Chapter 7

The Multiple Regression Model

- When we turn an economic model with more than one explanatory variable into its corresponding statistical model, we refer to it as a *multiple regression model*.
- Most of the results we developed for the simple regression model in Chapters 3–6 can be extended naturally to this general case. There are slight changes in the interpretation of the β parameters, the degrees of freedom for the *t*-distribution will change, and we will need to modify the assumption concerning the characteristics of the explanatory (x) variables.

7.1 Model Specification and the Data

7.1.1 The Economic Model

- Each week the management of a Bay Area Rapid Food hamburger chain must decide how much money should be spent on advertising their products, and what specials (lower prices) should be introduced for that week.
- How does total revenue change as the level of advertising expenditure changes? Does an increase in advertising expenditure lead to an increase in total revenue? If so, is the increase in total revenue sufficient to justify the increased advertising expenditure?
- Management is also interested in pricing strategy. Will reducing prices lead to an increase or decrease in total revenue? If a reduction in price leads a decrease in total revenue then demand is price inelastic; If a price reduction leads to an increase in total revenue then demand is price elastic.

- We initially hypothesize that total revenue, *tr*, is linearly related to price, *p*, and advertising expenditure, *a*. Thus the economic model is:

$$tr = \beta_1 + \beta_2 p + \beta_3 a \tag{7.1.1}$$

where *tr* represents total revenue for a given week, *p* represents price in that week and *a* is the level of advertising expenditure during that week. Both *tr* and *a* are measured in terms of thousands of dollars.

- Let us assume that management has constructed a single weekly price series, *p*, measured in dollars and cents, that describes overall prices.
- The remaining items in (7.1.1) are the unknown parameters β_1, β_2 and β_3 that describe the dependence of revenue (*tr*) on price (*p*) and advertising (*a*).

- In the multiple regression model the intercept parameter, β_1, is the value of the dependent variable when each of the independent, explanatory variables takes the value zero. In many cases this parameter has no clear economic interpretation, but it is almost always included in the regression model. It helps in the overall estimation of the model and in prediction.
- The other parameters in the model measure the change in the value of the dependent variable given a unit change in an explanatory variable, *all other variables held constant.* For example, in (7.1.1)

β_2 = the change in tr ($1000) when p is increased by one unit ($1), and a is held constant, or

$$\beta_2 = \left.\frac{\Delta tr}{\Delta p}\right|_{(a \text{ held constant})} = \frac{\partial tr}{\partial p}$$

- The sign of β_2 could be positive or negative. If an increase in price leads to an increase in revenue, then $\beta_2 > 0$, and the demand for the chain's products is price inelastic. Conversely, a price elastic demand exists if an increase in price leads to a decline in revenue, in which case $\beta_2 < 0$.
- The parameter β_3 describes the response of revenue to a change in the level of advertising expenditure. That is,

β_3 = the change in *tr* ($1000) when *a* is increased by one unit ($1000), and *p* is held constant

$$\beta_3 = \frac{\Delta tr}{\Delta a}_{(p \text{ held constant})} = \frac{\partial tr}{\partial a}$$

- We expect the sign of β_3 to be positive.

7.1.2 The Econometric Model

- The economic model (7.1.1) describes the expected behavior of many individual franchises. As such we should write it as $E(tr) = \beta_1 + \beta_2 p + \beta_3 a$, where $E(tr)$ is the "expected value" of total revenue.
- Weekly data for total revenue, price and advertising will not follow a exact linear relationship. The equation 7.1.1 describes, not a line as in Chapters 3-6, but a *plane*.
- The plane intersects the vertical axis at β_1. The parameters β_2 and β_3 measure the slope of the plane in the directions of the "price axis" and the "advertising axis," respectively.

[Figure 7.1 goes here]

- To allow for a difference between observable total revenue and the expected value of total revenue we add a *random error term*, $e = tr - E(tr)$.

- This random error represents all the factors that cause weekly total revenue to differ from its expected value. These factors might include the weather, the behavior of competitors, a new Surgeon General's report on the deadly effects of fat intake, etc.
- Denoting the t'th weekly observation by the subscript t, we have

$$tr_t = E(tr_t) + e_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + e_t \tag{7.1.2}$$

7.1.2a The General Model

- In a general multiple regression model a dependent variable y_t is related to a number of *explanatory variables* $x_{t2}, x_{t3}, \ldots, x_{tK}$ through a linear equation that can be written as

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \ldots + \beta_K x_{tK} + e_t \qquad (7.1.3)$$

- The coefficients β_1, β_2,…, β_K are unknown parameters.
- The parameter β_k measures the effect of a change in the variable x_{tk} upon the expected value of y_t, $E(y_t)$, all other variables held constant.
- The parameter β_1 is the intercept term. The "variable" $x_{t1} = 1$.
- The equation for total revenue can be viewed as a special case of (7.1.3) where $K = 3$, $y_t = tr_t$, $x_{t1} = 1$, $x_{t2} = p_t$ and $x_{t3} = a_t$. Thus we rewrite (7.1.2) as

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t \qquad (7.1.4)$$

7.1.2b The Assumptions of the Model

To make the statistical model in (7.1.4) complete, assumptions about the probability distribution of the random errors, e_t, need to be made.

1. $E[e_t] = 0$. Each random error has a probability distribution with zero mean. We are asserting that our model is on average correct.
2. $\text{var}(e_t) = \sigma^2$. The variance σ^2 is an unknown parameter and it measures the uncertainty in the statistical model. It is the same for each observation. Errors with this property are said to be *homoskedastic*.
3. $\text{cov}(e_t, e_s) = 0$. The covariance between the two random errors corresponding to any two different observations is zero. Thus, any pair of errors is uncorrelated.
4. We will sometimes further assume that the random errors e_t have normal probability distributions. That is, $e_t \sim N\left(0, \sigma^2\right)$.

The statistical properties of y_t follow from those of e_t.

1. $E(y_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}$. This assumption says that the average value of y_t changes for each observation and is given by the *regression function* $E(y_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}$.
2. $\text{var}(y_t) = \text{var}(e_t) = \sigma^2$.
3. $\text{cov}(y_t, y_s) = \text{cov}(e_t, e_s) = 0$.
4. $y_t \sim N\left[(\beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}), \sigma^2\right]$ is equivalent to assuming that $e_t \sim N\left(0, \sigma^2\right)$.

- In addition to the above assumptions about the error term (and hence about the dependent variable), we make two assumptions about the explanatory variables.
- The first is that the explanatory variables are not random variables.
- The second assumption is that any one of the explanatory variables is not an exact linear function of any of the others. This assumption is equivalent to assuming that no variable is redundant. As we will see, if this assumption is violated, a condition called "exact multicollinearity," the least squares procedure fails.

Assumptions of the Multiple Regression Model

MR1. $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + e_t,\ t = 1, \ldots, T$

MR2. $E(y_t) = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} \Leftrightarrow E(e_t) = 0$.

MR3. $\text{var}(y_t) = \text{var}(e_t) = \sigma^2$.

MR4. $\text{cov}(y_t, y_s) = \text{cov}(e_t, e_s) = 0$

MR5. The values of x_{tk} are not random and are not exact linear functions of the other explanatory variables.

MR6. $y_t \sim N\left[(\beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK}), \sigma^2\right] \Leftrightarrow e_t \sim N(0, \sigma^2)$,

7.2 Estimating the Parameters of the Multiple Regression Model

We will discuss estimation in the context of the model in equation 7.1.4, which is

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t \tag{7.2.1}$$

7.2.1 Least Squares Estimation Procedure

- With the least squares principle we minimize the sum of squared differences between the observed values of y_t and its expected value $E[y_t] = \beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3$.
- Mathematically, we minimize the sum of squares function $S(\beta_1, \beta_2, \beta_3)$, which is a function of the unknown parameters, given the data,

$$
\begin{aligned}
S(\beta_1, \beta_2, \beta_3) &= \sum_{t=1}^{T} \left(y_t - E[y_t]\right)^2 \\
&= \sum_{t=1}^{T} \left(y_t - \beta_1 - \beta_2 x_{t2} - \beta_3 x_{t3}\right)^2
\end{aligned}
\qquad (7.2.2)
$$

- Given the sample observations y_t, minimizing the sum of squares function is a straightforward exercise in calculus.
- In order to give expressions for the least squares estimates it is convenient to express each of the variables as deviations from their means. That is, let

$$
y_t^* = y_t - \bar{y}, \quad x_{t2}^* = x_{t2} - \bar{x}_2, \quad x_{t3}^* = x_{t3} - \bar{x}_3
$$

- The least squares estimates b_1, b_2 and b_3 are:

$$b_1 = \bar{y} - b_2\bar{x}_2 - b_3\bar{x}_3$$

$$b_2 = \frac{\left(\sum y_t^* x_{t2}^*\right)\left(\sum x_{t3}^{*2}\right) - \left(\sum y_t^* x_{t3}^*\right)\left(\sum x_{t2}^* x_{t3}^*\right)}{\left(\sum x_{t2}^{*2}\right)\left(\sum x_{t3}^{*2}\right) - \left(\sum x_{t2}^* x_{t3}^*\right)^2} \tag{7.2.3}$$

$$b_3 = \frac{\left(\sum y_t^* x_{t3}^*\right)\left(\sum x_{t2}^{*2}\right) - \left(\sum y_t^* x_{t2}^*\right)\left(\sum x_{t3}^* x_{t2}^*\right)}{\left(\sum x_{t2}^{*2}\right)\left(\sum x_{t3}^{*2}\right) - \left(\sum x_{t2}^* x_{t3}^*\right)^2}$$

- Looked at as a general way to use sample data, the formulas in (7.2.3) are referred to as estimation rules or procedures and are called the ***least squares estimators*** of the unknown parameters.

- Since their values are not known until the data are observed and the estimates calculated, the ***least squares estimators are random variables***.
- When applied to a specific sample of data, the rules produce the least squares estimates, which are numeric values.

7.2.2 Least Squares Estimates Using Hamburger Chain Data

For Bay Area Rapid Food data we obtain the following least squares estimates:

$$\begin{aligned} b_1 &= 104.79 \\ b_2 &= -6.642 \\ b_3 &= \ \ 2.984 \end{aligned} \qquad \text{(R7.1)}$$

The regression function that we are estimating is

$$E\left[y_t\right] = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} \qquad (7.2.4)$$

The fitted regression line is

$$\begin{aligned}\hat{y}_t &= b_1 + b_2 x_{t2} + b_3 x_{t3} \\ &= 104.79 - 6.642 x_{t2} + 2.984 x_{t3}\end{aligned} \tag{R7.2}$$

In terms of the original economic variables

$$\widehat{tr}_t = 104.79 - 6.642\, p_t + 2.984 a_t \tag{R7.3}$$

Based on these results, what can we say?

1. The negative coefficient of p_t suggests that demand is price elastic and we estimate that an increase in price of $1 will lead to a fall in weekly revenue of $6,642. Or, stated positively, a reduction in price of $1 will lead to an increase in revenue of $6,642.
2. The coefficient of advertising is positive, and we estimate that an increase in advertising expenditure of $1,000 will lead to an increase in total revenue of $2,984.
3. The estimated intercept implies that if both price and advertising expenditure were zero the total revenue earned would be $104,790. This is obviously not correct. In this model, as in many others, the intercept is included in the model for mathematical completeness and to improve the model's predictive ability.

- The estimated equation can also be used for prediction. Suppose management is interested in predicting total revenue for a price of \$2 and an advertising expenditure of \$10,000. This prediction is given by

$$\begin{aligned}\hat{tr}_t &= 104.785 - 6.6419(2) + 2.9843(10) \\ &= 121.34\end{aligned} \qquad \text{(R7.4)}$$

Thus, the predicted value of total revenue for the specified values of p and a is approximately \$121,340.

Remark: A word of caution is in order about interpreting regression results. The negative sign attached to price implies that reducing the price will increase total revenue. If taken literally, why should we not keep reducing the price to zero? Obviously that would not keep increasing total revenue. This makes the following important point: estimated regression models describe the relationship between the economic variables for values *similar* to those found in the sample data. Extrapolating the results to extreme values is generally not a good idea. In general, predicting the value of the dependent variable for values of the explanatory variables far from the sample values invites disaster.

7.2.3 Estimation of the Error Variance σ^2

- The least squares residuals for the model in equation 7.2.1 are:

$$\hat{e}_t = y_t - \hat{y}_t = y_t - \left(b_1 + x_{t2}b_2 + x_{t3}b_3\right) \tag{7.2.5}$$

One estimator of σ^2, and the one we will use, is

$$\hat{\sigma}^2 = \frac{\Sigma \hat{e}_t^2}{T - K} \tag{7.2.6}$$

where K is the number of parameters being estimated in the multiple regression model.

- In the hamburger chain example we have $K = 3$. The estimate for our sample of data in Table 7.1 is

$$\hat{\sigma}^2 = \frac{\Sigma \hat{e}_t^2}{T-K} = \frac{1805.168}{52-3} = 36.84 \tag{R7.5}$$

7.3 Sampling Properties of the Least Squares Estimator

- The sampling properties of a least squares estimator tell us how the estimates vary from sample to sample.

The Gauss-Markov Theorem: For the multiple regression model, if assumptions MR1-MR5 hold, then the least squares estimators are the Best Linear Unbiased Estimators (BLUE) of the parameters in a multiple regression model.

- If we are able to assume that the errors are *normally distributed*, then y_t will also be a normally distributed random variable.
- The least squares estimators will also have normal probability distributions, since they are linear functions of y_t.

- If the errors are not normally distributed, then the least squares estimators are approximately normally distributed in large samples, in which $T – K$ is greater than, perhaps, 50.

7.3.1 The Variances and Covariances of the Least Squares Estimators

- Since the least squares estimators are unbiased, the smaller their variances the higher is the probability that they will produce estimates “near” the true parameter values.
- For $K = 3$ we can express the variances and covariances in an algebraic form that provides useful insights into the behavior of the least squares estimator. For example, we can show that:

$$\operatorname{var}(b_2) = \frac{\sigma^2}{\sum\left(x_{t2} - \bar{x}_2\right)(1 - r_{23}^2)} \tag{7.3.1}$$

Where r_{23} is the sample correlation coefficient between the T values of x_{t2} and x_{t3},

$$r_{23} = \frac{\sum \left(x_{t2} - \overline{x}_2\right)\left(x_{t3} - \overline{x}_3\right)}{\sqrt{\sum \left(x_{t2} - \overline{x}_2\right)^2 \sum \left(x_{t3} - \overline{x}_3\right)^2}} \tag{7.3.2}$$

For the other variances and covariances there are formulas of a similar nature. It is important to understand the factors affecting the variance of b_2:

1. The larger σ^2 the larger the variance of the least squares estimators. This is to be expected since σ^2 measures the overall uncertainty in the model specification. If σ^2 is large, then data values may be widely spread about the regression function $E\left[y_t\right] = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}$ and there is less information in the data about the parameter values.

2. The sample size T the smaller the variances. The sum in the denominator is

$$\sum_{t=1}^{T}(x_{t2}-\bar{x}_2)^2$$

The larger is the sample size T the larger is this sum and thus the smaller is the variance. <u>More observations yield more precise parameter estimation.</u>

3. In order to estimate β_2 precisely we would like there to be a large amount of variation in x_{t2}, $\sum_{t=1}^{T}(x_{t2}-\bar{x}_2)^2$. The intuition here is that it is easier to measure β_2, the change in y we expect given a change in x_2, the more sample variation (change) in the values of x_2 that we observe.

4. In the denominator of var(b_2) is the term $1-r_{23}^2$, where r_{23} is the correlation between the sample values of x_{t2} and $x_{t3.}$ Recall that the correlation coefficient measures the linear association between two variables. If the values of x_{t2} and x_{t3} are correlated then $1-r_{23}^2$ is a fraction that is less than 1.

- The larger the correlation between x_{t2} and x_{t3} the larger is the variance of the least squares estimator b_2. The reason for this fact is that variation in x_{t2} adds most to the precision of estimation when it is not connected to variation in the other explanatory variables.

- “Independent” variables ideally exhibit variation that is “independent” of the variation in other explanatory variables. When the variation in one explanatory variable is connected to variation in another explanatory variable, it is difficult to disentangle their separate effects. In Chapter 8 we discuss “multicollinearity,” which is the situation when the independent variables are correlated with one another. Multicollinearity leads to increased variances of the least squares estimators.

- It is customary to arrange the estimated variances and covariances of the least squares estimators in an array, or matrix, with variances on the diagonal and covariances in the off-diagonal positions. For the K=3 model the arrangement is

$$\operatorname{cov}(b_1, b_2, b_3) = \begin{bmatrix} \operatorname{var}(b_1) & \operatorname{cov}(b_1, b_2) & \operatorname{cov}(b_1, b_3) \\ \operatorname{cov}(b_1, b_2) & \operatorname{var}(b_2) & \operatorname{cov}(b_2, b_3) \\ \operatorname{cov}(b_1, b_3) & \operatorname{cov}(b_2, b_3) & \operatorname{var}(b_3) \end{bmatrix} \tag{7.3.3}$$

Using the estimate $\hat{\sigma}^2$ = 36.84 the estimated variances and covariances for b_1, b_2, b_3, in the Bay Area Rapid Food Burger example are

$$\widehat{\text{cov}}(b_1, b_2, b_3) = \begin{bmatrix} 42.026 & -19.863 & -0.16111 \\ -19.863 & 10.184 & -0.05402 \\ -0.16111 & -0.05402 & 0.02787 \end{bmatrix} \tag{7.3.4}$$

Thus, we have

$$\begin{aligned} \widehat{\text{var}}(b_1) &= 42.026 & \widehat{\text{cov}}(b_1, b_2) &= -19.863 \\ \widehat{\text{var}}(b_2) &= 10.184 & \widehat{\text{cov}}(b_1, b_3) &= -0.16111 \\ \widehat{\text{var}}(b_3) &= 0.02787 & \widehat{\text{cov}}(b_2, b_3) &= -0.05402 \end{aligned}$$

7.3.2 The Properties of the Least Squares Estimators Assuming Normally Distributed Errors

If we add assumption MR6, that the random errors e_t have normal probability distributions, then the dependent variable y_t is normally distributed,

$$y_t \sim N\left[(\beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK}), \sigma^2\right] \Leftrightarrow e_t \sim N(0, \sigma^2)$$

Since the least squares estimators are linear functions of dependent variables it follows that the least squares estimators are normally distributed also,

$$b_k \sim N\left[\beta_k, \operatorname{var}(b_k)\right] \tag{7.3.6}$$

- We can transform the normal random variable b_k into the <u>standard normal variable</u> z,

$$z = \frac{b_k - \beta_k}{\sqrt{\operatorname{var}(b_k)}} \sim N(0,1), \text{ for } k = 1, 2, \ldots, K \tag{7.3.7}$$

- When we replace σ^2 by its estimator $\hat{\sigma}^2$, we obtain the estimated var(b_k) which we denote as $\widehat{\operatorname{var}}(b_k)$. When var($b_k$) is replaced by $\widehat{\operatorname{var}}(b_k)$ in (7.3.7) we obtain a t-random variable instead of the normal variable. That is,

$$t = \frac{b_k - \beta_k}{\sqrt{\widehat{\operatorname{var}}(b_k)}} \sim t_{(T-K)} \tag{7.3.8}$$

- In this chapter there are K unknown coefficients in the general model and *the number of degrees of freedom for t-statistics is* (T K).

- The square root of the variance estimator $\hat{\text{var}}\left(b_k\right)$ is called the "standard error" of b_k, which is written as

$$\text{se}(b_k) = \sqrt{\hat{\text{var}}(b_k)} \tag{7.3.9}$$

Consequently, we will usually express the t random variable as

$$t = \frac{b_k - \beta_k}{\text{se}(b_k)} \sim t_{(T-K)} \tag{7.3.10}$$

7.4 Interval Estimation

- Interval estimates of unknown parameters are based on the probability statement that

$$P\left(-t_c \le \frac{b_k - \beta_k}{\mathrm{se}(b_k)} \le t_c\right) = 1 - \alpha \qquad (7.4.1)$$

Where t_c is the critical value for the t-distribution with $(T–K)$ degrees of freedom, such that $P(t \ge t_c) = \alpha/2$.

- Rearranging equation 7.4.1 we obtain

$$P\left[b_k - t_c\mathrm{se}(b_k) \le \beta_k \le b_k + t_c\mathrm{se}(b_k)\right] = 1 - \alpha \qquad (7.4.2)$$

- The interval endpoints, define a 100(1-α)% confidence interval estimator of β_k.

$$[b_k - t_c \mathrm{se}(b_k),\ b_k + t_c \mathrm{se}(b_k)] \tag{7.4.3}$$

- If this interval estimator is used in many samples from the population, then 95% of them will contain the true parameter β_k.
- Returning to the equation used to describe how the hamburger chain's revenue depends on price and advertising expenditure, we have

$$\begin{array}{ll} T = 52 & K = 3 \\ b_1 = 104.79 & \mathrm{se}(b_1) = \sqrt{\widehat{\mathrm{var}}(b_1)} = 6.483 \\ b_2 = -6.642 & \mathrm{se}(b_2) = \sqrt{\widehat{\mathrm{var}}(b_2)} = 3.191 \\ b_3 = 2.984 & \mathrm{se}(b_3) = \sqrt{\widehat{\mathrm{var}}(b_3)} = 0.1669 \end{array}$$

We will use this information to construct interval estimates for

$$\beta_2 = \text{ the response of revenue to a price change}$$
$$\beta_3 = \text{ the response of revenue to a change in advertising expenditure}$$

- The degrees of freedom are given by (T K) = (52 3) = 49.
- The critical value t_c = 2.01.
- A 95% interval estimate for β_2 is given by

$$(-13.06, -0.23) \qquad (7.4.4)$$

- This interval estimate suggests that decreasing price by \$1 will lead to an increase in revenue somewhere between \$230 and \$13,060.
- This is a wide interval and it is not very informative.

- Another way of describing this situation is to say that the point estimate of b_2 = 6.642 is not very reliable.
- A narrower interval can only be obtained by reducing the variance of the estimator. one way is to obtain more and better data. Alternatively we might introduce some kind of nonsample information on the coefficients.
- The 95% interval estimate for β_3, the response of revenue to advertising is

$$(2.65, 3.32) \tag{7.4.5}$$

- This interval is relatively narrow and informative. We estimate that an increase in advertising expenditure of $1000 leads to an increase in total revenue that is somewhere between $2,650 and $3,320.

7.5 Hypothesis Testing for a Single Coefficient

7.5.1 Testing the Significance of a Single Coefficient

- To find whether the data contain any evidence suggesting y is related to x_k we test the null hypothesis

$$H_0 : \beta_k = 0$$

against the alternative hypothesis

$$H_1 : \beta_k \neq 0$$

- To carry out the test we use the test statistic (7.3.10), which, if the null hypothesis is true, is

$$t = \frac{b_k}{\mathrm{se}\left(b_k\right)} \sim t_{(T-K)}$$

- For the alternative hypothesis “not equal to” we use a two-tailed test, and reject H_0 if the computed t-value is greater than or equal to t_c, or less than or equal to $-t_c$.
- In the Bay Area Burger example we test, following our standard testing format, whether revenue is related to price:

1. $H_0 : \beta_2 = 0$
2. $H_1 : \beta_2 \neq 0$
3. The test statistic, if the null hypothesis is true, is

$$t = \frac{b_2}{\mathrm{se}(b_2)} \sim t_{(T-K)}$$

4. With 49 degrees of freedom and a 5% significance level, the critical values that lead to a probability of 0.025 in each tail of the distribution are $t_c = 2.01$ and $-t_c = -2.01$. Thus we reject the null hypothesis if $t \geq 2.01$ or if $t \leq -2.01$. In shorthand notation, we reject the null hypothesis if $|t| \geq 2.01$.

5. The computed value of the *t*-statistic is

$$t = \frac{-6.642}{3.191} = -2.08$$

Since –2.08 < –2.01 we reject the null hypothesis and accept the alternative.

- The *p*-value in this case is given by $P[|t_{(49)}| > 2.08] = 2 \times 0.021 = 0.042$. Using this procedure we reject H_0 because $0.042 < 0.05$.

- For testing whether revenue is related to advertising expenditure, we have

1. $H_0 : \beta_3 = 0$
2. $H_1 : \beta_3 \neq 0$
3. The test statistic, if the null hypothesis is true, is

$$t = \frac{b_3}{\mathrm{se}\left(b_3\right)} \sim t_{(T-K)}$$

4. We reject the null hypothesis if $|t| \geq 2.01$.

5. The value of the test statistic is

$$t = \frac{2.984}{0.1669} = 17.88$$

Because $17.88 > t_c = 2.01$, the data support the conjecture that revenue is related to advertising expenditure.

7.5.2 One-Tailed Hypothesis Testing for a Single Coefficient

7.5.2a Testing for Elastic Demand

- With respect to demand elasticity we wish to know if:
- $\beta_2 \geq 0$: a decrease in price leads to a decrease in total revenue (demand is price inelastic)
- $\beta_2 < 0$: a decrease in price leads to an increase in total revenue (demand is price elastic)

1. $H_0 : \beta_2 \geq 0$ (demand is unit elastic or inelastic)

2. $H_1 : \beta_2 < 0$ (demand is elastic)

3. To create a test statistic we act as if the null hypothesis were the equality $\beta_2 = 0$. If this null hypothesis is true, then the t-statistic is $t = \dfrac{b_k}{\mathrm{se}(b_k)} \sim t_{(T-K)}$.

4. The rejection region consists of values from the t-distribution that are unlikely to occur if the null hypothesis is true. If we define "unlikely" in terms of a 5% significance level, we answer this question by finding a critical value t_c such that $P[t_{(T-K)} \leq t_c] = 0.05$. Then, we reject H_0 if $t \leq t_c$. Given a sample of T=52 data observations, the degrees of freedom are $T - K = 49$ and the t-critical value is t_c = 1.68.

5. The value of the test statistic is

$$t = \frac{b_2}{\mathrm{se}(b_2)} = \frac{-6.642}{3.191} = -2.08 \qquad \text{(R7.10)}$$

Since $t = -2.08 < t_c = -1.68$, we reject H_0: $\beta_2 \geq 0$ and conclude H_1: $\beta_2 < 0$ (demand is elastic) is more compatible with the data. The sample evidence supports the proposition that a reduction in price will bring about an increase in total revenue.

- The p-value that is given by $P[t_{(49)} < -2.08]$ and to reject H_0 if this p-value is less than 0.05. Using our computer software, we find that $P[t_{(49)} < -2.08] = 0.021$. Since $0.021 < 0.05$, the same conclusion is reached.

- The other hypothesis of interest is whether an increase in advertising expenditure will bring an increase in total revenue that is sufficient to cover the increased cost of advertising. This will occur if $\beta_3 > 1$. Setting up the test, we have:

1. $H_0 : \beta_3 \leq 1$
2. $H_1 : \beta_3 > 1$
3. We compute a value for the t-statistic as if the null hypothesis were $\beta_3 = 1$. Using (7.3.10) we have, if the null hypothesis is true,

$$t = \frac{b_3 - 1}{\mathrm{se}(b_3)} \sim t_{(T-K)}$$

4. In this case, if the level of significance is $\alpha = .05$, we reject H_0 if $t \geq t_c = 1.68$

5. The value of the test statistic is:

$$t = \frac{b_3 - \beta_3}{\mathrm{se}(b_3)} = \frac{2.984 - 1}{0.1669} = 11.89$$

Since 11.89 is much greater than 1.68, we do indeed reject H_0 and accept the alternative $\beta_3 > 1$ as more compatible with the data.

- Also, the *p*-value in this case is essentially zero (less than 10^{-12}). Thus, we have *statistical evidence* that an increase in advertising expenditure will be justified by the increase in revenue.

7.6 Measuring Goodness of Fit

- The coefficient of determination is

$$
\begin{aligned}
R^2 &= \frac{SSR}{SST} = \frac{\Sigma\left(\hat{y}_t - \overline{y}\right)^2}{\Sigma\left(y_t - \overline{y}\right)^2} \\
&= 1 - \frac{SSE}{SST} = 1 - \frac{\Sigma \hat{e}_t^2}{\Sigma\left(y_t - \overline{y}\right)^2}
\end{aligned}
\qquad (7.6.1)
$$

- For the Bay Area Burger example the Analysis of Variance table includes the following information:

Table 7.4 Partial ANOVA Table

Source	DF	Sum of Squares
Explained	2	11776.18
Unexplained	49	1805.168
Total	51	13581.35

- Using these sums of square we have

$$R^2 = 1 - \frac{\Sigma \hat{e}_t^2}{\Sigma\left(y_t - \bar{y}\right)^2} = 1 - \frac{1805.168}{13581.35} = 0.867 \qquad \text{(R7.12)}$$

- The interpretation of R^2 is that 86.7% of the variation in total revenue is explained by the variation in price and by the variation in the level of advertising expenditure.

- One difficulty with R^2 is that it can be made large by adding more and more variables, even if the variables added have no economic justification. Algebraically it is a fact that as variables are added the sum of squared errors *SSE* goes down (it can remain unchanged but this is rare) and thus R^2 goes up. If the model contains T–1 variables, the $R^2 = 1$.
- An alternative measure of goodness-of-fit called the adjusted-R^2, and often symbolized as $\bar{R}^2$, is usually reported by regression programs; it is computed as

$$\bar{R}^2 = 1 - \frac{SSE/(T-K)}{SST/(T-1)}$$

- For the Bay Area Burger data the value of this descriptive measure is $\bar{R}^2 = .8617$.
- This measure does not always go up when a variable is added.
- While solving one problem, this corrected measure of goodness of fit unfortunately introduces another one. It loses its interpretation; $\bar{R}^2$ is no longer the percent of variation explained.

- One final note is in order. The intercept parameter β_1 is the is the *y*-intercept of the regression "plane," as shown in Figure 7.1. If, for theoretical reasons, you are certain that the regression plane passes through the origin, then $\beta_1 = 0$ and can be omitted from the model.
- If the model does not contain an intercept parameter, then the measure R^2 given in (7.6.1) is no longer appropriate. The reason it is no longer appropriate is that, without an intercept term in the model,

$$\Sigma\left(y_t - \overline{y}\right)^2 \neq \Sigma\left(\hat{y}_t - \overline{y}\right)^2 + \Sigma e_t^2$$

$$SST \neq SSR + SSE$$

- Under these circumstances it does not make sense to talk of the proportion of total variation that is explained by the regression.