6 Simple Linear Regression

단순 선형 회귀.

6.1 THE MODEL

By (1.1), the *simple linear regression* model for n observations can be written as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{6.1}$$

The designation *simple* indicates that there is only one x to predict the response y, and *linear* means that the model (6.1) is linear in β_0 and β_1. [Actually, it is the assumption $E(y_i) = \beta_0 + \beta_1 x_i$ that is linear; see assumption 1 below.] For example, a model such as $y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$ is linear in β_0 and β_1, whereas the model $y_i = \beta_0 + e^{\beta_1 x_i} + \varepsilon_i$ is not linear.

선형 비선형

In this chapter, we assume that y_i and ε_i are random variables and that the values of x_i are known constants, which means that the same values of $x_1, x_2, \ldots, x_n$ would be used in repeated sampling. The case in which the x variables are random variables is treated in Chapter 10. x가 관찰된 값이 아닌 변수일 때를 다룰거임.

To complete the model in (6.1), we make the following additional assumptions:

가정.

1. $E(\varepsilon_i) = 0$ for all $i = 1, 2, \ldots, n$, or, equivalently, $E(y_i) = \beta_0 + \beta_1 x_i$.
2. $\text{var}(\varepsilon_i) = \sigma^2$ for all $i = 1, 2, \ldots, n$, or, equivalently, $\text{var}(y_i) = \sigma^2$. → 등분산성
3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$, or, equivalently, $\text{cov}(y_i, y_j) = 0$.

Assumption 1 states that the model (6.1) is correct, implying that y_i depends only on x_i and that all other variation in y_i is random. Assumption 2 asserts that the variance of ε or y does not depend on the values of x_i. (Assumption 2 is also known as the assumption of 등분산성 *homoscedasticity*, *homogeneous variance* or *constant variance*.) Under assumption 3, the ε variables (or the y variables) are uncorrelated with each other. In Section 6.3, we will add a normality assumption, and the y (or the ε) variables will thereby be independent as well as uncorrelated. Each assumption has been stated in terms of the ε's or the y's. For example, if $\text{var}(\varepsilon_i) = \sigma^2$, then $\text{var}(y_i) = E[y_i - E(y_i)]^2 = E(y_i - \beta_0 - \beta_1 x_i)^2 = E(\varepsilon_i^2) = \sigma^2$.

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

실제 데이터에서는 X.

Any of these assumptions may fail to hold with real data. A plot of the data will often reveal departures from assumptions 1 and 2 (and to a lesser extent assumption 3). Techniques for checking on the assumptions are discussed in Chapter 9.

위 가정들에 대한 확인 방법

6.2 ESTIMATION OF β_0, β_1, AND σ^2

Using a random sample of n observations $y_1, y_2, \ldots, y_n$ and the accompanying fixed values $x_1, x_2, \ldots, x_n$, we can estimate the parameters β_0, β_1, and σ^2. To obtain the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, we use the method of least squares, which does not require any distributional assumptions (for maximum likelihood estimators based on normality, see Section 7.6.2).

In the *least-squares* approach, we seek estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squares of the deviations $y_i - \hat{y}_i$ of the n observed y_i's from their predicted values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$:

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \tag{6.2}$$

이를 최소화하는 $\hat{\beta}_0, \hat{\beta}_1$

Note that the predicted value $\hat{y}_i$ estimates $E(y_i)$, not y_i; that is, $\hat{\beta}_0 + \hat{\beta}_1 x_i$ estimates $\beta_0 + \beta_1 x_i$, not $\beta_0 + \beta_1 x_i + \varepsilon_i$. A better notation would be $\widehat{E(y_i)}$, but $\hat{y}_i$ is commonly used.

To find the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ in (6.2), we differentiate with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and set the results equal to 0:

미분,

$$\frac{\partial \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \tag{6.3}$$

$$\frac{\partial \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0. \tag{6.4}$$

The solution to (6.3) and (6.4) is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{6.5}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{6.6}$$

To verify that $\hat{\beta}_0$ and $\hat{\beta}_1$ in (6.5) and (6.6) minimize $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ in (6.2), we can examine the second derivatives or simply observe that $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ has no maximum and therefore the first

2차 미분.

$\hat{\varepsilon}'\hat{\varepsilon}$이 최댓값을 가지지 않음을 보이면 됨.

↳ 최댓값이 없기에 1차 미분이 0이 되는 점이 최소값이 됨.

derivatives yield a minimum. For an algebraic proof that $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize (6.2), see (7.10) in Section 7.3.1.

Example 6.2. Students in a statistics class (taught by one of the authors) claimed that doing the homework had not helped prepare them for the midterm exam. The exam score y and homework score x (averaged up to the time of the midterm) for the 18 students in the class were as follows:

y	x	y	x	y	x
95	96	72	89	35	0
80	77	66	47	50	30
0	0	98	90	72	59
0	0	90	93	55	77
79	78	0	18	75	74
77	64	95	86	66	67

$$\hat{\beta}_1 = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$$

$$\Sigma(x_i y_i - \bar{x}y_i - x_i\bar{y} + \bar{x}\bar{y}) = \underline{\Sigma x_i y_i - n\bar{x}\bar{y}} - \cancel{n\bar{x}\bar{y}} + \cancel{n\bar{x}\bar{y}}$$

$$\Sigma(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \Sigma x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \Sigma x_i^2 - n\bar{x}^2$$

Using (6.5) and (6.6), we obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

$$= \frac{81{,}195 - 18(58.056)(61.389)}{80{,}199 - 18(58.056)^2} = .8726,$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 61.389 - .8726(58.056) = 10.73.$$

The prediction equation is thus given by

$$\hat{y} = 10.73 + .8726x.$$

This equation and the 18 points are plotted in Figure 6.1. It is readily apparent in the plot that the slope $\hat{\beta}_1$ is the rate of change of $\hat{y}$ as x varies and that the intercept $\hat{\beta}_0$ is the value of $\hat{y}$ at $x = 0$.

The apparent linear trend in Figure 6.1 does not establish cause and effect between homework and test results (for inferences that can be drawn, see Section 6.3). The assumption $\text{var}(\varepsilon_i) = \sigma^2$ (constant variance) for all $i = 1, 2, \ldots, 18$ appears to be reasonable. ■

Note that the three assumptions in Section 6.1 were not used in deriving the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ in (6.5) and (6.6). It is not necessary that $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be based on $E(y_i) = \beta_0 + \beta_1 x_i$; that is, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ can be fit to a set of data for which $E(y_i) \neq \beta_0 + \beta_1 x_i$. This is illustrated in Figure 6.2, where a straight line has been fitted to curved data.

최소제곱법이 반드시 선형 회귀 모델의 기댓값 $E(y_i) = \beta_0 + \beta_1 x_i$ 를 기반으로 할 필요는 X.

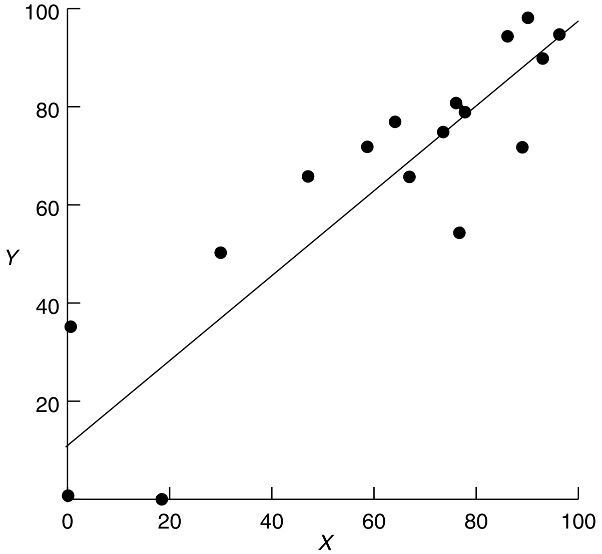


Figure 6.1 Regression line and data for homework and test scores.

세가지 가정이 성립하면, 최소제곱 법으로 구한 추정량 $\hat{\beta}_0, \hat{\beta}_1$은 ①편향되지 않으며 ② 선형 편향 없는 추정량중 최소 분산을 가짐!

However, if the three assumptions in Section 6.1 hold, then the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased and have minimum variance among all linear unbiased estimators (for the minimum variance property, see Theorem 7.3d in Section 7.3.2; note that $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear functions of $y_1, y_2, \ldots, y_n$). Using the three

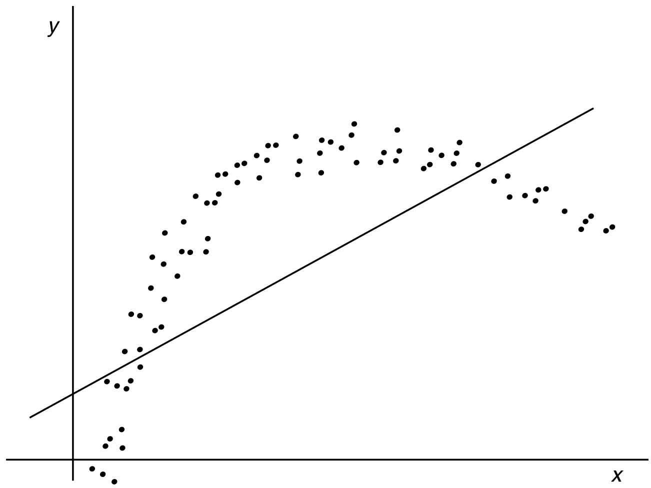


Figure 6.2 A straight line fitted to data with a curved trend.

assumptions, we obtain the following means and variances of $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$E(\hat{\beta}_1) = \beta_1 \tag{6.7}$$

$$E(\hat{\beta}_0) = \beta_0 \tag{6.8}$$

$$\downarrow \operatorname{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \uparrow \tag{6.9}$$

$\bar{x}=0$ 이면 $var(\hat{\beta}_0) \downarrow$

$$\operatorname{var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]. \tag{6.10}$$

Note that in discussing $E(\hat{\beta}_1)$ and $\operatorname{var}(\hat{\beta}_1)$, for example, we are considering random variation of $\hat{\beta}_1$ from sample to sample. It is assumed that the n values $x_1, x_2, \ldots, x_n$ would remain the same in future samples so that $\operatorname{var}(\hat{\beta}_1)$ and $\operatorname{var}(\hat{\beta}_0)$ are constant.

In (6.9), we see that $\operatorname{var}(\hat{\beta}_1)$ is minimized when $\sum_{i=1}^n (x_i - \bar{x})^2$ is maximized. If the x_i values have the range $a \le x_i \le b$, then $\sum_{i=1}^n (x_i - \bar{x})^2$ is maximized if half the x's are selected equal to a and half equal to b (assuming that n is even; see Problem 6.4). In (6.10), it is clear that $\operatorname{var}(\hat{\beta}_0)$ is minimized when $\bar{x} = 0$.

최소제곱법으로는 $var(y_i)=\sigma^2$는 못구해.

The method of least squares does not yield an estimator of $\operatorname{var}(y_i) = \sigma^2$; minimization of $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ yields only $\hat{\beta}_0$ and $\hat{\beta}_1$. To estimate σ^2, we use the definition in (3.6), $\sigma^2 = E[y_i - E(y_i)]^2$. By assumption 2 in Section 6.1, σ^2 is the same for each (등분산성.) y_i, $i = 1, 2, \ldots, n$. Using $\hat{y}_i$ as an estimator of $E(y_i)$, we estimate σ^2 by an average from the sample, that is

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} = \frac{\text{SSE}}{n-2}, \tag{6.11}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by (6.5) and (6.6) and $\text{SSE} = \sum_i (y_i - \hat{y}_i)^2$. The deviation 잔차. $\hat{\varepsilon}_i = y_i - \hat{y}_i$ is often called the *residual* of y_i, and SSE is called the *residual sum of squares* or *error sum of squares*. With $n-2$ in the denominator, s^2 is an unbiased estimator of σ^2:

↳ 최소제곱법 사용할 때, 두 개의 모수($\hat{\beta}_0, \hat{\beta}_1$)를 추정했으므로 자유도 2 차감
s^2가 σ^2의 불편향추정량이 되도록 하기 위함.

$$E(s^2) = \frac{E(\text{SSE})}{n-2} = \frac{(n-2)\sigma^2}{n-2} = \sigma^2. \tag{6.12}$$

Intuitively, we divide by $n - 2$ in (6.11) instead of $n - 1$ as in $s^2 = \sum_i (y_i - \bar{y})^2/(n-1)$ in (5.6), because $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ has two estimated parameters and should thereby be a better estimator of $E(y_i)$ than $\bar{y}$. Thus we

$E(y_i)$의 best 추정량이 되어야 함! $\bar{y}$보다!

즉, $\Sigma(y_i - \hat{y}_i)^2 < \Sigma(y_i - \bar{y})^2$.

expect SSE $= \sum_i (y_i - \hat{y}_i)^2$ to be less than $\sum_i (y_i - \bar{y})^2$. In fact, using (6.5) and (6.6), we can write the numerator of (6.11) in the form

$\hat\beta_0 + \hat\beta_1 x_i = \bar y - \hat\beta_1 \bar x + \hat\beta_1 x_i = \bar y + \hat\beta_1 (x_i - \bar x)$

SST SSR

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{6.13}$$

↳ $\Sigma\left((y_i - \bar y) - \hat\beta_1 (x_i - \bar x)\right)^2$ 에 $\hat\beta_1 = \frac{\Sigma(x_i - \bar x)(y_i - \bar y)}{\Sigma (x_i - \bar x)^2}$ 대입하고 정리하면

which shows that $\sum_i (y_i - \hat{y}_i)^2$ is indeed smaller than $\sum_i (y_i - \bar{y})^2$.

6.3 HYPOTHESIS TEST AND CONFIDENCE INTERVAL FOR β_1

Typically, hypotheses about β_1 are of more interest than hypotheses about β_0, since our first priority is to determine whether there is a linear relationship between y and x. (See Problem 6.9 for a test and confidence interval for β_0.) In this section, we consider the hypothesis H_0: $\beta_1 = 0$, which states that there is no linear relationship between y and x in the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. The hypothesis H_0:$\beta_1 = c$ (for $c \neq 0$) is of less interest.

In order to obtain a test for H_0: $\beta_1 = 0$, we assume that y_i is $N(\beta_0 + \beta_1 x_i, \sigma^2)$. Then $\hat{\beta}_1$ and s^2 have the following properties (these are special cases of results established in Theorem 7.6b in Section 7.6.3):

1. $\hat{\beta}_1$ is $N\left[\beta_1, \sigma^2 / \sum_i (x_i - \bar{x})^2\right]$. $\hat\beta_1 \sim N\left(\beta_1, \frac{\sigma^2}{\Sigma (x_i - \bar x)^2}\right)$
2. $(n-2)s^2/\sigma^2$ is $\chi^2(n-2)$. $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$
3. $\hat{\beta}_1$ and s^2 are independent.

From these three properties it follows by (5.~~29~~ 34) that

$$t = \frac{\hat{\beta}_1}{s / \sqrt{\sum_i (x_i - \bar{x})^2}} \tag{6.14}$$

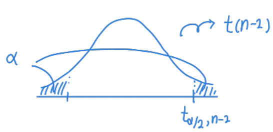
$\sim t(n-2, \delta)$

↳ $\delta = \frac{E(\hat\beta_1)}{\sqrt{Var(\hat\beta_1)}} = \frac{\beta_1}{\sigma / \sqrt{\Sigma (x_i - \bar x)^2}}$

is distributed as $t(n-2, \delta)$, the noncentral t with noncentrality parameter δ. By a comment following (5.~~29~~ 34), δ is given by $\delta = E(\hat{\beta}_1)/\sqrt{\mathrm{var}(\hat{\beta}_1)} = \beta_1 / [\sigma / \sqrt{\sum_i (x_i - \bar{x})^2}]$. If $\beta_1 = 0$, then by (5.~~28~~ 33), t is distributed as $t(n-2)$. For a two-sided alternative hypothesis $H_1 : \beta_1 \neq 0$, we reject $H_0 : \beta_1 = 0$ if $|t| \geq t_{\alpha/2,\, n-2}$, where $t_{\alpha/2,\, n-2}$ is the upper $\alpha/2$ percentage point of the central t distribution and α is the desired significance level of the test (probability of rejecting H_0 when it is true). Alternatively, we reject H_0 if $p \leq \alpha$, where p is the p value. For a two-sided test, the p value is defined as twice the probability that $t(n-2)$ exceeds the absolute value of the observed t.

$2P(T > |t|)$

$T \sim t(n-2)$



A $100(1-\alpha)\%$ confidence interval for β_1 is given by

$$\hat{\beta}_1 \pm t_{\alpha/2,\, n-2} \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}. \tag{6.15}$$

Confidence intervals are defined and discussed further in Section 8.6. A confidence interval for $E(y)$ and a prediction interval for y are also given in Section 8.6.

Example 6.3. We test the hypothesis H_0: $\beta_1 = 0$ for the grades data in Example 6.2. By (6.14), the t statistic is

$$t = \frac{\hat{\beta}_1}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{.8726}{(13.8547)/(139.753)} = 8.8025.$$

Since $t = 8.8025 > t_{.025,\, 16} = 2.120$, we reject H_0: $\beta_1 = 0$ at the $\alpha = .05$ level of significance. Alternatively, the p value is 1.571×10^{-7}, which is less than .05.

A 95% confidence interval for β_1 is given by (6.15) as

$$\hat{\beta}_1 \pm t_{.025,\, 16} \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$.8726 \pm 2.120(.09914)$$

$$.8726 \pm .2102$$

$$(.6624,\ 1.0828).$$

6.4 COEFFICIENT OF DETERMINATION 결정 계수.

The *coefficient of determination* r^2 is defined as

$$r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{6.16}$$

모델이 설명할 수 있는 변동량

총 변동량

where $\text{SSR} = \sum_i (\hat{y}_i - \bar{y})^2$ is the regression sum of squares and $\text{SST} = \sum_i (y_i - \bar{y})^2$ is the total sum of squares. The total sum of squares can be partitioned into SST = SSR + SSE, that is,

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{6.17}$$

Thus r^2 in (6.16) gives the proportion of variation in y that is explained by the model or, equivalently, accounted for by regression on x.

We have labeled (6.16) as r^2 because it is the same as the square of the *sample correlation coefficient* r between y and x

$$\sqrt{r^2} = r = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}, \tag{6.18}$$

where s_{xy} is given by 5.15 (see Problem 6.11). When x is a random variable, r estimates the population correlation in (3.19). The coefficient of determination r^2 is discussed further in Sections 7.7, 10.4, and 10.5.

상관계수

Example 6.4. For the grades data of Example 6.2, we have

$$r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{14,873.0}{17,944.3} = .8288.$$

The correlation between homework score and exam score is $r = \sqrt{.8288} = .910$.

The t statistic in (6.14) can be expressed in terms of r as follows:

$r^2 = \frac{\Sigma}{\quad}$

$$\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$$

$$t = \frac{\hat{\beta}_1}{s/\sqrt{\sum_i (x_i - \bar{x})^2}} \tag{6.19}$$

$$s \leftarrow \sqrt{\frac{SSE}{n-2}}$$

$$SST \times \frac{SST-SSR}{SST} = \Sigma(y_i-\bar{y})^2 \times (1-r^2)$$

$$\boxed{\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} \times \sqrt{\frac{n-2}{1-r^2}} \times \frac{\sqrt{\Sigma(x_i-\bar{x})^2}}{\sqrt{\Sigma(y_i-\bar{y})^2}}}$$

$r.$

$$= \frac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}. \tag{6.20}$$

If H_0: $\beta_1 = 0$ is true, then, as noted following (6.14), the statistic in (6.19) is distributed as $t(n-2)$ under the assumption that the x_i's are fixed and the y_i's are independently distributed as $N(\beta_0 + \beta_1 x_i, \sigma^2)$. If x is a random variable such that x and y have a bivariate normal distribution, then $t = \sqrt{n-2}\, r/\sqrt{1-r^2}$ in (6.20) also has the $t(n-2)$ distribution provided that $H_0 : \rho = 0$ is true, where ρ is the population correlation coefficient defined in (3.19) (see Theorem 10.5). However, (6.19) and (6.20) have different distributions if $H_0 : \beta_1 = 0$ and $H_0 : \rho = 0$ are false (see Section 10.4). If $\beta_1 \neq 0$, then (6.19) has a noncentral t distribution, but if $\rho \neq 0$, (6.20) does not have a noncentral t distribution.

PROBLEMS

6.1 Obtain the least-squares solutions (6.5) and (6.6) from (6.3) and (6.4).

6.2 **(a)** Show that $E(\hat{\beta}_1) = \beta_1$ as in (6.7).

(b) Show that $E(\hat{\beta}_0) = \beta_0$ as in (6.8).