1.6939$. The observed value of the test statistics is:

$$t = \frac{\hat{\beta}_2 - \beta_{20}}{S/\sqrt{SS_x}} \simeq \frac{91.397}{23.725} = 3.852,$$

and the null hypothesis is rejected; the GMAT scores increase along with increasing GPA scores.

Example 7. In reference to Example 4:

(i) Construct 95% confidence intervals for β_1 and β_2.
(ii) Test the hypothesis that crop yield increases with increasing compost fertilizer amounts.

Discussion. (i) In the discussion of Example 4, we have seen that: $n = 15, \bar{x} = 10.8, SS_x = 70.6, SS_y = 98.5, \hat{\beta}_1 \simeq 112.256$, and $\hat{\beta}_2 \simeq 0.967$. It follows that:

$$S = \left(\frac{98.5}{13}\right)^{1/2} \simeq 2.753 \quad \text{and} \quad S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}} = 2.753 \times \sqrt{\frac{1}{15} + \frac{(10.8)^2}{70.6}}$$
$$\simeq 2.753 \times 1.311 \simeq 3.609.$$

Since $t_{13\,;0.025} = 1.1604$, it follows that the required observed confidence interval for β_1 is: $[112.256 - 1.1604 \times 3.609, 112.256 + 1.1604 \times 3.609]$, or [108.068, 116.444].

Next, $\frac{S}{\sqrt{SS_x}} \simeq \frac{2.753}{\sqrt{70.6}} \simeq 0.328$, and $t_{13\,;0.025} \frac{S}{\sqrt{SS_x}} = 1.1604 \times 0.328 \simeq 0.381$, so that the required observed confidence interval for β_2 is: [0.586, 1.348].

(ii) The hypothesis to be tested is H_0: $\beta_2 = 0$ against the alternative H_A: $\beta_2 > 0$. Take $\alpha = 0.05$, so that $t_{13\,;0.05} = 1.7709$. The observed value of the test statistic is:

$$t = \frac{\hat{\beta}_2 - \beta_{20}}{S/\sqrt{SS_x}} \simeq \frac{0.967}{0.328} \simeq 2.948,$$

and therefore the null hypothesis is rejected. Consequently, crop yield increases with increasing amounts of compost fertilizer.

Example 8. In reference to Example 5:

(i) Construct 95% confidence intervals for β_1 and β_2.
(ii) Test the hypothesis that the duration of relief increases with higher dosages of the medication.

Discussion. (i) From the discussion of Example 5, we have: $n = 10, \bar{x} = 5.9, SS_x = 40.9, SS_E = 63.653, \hat{\beta}_1 = -1.072$, and $\hat{\beta}_2 \simeq 2.741$. Then $S = (\frac{63.653}{8})^{1/2} \simeq 2.821$. Also, $t_{8\,;0.025} = 3.3060$. Therefore:

$$t_{n-2\,;\alpha/2} S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}} = 3.306 \times 2.821 \times \sqrt{\frac{1}{10} + \frac{(5.9)^2}{40.9}} \simeq 9.326 \times 0.975$$
$$\simeq 9.093.$$

Hence the observed confidence interval for β_1 is: $[-1.072 - 9.093, -1.072 + 9.093]$, or $[-10.165, 8.021]$. Next,

$$t_{n-2\,;\alpha/2} \frac{S}{\sqrt{SS_x}} = 3.306 \times \frac{2.821}{\sqrt{40.9}} \simeq 3.306 \times 0.441 \simeq 1.458,$$

and therefore the observed confidence interval for β_2 is: $[2.741 - 1.458, 2.741 + 1.458]$, or $[1.283, 4.199]$.

(ii) The hypothesis to be tested is H_0: $\beta_2 = 0$ against H_A: $\beta_2 > 0$, and let us take $\alpha = 0.05$, so that $t_{8\,;0.05} = 1.8595$. The observed value of the test statistic is:

$$t = \frac{\hat{\beta}_2 - \beta_{20}}{S/\sqrt{SS_x}} \simeq \frac{2.741}{0.441} \simeq 6.215,$$

and the null hypothesis is rejected. Thus, increased dosages of medication provide longer duration of relief.

EXERCISES

4.1 Refer to Exercises 3.2 and 3.4, and compute 95% confidence intervals for β_1 and β_2.

4.2 Refer to Exercises 3.3 and 4.1, and test the hypotheses H_0: $\beta_1 = 300$ against H_A: $\beta_1 \neq 300$, and H_0: $\beta_2 = 60$ against H_A: $\beta_2 \neq 60$, each at level of significance $\alpha = 0.05$.

4.3 Refer to Example 5 and:

(i) Derive 95% confidence intervals for β_1 and β_2.

(ii) Test the hypothesis H_0: $\beta_1 = -1$ against the alternative H_A: $\beta_1 \neq -1$ at level of significance $\alpha = 0.05$.

(iii) Do the same for the hypothesis H_0: $\beta_2 = 3$ against the alternative H_A:$\beta_2 \neq 3$ at the same level $\alpha = 0.05$.

4.4 Suppose the observations $Y_1, \ldots, Y_n$ are of the following structure: $Y_i = \beta + \gamma(x_i - \bar{x}) + e_i$, where β and γ are parameters and the e_i's are independent r.v.'s with mean 0 and unknown variance σ^2.

(i) Set $t_i = x_i - \bar{x}, i = 1, \ldots, n$, and observe that the model $Y_i = \beta + \gamma t_i + e_i$ is of the standard form (1) with $\beta_1 = \beta, \beta_2 = \gamma$, and the additional property that $\sum_{i=1}^n t_i = 0$, or $\bar{t} = 0$.

(ii) Use expressions (5) and (8) to conclude that the LSE's of β and γ are given by:

$$\hat{\beta} = \bar{Y}, \quad \hat{\gamma} = \frac{\sum_{i=1}^{n} t_i Y_i}{\sum_{i=1}^{n} t_i^2}.$$

(iii) Employ Theorem 4 in order to conclude that:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{n}\right) \quad \text{and} \quad \hat{\gamma} \sim N\left(\gamma, \frac{\sigma^2}{SS_t}\right),$$

where (by (9)) $SS_t = \sum_{i=1}^{n} t_i^2$.

(iv) Determine the form of the confidence intervals for β and γ from relations (22) and (23).

(v) Determine the expression of the test statistics by means of relations (24) and (25).

(vi) How do the confidence intervals in relation (29) and in Theorem 9 (iii) become here?

4.5 Consider the linear regression models: $Y_i = \beta_1 + \beta_2 x_i + e_i, i = 1, \ldots, m$ and $Y_j^* = \beta_1^* + \beta_2^* x_j^* + e_j^*, j = 1, \ldots, n$, where the random errors $e_1, \ldots, e_m$ and $e_1^*, \ldots, e_n^*$ are i.i.d. r.v.'s distributed as $N(0, \sigma^2)$.

(i) The independence of $e_1, \ldots, e_m$ and $e_1^*, \ldots, e_n^*$ implies independence of $Y_1, \ldots, Y_m$ and $Y_1^*, \ldots, Y_n^*$. Then write down, the joint likelihood of the Y_i's and the Y_j^*'s and observe that the MLE's of $\beta_1, \beta_2, \beta_1^*, \beta_2^*$, and σ^2, in obvious notation, are given by:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x}, \quad \hat{\beta}_2 = \frac{m \sum_{i=1}^{m} x_i Y_i - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} Y_i\right)}{m \sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2},$$

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{x}^*, \quad \hat{\beta}_2^* = \frac{n \sum_{j=1}^{n} x_j^* Y_j^* - \left(\sum_{j=1}^{n} x_j^*\right)\left(\sum_{j=1}^{n} Y_j^*\right)}{n \sum_{j=1}^{n} x_j^{2*} - \left(\sum_{j=1}^{n} x_j\right)^2},$$

$$\hat{\sigma}^2 = (SS_E + SS_E^*)/(m+n),$$

where

$$\begin{aligned}
SS_E &= \sum_{i=1}^{m} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 = SS_y - \frac{SS_{xy}^2}{SS_x}, \\
SS_x &= \sum_{i=1}^{m} x_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_i\right)^2, \\
SS_y &= \sum_{i=1}^{m} Y_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} Y_i\right)^2, \\
SS_{xy} &= \sum_{i=1}^{m} x_i Y_i - \frac{1}{m}\left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} Y_i\right),
\end{aligned}$$

and

$$
\begin{aligned}
SS_E^* &= \sum_{j=1}^{n}(Y_j^* - \hat{\beta}_1^* - \hat{\beta}_2^* x_j^*)^2 = SS_y^* - \frac{SS_{xy}^{*2}}{SS_x^*},\\
SS_x^* &= \sum_{j=1}^{n} x_j^{2*} - \frac{1}{n}\left(\sum_{j=1}^{n} x_j^*\right)^2,\\
SS_y^* &= \sum_{j=1}^{n} Y_j^{2*} - \frac{1}{n}\left(\sum_{j=1}^{n} Y_j^*\right)^2,\\
SS_{xy}^* &= \sum_{j=1}^{n} x_j^* Y_j^* - \frac{1}{n}\left(\sum_{j=1}^{n} x_j^*\right)\left(\sum_{j=1}^{n} Y_j^*\right).
\end{aligned}
$$

(ii) In accordance with Theorem 4, observe that

$$
\begin{aligned}
&\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2\left(\frac{1}{m} + \frac{\bar{x}^2}{SS_x}\right)\right), \quad \hat{\beta}_2 \sim N\left(\beta_2, \frac{\sigma^2}{SS_x}\right),\\
&\hat{\beta}_1^* \sim N\left(\beta_1^*, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^{*2}}{SS_x^*}\right)\right), \quad \hat{\beta}_2^* \sim N\left(\beta_2^*, \frac{\sigma^2}{SS_x^*}\right),
\end{aligned}
$$

and

$$
\frac{SS_E + SS_E^*}{\sigma^2} \sim \chi^2_{m+n-4}.
$$

(iii) From part (ii) and Theorem 5, conclude that

$$
\frac{\sqrt{m+n-4}\left[(\hat{\beta}_1 - \hat{\beta}_1^*) - (\beta_1 - \beta_1^*)\right]}{\sqrt{\left(SS_E + SS_E^*\right)\left(\frac{1}{m} + \frac{1}{n} + \frac{\bar{x}^2}{SS_x} + \frac{\bar{x}^{*2}}{SS_x^*}\right)}} \sim t_{m+n-4},
$$

and

$$
\frac{\sqrt{m+n-4}\left[(\hat{\beta}_2 - \hat{\beta}_2^*) - (\beta_2 - \beta_2^*)\right]}{\sqrt{\left(SS_E + SS_E^*\right)\left(\frac{1}{SS_x} + \frac{1}{SS_x^*}\right)}} \sim t_{m+n-4}.
$$

(iv) From part (iii), observe that the two regression lines can be compared through the test of the hypotheses H_0: $\beta_1 = \beta_1^*$ against the alternative H_A: $\beta_1 \neq \beta_1^*$, and H_0': $\beta_2 = \beta_2^*$ against the alternative H_A': $\beta_2 \neq \beta_2^*$ by using the respective test statistics:

$$
\begin{aligned}
t &= \frac{\sqrt{m+n-4}\,(\hat{\beta}_1 - \hat{\beta}_1^*)}{\sqrt{(SS_E + SS_E^*)\left(\frac{1}{m} + \frac{1}{n} + \frac{\bar{x}^2}{SS_x} + \frac{\bar{x}^{*2}}{SS_x^*}\right)}},\\
t' &= \frac{\sqrt{m+n-4}\,(\hat{\beta}_2 - \hat{\beta}_2^*)}{\sqrt{(SS_E + SS_E^*)\left(\frac{1}{SS_x} + \frac{1}{SS_x^*}\right)}}.
\end{aligned}
$$

At level of significance α, the hypothesis H_0 is rejected when $|t| > t_{m+n-4;\frac{\alpha}{2}}$, and the hypothesis H_0' is rejected when $|t'| > t_{m+n-4\,;\frac{\alpha}{2}}$.

(v) Again from part (iii), observe that 95% confidence intervals for $\beta_1 - \beta_1^*$ and $\beta_2 - \beta_2^*$ are given by:

$$(\hat{\beta}_1 - \hat{\beta}_1^*) \pm t_{m+n-4;\frac{\alpha}{2}} \sqrt{\frac{SS_E + SS_E^*}{m+n-4}\left(\frac{1}{m} + \frac{1}{n} + \frac{\bar{x}^2}{SS_x} + \frac{\bar{x}^{*2}}{SS_x^*}\right)},$$

and

$$(\hat{\beta}_2 - \hat{\beta}_2^*) \pm t_{m+n-4;\frac{\alpha}{2}} \sqrt{\frac{SS_E + SS_E^*}{m+n-4}\left(\frac{1}{SS_x} + \frac{1}{SS_x^*}\right)},$$

respectively.

(vi) Finally, from part (ii) conclude that a 95% confidence interval for σ^2 is given by:

$$\left[\frac{SS_E + SS_E^*}{\chi^2_{m+n-4;\frac{\alpha}{2}}}, \quad \frac{SS_E + SS_E^*}{\chi^2_{m+n-4;1-\frac{\alpha}{2}}}\right].$$

13.5 SOME PREDICTION PROBLEMS

According to model (18), the expectation of the observation Y_i at x_i is $EY_i = \beta_1 + \beta_2 x_i$. Now, suppose x_0 is a point distinct from all x_i's, but lying in the range that the x_i's span, and we wish to predict the expected value of the observation Y_0 at x_0; i.e., $EY_0 = \beta_1 + \beta_2 x_0$. An obvious predictor for EY_0 is the statistic $\hat{y}_0$ given by the expression below and modified as indicated:

$$\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0 = (\bar{Y} - \hat{\beta}_2 \bar{x}) + \hat{\beta}_2 x_0 = \bar{Y} + (x_0 - \bar{x})\hat{\beta}_2. \tag{26}$$

The result below gives the distribution of $\hat{y}_0$, which also provides for the construction of a confidence interval for $\beta_1 + \beta_2 x_0$.

Theorem 8. *Under model (18) and with $\hat{y}_0$ given by (26), we have:*

(i)

$$\frac{\hat{y}_0 - (\beta_1 + \beta_2 x_0)}{\sigma\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_x}}} \sim N(0, 1). \tag{27}$$

(ii)

$$\frac{\hat{y}_0 - (\beta_1 + \beta_2 x_0)}{S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_x}}} \sim t_{n-2}. \tag{28}$$

(iii) *A $100(1-\alpha)\%$ confidence interval for $\beta_1 + \beta_2 x_0$ is given by:*

$$\left[\hat{y}_0 - t_{n-2\,;\alpha/2} S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_x}}, \quad \hat{y}_0 + t_{n-2\,;\alpha/2} S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_x}}\right]. \tag{29}$$

It is recalled that $S = \sqrt{SS_E/(n-2)}, SS_E = SS_y - \frac{SS_{xy}^2}{SS_x}$, *and* SS_y *and* SS_x *are given in (9).*

Proof. (i) The assumption that $Y_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2), i = 1, \ldots, n$ independent implies that $\sum_i Y_i \sim N(n\beta_1 + \beta_2 \sum_i x_i, n\sigma^2)$ and hence

$$\bar{Y} \sim N(\beta_1 + \beta_2 \bar{x}, \sigma^2/n). \tag{30}$$

By Theorem 4(iii), $\hat{\beta}_2 \sim N(\beta_2, \sigma^2/SS_x)$, so that

$$(x_0 - \bar{x})\hat{\beta}_2 \sim N\left((x_0 - \bar{x})\beta_2, \quad \frac{\sigma^2(x_0 - \bar{x})^2}{SS_x}\right). \tag{31}$$

Furthermore, by Theorem 5(ii)(b), $\bar{Y}$ and $\hat{\beta}_2$ are independent. Then, relations (26), (30), and (31) yield:

$$\hat{y}_0 = \bar{Y} + (x_0 - \bar{x})\hat{\beta}_2 \sim N\left(\beta_1 + \beta_2 x_0, \quad \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_x}\right)\right), \tag{32}$$

and therefore (27) follows by standardization.

(ii) By Theorem 5(ii)(c), SS_E is independent of $\bar{Y}$ and $\hat{\beta}_2$ and hence independent of $\hat{y}_0$ because of (26). Furthermore, by Theorem 5(i),

$$\frac{SS_E}{\sigma^2} = \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}. \tag{33}$$

Therefore,

$$\frac{[\hat{y}_0 - (\beta_1 + \beta_2 x_0)]/\sigma\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}}}{\sqrt{\frac{(n-2)S^2}{\sigma^2}/(n-2)}} = \frac{\hat{y}_0 - (\beta_1 + \beta_2 x_0)}{S\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}}} \sim t_{n-2},$$

which is relation (28).

(iii) This part follows immediately from part (ii) and the standard procedure of setting up confidence intervals. ■

Finally, we would like to consider the problem of predicting a *single response* at a given point x_0 rather than its *expected value*. Call Y_0 the response corresponding to x_0 and, reasonably enough, assume that Y_0 is independent of the Y_i's. The predictor for Y_0 is $\hat{y}_0$, the same as the one given in (26). The objective here is to construct a prediction interval for Y_0. This is done indirectly in the following result.

Theorem 9. *Under model (18), let* Y_0 *be the (unobserved) observation at* x_0, *and assume that* Y_0 *is independent of the* Y_i*'s. Predict* Y_0 *by* $\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0$. *Then:*

(i)

$$\frac{\hat{y}_0 - Y_0}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}}} \sim N(0, 1). \tag{34}$$

(ii)

$$\frac{\hat{y}_0 - Y_0}{S\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}}} \sim t_{n-2}. \tag{35}$$

(iii) *A $100(1-\alpha)\%$ prediction interval for Y_0 is given by:*

$$\left[\hat{y}_0 - t_{n-2\,;\alpha/2} S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{SS_x}},\ \hat{y}_0 + t_{n-2\,;\alpha/2} S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{SS_x}}\right],$$

where S and SS_x are as in Theorem 8(ii), (iii).

Proof. First, we compute the expectation and the variance of $\hat{y}_0 - Y_0$. To this end, we have: $Y_0 = \beta_1 + \beta_2 x_0 + e_0$, predicted by $\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0$. Then $EY_0 = \beta_1 + \beta_2 x_0$ and $E\hat{y}_0 = \beta_1 + \beta_2 x_0$, so that $E(\hat{y}_0 - Y_0) = 0$. In deriving the distribution of $\hat{y}_0 - Y_0$, we need its variance. By (26), we have:

$$Var(\hat{y}_0 - Y_0) = Var(\bar{Y} + (x_0-\bar{x})\hat{\beta}_2 - Y_0) = Var(\bar{Y}) + (x_0-\bar{x})^2 Var(\hat{\beta}_2) + Var(Y_0)$$

(since all three r.v.'s, $\bar{Y}$, $\hat{\beta}_2$, and Y_0, are independent)

$$= \frac{\sigma^2}{n} + (x_0-\bar{x})^2 \times \frac{\sigma^2}{SS_x} + \sigma^2 \quad \text{(by Theorem 2)}$$

$$= \sigma^2\left\{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}\right\}; \quad \text{i.e.,}$$

$$E(\hat{y}_0 - Y_0) = 0 \quad \text{and} \quad Var(\hat{y}_0 - Y_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}\right].$$

We now proceed with parts (i) and (ii).

(i) Since $\hat{y}_0$ and Y_0 are independent and $Y_0 \sim N(\beta_1 + \beta_2 x_0, \sigma^2)$, then these facts along with (32) yield:

$$\hat{y}_0 - Y_0 \sim N\left(0,\quad \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{SS_x}\right]\right).$$

Relation (34) follows by standardizing $\hat{y}_0 - Y_0$.

(ii) It has been argued in the proof of Theorem 8(ii) that S and $\hat{y}_0$ are independent. It follows that S and $\hat{y}_0 - Y_0$ are also independent. Then, dividing the expression on the left-hand side in (34) by $\sqrt{\frac{(n-2)S^2}{\sigma^2}/(n-2)} = \frac{S}{\sigma}$ in (33), we obtain the result in (35), after some simplifications.

(iii) This part follows from part (ii) through the usual procedure of setting up confidence intervals. ■

EXERCISES

5.1 Refer to Exercises 3.1, 3.3, 4.1, and:

(i) Predict EY_0 at $x_0 = 3.25$, and construct a 95% confidence interval of EY_0.

(ii) Predict the response Y_0 at $x_0 = 3.25$, and construct a 95% prediction interval for Y_0.

5.2 In reference to Example 22 in Chapter 1 (see also scatter diagram in Figure 13.1 and Examples 1, 2, 3, and 6 here), do the following:
- **(i)** Predict the EY_0, where Y_0 is the response at $x_0 = 3.25$.
- **(ii)** Construct a 95% confidence interval for $EY_0 = \beta_1 + \beta_2 x_0 = \beta_1 + 3.25\beta_2$.
- **(iii)** Predict the response Y_0 at $x_0 = 2.5$.
- **(iv)** Construct a 90% prediction interval for Y_0.

5.3 In reference to Example 23 in Chapter 1 (see also Examples 4 and 7 here), do the following:
- **(i)** Predict the EY_0, where Y_0 is the response at $x_0 = 12$.
- **(ii)** Construct a 95% confidence interval for $EY_0 = \beta_1 + \beta_2 x_0 = \beta_1 + 12\beta_2$.
- **(iii)** Predict the response Y_0 at $x_0 = 12$.
- **(iv)** Construct a 95% prediction interval for Y_0.

5.4 Refer to Example 5 and:
- **(i)** Predict the EY_0 at $x_0 = 6$.
- **(ii)** Construct a 95% confidence interval for $EY_0 = \beta_1 + 6\beta_2$.
- **(iii)** Predict the response Y_0 at $x_0 = 6$.
- **(iv)** Construct a 95% prediction interval for Y_0.

5.5 Suppose that the data given in the table below follow model (18).

x	5	10	15	20	25	30
y	0.10	0.21	0.30	0.35	0.44	0.62

- **(i)** Determine the MLE's (LSE's) of β_1, β_2, and σ^2.
- **(ii)** Construct 95% confidence intervals for β_1, β_2, and σ^2.
- **(iii)** At $x_0 = 17$, predict both EY_0 and Y_0 (the respective observation at x_0), and construct a 95% confidence interval and prediction interval, respectively, for them.

Hint. For a confidence interval for σ^2, see Exercise 3.5.

5.6 The following table gives the reciprocal temperatures x and the corresponding observed solubilities of a certain chemical substance, and assume that they follow model (18).

x	3.80	3.72	3.67	3.60	3.54
	1.27	1.20	1.10	0.82	0.65
y	1.32	1.26	1.07	0.84	0.57
	1.50			0.80	0.62

- **(i)** Determine the MLE's (LSE's) of β_1, β_2, and σ^2.
- **(ii)** Construct 95% confidence intervals for β_1, β_2, and σ^2.
- **(iii)** At $x_0 = 3.77$, predict both EY_0 and Y_0 (the respective observation at x_0), and construct a 95% confidence interval and prediction interval, respectively, for them.

Hint. Here $n = 13$ and $x_1 = x_2 = x_3,\ x_4 = x_5,\ x_6 = x_7, x_8 = x_9 = x_{10}$, and $x_{11} = x_{12} = x_{13}$.

13.6 PROOF OF THEOREM 5

This section is solely devoted to justifying Theorem 5. Its proof is presented in considerable detail, and it makes use of some linear algebra results. The sources of those results are cited.

Proof of Theorem 5. For later use, let us set

$$U_i = Y_i - \beta_1 - \beta_2 x_i, \quad \text{so that } \bar{U} = \bar{Y} - \beta_1 - \beta_2 \bar{x}, \tag{36}$$

and

$$U_i - \bar{U} = (Y_i - \bar{Y}) - \beta_2(x_i - \bar{x}) \quad \text{and} \quad Y_i - \bar{Y} = (U_i - \bar{U}) + \beta_2(x_i - \bar{x}). \tag{37}$$

Then, by (10),

$$\hat{\beta}_2 SS_x = SS_{xy} = \sum_i (x_i - \bar{x})(Y_i - \bar{Y}), \tag{38}$$

so, that

$$\begin{aligned}
(\hat{\beta}_2 - \beta_2)SS_x &= \sum_i (x_i - \bar{x})(Y_i - \bar{Y}) - \beta_2 SS_x \\
&= \sum_i (x_i - \bar{x})[(U_i - \bar{U}) + \beta_2(x_i - \bar{x})] - \beta_2 SS_x \\
&= \sum_i (x_i - \bar{x})(U_i - \bar{U}) + \beta_2 SS_x - \beta_2 SS_x \\
&= \sum_i (x_i - \bar{x})(U_i - \bar{U}).
\end{aligned} \tag{39}$$

Next,

$$\begin{aligned}
SS_E &= \sum_i (Y_i - \hat{y}_i)^2 = \sum_i (Y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \\
&= \sum_i (Y_i - \bar{Y} + \hat{\beta}_2 \bar{x} - \hat{\beta}_2 x_i)^2 \quad (\text{ by } (8)) \\
&= \sum_i [(Y_i - \bar{Y}) - \hat{\beta}_2(x_i - \bar{x})]^2 \\
&= \sum_i [(Y_i - \bar{Y}) - \hat{\beta}_2(x_i - \bar{x}) + \beta_2(x_i - \bar{x}) - \beta_2(x_i - \bar{x})]^2 \\
&= \sum_i \{[(Y_i - \bar{Y}) - \beta_2(x_i - \bar{x})] - (\hat{\beta}_2 - \beta_2)(x_i - \bar{x})\}^2 \\
&= \sum_i [(U_i - \bar{U}) - (\hat{\beta}_2 - \beta_2)(x_i - \bar{x})]^2 \quad (\text{ by } (37))
\end{aligned}$$

$$
\begin{aligned}
&= \sum_i (U_i - \bar{U})^2 + (\hat{\beta}_2 - \beta_2)^2 SS_x - 2(\hat{\beta}_2 - \beta_2) \sum_i (x_i - \bar{x})(U_i - \bar{U}) \\
&= \sum_i (U_i - \bar{U})^2 + (\hat{\beta}_2 - \beta_2)^2 SS_x - 2(\hat{\beta}_2 - \beta_2)^2 SS_x \quad \text{(by (39))} \\
&= \sum_i (U_i - \bar{U})^2 - (\hat{\beta}_2 - \beta_2)^2 SS_x \\
&= \sum_i U_i{}^2 - n\bar{U}^2 - (\hat{\beta}_2 - \beta_2)^2 SS_x; \text{ i.e.,}
\end{aligned}
$$

$$
SS_E = \sum_i U_i^2 - n\bar{U}^2 - (\hat{\beta}_2 - \beta_2)^2 SS_x. \tag{40}
$$

From (18) and (36), we have that the r.v.'s $U_1, \ldots, U_n$ are independent and distributed as $N(0, \sigma^2)$. Transform them into the r.v.'s $V_1, \ldots, V_n$ by means of an orthogonal transformation $\mathbf{C}$ as described below (see also Remark 4):

$$
\mathbf{C} = \begin{pmatrix}
\frac{x_1 - \bar{x}}{\sqrt{SS_x}} & \frac{x_2 - \bar{x}}{\sqrt{SS_x}} & \cdots & \frac{x_n - \bar{x}}{\sqrt{SS_x}} \\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\
\multicolumn{4}{c}{\text{(whatever, subject to the res-}} \\
\multicolumn{4}{c}{\text{triction that } \mathbf{C} \text{ is orthogonal)}}
\end{pmatrix}
$$

That is, with "′" standing for transpose, we have:

$$
(V_1, V_2, \ldots, V_n)' = \mathbf{C}(U_1, U_2, \ldots, U_n)'. \tag{41}
$$

Then, by Theorem 8 in Chapter 6, the r.v.'s $V_1, \ldots, V_n$ are independent and distributed as $N(0, \sigma^2)$, whereas by relation (21) in the same chapter

$$
\sum_i V_i^2 = \sum_i U_i^2. \tag{42}
$$

From (41),

$$
V_1 = \frac{1}{\sqrt{SS_x}} \sum_i (x_i - \bar{x}) U_i, \quad V_2 = \frac{1}{\sqrt{n}} \sum_i U_i = \sqrt{n} \times \frac{1}{n} \sum_i U_i = \sqrt{n}\bar{U}. \tag{43}
$$

But

$$
\sum_i (x_i - \bar{x}) U_i = \sum_i (x_i - \bar{x})(U_i - \bar{U}) = (\hat{\beta}_2 - \beta_2) SS_x, \quad \text{(by (39))},
$$

so that

$$
V_1 = (\hat{\beta}_2 - \beta_2)\sqrt{SS_x}, \quad V_1^2 = (\hat{\beta}_2 - \beta_2)^2 SS_x, \quad \text{and} \quad V_2^2 = n\bar{U}^2. \tag{44}
$$

Then, from relations (40), (42), and (44), it follows that

$$
SS_E = \sum_{i=1}^n V_i^2 - V_1^2 - V_2^2 = \sum_{i=3}^n V_i^2. \tag{45}
$$

We now proceed with the justifications of parts (i) and (ii) of the theorem.

(i) From (45), $\frac{SS_E}{\sigma^2} = \sum_{i=3}^{n} \left(\frac{V_i}{\sigma}\right)^2 \sim \chi^2_{n-2}$, since $\frac{V_i}{\sigma}, i = 1, \ldots, n$ are independent and distributed as $N(0, 1)$.

(ii) **(a)** From (44) and (45), $\hat{\beta}_2$ and SS_E are functions of nonoverlapping V_i's (of V_1 the former, and of $V_3, \ldots, V_n$ the latter). Thus, SS_E and $\hat{\beta}_2$ are independent.

(b) By (36) and (43), $\bar{Y} = \bar{U} + (\beta_1 + \beta_2\bar{x}) = \frac{V_2}{\sqrt{n}} + \beta_1 + \beta_2\bar{x}$, so that $\bar{Y}$ is a function of V_2 and recall that $\hat{\beta}_2$ is a function of V_1. Then the independence of $\bar{Y}$ and $\hat{\beta}_2$ follows.

(c) As was seen in (a) and (b), SS_E is a function of $V_3, \ldots, V_n$; $\bar{Y}$ is a function of V_2; and $\hat{\beta}_2$ is a function of V_1; i.e., they are functions of nonoverlapping V_i's, and therefore independent.

(d) By (8), $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{x}$ and the right-hand side is a function of V_1 and V_2 alone, by (44) and part (b). Since SS_E is a function of $V_3, \ldots, V_n$, by (45), the independence of SS_E and $\hat{\beta}_1$ follows. ■

Remark 4. There is always an orthogonal matrix **C** with the first two rows as given above. Clearly, the vectors $\boldsymbol{r}_1 = (x_1 - \bar{x}, \ldots, x_n - \bar{x})'$ and $\boldsymbol{r}_2 = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})'$ are linearly independent. Then supplement them with $n - 2$ vectors $\boldsymbol{r}_3, \ldots, \boldsymbol{r}_n$, so that the vectors $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ are linearly independent. Finally, use the Gram-Schmidt orthogonalization process (which leaves $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ intact) to arrive at an orthogonal matrix **C**. (See, e.g., Theorem 1.16 and the discussion following it, in pages 33–34, of the book *Linear Algebra for Undergraduates* (1957), John Wiley & Sons, by D. C. Murdoch.)

13.7 CONCLUDING REMARKS

In this chapter, we studied the simplest linear regression model, according to which the response Y at a point x is given by $Y = \beta_1 + \beta_2 x + e$. There are extensions of this model in different directions. First, the model may *not* be linear in the parameters involved; i.e., the expectation $\eta = EY$ is not linear. Here are some such examples.

$$\text{(i)}\eta = ae^{bx}; \quad \text{(ii)}\eta = ax^b; \quad \text{(iii) } \eta = \frac{1}{a + bx}; \quad \text{(iv) } \eta = a + b\sqrt{x}.$$

It happens that these particular nonlinear models can be reduced to linear ones by suitable transformations. Thus, in (i), taking the logarithms (always with base e), we have:

$$\log \eta = \log a + bx, \quad \text{or} \quad \eta' = \beta_1 + \beta_2 x',$$

where $\eta' = \log \eta$, $\beta_1 = \log a, \beta_2 = b$, and $x' = x$, and the new model is linear. Likewise, in (ii):

$$\log \eta = \log a + b \log x, \quad \text{or} \quad \eta' = \beta_1 + \beta_2 x',$$

where $\eta' = \log \eta, \beta_1 = \log a, \beta_2 = b$, and $x' = \log x$, and the transformed model is linear. In (iii), simply set $\eta' = \frac{1}{\eta}$ to get $\eta' = a + bx$, or $\eta' = \beta_1 + \beta_2 x'$, where $\beta_1 = a$, $\beta_2 = b$ and $x' = x$. Finally, in (iv), let $x' = \sqrt{x}$ in order to get the linear model $\eta' = \beta_1 + \beta_2 x'$, with $\eta' = \eta, \beta_1 = a$, and $\beta_2 = b$.

Another direction of a generalization is the consideration of the so-called *multiple regression* linear models. In such models, there is more than one input variable x and more than two parameters β_1 and β_2. This simply reflects the fact that the response is influenced by more than one factor each time. For example, the observation may be the systolic blood pressure of the individual in a certain group, and the influencing factors may be weight and age. The general form of a multiple regression linear model is as follows:

$$Y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + x_{pi}\beta_p + e_i, \quad i = 1, \ldots, n,$$

and the assumptions attached to it are similar to those used in model (18). The analysis of such a model can be done, in principle, along the same lines as those used in analyzing model (18). However, the analysis becomes unwieldy and one has to employ, most efficiently, linear algebra methodology. Such models are referred to as *general linear models* in the statistical literature, and they have proved very useful in a host of applications. The theoretical study of such models can be found, e.g., in Chapter 16 of the book *A Course in Mathematical Statistics*, 2nd edition (1997), Academic Press, by G. G. Roussas.