

9-8 ANALYSIS OF VARIANCE TABLE AND AN F TEST OF THE REGRESSION MODEL:

We know from our discussion of the t test for the existence of a linear relationship that the degrees of freedom for *error* in simple linear regression are $n - 2$. For the *regression*, we have 1 degree of freedom because there is one independent X variable in the regression. The *total* degrees of freedom are $n - 1$ because here we only consider the mean of Y , to which one degree of freedom is lost. These are similar to the degrees of freedom for ANOVA in the last chapter. Mean squares are obtained, as usual, by dividing the sums of squares by their corresponding degrees of freedom. This gives us the mean square regression, MSR, and mean square error (which we encountered earlier), MSE. Further dividing MSR by MSE gives us an F ratio with degrees of freedom 1 and $n - 2$. All these can be put in an ANOVA table for regression. This has been done in Table 9-4.

In regression, there are three sources of variation (see Figure 9-19). These are: *regression*—the explained variation; *error*—the unexplained variation; and their sum, the *total* variation. We know how to obtain the sums of squares and the degrees of freedom, and from them the mean squares. Dividing the mean square regression by the mean square error should give us another measure of the accuracy of our regression because MSR is the average squared explained deviation and MSE is the average squared error (where averaging is done using the appropriate degrees of freedom). The ratio of the two has an F distribution with 1 and $n - 2$ degrees of freedom *when there is no regression relationship between X and Y* . This suggests an F test for the existence of a linear relationship between X and Y . *In simple linear regression, this test is equivalent to the t test*. In multiple regression, as we will see in the next chapter, the F test serves a general role, and separate t tests are used to evaluate the significance of different variables. In simple linear regression, we may conduct either an F test or a t test; the results of the two tests will be the same. The

TABLE 9-4 ANOVA Table for Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Regression	SSR	1	$MSR = \frac{SSR}{1}$	$F_{(1, n-2)} = \frac{MSR}{MSE}$
Error	SSE	$n - 2$	$MSE = \frac{SSE}{n - 2}$	
Total	SST	$n - 1$		

TABLE 9-5 ANOVA Table for American Express Example

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Regression	64,527,736.8	1	64,527,736.8	637.47
Error	2,328,161.2	23	101,224.4	
Total	66,855,898.0	24		

hypothesis test is as given in Equation 9-25; the test is carried on the right-hand tail of the F distribution with 1 and $n - 2$ degrees of freedom. We illustrate the analysis with data from Example (a). The ANOVA results are given in Table 9-5.

To carry out the test for the existence of a linear relationship between miles traveled and dollars charged on the card, we compare the computed F ratio of 637.47 with a critical point of the F distribution with 1 degree of freedom for the numerator and 23 degrees of freedom for the denominator. Using $\alpha = 0.01$, the critical point from Appendix C, Table 5 is found to be 7.88. Clearly, the computed value is far in the rejection region, and the p -value is very small. We conclude, again, that there is evidence of a linear relationship between the two variables.

Recall from Chapter 7 that an F distribution with 1 degree of freedom for the numerator and k degrees of freedom for the denominator is the *square* of a t distribution with k degrees of freedom. In Example (a), our computed F statistic value is 637.47, which is the square of our obtained t statistic, 25.25 (to within rounding error). The same relationship holds for the critical points: for $\alpha = 0.01$, we have a critical point for $F_{(1,23)}$ equal to 7.88, and the (right-hand) critical point of a two-tailed test at $\alpha = 0.01$ for t with 23 degrees of freedom is $2.807 = \sqrt{7.88}$.