The ideas from economic theory must be tested and measured according to the standards of real-world data. Statistical techniques applied to economics are collectively called econometrics. In Chapter 2, we discussed supply and demand, as well as the importance of price and income elasticities. Economic theory predicts that demand curves will slope downward, but it does not predict the degree of responsiveness of demand to price and other variables; it is the task of statistical analysis to estimate these magnitudes. When close substitutes are available for a good, theory predicts that demand will be more sensitive to price than if no close substitutes are available. Yet it is hard to know whether a 1 percent increase in price will decrease the quantity demanded of the good by 10 percent, 1 percent, or 1/10 of a percent, yielding elasticities of \(-10.0\), \(-1.0\), or \(-0.1\), respectively. Measurements of the economic behaviors of both people and firms may be crucial in analyzing whether drug companies raise drug prices, whether higher insurance copayments will lead people to use less treatment, or whether mandated levels of health care are economically efficient.

This chapter considers statistical methods that econometricians use to draw inferences from data that are collected. Many students with natural science backgrounds are familiar with laboratory experiments, where the environment is held as constant as possible and treatments are administered to experimental groups. The results are then compared to those of untreated control groups. One form of this design is called the dose-response model because the results or responses are generally related to the experimental treatment, or the dose. If statisticians determine the differences are significant, a term we will discuss in this chapter, then the dose is considered to be effective.

Social science analysis of human behavior is rarely so fortunate as to find an experimental group that can be matched with a convenient control group. Social scientists, economists among them, must usually collect
information from people doing day-to-day activities. Using statistical methods, they try to control for the confounding differences among the people that they are analyzing. The more successful they are in controlling for such differences, the more reliable the analysis will be.

This chapter begins with discussions on how we form hypotheses. It then considers difference of means analysis as a way of introducing statistical inference. Most of the rest of the chapter concentrates on simple regression and multiple regression analyses that are most often used in economic and econometric analysis.

**HYPOTHESIS TESTING**

Economists who study health care have been confronted on occasion by statements that, while plausible, demand some validation:


demands some validation:

“Men and women don’t smoke the same numbers of cigarettes.”
“Rich people spend more on health care than do poor people.”
“Monopolists make more money than competitive firms.”

These are all statements that either logic or casual observation would suggest to be true.

It would be useful, however, to have a rigorous method of determining whether the assertions are correct. Statistical methods suggest formulating these statements as hypotheses and collecting data to determine whether they are correct.

Take, for example, the first assertion about smoking levels. We state clearly both the hypothesis we wish to disprove (the null hypothesis), as well as the hypothesis the theory suggests to be the case (the alternative hypothesis). The null hypothesis here, $H_0$, is that men’s levels ($c_m$) equal women’s levels ($c_w$), or

$$H_0: c_m = c_w \quad (3.1)$$

The alternative hypothesis $H_1$ is that $c_m$ does not equal $c_w$:

$$H_1: c_m \neq c_w \quad (3.2)$$

It is necessary to show convincing evidence that $c_m$ differs from $c_w$. Hypotheses that are designed to test for equality among two or more items are sometimes called *simple* hypotheses.

Consider the second hypothesis, which asserts that rich people spend more on health care than do poor people. If we define health care expenditures of the rich as $E_r$ and the poor as $E_p$, then the null hypothesis is:

$$H_0: E_r = E_p \quad (3.3)$$

The alternative is:

$$H_1: E_r > E_p \quad (3.4)$$

In this analysis, it may not be enough just to show that $E_r$ differs from $E_p$. Certainly, even convincing evidence that $E_p$ is greater than $E_r$ would not validate the hypothesis. Hypotheses that are used to test whether two or more items are greater (or less) than each other are sometimes called *composite* hypotheses. Having seen how one might construct the hypotheses in question, we now discuss how to test them.
DIFFERENCE OF MEANS

Return to the hypothesis about men’s and women’s smoking. Smoking is the single most preventable health risk factor in all societies. People have been smoking for centuries. By the mid-1960s it was clear that smoking had numerous adverse public health impacts. Public health initiatives have sought to eliminate, or at least reduce, people’s smoking.

Smoking depends on many factors. Younger people often start in order to look more mature. People living in certain cultures (often where tobacco is raised) smoke more, and men have traditionally smoked more than women (although women appear to be catching up). People with more education may better recognize the health impacts of smoking, and hence smoke less. Economists hypothesize that, like most goods, smoking is negatively related to cigarette price (as price increases, smoking decreases). The impact of income is unclear.

To compare men’s and women’s smoking rates, we can sample the population, yet we know that there are lots of different types of people, and we would like to avoid the confounding influences of age, education, or location. We could attempt to avoid this sort of distortion by drawing samples randomly from among all possible 20-year-olds, called the universe of data. Alternatively, we may try to choose samples of 20-year-old men and women from the same general income group. A sample of college sophomores, for example, from the same location and with similar socioeconomic status, may be a good group for holding many factors constant. Even this example shows how difficult things may be to control. People of the same age (at the same college) may come from different locations and different types of families. Some may have parents or other family members who smoke.

We need a test to determine the differences between two distributions of continuous data. Continuous data are natural measures that in principle could take on different values for each observation. Examples include height, weight, income, or price. Categorical data refer to arbitrary categories such as gender (male or female), race (black, white, or other), or location (urban or rural). In this chapter, unless we specify categorical data, our methods will refer exclusively to the analysis of continuous data.

Lots of people do not smoke, so let us concentrate only on smokers (looking at the decision whether or not to smoke is an important policy question, but is far beyond the scope of this example). The econometrician asks one woman and finds that she smokes 10 cigarettes per day ($c_w = 10$). The first man asked smokes 15 cigarettes per day ($c_m = 15$). This provides evidence that men smoke more than women, because $c_m > c_w$, or 15 > 10. It is not very convincing evidence, however. The man, or woman, or either, may not be typical of the entire group. What if a different man and/or woman had been selected? Would the answer have been different?

It seems logical to test several men and to compute the mean or average level by summing the levels and dividing them by the total number of men tested. The National Institutes of Health (in 2001 and 2002) collected a database of over 43,000 individuals called the National Epidemiologic Survey on Alcohol and Related Conditions, or NESARC. They focused on potentially substance abusive activities including smoking, drinking of alcoholic beverages, and the taking of recreational (and harder) drugs. They asked a number of questions about smoking, and from the analysis of smokers the textbook authors found that

For 4,714 men, the mean, or average level, $\bar{c}_m$, was 15.60 cigarettes per day.

For 4,841 women, the average level, $\bar{c}_w$, was 13.47 cigarettes per day.

The difference, $d = \bar{c}_m - \bar{c}_w$, then, is 2.13 cigarettes per day.

The Variance of a Distribution

Although a difference of the two means is improved evidence, the econometrician desires a more rigorous criterion. It could be that the true level for both men and women is 14 cigarettes per day,
but our sample randomly drew a higher average level for men (15.60) than for women (13.47). Figure 3-1 plots the distributions in percentage terms. Almost 25 percent of the women smoke between 1 and 5 cigarettes per day, compared to a slightly smaller percentage of men; in contrast, while about 28 percent of the women smoke between 16 and 20 cigarettes per day, about 33 percent of the men smoke at this level. Although the mean levels differ (men are higher than women), some groups of women (those who smoke less) have higher percentages than some groups of men. Statisticians have found the variance of a distribution to be a useful way to summarize its dispersion. To calculate the variance of women’s levels, we subtract each observation from the mean (13.47), square that term, sum the total, and divide that total by the number of observations, N. Hence, variance, $V_w$, equals:

$$V_w = \frac{N_1(1 - 13.47)^2 + N_2(2 - 13.47)^2 + \ldots + N_{80}(80 - 13.47)^2}{4,841}$$ (3.5)

Here $N_1$ is the number of women who smoke 1 cigarette per day, $N_2$ is the number who smoke 2 cigarettes per day, and so on (and yes, there are some women who smoke 80 cigarettes per day!). $V_w$ reflects the variance of any individual term in the distribution. If $V$ is large, then the dispersion around the mean is wide and another woman tested might be far from our mean. If $V$ is small, then the dispersion around the mean is narrow and another observation might be close to the mean.

**Standard Error of the Mean**

The variance is often deflated by taking the square root to get the standard deviation, $s$, yielding:

$$s_w = \sqrt{\frac{N_1(1 - 13.47)^2 + N_2(2 - 13.47)^2 + \ldots + N_{80}(80 - 13.47)^2}{4,841}}$$ (3.6)

As with $V$, a large (small) value of $s$ indicates a large (small) dispersion around the mean. Statisticians have shown that we can calculate the standard error of the mean itself by dividing $s$ by the square root of the number of observations. In this sample, the standard deviation of the distribution for...
women equals 9.71. The standard error of the mean of the women’s distribution would then equal \( s_w \), divided by the square root of 4,841, or \((9.71 \div 69.6)\), which equals 0.14.¹

A powerful theorem in statistics, the Central Limit Theorem, states that no matter what the underlying distribution, the means of that distribution are distributed like a normal, or bell-shaped, curve. Hence, we can plot the normal distribution of means of women’s levels with a mean of 13.47 and a standard error of 0.14.

Statisticians have also shown that a little more than 68 percent of the area under the curve would be within one standard error, or between levels of 13.33 (that is, 13.47 – 0.14) and 13.61 (i.e., 13.47 + 0.14), and that 95.4 percent would be within two standard errors. This means that we could be about 95 percent sure that the true mean quantity of cigarettes smoked for women was between 13.19, [that is, 13.47 – (2 × 0.14)], and 13.75, [i.e., 13.47 + (2 × 0.14)]. A similar calculation can be done for men, yielding a similar measurement. Intuitively, the further apart the means and the smaller the dispersions (standard errors), the more likely we are to determine that the average level for men is smaller than that for women. To test the hypothesis formally, we then construct a “difference of means” test. We wish to compare the measurement \( d = \overline{c}_m - \overline{c}_w \), to zero, which was the original hypothesis.

Here \( d = 2.13 \). The variance of the difference is defined as the sum of the variances of the standard errors. If the standard error for women was 0.14 as we calculated it, and the standard error for men given the sample in Figure 3-2A was 0.17, then the standard error of the difference would be:

\[
s_d = \sqrt{0.14^2 + 0.17^2} = 0.216
\]

The difference and its distribution also can be plotted.

The most probable value of the difference, as noted in Figure 3-2B, is 2.13. About 68 percent of the distribution lies between 1.91 (i.e., 2.13 – 1 × 0.216) and 2.35 (i.e., 2.13 + 1 × 0.216). About 95.4 percent of the distribution lies between 1.69 (i.e., 2.13 – 2 × 0.216) and 2.56 (i.e., 2.13 + 2 × 0.216).

This experiment would find very good evidence that among smokers women smoke fewer cigarettes than men. The males have higher levels than the females, and the probability is well over 95 percent that this difference is statistically significant.

Alternatively, the \( t \)-statistic, comparing the numbers 2.13 and 0.0, equals \( 2.13 \div 0.216 \), or approximately 9.86. Statisticians calculate tables of \( t \)-statistics, whose critical values are related to the size of the sample. With a sample of over 8,000, a \( t \)-statistic of nearly 10 is statistically significant at well over the 99 percent level. In other words, we can be 99+ percent certain that men smoke more cigarettes than women.

### Hypotheses and Inferences

This process illustrates the steps that are necessary to test hypotheses appropriately. The econometrician must:

1. State the hypotheses clearly
   
   \( H_0: c_m = c_w \), against
   
   \( H_1: c_m \neq c_w \).

2. Choose a sample that is suitable to the task of testing.

3. Calculate the appropriate measures of central tendency and dispersion: the mean and the standard error of the mean for both men and women, leading to the difference of the two means.

4. Draw the appropriate inferences: men smoke more than women.

¹ Formally, in a sample (as opposed to the entire population), we calculate the standard error by dividing by \( n - 1 \). All calculations are rounded to the nearest hundredth.
No matter how sophisticated the method used, good statistical analysis depends on the ability to address these four criteria and stands (or falls) on the success in fulfilling them. Box 3-1 provides a particularly good example of how analysts have examined the impacts of electromagnetic fields (EMFs) on children living near power lines.

There are, of course, measures of central tendency other than the mean (or average). Someone who smokes 4 packs per day (80 cigarettes) may unduly influence the mean. A different measure, the median, calculates a statistic such that half of the observations are greater than the median and half are less. Thus, a median smoking level of 15 cigarettes would imply that half of the people smoked more than 15, and half smoked less. The median is less sensitive to extreme values in the data (e.g., someone who smokes 60 cigarettes per day would have no more effect on a median of 15 than does someone who smokes 20). However, the median can present mathematical problems in hypothesis testing. Simple formulas for standard errors of medians have not been available,
although popular numerical “bootstrapping” methods now provide intuitive and accurate standard errors. For a good discussion of bootstrapping, see Effron and Tibshirani (1993).

REGRESSION ANALYSIS

The difference of means analysis is extremely useful in treating continuous data that can be broken up by categories, such as gender, race, or location. Yet many interesting economic variables occur naturally as continuous variables. Health care expenses, physician visits, firm profits, as well as prices and/or incomes could take large numbers of values naturally, and are grouped into categories only with serious loss of information. If we have information on income, in terms of dollars per year, we can distinguish among households with incomes of $10,000, $20,000, $30,000, and so on. If we were to define high income as greater than $40,000, for example, separating all of the different incomes into

---

2 Although difference of means considers only two categories, analysis of variance methods allow the consideration of three or more categories. Newbold, Carlson, and Thorne (2007) present good discussions on this and other statistical topics.
two categories, we would then have no way of distinguishing between households with incomes of $10,000 and $20,000 (or, for that matter, between households with incomes of $50,000 and $100,000).

Regression analysis allows the econometrician to fit a straight line through a set of data points. How might this be used for policy analyses? Economists and health policy makers alike have long sought to determine how responsive cigarette consumption is to excise taxes. Recognizing that cigarette smoking is a dangerous habit, economists have reasoned that if the price is raised by taxes, consumption is likely to decline, but the question is “how much?”

Cigarettes are produced nationally and are subject to the same federal taxes but, as of 2008, excise tax rates vary from $0.07 per pack in South Carolina to over $2.50 per pack in New Jersey. The NESARC smokers, in addition to number of cigarettes smoked, were also asked a wide range of questions focusing on addictive behaviors, including smoking, drinking, legal and illegal drugs, as well as numerous questions about occupation, education, and income. Because the state of residence was known, an analyst could append readily available data on state level excise taxes.

Suppose that we wish to relate the amount of cigarettes smoked per day to the tax price of the cigarettes. Since cigarettes must be purchased in order to be consumed, we would like to know how responsive the quantity demanded is to the tax price (the tax price elasticity). Recall that the price elasticity relates the percent change in quantity to the percent change in price. It would be useful to draw a straight line that would summarize the relationship.

**Ordinary Least Squares (OLS) Regressions**

Two rules are commonly used to determine this line. The first rule is that the deviations (unless the line fits perfectly) from the line must sum to zero. Positive deviations must be offset by negative deviations. We can show, however (see Figure 3-3), that many lines have this characteristic (for example, dashed lines $R_1$ and $R_2$). It is necessary to have a second criterion by which to distinguish among the large number of lines where the sum of the deviations equals zero.

The second criterion is to minimize the sum of the squared deviations of the actual data points from the line that is fitted. Even though the sum of the deviations equals zero, the sum of the squared deviations must be positive (any number multiplied by itself is either zero or positive). Hence, one can choose among the many lines that have sums of zero deviations by picking the one line with the minimum or least sum of the squared deviations. Such analyses are called ordinary least squares (OLS) analyses.
The resulting equation would have the following form:

\[ Q = a + bP + \varepsilon \]  
(3.8)

where \( P \) and \( Q \) refer to price and quantity, and \( a \) and \( b \) are the parameters to be estimated. Parameter \( a \) is sometimes referred to as the constant, or the intercept. It might refer to the demand for \( Q \) in the unlikely event that the tax price of cigarettes was zero.

Parameter \( b \) refers to the slope of the line and shows the direction and magnitude of the impact of a change in \( P \) on the quantity demanded. Because we expect a higher level of \( P \) to decrease the amount of cigarettes purchased (assuming that the subjects purchase the cigarettes that they smoke), we expect \( b \) to be negative.

The last parameter is the error term \( \varepsilon \). No regression analysis will fit the data exactly. Errors are likely and may reflect several causes. We may have omitted a variable, such as age (older people may smoke more) or education (more educated people may be more aware of the dangers of smoking and smoke less). We may have measured one or more of the explanatory variables or the dependent variable (the amount of cigarettes) inaccurately. All of these may stand in the way of our predicting the amount demanded exactly. In advanced econometric work, understanding \( \varepsilon \) is crucial for ensuring that the estimated parameters are accurate. Our exposition here will assume that \( \varepsilon \) obeys the rules to allow us to make appropriate inferences with OLS analyses. We will examine some exceptions later in the chapter.

**A Demand Regression**

Table 3-1a shows the result of a simple regression of cigarette consumption against the tax price of cigarettes.

\[ Q = 16.83 - 3.24 \times \text{tax per pack}, \quad R^2 = 0.01 \]  
(3.9)

This equation indicates that a $1 increase in the tax per pack of cigarettes is correlated with a decrease in quantity demand of 3.24 fewer cigarettes per day among those who smoke. The *standard error of...*
estimate for the coefficient of tax is 0.34. This term is similar to the standard error of the estimated mean in the example of cigarette smoking in men and women earlier in the chapter. As before, the smaller the standard error is relative to the estimated value of \( b \) (in this case, \(-3.24\)), the better the estimate. In this regression, the standard error of 0.34 is relatively small compared to the coefficient, \(-3.24\); hence, the coefficient is significantly different from zero. The expression \( R^2 \) is used to measure the fraction of the variation of the quantity of cigarettes explained by the price alone. An \( R^2 \) of 0.01 implies that about 1 percent of the variation was explained.

It is useful to examine this simple regression in detail because it has many features that occur in more complex analyses. Consider the following hypothesis:

- \( H_0 \): Tax price doesn’t matter; that is, \( b = 0 \) against the alternative hypothesis, which is:
- \( H_1 \): Tax price is inversely related to quantity consumed; that is, \( b < 0 \). The test of the hypothesis is similar to a difference of means test. In particular, we are testing the difference between \(-3.24\) (estimated with standard error 0.34) and 0.

Remember that because demand is downward sloping, the coefficient will be negative. The \( t \)-statistic here is 9.5; that is, the value of the coefficient, 3.24, divided by the standard error of 0.34. The value of 9.5 suggests that we can be more than 99 percent sure that the tax price has an effect on quantity of cigarettes consumed. This term is statistically significant in its difference from zero.

If 1 percent of the variation of the quantity of cigarette demand is explained, then 99 percent is unexplained. In part, this occurs because the regression does not include some variables that are likely to be important. We have noted earlier that there are several other variables that might help explain the consumption of cigarettes. If these are included in the analysis, we are likely to explain more of the variation in cigarette consumption. The inclusion of more variables in a multiple regression is explained later.

This example illustrates cross-sectional analysis, which provides snapshots of a slice of the population at one period in time. Because 2001–2002 was the first time that the NESARC was collected, it could not yet follow the people in the sample over time, and we could not be aware of continuing health problems, changes in wealth or income, or systematic differences in ability that cannot be measured and that cross-sectional models treat as “random noise.” As a result, cross-sectional regressions often explain less variance than panel data, which follow households over time, or time-series data, which calculate aggregates over time.

**Estimating Elasticities**

Regressions also are used to estimate elasticities. Recall that the definition of the price elasticity of demand \( (E_p) \) is the percentage change in quantity demanded, elicited by a 1 percent change in price. This is written as

\[
E_p = \frac{\text{% change in quantity}}{\text{% change in price}} = \frac{\frac{\Delta Q}{Q}}{\frac{\Delta P}{P}}, \quad \text{or} \quad E_p = \left( \frac{\Delta Q}{\Delta P} \right) \left( \frac{P}{Q} \right)
\]

The term \( \Delta P \) stands for a one-unit change in the price, while the term \( \Delta Q \) represents the resulting change in the quantity demanded. It follows that \( \Delta Q/Q \) is a measure of the percentage change in quantity, whereas \( \Delta P/P \) is a measure of the percentage change in price. In rearranging terms at the right, the term \( \Delta Q/\Delta P \) represents the ratio of changes and is the reciprocal of the slope of the demand curve. With the linear function here, this is \(-3.24\).

In calculating an elasticity from the coefficients estimated in a regression, a different elasticity could be calculated for each different starting price that is assumed. Therefore, it is also necessary to have reference values for \( P \) and \( Q \), and the mean (or average) values are often used. In our sample, the mean number of cigarettes smoked per day is 15.3 (about three-quarters of a pack), and the mean tax price is $0.454 (about $0.45 per pack).
Hence, calculated at the mean,

\[ E_p = -3.24 \times (0.45 \div 15.3), \text{ or } -0.10 \]  \hspace{1cm} (3.10)

This says that a 10 percent increase in the tax price of cigarettes would lead to a 1.0 percent decrease in quantity demanded. Does this make sense? Cigarettes, after all, are an addictive substance, and many people find it difficult to reduce their demand. Yet, there is clearly a negative effect, and one could also argue that doubling the tax (say from $.50 per pack to $1 per pack—a 100% increase) could reduce demand by 10%, a sizable amount.\(^3\)

**MULTIPLE REGRESSION ANALYSIS**

Real-world relationships are seldom two-dimensional, as useful as this situation would be in drawing graphs. As noted, demand for cigarettes may be related not only to the price, but to income, \(Y\). Older people (variable \(A\) for age) may smoke more, having been addicted for longer, and more educated people, \(E\), may recognize the dangers of smoking and smoke less. In addition, women have traditionally smoked less, and various groups, variable \(G\), may have differing tastes toward smoking cigarettes. Indeed, the omission of important variables may lead to particular behavior in the error term, \(\varepsilon\).

If each relationship could be graphed, assuming that nothing else was changing, then simple regression would work fine. Fortunately, the mathematics necessary to estimate the appropriate relationship can accommodate more than two dimensions. It is easy to write the following multiple regression:

\[ Q = a + bP + cY + dA + eE + fG + \varepsilon \]  \hspace{1cm} (3.11)

Although the example presented in Table 3-1 will summarize eight dimensions now rather than two dimensions, we use exactly the same least squares criteria as before. The interpretation of the coefficients is similar to before but is done with more confidence. With the simple regression, relating \(Q\) only to \(P\), the econometrician would not know whether income, \(Y\), or age, or education, was varying as well. Including them in this regression allows us to “hold constant” these other variables and reduce the error. As a result, elasticities can now be calculated under the condition that “all else is equal.” The \(R^2\) measure of variation explained earlier also is available here.

\(R^2\) will always rise with more variables. (If you add variables, you can’t explain less of the variation!) Several methods can be used to interpret \(R^2\), and some statisticians wish to maximize \(R^2\); that is, to explain as much variation as possible.\(^4\) While this may be desirable, most econometricians are at least as interested in the values of the parameters that are estimated.

**Interpreting Regression Coefficients**

Table 3-1 shows both the original simple regression (a), and more complex multiple regressions (b) and (c) with standard errors of coefficients and \(t\)-statistics also included. Multiple regression (b) shows that

---

\(^3\) This regression (and others in the chapter), and the estimated elasticities are provided primarily to illustrate how to read and use regression analyses. For a number of reasons they probably represent short-term responsiveness to price changes (we do not examine those who either start or stop smoking, and we do not see individuals over time, both of which could provide a long-run elasticity). For more development of cigarette issues, see Chapter 24, “The Economics of Bads.”

\(^4\) Often, \(R^2\) is adjusted for the number of explanatory variables and the number of observations, giving:

\[ 1 - \overline{R^2} = \frac{n - 1}{n - k - 1}(1 - R^2) \]

where \(n\) is the number of observations and \(k\) is the number of explanatory variables. Adding more variables, \(k\) will always raise \(R^2\) but it will not necessarily raise \(\overline{R^2}\); termed the “adjusted \(R^2\).”
a $1 increase in the tax price of cigarettes, \( P \), implies a decrease in quantity of tax demanded of 2.28 cigarettes per day. Income, measured in thousands of dollars per year, has a slightly negative effect, but that effect does not differ statistically from zero. Older people smoke slightly more, and more educated people smoke slightly less. Women, African Americans, and Hispanics all smoke significantly fewer cigarettes than do white males.

As was done with the simple regression, predicted values of the dependent variable and values of the elasticities can be computed. As before, hypotheses can be tested. The addition of more variables leads to a fall in the price elasticity from about \(-0.10\) to about \(-0.07\), but it is still statistically significant from 0.

Most often, again, econometricians are interested in whether coefficients are positive or negative and whether they differ significantly from zero. In the cigarette example, excise tax increases clearly accompanied decreases in cigarette consumption. In a now famous example, Box 3-2 presents the surprising results of a detailed multivariate analysis looking at the impacts of hormone replacement therapy on postmenopausal women.

### BOX 3-2

**Hormone Replacement Therapy—Rigorous Statistics Reveal Surprising Results**

As of July 2002, approximately 38 percent of postmenopausal women in the United States used hormone replacement therapy (HRT). While the U.S. Food and Drug Administration-approved indications for HRT included relief of menopausal symptoms (hot flashes, night sweats, and vaginal dryness) and prevention of osteoporosis, long-term use had been common to prevent a range of chronic conditions, especially heart disease. Advertisements by drug companies urged women to take HRT so they would stay “forever feminine.”

Although the drugs were widely used, many scientists had expressed concern that studies finding benefits of HRT were based on nonrandom samples of women who sought out the hormone therapy. The users of HRT were better educated and healthier than postmenopausal women who did not take HRT. Thus, some researchers felt that “selection bias” could account for the effectiveness of HRT because those women for whom it was not effective and those who found the side effects bothersome or harmful, as well as less educated and sicker women, were not included in the studies.

Between 1993 and 1998, a randomized clinical trial called the Women’s Health Initiative (WHI) studied 16,608 postmenopausal women aged 50 to 79. Roughly half of the participants were randomly assigned to the experimental group and received HRT, a daily tablet containing conjugated equine estrogen and medroxy progesterone acetate (progestin). The other half was randomly assigned to the control group and received a placebo (an inert pill with no medical properties). Study participants were contacted by telephone six weeks after randomization to assess symptoms and reinforce adherence. Follow-up for clinical events occurred every six months with annual in-clinic visits required.

A multitude of health outcomes related to cardiovascular disease, stroke, cancer, fractures, and death were measured. The statistical analysis was complex and compared the health outcomes for the experimental group who took the estrogen/progestin tablet to the control group who took the placebo. Formal monitoring began in the fall of 1997, with the expectation of final analysis in 2005 after an average of approximately 8.5 years of follow-up. An independent data and safety monitoring board (DSMB) examined interim results to determine whether the trial should be stopped early, in particular if the treatment proved either significantly beneficial or harmful to the experimental group, relative to the control group.

By May 2002, an average of 5.2 years into the analysis (recall that women had entered the study over a five-year period), the DSMB determined that there were significantly higher risks of breast cancer, coronary heart disease, stroke, and pulmonary embolism in the experimental group, and that these increased probabilities outweighed some evidence of reduced risk of fractures and colon cancer. Therefore, the DSMB recommended an early stopping of the estrogen plus progestin component of the trial because it would be unethical to put more women at risk for adverse events by continuing HRT. The results were (continued)
Dummy Variables

In health care research, econometricians often are interested in whether particular groups of patients or subjects differ from others. For example, Table 3-1 denoted men, African Americans, and Hispanics by using 1 if the person was a member of such a group and zero otherwise. These groups were indicated by using binary, or dummy, variables. For example, an econometrician may wish to indicate whether a research subject is white (white = 1), or not (white = 0), or whether the subject is a woman (female = 1) or not (female = 0).

Figure 3-4 shows how the estimated coefficients of these two variables could be treated, using the regression from Table 3-1. The northwest box indicates an observation for which both male and African American equal 0. If male = 1, then the coefficient $b_m$ for both lower boxes indicates whether the person smokes more or less (and whether this is significant). This is a row effect. If African American = 1, then the coefficient $b_a$ indicates whether white households purchase more or less cigarettes (and whether that is significant). This is a column effect. If the household is both white and female-headed, then the combined effect is $(b_m + b_a)$.

In fact, if one felt that black males may have particular attitudes or preferences for smoking, one could estimate a coefficient that addresses the interaction of the two effects (1 if African American and male, 0 otherwise). This would be coefficient $b_{am}$. Here, the impact of being black and male would be $(b_a + b_m + b_{am})$. This is noted in Table 3-1 (c). Compared with white women (the “northwest” quadrant of the diagram), white men smoke 2.38 more cigarettes, African American women smoke 4.29 fewer cigarettes, and African American men smoke (2.38 – 4.29 – 1.43), or 3.34 fewer cigarettes.

STATISTICAL INFERENCE IN THE SCIENCES AND SOCIAL SCIENCES

Natural scientists attempt, not always successfully, to control experimentally for all of the other possible sorts of variation other than the relation being studied. By contrast, econometricians are seldom so fortunate. Occasionally, experimental economic studies are done, but such projects are expensive. One such study was the multimillion-dollar health insurance experiment conducted by the RAND Corporation in the late 1970s and early 1980s, funded by the federal government, and we discuss parts of that study in several later sections of the book. (There are excellent reviews by Manning and collaborators, 1987, and Newhouse and collaborators, 1993.) Even with the careful planning that went into the experimental design, this study could not avoid some major analytical issues.

Other fields have similar problems. A 1988 report from the Panel of the Institute of Mathematical Statistics referred to analytical problems in chemistry:

The data are frequently complex with a large number of dimensions, may sometimes have a time element, and can be further complicated because of missing values. In some instances, standard multivariate or time series methods may suffice for analysis, but, more commonly, novel developments are required, for example, to handle the problem of multivariate calibration.... (Olkin and Sacks, 1988, p. II-1)

Econometricians must most often use natural experiments and must seek ways to account for the other variations. Because many policies, such as the provision of public health services or the regulation of the prescription drug industry, depend on accurate measurement of economic phenomena, it is essential that the measurements be accomplished carefully and scientifically.

CONCLUSIONS

This chapter has provided a “taste” of the statistical methods necessary to address questions that occur in health economics and to clarify the analyses where statistical material is presented later in the text. To understand the text, it is important to be able to formulate questions in terms of hypotheses, read statistical test results to determine if the result is significant, understand statistical significance, and interpret reported regression results. The emphasis on problems to watch for in statistical analysis is not meant to generate undue skepticism over the statistical data to be reported. On the contrary, the discussion is meant to help distinguish the better studies where confidence can best be placed.

Summary

1. Economists usually must collect information from people doing day-to-day activities and use statistical methods to control for the confounding differences among the people that they are analyzing. The more successful they are in controlling for such differences, the more reliable the analysis will be.

2. Statistical methods suggest formulating economic assertions as hypotheses, and collecting data to determine whether the hypotheses are correct.

3. Hypotheses that test for equality among two or more items are called simple hypotheses. Hypotheses that test whether two or more items are greater (or less) than each other are called composite hypotheses.

4. Several steps are necessary to test hypotheses appropriately. The econometrician must:
   • state the hypothesis clearly,
   • choose a sample that is suitable to the task of testing,
   • calculate the appropriate measures of central tendency and dispersion, and
   • draw the appropriate inferences.

5. Regression analysis allows the econometrician to fit a straight line through a set of data points. In ordinary least squares regression, the sum of the squared deviations of the actual data points from the line is minimized.
Discussion Questions

1. List at least three ways in which natural experiments differ from laboratory experiments.
2. What is the difference between a simple hypothesis and a composite hypothesis? Why might economists choose one over another?
3. In considering the difference in smoking between men and women, what is the null hypothesis? What is the alternative hypothesis? Is the alternative hypothesis simple or composite?
4. Suppose that we wish to compare the health status of two groups of people. What variable might we use to measure the status? What variables might we wish to control in order to draw the appropriate inferences?
5. If someone reports that the mean weight for fourth-grade boys is 80 pounds and for fourth-grade girls is 78 pounds, what must you know to test hypotheses using the difference of means?
6. If we are trying to relate output to labor inputs and capital inputs using regression analysis, would we expect the coefficients of the regressions to be positive or negative? Why?
7. What are dummy variables? How are they useful in identifying differences among groups?
8. Suppose that you used regression methods to estimate the demand curve for physician visits and found a positive relationship; that is, you found that the higher the price is, the more visits are demanded. What problem has likely arisen? Explain the problem in words. Why might it make statistical inference difficult?
9. Rich people consume more health care services than poor people. Explain two ways one might test this hypothesis.

Exercises

(For students with access to spreadsheet computer programs.)
Consider the following data for a cross section of individuals in the population, in which

\[ Q = \text{Quantity (in 100s) of aspirin purchased in a year} \]
\[ P = \text{Average price of aspirin in that year} \]
\[ Y = \text{Annual income} \]
\[ A = \text{Age of buyer} \]

Now consider questions 1 to 4:

1. If we divide the population into two groups, up to age 35 and over age 35, which group purchases more aspirin?
2. Divide the population into three groups—up to age 30, over 30 and up to 45, and over 45. Do the purchases vary by age?
3. What is the relationship in a regression analysis between \( Q \) and \( P \)? Between \( Q \) and \( Y \)? Between \( Q \) and \( A \)?

Now consider questions 1 to 4:

<table>
<thead>
<tr>
<th>Observation</th>
<th>( Q )</th>
<th>( P )</th>
<th>( Y )</th>
<th>( A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.5</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>3.5</td>
<td>1.5</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>1</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>3.5</td>
<td>2</td>
<td>15</td>
<td>36</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>4</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>2</td>
<td>15</td>
<td>55</td>
</tr>
<tr>
<td>18</td>
<td>9</td>
<td>1</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>1.5</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>21</td>
<td>6</td>
<td>1.5</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>1</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>1</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>26</td>
<td>4</td>
<td>3</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>4</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>4</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>5</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>1</td>
<td>30</td>
<td>32</td>
</tr>
</tbody>
</table>
4. Calculate the multiple regression that relates $Q$ with $P$, $Y$, and $A$. Which variables are statistically significant? What is the elasticity of $Q$ with respect to $P$, to $Y$, and/or to $A$?

5. From Table 3-1, column b, suppose income is $20,000, the excise tax on cigarettes is $1, and the person is a 40-year-old white, non-Hispanic male who completed high school (education level = 9). Calculate the elasticities of demand for aspirin with respect to excise tax, income, and age.

6. Consider demand curves for aspirin, estimated for two different sets of consumers:
   (a) $Q = 20 - 5P + 0.2Y$
   (b) $Q = 30 - 5P + 0.2Y$
If $Y = $20 and $P = $1, calculate the price and income elasticities for group (a) and group (b). Whose elasticities will be higher? Why?

7. Given the regression estimate of the demand equation of
   $$Q_x = 1,000 - 3.3P_x + 0.2P_z + 0.001Y$$
   where $Y$ is income, what is the change in demand if price rises by $1, holding income constant? What is the percentage change in demand if price rises by $1 from an initial price of $P_x = 200 given $Y = 10,000? What is the effect on demand of a $1 increase in income, holding price constant?

8. Consider the estimate demand equation of
   $$Q_x = 1,000 - 3.3P_x - 0.2P_z + 0.001Y$$
   with $t$ values in parentheses, where $P_z$ is the price of another good $Z$, and $Y$ is income. Is good $Z$ a substitute or a complement? Can we say confidently whether good $X$ is a normal good or an inferior good?

9. Look at Regression (b) in Table 3-1, and consider the following questions:
   (a) Does cigarette consumption increase as income rises? Are cigarettes a “necessity” or a “luxury”?
   (b) For the variable “Educational Level” a high school graduate is coded with level 9, and a college graduate is coded with level 12. What is the predicted difference in cigarette consumption between the two levels of education?