Chapter 7

Multiple Regression II

In this chapter, we take up some specialized topics that are unique to multiple regression. These include extra sums of squares, which are useful for conducting a variety of tests about the regression coefficients, the standardized version of the multiple regression model, and multicollinearity, a condition where the predictor variables are highly correlated.

7.1 Extra Sums of Squares

Basic Ideas

An extra sum of squares measures the marginal reduction in the error sum of squares when one or several predictor variables are added to the regression model, given that other predictor variables are already in the model. Equivalently, one can view an extra sum of squares as measuring the marginal increase in the regression sum of squares when one or several predictor variables are added to the regression model.

We first utilize an example to illustrate these ideas, and then we present definitions of extra sums of squares and discuss a variety of uses of extra sums of squares in tests about regression coefficients.

Example

Table 7.1 contains a portion of the data for a study of the relation of amount of body fat (Y) to several possible predictor variables, based on a sample of 20 healthy females 25–34 years old. The possible predictor variables are triceps skinfold thickness (X_1), thigh circumference (X_2), and midarm circumference (X_3). The amount of body fat in Table 7.1 for each of the 20 persons was obtained by a cumbersome and expensive procedure requiring the immersion of the person in water. It would therefore be very helpful if a regression model with some or all of these predictor variables could provide reliable estimates of the amount of body fat since the measurements needed for the predictor variables are easy to obtain.

Table 7.2 contains some of the main regression results when body fat (Y) is regressed (1) on triceps skinfold thickness (X_1) alone, (2) on thigh circumference (X_2) alone, (3) on X_1 and X_2 only, and (4) on all three predictor variables. To keep track of the regression model that is fitted, we shall modify our notation slightly. The regression sum of squares when X_1 only is in the model is, according to Table 7.2a, 352.27. This sum of squares will be denoted by $SSR(X_1)$. The error sum of squares for this model will be denoted by $SSE(X_1)$; according to Table 7.2a it is $SSE(X_1) = 143.12$.

Similarly, Table 7.2c indicates that when X_1 and X_2 are in the regression model, the regression sum of squares is $SSR(X_1, X_2) = 385.44$ and the error sum of squares is $SSE\ (X_1, X_2) = 109.95$.

Notice that the error sum of squares when X_1 and X_2 are in the model, $SSE(X_1, X_2) = 109.95$, is smaller than when the model contains only X_1, $SSE(X_1) = 143.12$. The difference is called an *extra sum of squares* and will be denoted by $SSR(X_2|X_1)$:

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2)$$
$$= 143.12 - 109.95 = 33.17$$

TABLE 7.1 **Basic Data—Body Fat Example.**

Subject i	Triceps Skinfold Thickness X_{i1}	Thigh Circumference X_{i2}	Midarm Circumference X_{i3}	Body Fat Y_i
1	19.5	43.1	29.1	11.9
2	24.7	49.8	28.2	22.8
3	30.7	51.9	37.0	18.7
...	...	...	...	...
18	30.2	58.6	24.6	25.4
19	22.7	48.2	27.1	14.8
20	25.2	51.0	27.5	21.1

TABLE 7.2 **Regression Results for Several Fitted Models—Body Fat Example.**

(a) Regression of Y on X_1
$\hat{Y} = -1.496 + .8572X_1$

Source of Variation	SS	df	MS
Regression	352.27	1	352.27
Error	143.12	18	7.95
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_1	$b_1 = .8572$	$s\{b_1\} = .1288$	6.66

(b) Regression of Y on X_2
$\hat{Y} = -23.634 + .8565X_2$

Source of Variation	SS	df	MS
Regression	381.97	1	381.97
Error	113.42	18	6.30
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_2	$b_2 = .8565$	$s\{b_2\} = .1100$	7.79

(continued)

TABLE 7.2 (Continued).

(c) Regression of Y on X_1 and X_2
$\hat{Y} = -19.174 + .2224X_1 + .6594X_2$

Source of Variation	*SS*	*df*	*MS*
Regression	385.44	2	192.72
Error	109.95	17	6.47
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_1	$b_1 = .2224$	$s\{b_1\} = .3034$	.73
X_2	$b_2 = .6594$	$s\{b_2\} = .2912$	2.26

(d) Regression of Y on X_1, X_2, and X_3
$\hat{Y} = 117.08 + 4.334X_1 - 2.857X_2 - 2.186X_3$

Source of Variation	*SS*	*df*	*MS*
Regression	396.98	3	132.33
Error	98.41	16	6.15
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_1	$b_1 = 4.334$	$s\{b_1\} = 3.016$	1.44
X_2	$b_2 = -2.857$	$s\{b_2\} = 2.582$	−1.11
X_3	$b_3 = -2.186$	$s\{b_3\} = 1.596$	−1.37

This reduction in the error sum of squares is the result of adding X_2 to the regression model when X_1 is already included in the model. Thus, the extra sum of squares $SSR(X_2|X_1)$ measures the marginal effect of adding X_2 to the regression model when X_1 is already in the model. The notation $SSR(X_2|X_1)$ reflects this additional or extra reduction in the error sum of squares associated with X_2, given that X_1 is already included in the model.

The extra sum of squares $SSR(X_2|X_1)$ equivalently can be viewed as the marginal increase in the regression sum of squares:

$$\begin{aligned} SSR(X_2|X_1) &= SSR(X_1, X_2) - SSR(X_1) \\ &= 385.44 - 352.27 = 33.17 \end{aligned}$$

The reason for the equivalence of the marginal reduction in the error sum of squares and the marginal increase in the regression sum of squares is the basic analysis of variance identity (2.50):

$$SSTO = SSR + SSE$$

Since $SSTO$ measures the variability of the Y_i observations and hence does not depend on the regression model fitted, any reduction in SSE implies an identical increase in SSR.

We can consider other extra sums of squares, such as the marginal effect of adding X_3 to the regression model when X_1 and X_2 are already in the model. We find from Tables 7.2c and 7.2d that:

$$\begin{aligned} SSR(X_3|X_1, X_2) &= SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \\ &= 109.95 - 98.41 = 11.54 \end{aligned}$$

or, equivalently:

$$\begin{aligned} SSR(X_3|X_1, X_2) &= SSR(X_1, X_2, X_3) - SSR(X_1, X_2) \\ &= 396.98 - 385.44 = 11.54 \end{aligned}$$

We can even consider the marginal effect of adding several variables, such as adding both X_2 and X_3 to the regression model already containing X_1 (see Tables 7.2a and 7.2d):

$$\begin{aligned} SSR(X_2, X_3|X_1) &= SSE(X_1) - SSE(X_1, X_2, X_3) \\ &= 143.12 - 98.41 = 44.71 \end{aligned}$$

or, equivalently:

$$\begin{aligned} SSR(X_2, X_3|X_1) &= SSR(X_1, X_2, X_3) - SSR(X_1) \\ &= 396.98 - 352.27 = 44.71 \end{aligned}$$

Definitions

We assemble now our earlier definitions of extra sums of squares and provide some additional ones. As we noted earlier, an extra sum of squares always involves the difference between the error sum of squares for the regression model containing the X variable(s) already in the model and the error sum of squares for the regression model containing both the original X variable(s) and the new X variable(s). Equivalently, an extra sum of squares involves the difference between the two corresponding regression sums of squares.

Thus, we define:

$$SSR(X_1|X_2) = SSE(X_2) - SSE(X_1, X_2) \quad \textbf{(7.1a)}$$

or, equivalently:

$$SSR(X_1|X_2) = SSR(X_1, X_2) - SSR(X_2) \quad \textbf{(7.1b)}$$

If X_2 is the extra variable, we define:

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) \quad \textbf{(7.2a)}$$

or, equivalently:

$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1) \quad \textbf{(7.2b)}$$

Extensions for three or more variables are straightforward. For example, we define:

$$SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \quad \textbf{(7.3a)}$$

or:

$$SSR(X_3|X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2) \quad \textbf{(7.3b)}$$

and:

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) \quad \textbf{(7.4a)}$$

or:

$$SSR(X_2, X_3|X_1) = SSR(X_1, X_2, X_3) - SSR(X_1) \quad \textbf{(7.4b)}$$

Decomposition of *SSR* into Extra Sums of Squares

In multiple regression, unlike simple linear regression, we can obtain a variety of decompositions of the regression sum of squares *SSR* into extra sums of squares. Let us consider the case of two X variables. We begin with the identity (2.50) for variable X_1:

$$SSTO = SSR(X_1) + SSE(X_1) \quad \textbf{(7.5)}$$

where the notation now shows explicitly that X_1 is the X variable in the model. Replacing $SSE(X_1)$ by its equivalent in (7.2a), we obtain:

$$SSTO = SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2) \quad \textbf{(7.6)}$$

We now make use of the same identity for multiple regression with two X variables as in (7.5) for a single X variable, namely:

$$SSTO = SSR(X_1, X_2) + SSE(X_1, X_2) \quad \textbf{(7.7)}$$

Solving (7.7) for $SSE(X_1, X_2)$ and using this expression in (7.6) lead to:

$$SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1) \quad \textbf{(7.8)}$$

Thus, we have decomposed the regression sum of squares $SSR(X_1, X_2)$ into two marginal components: (1) $SSR(X_1)$, measuring the contribution by including X_1 alone in the model, and (2) $SSR(X_2|X_1)$, measuring the additional contribution when X_2 is included, given that X_1 is already in the model.

Of course, the order of the X variables is arbitrary. Here, we can also obtain the decomposition:

$$SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2) \quad \textbf{(7.9)}$$

We show in Figure 7.1 schematic representations of the two decompositions of $SSR(X_1, X_2)$ for the body fat example. The total bar on the left represents *SSTO* and presents decomposition (7.9). The unshaded component of this bar is $SSR(X_2)$, and the combined shaded area represents $SSE(X_2)$. The latter area in turn is the combination of the extra sum of squares $SSR(X_1|X_2)$ and the error sum of squares $SSE(X_1, X_2)$ when both X_1 and X_2 are included in the model. Similarly, the bar on the right in Figure 7.1 shows decomposition (7.8). Note in both cases how the extra sum of squares can be viewed either as a reduction in the error sum of squares or as an increase in the regression sum of squares when the second predictor variable is added to the regression model.

When the regression model contains three X variables, a variety of decompositions of $SSR(X_1, X_2, X_3)$ can be obtained. We illustrate three of these:

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \quad \textbf{(7.10a)}$$

$$SSR(X_1, X_2, X_3) = SSR(X_2) + SSR(X_3|X_2) + SSR(X_1|X_2, X_3) \quad \textbf{(7.10b)}$$

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2, X_3|X_1) \quad \textbf{(7.10c)}$$

FIGURE 7.1 **Schematic Representation of Extra Sums of Squares—Body Fat Example.**

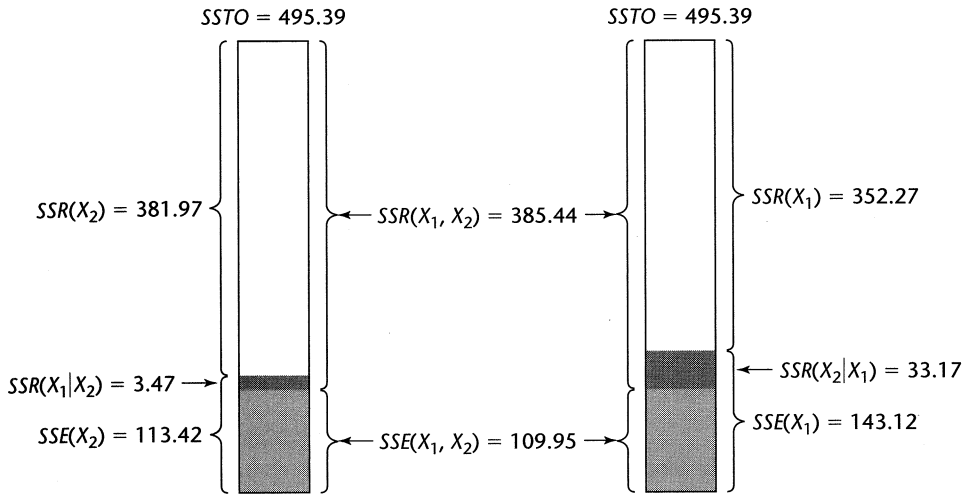

TABLE 7.3 **Example of ANOVA Table with Decomposition of SSR for Three X Variables.**

Source of Variation	*SS*	*df*	*MS*
Regression	$SSR(X_1, X_2, X_3)$	3	$MSR(X_1, X_2, X_3)$
X_1	$SSR(X_1)$	1	$MSR(X_1)$
$X_2 \mid X_1$	$SSR(X_2 \mid X_1)$	1	$MSR(X_2 \mid X_1)$
$X_3 \mid X_1, X_2$	$SSR(X_3 \mid X_1, X_2)$	1	$MSR(X_3 \mid X_1, X_2)$
Error	$SSE(X_1, X_2, X_3)$	$n-4$	$MSE(X_1, X_2, X_3)$
Total	$SSTO$	$n-1$	

It is obvious that the number of possible decompositions becomes vast as the number of X variables in the regression model increases.

ANOVA Table Containing Decomposition of *SSR*

ANOVA tables can be constructed containing decompositions of the regression sum of squares into extra sums of squares. Table 7.3 contains the ANOVA table decomposition for the case of three X variables often used in regression packages, and Table 7.4 contains this same decomposition for the body fat example. The decomposition involves single extra X variables.

Note that each extra sum of squares involving a single extra X variable has associated with it one degree of freedom. The resulting mean squares are constructed as usual. For example, $MSR(X_2|X_1)$ in Table 7.3 is obtained as follows:

$$MSR(X_2|X_1) = \frac{SSR(X_2|X_1)}{1}$$

Extra sums of squares involving two extra X variables, such as $SSR(X_2, X_3|X_1)$, have two degrees of freedom associated with them. This follows because we can express such an extra sum of squares as a sum of two extra sums of squares, each associated with one

TABLE 7.4 ANOVA Table with Decomposition of *SSR*—Body Fat Example with Three Predictor Variables.

Source of Variation	*SS*	*df*	*MS*
Regression	396.98	3	132.33
X_1	352.27	1	352.27
$X_2 \mid X_1$	33.17	1	33.17
$X_3 \mid X_1, X_2$	11.54	1	11.54
Error	98.41	16	6.15
Total	495.39	19	

degree of freedom. For example, by definition of the extra sums of squares, we have:

$$SSR(X_2, X_3|X_1) = SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \tag{7.11}$$

The mean square $MSR(X_2, X_3|X_1)$ is therefore obtained as follows:

$$MSR(X_2, X_3|X_1) = \frac{SSR(X_2, X_3|X_1)}{2}$$

Many computer regression packages provide decompositions of *SSR* into single-degree-of-freedom extra sums of squares, usually in the order in which the X variables are entered into the model. Thus, if the X variables are entered in the order X_1, X_2, X_3, the extra sums of squares given in the output are:

$$SSR(X_1)$$
$$SSR(X_2|X_1)$$
$$SSR(X_3|X_1, X_2)$$

If an extra sum of squares involving several extra X variables is desired, it can be obtained by summing appropriate single-degree-of-freedom extra sums of squares. For instance, to obtain $SSR(X_2, X_3|X_1)$ in our earlier illustration, we would utilize (7.11) and simply add $SSR(X_2|X_1)$ and $SSR(X_3|X_1, X_2)$.

If the extra sum of squares $SSR(X_1, X_3|X_2)$ were desired with a computer package that provides single-degree-of-freedom extra sums of squares in the order in which the X variables are entered, the X variables would need to be entered in the order X_2, X_1, X_3 or X_2, X_3, X_1. The first ordering would give:

$$SSR(X_2)$$
$$SSR(X_1|X_2)$$
$$SSR(X_3|X_1, X_2)$$

The sum of the last two extra sums of squares will yield $SSR(X_1, X_3|X_2)$.

The reason why extra sums of squares are of interest is that they occur in a variety of tests about regression coefficients where the question of concern is whether certain X variables can be dropped from the regression model. We turn next to this use of extra sums of squares.

7.2 Uses of Extra Sums of Squares in Tests for Regression Coefficients

Test whether a Single $\beta_k = 0$

When we wish to test whether the term $\beta_k X_k$ can be dropped from a multiple regression model, we are interested in the alternatives:

$$H_0: \beta_k = 0$$
$$H_a: \beta_k \neq 0$$

We already know that test statistic (6.51b):

$$t^* = \frac{b_k}{s\{b_k\}}$$

is appropriate for this test.

Equivalently, we can use the general linear test approach described in Section 2.8. We now show that this approach involves an extra sum of squares. Let us consider the first-order regression model with three predictor variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \text{Full model} \qquad \textbf{(7.12)}$$

To test the alternatives:

$$H_0: \beta_3 = 0$$
$$H_a: \beta_3 \neq 0 \qquad \textbf{(7.13)}$$

we fit the full model and obtain the error sum of squares $SSE(F)$. We now explicitly show the variables in the full model, as follows:

$$SSE(F) = SSE(X_1, X_2, X_3)$$

The degrees of freedom associated with $SSE(F)$ are $df_F = n - 4$ since there are four parameters in the regression function for the full model (7.12).

The reduced model when H_0 in (7.13) holds is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad \text{Reduced model} \qquad \textbf{(7.14)}$$

We next fit this reduced model and obtain:

$$SSE(R) = SSE(X_1, X_2)$$

There are $df_R = n - 3$ degrees of freedom associated with the reduced model.

The general linear test statistic (2.70):

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

here becomes:

$$F^* = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{(n-3) - (n-4)} \div \frac{SSE(X_1, X_2, X_3)}{n-4}$$

Note that the difference between the two error sums of squares in the numerator term is the extra sum of squares (7.3a):

$$SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = SSR(X_3|X_1, X_2)$$

Hence the general linear test statistic here is:

$$F^* = \frac{SSR(X_3|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{MSR(X_3|X_1, X_2)}{MSE(X_1, X_2, X_3)} \qquad (7.15)$$

We thus see that the test whether or not $\beta_3 = 0$ is a marginal test, given that X_1 and X_2 are already in the model. We also note that the extra sum of squares $SSR(X_3|X_1, X_2)$ has one degree of freedom associated with it, just as we noted earlier.

Test statistic (7.15) shows that we do not need to fit both the full model and the reduced model to use the general linear test approach here. A single computer run can provide a fit of the full model and the appropriate extra sum of squares.

Example

In the body fat example, we wish to test for the model with all three predictor variables whether midarm circumference (X_3) can be dropped from the model. The test alternatives are those of (7.13). Table 7.4 contains the ANOVA results from a computer fit of the full regression model (7.12), including the extra sums of squares when the predictor variables are entered in the order X_1, X_2, X_3. Hence, test statistic (7.15) here is:

$$F^* = \frac{SSR(X_3|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-4}$$

$$= \frac{11.54}{1} \div \frac{98.41}{16} = 1.88$$

For $\alpha = .01$, we require $F(.99; 1, 16) = 8.53$. Since $F^* = 1.88 \leq 8.53$, we conclude H_0, that X_3 can be dropped from the regression model that already contains X_1 and X_2.

Note from Table 7.2d that the t^* test statistic here is:

$$t^* = \frac{b_3}{s\{b_3\}} = \frac{-2.186}{1.596} = -1.37$$

Since $(t^*)^2 = (-1.37)^2 = 1.88 = F^*$, we see that the two test statistics are equivalent, just as for simple linear regression.

Comment

The F^* test statistic (7.15) to test whether or not $\beta_3 = 0$ is called a *partial F test* statistic to distinguish it from the F^* statistic in (6.39b) for testing whether *all* $\beta_k = 0$, i.e., whether or not there is a regression relation between Y and the set of X variables. The latter test is called the *overall F test*. ■

Test whether Several $\beta_k = 0$

In multiple regression we are frequently interested in whether several terms in the regression model can be dropped. For example, we may wish to know whether both $\beta_2 X_2$ and $\beta_3 X_3$ can be dropped from the full model (7.12). The alternatives here are:

$$H_0: \beta_2 = \beta_3 = 0$$
$$H_a: \text{not both } \beta_2 \text{ and } \beta_3 \text{ equal zero} \qquad (7.16)$$

With the general linear test approach, the reduced model under H_0 is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \qquad \text{Reduced model} \tag{7.17}$$

and the error sum of squares for the reduced model is:

$$SSE(R) = SSE(X_1)$$

This error sum of squares has $df_R = n - 2$ degrees of freedom associated with it.

The general linear test statistic (2.70) thus becomes here:

$$F^* = \frac{SSE(X_1) - SSE(X_1, X_2, X_3)}{(n-2) - (n-4)} \div \frac{SSE(X_1, X_2, X_3)}{n-4}$$

Again the difference between the two error sums of squares in the numerator term is an extra sum of squares, namely:

$$SSE(X_1) - SSE(X_1, X_2, X_3) = SSR(X_2, X_3|X_1)$$

Hence, the test statistic becomes:

$$F^* = \frac{SSR(X_2, X_3|X_1)}{2} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{MSR(X_2, X_3|X_1)}{MSE(X_1, X_2, X_3)} \tag{7.18}$$

Note that $SSR(X_2, X_3|X_1)$ has two degrees of freedom associated with it, as we pointed out earlier.

Example

We wish to test in the body fat example for the model with all three predictor variables whether both thigh circumference (X_2) and midarm circumference (X_3) can be dropped from the full regression model (7.12). The alternatives are those in (7.16). The appropriate extra sum of squares can be obtained from Table 7.4, using (7.11):

$$\begin{aligned} SSR(X_2, X_3|X_1) &= SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \\ &= 33.17 + 11.54 = 44.71 \end{aligned}$$

Test statistic (7.18) therefore is:

$$\begin{aligned} F^* &= \frac{SSR(X_2, X_3|X_1)}{2} \div MSE(X_1, X_2, X_3) \\ &= \frac{44.71}{2} \div 6.15 = 3.63 \end{aligned}$$

For $\alpha = .05$, we require $F(.95; 2, 16) = 3.63$. Since $F^* = 3.63$ is at the boundary of the decision rule (the P-value of the test statistic is .05), we may wish to make further analyses before deciding whether X_2 and X_3 should be dropped from the regression model that already contains X_1.

Comments

1. For testing whether a single β_k equals zero, two equivalent test statistics are available: the t^* test statistic and the F^* general linear test statistic. When testing whether several β_k equal zero, only the general linear test statistic F^* is available.

2. General linear test statistic (2.70) for testing whether several X variables can be dropped from the general linear regression model (6.7) can be expressed in terms of the coefficients of

multiple determination for the full and reduced models. Denoting these by R_F^2 and R_R^2, respectively, we have:

$$F^* = \frac{R_F^2 - R_R^2}{df_R - df_F} \div \frac{1 - R_F^2}{df_F} \tag{7.19}$$

Specifically for testing the alternatives in (7.16) for the body fat example, test statistic (7.19) becomes:

$$F^* = \frac{R_{Y|123}^2 - R_{Y|1}^2}{(n-2) - (n-4)} \div \frac{1 - R_{Y|123}^2}{n-4} \tag{7.20}$$

where $R_{Y|123}^2$ denotes the coefficient of multiple determination when Y is regressed on X_1, X_2, and X_3, and $R_{Y|1}^2$ denotes the coefficient when Y is regressed on X_1 alone.

We see from Table 7.4 that $R_{Y|123}^2 = 396.98/495.39 = .80135$ and $R_{Y|1}^2 = 352.27/495.39 = .71110$. Hence, we obtain by substituting in (7.20):

$$F^* = \frac{.80135 - .71110}{(20-2) - (20-4)} \div \frac{1 - .80135}{16} = 3.63$$

This is the same result as before. Note that $R_{Y|1}^2$ corresponds to the coefficient of simple determination R^2 between Y and X_1.

Test statistic (7.19) is not appropriate when the full and reduced regression models do not contain the intercept term β_0. In that case, the general linear test statistic in the form (2.70) must be used. ■

7.3 Summary of Tests Concerning Regression Coefficients

We have already discussed how to conduct several types of tests concerning regression coefficients in a multiple regression model. For completeness, we summarize here these tests as well as some additional types of tests.

Test whether All $\beta_k = 0$

This is the *overall F test* (6.39) of whether or not there is a regression relation between the response variable Y and the set of X variables. The alternatives are:

$$\begin{aligned} &H_0\text{: } \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \\ &H_a\text{: not all } \beta_k \ (k = 1, \ldots, p-1) \text{ equal zero} \end{aligned} \tag{7.21}$$

and the test statistic is:

$$\begin{aligned} F^* &= \frac{SSR(X_1, \ldots, X_{p-1})}{p-1} \div \frac{SSE(X_1, \ldots, X_{p-1})}{n-p} \\ &= \frac{MSR}{MSE} \end{aligned} \tag{7.22}$$

If H_0 holds, $F^* \sim F(p-1, n-p)$. Large values of F^* lead to conclusion H_a.

Test whether a Single $\beta_k = 0$

This is a *partial F test* of whether a particular regression coefficient β_k equals zero. The alternatives are:

$$H_0: \beta_k = 0$$
$$H_a: \beta_k \neq 0 \tag{7.23}$$

and the test statistic is:

$$F^* = \frac{SSR(X_k|X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{p-1})}{1} \div \frac{SSE(X_1, \ldots, X_{p-1})}{n-p}$$
$$= \frac{MSR(X_k|X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{p-1})}{MSE} \tag{7.24}$$

If H_0 holds, $F^* \sim F(1, n-p)$. Large values of F^* lead to conclusion H_a. Statistics packages that provide extra sums of squares permit use of this test without having to fit the reduced model.

An equivalent test statistic is (6.51b):

$$t^* = \frac{b_k}{s\{b_k\}} \tag{7.25}$$

If H_0 holds, $t^* \sim t(n-p)$. Large values of $|t^*|$ lead to conclusion H_a.

Since the two tests are equivalent, the choice is usually made in terms of available information provided by the regression package output.

Test whether Some $\beta_k = 0$

This is another *partial F test*. Here, the alternatives are:

$$H_0: \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$$
$$H_a: \text{not all of the } \beta_k \text{ in } H_0 \text{ equal zero} \tag{7.26}$$

where for convenience, we arrange the model so that the last $p-q$ coefficients are the ones to be tested. The test statistic is:

$$F^* = \frac{SSR(X_q, \ldots, X_{p-1}|X_1, \ldots, X_{q-1})}{p-q} \div \frac{SSE(X_1, \ldots, X_{p-1})}{n-p}$$
$$= \frac{MSR(X_q, \ldots, X_{p-1}|X_1, \ldots, X_{q-1})}{MSE} \tag{7.27}$$

If H_0 holds, $F^* \sim F(p-q, n-p)$. Large values of F^* lead to conclusion H_a.

Note that test statistic (7.27) actually encompasses the two earlier cases. If $q = 1$, the test is whether all regression coefficients equal zero. If $q = p-1$, the test is whether a single regression coefficient equals zero. Also note that test statistic (7.27) can be calculated without having to fit the reduced model if the regression package provides the needed extra sums of squares:

$$SSR(X_q, \ldots, X_{p-1}|X_1, \ldots, X_{q-1})$$
$$= SSR(X_q|X_1, \ldots, X_{q-1}) + \cdots + SSR(X_{p-1}|X_1, \ldots, X_{p-2}) \tag{7.28}$$

Test statistic (7.27) can be stated equivalently in terms of the coefficients of multiple determination for the full and reduced models when these models contain the intercept term β_0, as follows:

$$F^* = \frac{R^2_{Y|1\cdots p-1} - R^2_{Y|1\cdots q-1}}{p-q} \div \frac{1 - R^2_{Y|1\cdots p-1}}{n-p} \tag{7.29}$$

where $R^2_{Y|1\cdots p-1}$ denotes the coefficient of multiple determination when Y is regressed on all X variables, and $R^2_{Y|1\cdots q-1}$ denotes the coefficient when Y is regressed on $X_1, \ldots, X_{q-1}$ only.

Other Tests

When tests about regression coefficients are desired that do not involve testing whether one or several β_k equal zero, extra sums of squares cannot be used and the general linear test approach requires separate fittings of the full and reduced models. For instance, for the full model containing three X variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \text{Full model} \tag{7.30}$$

we might wish to test:

$$\begin{aligned} H_0&: \beta_1 = \beta_2 \\ H_a&: \beta_1 \neq \beta_2 \end{aligned} \tag{7.31}$$

The procedure would be to fit the full model (7.30), and then the reduced model:

$$Y_i = \beta_0 + \beta_c(X_{i1} + X_{i2}) + \beta_3 X_{i3} + \varepsilon_i \qquad \text{Reduced model} \tag{7.32}$$

where β_c denotes the common coefficient for β_1 and β_2 under H_0 and $X_{i1} + X_{i2}$ is the corresponding new X variable. We then use the general F^* test statistic (2.70) with 1 and $n - 4$ degrees of freedom.

Another example where extra sums of squares cannot be used is in the following test for regression model (7.30):

$$\begin{aligned} H_0&: \beta_1 = 3, \beta_3 = 5 \\ H_a&: \text{not both equalities in } H_0 \text{ hold} \end{aligned} \tag{7.33}$$

Here, the reduced model would be:

$$Y_i - 3X_{i1} - 5X_{i3} = \beta_0 + \beta_2 X_{i2} + \varepsilon_i \qquad \text{Reduced model} \tag{7.34}$$

Note the new response variable $Y - 3X_1 - 5X_3$ in the reduced model, since $\beta_1 X_1$ and $\beta_3 X_3$ are known constants under H_0. We then use the general linear test statistic F^* in (2.70) with 2 and $n - 4$ degrees of freedom.

7.4 Coefficients of Partial Determination

Extra sums of squares are not only useful for tests on the regression coefficients of a multiple regression model, but they are also encountered in descriptive measures of relationship called coefficients of partial determination. Recall that the coefficient of multiple determination, R^2, measures the proportionate reduction in the variation of Y achieved by the introduction

of the entire set of X variables considered in the model. A *coefficient of partial determination,* in contrast, measures the marginal contribution of one X variable when all others are already included in the model.

Two Predictor Variables

We first consider a first-order multiple regression model with two predictor variables, as given in (6.1):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

$SSE(X_2)$ measures the variation in Y when X_2 is included in the model. $SSE(X_1, X_2)$ measures the variation in Y when both X_1 and X_2 are included in the model. Hence, the relative marginal reduction in the variation in Y associated with X_1 when X_2 is already in the model is:

$$\frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1|X_2)}{SSE(X_2)}$$

This measure is the coefficient of partial determination between Y and X_1, given that X_2 is in the model. We denote this measure by $R^2_{Y1|2}$:

$$R^2_{Y1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1|X_2)}{SSE(X_2)} \tag{7.35}$$

Thus, $R^2_{Y1|2}$ measures the proportionate reduction in the variation in Y remaining after X_2 is included in the model that is gained by also including X_1 in the model.

The coefficient of partial determination between Y and X_2, given that X_1 is in the model, is defined correspondingly:

$$R^2_{Y2|1} = \frac{SSR(X_2|X_1)}{SSE(X_1)} \tag{7.36}$$

General Case

The generalization of coefficients of partial determination to three or more X variables in the model is immediate. For instance:

$$R^2_{Y1|23} = \frac{SSR(X_1|X_2, X_3)}{SSE(X_2, X_3)} \tag{7.37}$$

$$R^2_{Y2|13} = \frac{SSR(X_2|X_1, X_3)}{SSE(X_1, X_3)} \tag{7.38}$$

$$R^2_{Y3|12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} \tag{7.39}$$

$$R^2_{Y4|123} = \frac{SSR(X_4|X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)} \tag{7.40}$$

Note that in the subscripts to R^2, the entries to the left of the vertical bar show in turn the variable taken as the response and the X variable being added. The entries to the right of the vertical bar show the X variables already in the model.

Example

For the body fat example, we can obtain a variety of coefficients of partial determination. Here are three (Tables 7.2 and 7.4):

$$R^2_{Y2|1} = \frac{SSR(X_2|X_1)}{SSE(X_1)} = \frac{33.17}{143.12} = .232$$

$$R^2_{Y3|12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.54}{109.95} = .105$$

$$R^2_{Y1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = \frac{3.47}{113.42} = .031$$

We see that when X_2 is added to the regression model containing X_1 here, the error sum of squares $SSE(X_1)$ is reduced by 23.2 percent. The error sum of squares for the model containing both X_1 and X_2 is only reduced by another 10.5 percent when X_3 is added to the model. Finally, if the regression model already contains X_2, adding X_1 reduces $SSE(X_2)$ by only 3.1 percent.

Comments

1. The coefficients of partial determination can take on values between 0 and 1, as the definitions readily indicate.

2. A coefficient of partial determination can be interpreted as a coefficient of simple determination. Consider a multiple regression model with two X variables. Suppose we regress Y on X_2 and obtain the residuals:

$$e_i(Y|X_2) = Y_i - \hat{Y}_i(X_2)$$

where $\hat{Y}_i(X_2)$ denotes the fitted values of Y when X_2 is in the model. Suppose we further regress X_1 on X_2 and obtain the residuals:

$$e_i(X_1|X_2) = X_{i1} - \hat{X}_{i1}(X_2)$$

where $\hat{X}_{i1}(X_2)$ denotes the fitted values of X_1 in the regression of X_1 on X_2. The coefficient of simple determination R^2 between these two sets of residuals equals the coefficient of partial determination $R^2_{Y1|2}$. Thus, this coefficient measures the relation between Y and X_1 when both of these variables have been adjusted for their linear relationships to X_2.

3. The plot of the residuals $e_i(Y|X_2)$ against $e_i(X_1|X_2)$ provides a graphical representation of the strength of the relationship between Y and X_1, adjusted for X_2. Such plots of residuals, called *added variable plots* or *partial regression plots,* are discussed in Section 10.1. ■

Coefficients of Partial Correlation

The square root of a coefficient of partial determination is called a *coefficient of partial correlation.* It is given the same sign as that of the corresponding regression coefficient in the fitted regression function. Coefficients of partial correlation are frequently used in practice, although they do not have as clear a meaning as coefficients of partial determination. One use of partial correlation coefficients is in computer routines for finding the best predictor variable to be selected next for inclusion in the regression model. We discuss this use in Chapter 9.

Example

For the body fat example, we have:

$$r_{Y2|1} = \sqrt{.232} = .482$$
$$r_{Y3|12} = -\sqrt{.105} = -.324$$
$$r_{Y1|2} = \sqrt{.031} = .176$$

Note that the coefficients $r_{Y2|1}$ and $r_{Y1|2}$ are positive because we see from Table 7.2c that $b_2 = .6594$ and $b_1 = .2224$ are positive. Similarly, $r_{Y3|12}$ is negative because we see from Table 7.2d that $b_3 = -2.186$ is negative.

Comment

Coefficients of partial determination can be expressed in terms of simple or other partial correlation coefficients. For example:

$$R^2_{Y2|1} = [r_{Y2|1}]^2 = \frac{(r_{Y2} - r_{12}r_{Y1})^2}{(1 - r^2_{12})(1 - r^2_{Y1})} \tag{7.41}$$

$$R^2_{Y2|13} = [r_{Y2|13}]^2 = \frac{(r_{Y2|3} - r_{12|3}r_{Y1|3})^2}{(1 - r^2_{12|3})(1 - r^2_{Y1|3})} \tag{7.42}$$

where r_{Y1} denotes the coefficient of simple correlation between Y and X_1, r_{12} denotes the coefficient of simple correlation between X_1 and X_2, and so on. Extensions are straightforward. ■

7.5 Standardized Multiple Regression Model

A standardized form of the general multiple regression model (6.7) is employed to control roundoff errors in normal equations calculations and to permit comparisons of the estimated regression coefficients in common units.

Roundoff Errors in Normal Equations Calculations

The results from normal equations calculations can be sensitive to rounding of data in intermediate stages of calculations. When the number of X variables is small—say, three or less—roundoff effects can be controlled by carrying a sufficient number of digits in intermediate calculations. Indeed, most computer regression programs use double-precision arithmetic in all computations to control roundoff effects. Still, with a large number of X variables, serious roundoff effects can arise despite the use of many digits in intermediate calculations.

Roundoff errors tend to enter normal equations calculations primarily when the inverse of $\mathbf{X}'\mathbf{X}$ is taken. Of course, any errors in $(\mathbf{X}'\mathbf{X})^{-1}$ may be magnified in calculating $\mathbf{b}$ and other subsequent statistics. The danger of serious roundoff errors in $(\mathbf{X}'\mathbf{X})^{-1}$ is particularly great when (1) $\mathbf{X}'\mathbf{X}$ has a determinant that is close to zero and/or (2) the elements of $\mathbf{X}'\mathbf{X}$ differ substantially in order of magnitude. The first condition arises when some or all of the X variables are highly intercorrelated. We shall discuss this situation in Section 7.6.

The second condition arises when the X variables have substantially different magnitudes so that the entries in the $\mathbf{X}'\mathbf{X}$ matrix cover a wide range, say, from 15 to 49,000,000. A solution for this condition is to transform the variables and thereby reparameterize the regression model into the standardized regression model.

The transformation to obtain the standardized regression model, called the *correlation transformation,* makes all entries in the $\mathbf{X}'\mathbf{X}$ matrix for the transformed variables fall between -1 and 1 inclusive, so that the calculation of the inverse matrix becomes much less subject to roundoff errors due to dissimilar orders of magnitudes than with the original variables.

Comment

In order to avoid the computational difficulties inherent in inverting the $\mathbf{X}'\mathbf{X}$ matrix, many statistical packages use an entirely different computational approach that involves decomposing the $\mathbf{X}$ matrix into a product of several matrices with special properties. The $\mathbf{X}$ matrix is often first modified by centering each of the variables (i.e., using the deviations around the mean) to further improve computational accuracy. Information on decomposition strategies may be found in texts on statistical computing, such as Reference 7.1. ■

Lack of Comparability in Regression Coefficients

A second difficulty with the nonstandardized multiple regression model (6.7) is that ordinarily regression coefficients cannot be compared because of differences in the units involved. We cite two examples.

1. When considering the fitted response function:

$$\hat{Y} = 200 + 20{,}000X_1 + .2X_2$$

one may be tempted to conclude that X_1 is the only important predictor variable, and that X_2 has little effect on the response variable Y. A little reflection should make one wary of this conclusion. The reason is that we do not know the units involved. Suppose the units are:

Y in dollars
X_1 in thousand dollars
X_2 in cents

In that event, the effect on the mean response of a \$1,000 increase in X_1 (i.e., a 1-unit increase) when X_2 is constant would be an increase of \$20,000. This is exactly the same as the effect of a \$1,000 increase in X_2 (i.e., a 100,000-unit increase) when X_1 is constant, despite the difference in the regression coefficients.

2. In the Dwaine Studios example of Figure 6.5, we cannot make any comparison between b_1 and b_2 because X_1 is in units of thousand persons aged 16 or younger, whereas X_2 is in units of thousand dollars of per capita disposable income.

Correlation Transformation

Use of the correlation transformation helps with controlling roundoff errors and, by expressing the regression coefficients in the same units, may be of help when these coefficients are compared. We shall first describe the correlation transformation and then the resulting standardized regression model.

The correlation transformation is a simple modification of the usual standardization of a variable. Standardizing a variable, as in (A.37), involves centering and scaling the variable. *Centering* involves taking the difference between each observation and the mean of all observations for the variable; *scaling* involves expressing the centered observations in units of the standard deviation of the observations for the variable. Thus, the usual standardizations

of the response variable Y and the predictor variables $X_1, \ldots, X_{p-1}$ are as follows:

$$\frac{Y_i - \bar{Y}}{s_Y} \tag{7.43a}$$

$$\frac{X_{ik} - \bar{X}_k}{s_k} \qquad (k = 1, \ldots, p-1) \tag{7.43b}$$

where $\bar{Y}$ and $\bar{X}_k$ are the respective means of the Y and the X_k observations, and s_Y and s_k are the respective standard deviations defined as follows:

$$s_Y = \sqrt{\frac{\sum_i (Y_i - \bar{Y})^2}{n-1}} \tag{7.43c}$$

$$s_k = \sqrt{\frac{\sum_i (X_{ik} - \bar{X}_k)^2}{n-1}} \qquad (k = 1, \ldots, p-1) \tag{7.43d}$$

The correlation transformation is a simple function of the standardized variables in (7.43a, b):

$$Y_i^* = \frac{1}{\sqrt{n-1}} \left(\frac{Y_i - \bar{Y}}{s_Y} \right) \tag{7.44a}$$

$$X_{ik}^* = \frac{1}{\sqrt{n-1}} \left(\frac{X_{ik} - \bar{X}_k}{s_k} \right) \qquad (k = 1, \ldots, p-1) \tag{7.44b}$$

Standardized Regression Model

The regression model with the transformed variables Y^* and X_k^* as defined by the correlation transformation in (7.44) is called a *standardized regression model* and is as follows:

$$Y_i^* = \beta_1^* X_{i1}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^* \tag{7.45}$$

The reason why there is no intercept parameter in the standardized regression model (7.45) is that the least squares or maximum likelihood calculations always would lead to an estimated intercept term of zero if an intercept parameter were present in the model.

It is easy to show that the parameters $\beta_1^*, \ldots, \beta_{p-1}^*$ in the standardized regression model and the original parameters $\beta_0, \beta_1, \ldots, \beta_{p-1}$ in the ordinary multiple regression model (6.7) are related as follows:

$$\beta_k = \left(\frac{s_Y}{s_k} \right) \beta_k^* \qquad (k = 1, \ldots, p-1) \tag{7.46a}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \cdots - \beta_{p-1} \bar{X}_{p-1} \tag{7.46b}$$

We see that the standardized regression coefficients β_k^* and the original regression coefficients β_k $(k = 1, \ldots, p-1)$ are related by simple scaling factors involving ratios of standard deviations.

X′X Matrix for Transformed Variables

In order to be able to study the special nature of the $\mathbf{X}'\mathbf{X}$ matrix and the least squares normal equations when the variables have been transformed by the correlation transformation, we need to decompose the correlation matrix in (6.67) containing all pairwise correlation coefficients among the response and predictor variables $Y, X_1, X_2, \ldots, X_{p-1}$ into two matrices.

1. The first matrix, denoted by $\mathbf{r}_{XX}$, is called the *correlation matrix of the X variables*. It has as its elements the coefficients of simple correlation between all pairs of the X variables. This matrix is defined as follows:

$$\underset{(p-1)\times(p-1)}{\mathbf{r}_{XX}} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{21} & 1 & \cdots & r_{2,p-1} \\ \vdots & \vdots & & \vdots \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix} \tag{7.47}$$

Here, r_{12} again denotes the coefficient of simple correlation between X_1 and X_2, and so on. Note that the main diagonal consists of 1s because the coefficient of simple correlation between a variable and itself is 1. The correlation matrix $\mathbf{r}_{XX}$ is symmetric; remember that $r_{kk'} = r_{k'k}$. Because of the symmetry of this matrix, computer printouts frequently omit the lower or upper triangular block of elements.

2. The second matrix, denoted by $\mathbf{r}_{YX}$, is a vector containing the coefficients of simple correlation between the response variable Y and each of the X variables, denoted again by r_{Y1}, r_{Y2}, etc.:

$$\underset{(p-1)\times 1}{\mathbf{r}_{YX}} = \begin{bmatrix} r_{Y1} \\ r_{Y2} \\ \vdots \\ r_{Y,p-1} \end{bmatrix} \tag{7.48}$$

Now we are ready to consider the $\mathbf{X}'\mathbf{X}$ matrix for the transformed variables in the standardized regression model (7.45). The $\mathbf{X}$ matrix here is:

$$\underset{n\times(p-1)}{\mathbf{X}} = \begin{bmatrix} X_{11}^* & \cdots & X_{1,p-1}^* \\ X_{21}^* & \cdots & X_{2,p-1}^* \\ \vdots & & \vdots \\ X_{n1}^* & \cdots & X_{n,p-1}^* \end{bmatrix} \tag{7.49}$$

Remember that the standardized regression model (7.45) does not contain an intercept term; hence, there is no column of 1s in the $\mathbf{X}$ matrix. It can be shown that the $\mathbf{X}'\mathbf{X}$ matrix for the transformed variables is simply the correlation matrix of the X variables defined in (7.47):

$$\underset{(p-1)\times(p-1)}{\mathbf{X}'\mathbf{X}} = \mathbf{r}_{XX} \tag{7.50}$$

Since the $\mathbf{X}'\mathbf{X}$ matrix for the transformed variables consists of coefficients of correlation between the X variables, all of its elements are between -1 and 1 and thus are of the same order of magnitude. As we pointed out earlier, this can be of great help in controlling roundoff errors when inverting the $\mathbf{X}'\mathbf{X}$ matrix.

Comment

We illustrate that the $\mathbf{X}'\mathbf{X}$ matrix for the transformed variables is the correlation matrix of the X variables by considering two entries in the matrix:

1. In the upper left corner of $\mathbf{X}'\mathbf{X}$ we have:

$$\sum (X_{i1}^*)^2 = \sum \left(\frac{X_{i1} - \bar{X}_1}{\sqrt{n-1}\, s_1} \right)^2 = \frac{\sum (X_{i1} - \bar{X}_1)^2}{n-1} \div s_1^2 = 1$$

2. In the first row, second column of $\mathbf{X}'\mathbf{X}$, we have:

$$\begin{aligned} \sum X_{i1}^* X_{i2}^* &= \sum \left(\frac{X_{i1} - \bar{X}_1}{\sqrt{n-1}\, s_1} \right) \left(\frac{X_{i2} - \bar{X}_2}{\sqrt{n-1}\, s_2} \right) \\ &= \frac{1}{n-1} \frac{\sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{s_1 s_2} \\ &= \frac{\sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\left[\sum (X_{i1} - \bar{X}_1)^2 \sum (X_{i2} - \bar{X}_2)^2 \right]^{1/2}} \end{aligned}$$

But this equals r_{12}, the coefficient of correlation between X_1 and X_2, by (2.84). ■

Estimated Standardized Regression Coefficients

The least squares normal equations (6.24) for the ordinary multiple regression model:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

and the least squares estimators (6.25):

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

can be expressed simply for the transformed variables. It can be shown that for the transformed variables, $\mathbf{X}'\mathbf{Y}$ becomes:

$$\underset{(p-1)\times 1}{\mathbf{X}'\mathbf{Y}} = \mathbf{r}_{YX} \tag{7.51}$$

where $\mathbf{r}_{YX}$ is defined in (7.48) as the vector of the coefficients of simple correlation between Y and each X variable. It now follows from (7.50) and (7.51) that the least squares normal equations and estimators of the regression coefficients of the standardized regression model (7.45) are as follows:

$$\mathbf{r}_{XX}\mathbf{b} = \mathbf{r}_{YX} \tag{7.52a}$$

$$\mathbf{b} = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX} \tag{7.52b}$$

where:

$$\underset{(p-1)\times 1}{\mathbf{b}} = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_{p-1}^* \end{bmatrix} \tag{7.52c}$$

The regression coefficients $b_1^*, \ldots, b_{p-1}^*$ are often called *standardized regression coefficients*.

The return to the estimated regression coefficients for regression model (6.7) in the original variables is accomplished by employing the relations:

$$b_k = \left(\frac{s_Y}{s_k}\right) b_k^* \qquad (k = 1, \ldots, p-1) \tag{7.53a}$$

$$b_0 = \bar{Y} - b_1\bar{X}_1 - \cdots - b_{p-1}\bar{X}_{p-1} \tag{7.53b}$$

Comment

When there are two X variables in the regression model, i.e., when $p - 1 = 2$, we can readily see the algebraic form of the standardized regression coefficients. We have:

$$\mathbf{r}_{XX} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \tag{7.54a}$$

$$\mathbf{r}_{YX} = \begin{bmatrix} r_{Y1} \\ r_{Y2} \end{bmatrix} \tag{7.54b}$$

$$\mathbf{r}_{XX}^{-1} = \frac{1}{1 - r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \tag{7.54c}$$

Hence, by (7.52b) we obtain:

$$\mathbf{b} = \frac{1}{1 - r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \begin{bmatrix} r_{Y1} \\ r_{Y2} \end{bmatrix} = \frac{1}{1 - r_{12}^2} \begin{bmatrix} r_{Y1} - r_{12}r_{Y2} \\ r_{Y2} - r_{12}r_{Y1} \end{bmatrix} \tag{7.55}$$

Thus:

$$b_1^* = \frac{r_{Y1} - r_{12}r_{Y2}}{1 - r_{12}^2} \tag{7.55a}$$

$$b_2^* = \frac{r_{Y2} - r_{12}r_{Y1}}{1 - r_{12}^2} \tag{7.55b}$$

■

Example

Table 7.5a repeats a portion of the original data for the Dwaine Studios example in Figure 6.5b, and Table 7.5b contains the data transformed according to the correlation transformation (7.44). We illustrate the calculation of the transformed data for the first case, using the means and standard deviations in Table 7.5a (differences in the last digit of the transformed data are due to rounding effects):

$$Y_1^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_1 - \bar{Y}}{s_Y}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{174.4 - 181.90}{36.191}\right) = -.04634$$

$$X_{11}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{11} - \bar{X}_1}{s_1}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{68.5 - 62.019}{18.620}\right) = .07783$$

$$X_{12}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{12} - \bar{X}_2}{s_2}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{16.7 - 17.143}{.97035}\right) = -.10208$$