Methodological appendix

1 Introduction

As we have seen throughout this text, economics is grounded in mathematics and statistics. This Appendix shows how mathematics and quantitative methods can help nonprofit managers and social entrepreneurs make more informed managerial decisions by applying techniques from differential calculus, statistics, econometrics, and experimental economics. (The presentation here is rudimentary; however, references are provided for those seeking further study.) Differential calculus allows us to think more precisely about the problem of "optimization," that is, finding the value of a decision variable that produces the best outcome. The strategy of "thinking at the margin" discussed in Chapter 5, which allowed us to determine the price for maximizing profits, for example, derives from the principles of calculus. Various other economic decisions, such as minimizing costs and maximizing social benefits can be framed as optimization problems solved by differential calculus, as well.

Statistics, based on the concept of probability, allows managers to analyze economic data in various ways to better inform their decisions. Recall that probability is a basic construct underlying the analysis of decisions under risk, as considered in Chapter 10. Statistical techniques also underlie the estimation of mathematical models to characterize the relationships between two or more economic variables, such as demand functions that relate price to quantity demanded and other variables such as income or age of consumers, or supply functions that relate price to quantity supplied and other variables such as size and age of firms. Econometrics is the specific application of statistical modeling to economic problems, which allows interpretation of results in terms of concepts such as price and income elasticities, as discussed in Chapter 9. Statistical and econometric methods are needed to model economic functions because real data involves uncertainty deriving from imprecise measurement, intrinsic randomness of the phenomenon of interest, and the impacts of various measured and unmeasured factors that may influence data values. Statistical techniques are designed to address these issues.

While statistical and econometric techniques are most often applied to data that is readily available, such as data collected by the government as part of a census or in connection with tax collection or regulation (such as IRS 990 data for nonprofit organizations), increasingly analysts and managers conduct original research by formulating experiments designed to collect the exact type of data they desire for making their particular decisions. In particular, they use the scientific method to conduct lab and field experiments with human subjects, an approach that helps to reduce uncertainty about the size and nature of behavioral responses. Chapter 12 reviews this approach to economic decision making in the nonprofit and social enterprise sectors. Here, in section 4, we review the basic concepts and rudiments of experimental design to help managers engage in this approach.

2 Calculus

Economics focuses substantially on how economic actors can obtain the best possible outcome given the constraints they face. For example, consumers wish to buy the best collection of goods, given the money available to them. Or a firm wishes to produce the optimal quantity of a good, given consumers' willingness to pay for them, which for typical for-profit firms is the quantity that maximizes its profits. In the nonprofit and social enterprise sectors, there is the question of what exactly the organization is trying to do. Does it want to serve the largest number of clients with its limited income, or the neediest clients? Does it wish to achieve a certain combination of its own profits and benefits to clients, bring about a particular change in society, or some weighted combination of such objectives? Any of these goals may be appropriate for a particular nonprofit organization or social enterprise. The techniques discussed in this section illustrate how to achieve a stated goal as fully as possible, whatever it may be.

Economists apply quantitative methods to solve decision problems of two types. Some decisions concern *whether to do one thing or another* (should we open a gift shop or not?). Other decisions require selecting a single alternative from a finite list of possibilities (should we build a new facility downtown, in the eastern suburb, or in the western suburb?). These are "discrete choice" or "whether" problems, solved by calculating the net benefits of each alternative and selecting the alternative with the highest net benefit. Other questions are about *how much of something to choose* (what size the facility should be), which requires picking a number from an infinite number of possible values. The latter are "continuous choice" or "how much" problems, solvable using the tools of calculus. For example, "how many clients should a counseling service see per week on average?" is a continuous choice decision, as the answer might be 1.37 or 12.6438 or any other positive number in a feasible particular range.

This section outlines some of these methods that apply to these decisions and how managers of nonprofit organizations and social enterprises might use them. We begin by discussing the generic concept of *optimization*.

Optimization

Generically, economics studies how consumers and producers make choices under constraints. This applies to nonprofit organizations and social enterprises, which evaluate choices by their impacts on mission-attainment and which face budget, regulatory and other constraints. The mission may require maximizing some quantity, like exposure to opera, or minimizing some quantity, like the number of clients denied service. Maximization or minimization of a variable is called "optimization." Before describing the mechanics of optimization, we need to discuss the mathematical formulation of mission, that is, the organization's "objective function," which translates possible outcomes into a number that can be used to place outcomes in rank order and identify the best outcomes.

This is not a trivial problem. Unlike the case of for-profit firms, there is no generally accepted, mathematically-based model predicting the nature of a given organization's objective. For-profit firms usually seek to maximize their profits because their owners (shareholders) want to receive the biggest dividend checks they can get and because for-profit firms that do not maximize profits are ripe for a takeover bid. Nonprofit organizations and social enterprises come in many forms, with diverse stated missions. The shared fact that nonprofits cannot distribute profits to those in control does not limit them to a single objective function. Indeed, nondistribution of profits makes diverse missions sustainable – there are no dividend-receiving shareholders or takeover threats that force nonprofit organizations to maximize profits. The situation for social enterprises is even more complex. These enterprises are designed to balance profits with social mission in a variety of ways, with diverse limitations on profit distribution, and different parameters of corporate governance and claims to organizational assets (Young, Searing, and Brewer, 2016). Again, feasible organizational objectives are highly diverse.

Mission statements are suggestive, but do not suffice to establish an organization's objective function. Mission statements are necessarily vague, in order to secure the widest possible stakeholder buy-in. Mission statements are also used to manipulate the impressions of external stakeholders. For mathematical purposes, mission statements are incomplete. They specify all the things the organization cares about, but rarely if ever detail exactly how the organization will weigh a gain in one objective against a loss in some other objective (Tuckman and Chang, 2006).

Steinberg (2006) summarizes possible objective functions appearing in the nonprofit literature. These include minimizing cost subject to serving a given number of consumers, maximizing the social benefit of their output, maximizing the total number of consumers served, maximizing the utility of those served, or even maximizing profit, subject to a given budget constraint. Before attempting to apply the quantitative methods discussed here, a nonprofit manager or social entrepreneur must give serious thought to what, exactly, her organization is attempting to optimize. Nonprofits will want to maximize profits from certain activities (e.g., unrelated income sales) in order to subsidize their mission-related activities, but profits in this case are the means to an end rather than the end itself. Profits, or financial surpluses however defined, enter the optimization problem as a constraint, not as an objective function. Simply put, to avoid bankruptcy profits cannot be persistently negative. Soup kitchens cannot provide free soup without donations or other sources of income.

Once an organization's objective is determined, it is useful to write that objective in the language of mathematics, in the forms of an "objective function," giving mathematical voice to desires of the organization. Then the calculus approach detailed below allows the organization to select optimal values of all its "choice variables" (variables that the organization controls). Some possible objective functions might be:

1. Maximize revenue: It is unlikely that this would be the goal of a nonprofit organization or social enterprise, although some have proposed this (James, 1998; Niskanen, 1971), but it provides an opportunity to illustrate the formal notation used in optimization problems:

$$\max_{Q} PQ$$

where: Q is the number of units sold, and
P is the price per unit

Note that we place a Q under the word "max" to indicate that we will maximize by choosing a value of Q. The choice variables are always listed below "max" or "min." For the moment, we assume that the firm is a perfect competitor, so the price is not a control variable.

A slightly different formulation applies if the organization has market power, allowing it to pick (P, Q) combinations. The set of available (P, Q) combinations is then constrained by market demand.[1] A profit-maximizing firm would always pick a point on the demand curve because given some value for Q, it is impossible to sell at a higher price and undesirable to sell at a lower price than the price on the demand curve at that Q. For example, for the straight-line demand curve $P^D = a - bQ$, total revenue PQ can be written as a function of one choice variable, Q:

$$\max_{Q} P^D(Q)Q \text{ or } (a - bQ)Q \text{ or } aQ - bQ^2$$

Note that $P^D(Q)$ denotes functional dependence rather than multiplication. The superscript D identifies that this is the demand function, not the supply function which is a different equation involving p and Q.

2. Minimize total cost:

$$\min_{q_1, q_2, \ldots, q_N} \sum_{i=1}^{N} q_i P_i$$

where: q_i is the quantity of the i^{th} input, $i = 1, \ldots, N$, and
P_i is the price of the i^{th} input, given by the market

By itself, this is a silly problem because the obvious way to minimize costs is to use no inputs and thereby produce plenty of nothing. But consider an orphanage serving exactly 10 children that is under great pressure to cut costs. Now the problem is cost minimization while producing 10 well-cared-for children. This creates at least two constraints on input choice. First, there must be enough inputs to care for these children. Second, there are one or more constraints protecting the quality of care (such as prohibiting substituting sedatives for staff).

3. Maximize (undistributed) profit:

$$\max_{Q, q_1, \ldots, q_N} PQ - \sum_{i=1}^{N} P_i q_i$$

Profits are simply total revenues minus total costs. However, this formulation ignores the constraint on the choice variables. The firm cannot select any old level of input quantities; rather, it must select an input combination that produces the chosen level of output. The relevant constraint is the production function, which specifies how much output is produced from each combination of inputs. Solving this problem requires a mathematical tool that is beyond this text, Lagrangian functions and Lagrange multipliers, which is covered well in Chiang and Wainwright (2004). Such a technique requires methods of linear algebra, which are also discussed in detail in this source. The problem becomes much simpler if we have already solved the input choice part. Then we can write total cost as a function of the quantity of output, and output becomes our only choice variable.

4. Maximize the number of consumers served:

$$\max_{Q} Z(Q):$$

where: Z(Q)is a function that gives the number of consumers served *as* a function of the quantity of output

For example, a local symphony may offer free concerts in the park because it wants to convert people to classical music lovers. The number of shows is Q, and from experience the nonprofit knows that the first five shows will each attract 200 listeners, the next will attract 150 listeners, and any additional shows attract 0 listeners. Although it is possible to produce 1.5 shows per season, because it is possible to produce 3 shows and spread them over 2 seasons, we restrict attention to integer answers in order to illustrate the approach. Then the Z function is best thought of as a table of values rather than a formula. Here are some values it takes: if Q is 3, Z(Q) is 600, if Q is 6, Z(Q) is 1,150. If the cost of putting on another show was zero, the solution is obvious – produce six shows. But if the musicians are paid by the hour, there is a bankruptcy constraint on Q. If the symphony had an endowment yielding \$10,000 per year and no other sources of revenue, the constraint would be:

$$\$10{,}000 - C(Q) \geq 0$$

where: C(Q) is the cost function showing the cost of performing Q shows

Because this is a discrete choice problem, calculus cannot be applied, but there are only 7 alternatives worth considering (0, 1, 2, 3, 4, 5, or 6 shows) so you can just figure out which one maximizes profits without causing bankruptcy by doing a bit of arithmetic.

Optimization using calculus

Imagine, for a moment, a hill that first goes up and then goes down. Where will that hill be the highest? At the point where it stops sloping up and begins to slope down. Thus, one can find the maximum point on a function by looking for the level at which its slope stops increasing and begins to decrease, taking on a value that is not positive or negative, but equal to zero in between these two regions.

In the same way, imagine walking into a hole in the ground. Where does that hole reach a minimum? At the point where it moves from getting deeper to getting shallower, where the slope, is, again, equal to zero. Thus, finding a maximum or finding a minimum is done the same way; by finding where the slope is equal to zero.

We can think of the objective function as a topographic map. Maximization problems concern finding the peaks of the mountains in that map (perhaps finding the top of Mount Profit, if that is your objective), and minimization problems concern finding the bottom of the valleys (perhaps Valley of the Nethermost Cost). The way to locate the top, if you are on the side of a hill or mountain, is to determine which direction leads you most rapidly uphill. You then take a step in that direction, determine the direction of most rapid ascent from your new starting point, and repeat. When you get to the point where there are no uphill directions, you have located the peak. You reach the bottom of the valley if you follow the direction of steepest descent until there is no direction that leads you further downward. This is the essential insight behind computer algorithms that maximize or minimize functions ("hill-climbing algorithms"). The idea that all optima are at flat points of the topographic map underlies the mathematics of optimization and constrained optimization. The rate of rise or fall is the slope, calculated between two points on the topographic map as "rise/run;" we are looking for places where this slope equals zero.

To illustrate, consider a two-dimensional map in which the horizontal axis exhibits the quantity of production of some good (cups of coffee) and the vertical axis shows the resulting profit. Suppose profits are \$20 when the firm produces 10 cups of coffee and \$23 when the firm produces 12 cups of coffee. The slope between these points is

$$\frac{\text{rise}}{\text{run}} = \frac{(\$23 - \$20)}{(12 - 10)} = \frac{\$3}{\$2} = \$$$

The slope is positive, meaning this step to the right allowed you to climb Mount Profit. The algorithm requires you to keep stepping to the right until the slope is zero, where you have then located the optimal quantity of production.

Now complicate the problem a little bit by moving to a three-dimensional map. On the east–west (X) axis, we include one choice variable (cups of coffee) and on the north–south (Y) axis we include a second choice-variable (packets of sugar) while measuring profit on the up-down (Z) axis. To climb Mount Profit, we might have to move northeast, so we need to define a new concept – directional slopes – before we know how to cross the map. The directional slope in the XZ direction is the change in profits when you increase the number of cups of coffee while holding sugar packets constant. The directional slope in the YZ direction is the change in profits when you increase the number of sugar packets while holding cups of coffee constant. There are two directional slopes at each point of the topographic map, and at the optimum, both must be zero.

One problem is that the function may squiggle up and down in-between the two points we used to calculate a slope, so we won't locate the truly flat place where the optimum occurs. Consider the following parabola, and calculate the slope between the points x = 1 and x = 5:

The slope of the line connecting these points is zero, but the parabola is flat only if the top of the mountain has been strip-mined out of existence. Calculus suggests you move these two points progressively closer together, calculating the slope after each move. The limit of this process is reached when the two points overlap, and then we have the "slope at a point" rather than the slope between two points. The limit slope is the same as the slope of a line that just touches, without crossing, the function, so we say that it is slope of the "tangent" line. In turn, the formula for the slope of the tangent line is called the "*derivative* of y with respect to x," written:

$$\lim_{x_2 \to x_1} \left[\frac{y_2 - y_1}{x_2 - x_1} \right] = \frac{dy}{dx}$$

The derivative is itself a function, as the slope of the tangent line depends on x. When there is only one choice variable, we can simplify the notation a little. If the original function is called f(x), then the derivative function can be written as f'(x). Calculus textbooks develop shortcuts for figuring out f'(x) known as the rules of differentiation. Here are some important rules:

1. Derivative of a constant: A constant value of y regardless of x graphs as a horizontal line. Its slope is therefore zero.

   If f(x) = c, then f'(x) = 0.

2. Derivative of a constant (m) multiplying x:

$$\text{If } f(x) = mx, \text{ then } f'(x) = m$$

3. Derivative of the sum or difference of two functions of x:

$$\text{If } f(x) = g(x) \pm h(x) \text{ then } f'(x) = g'(x) \pm h'(x)$$

Although this looks intimidating in functional notation, it simply says that you differentiate each term separately. For example, the formula for a straight line ($y = mx + b$) is differentiated term by term, where mx plays the role of g(x) and b plays the role of h(x), using the first two rules above:

$$\text{If } f(x) = mx + b, \text{ then } f'(x) = m + 0 = m$$

And in a particular example,

$$\text{If } y = 3x + 12, \text{ then } \frac{dy}{dx} = 3$$

verifying what we already know (that m indicates the slope of the line).

4. Derivative of a power of x:

$$\text{If } f(x) = x^n, \text{ the } f'(x) = nx^{n-1}$$

This rule covers a variety of cases. When n is a negative integer the function is interpreted as one divided by x^n:

$$\text{If } f(x) = x^{-n} = \frac{1}{x^n} \text{ then } f'(x) = -nx^{-n-1} = \frac{-n}{x^{n-1}}$$

And when n is a fraction (such as 0.5 indicating the square root or 0.33 indicating the cube root) the formula applies directly.

5. Derivative of a parabola. The general formula for any parabola is $y = a + bx + cx^2$, and its derivative is just what you would think using rules 1 – 4:

$$\text{If } f(x) = a + bx + cx^2, \text{ then } f'(x) = b + 2cx$$

A polynomial is any function of constants times powers of x, and the parabola formula is easily extended:

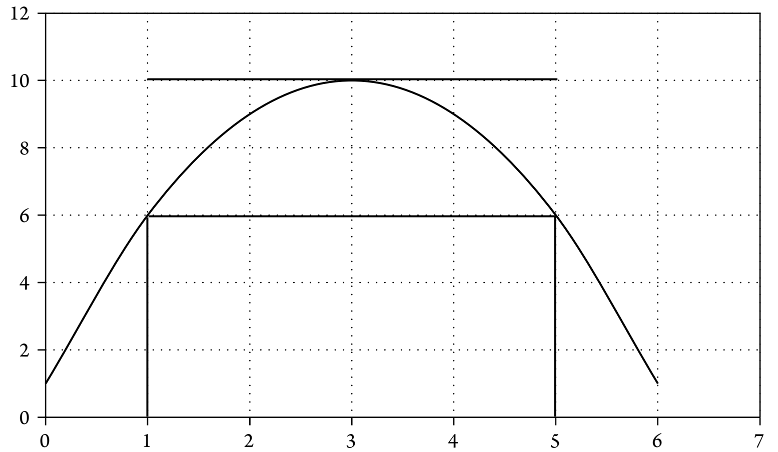

Figure A.1 Slope at a point and the optimum

$$\text{If } f(x) = a + bx + cx^2 + dx^3 + \ldots, \text{ then } f'(x) = b + 2cx + 3dx^2 + \ldots$$

Let's return to the parabola in Figure A.1 and find the top of the parabola using calculus rather than graph paper. The formula for that parabola is:

$$y = -x^2 + 6x + 1$$

We find the maximum or minimum of a function by looking for places where the slope of the function is zero. Calculus allows you to tell whether a particular point where slope is zero is a maximum, minimum, or something else, but this is a bit beyond the scope of this brief review, and we know from the graph that we will find a maximum here.[2] We find it by locating points where the slope is zero – that is, take the derivative, set it equal to zero, and solve for x. We label the resulting value with a* indicating that this is the particular value of x that is optimal:

$$f(x) = -x^2 + 6x + 1;$$

$$f'(x) = -2x + 6;$$

$$0 = -2x + 6;$$

$$6 = 2x;$$

$$x^* = 3.$$

One final rule is useful when working with elasticities. In the text, we defined elasticity between any two points on the demand (or supply) curve. There appeared to be arbitrariness in how we calculated percentage changes, solved by the midpoint formula. But calculus allows us to dispense with such arbitrariness by defining elasticity at a point. The point elasticity formula takes the limit of elasticity as the two points move together; here is the formula for the point price-elasticity of demand:

$$\varepsilon_P^D = \frac{P}{Q^D}\frac{dQ^D}{dP}$$

In this form, it is clear why the price elasticity varies from zero to infinity along a straight-line demand curve. For the moment, assume that we are talking about a downward-sloping demand curve – neither horizontal nor vertical. The derivative in the formula is the reciprocal of the slope. Because the slope is constant, the reciprocal is also constant and under our assumption it is finite and non-zero. However, the P/Q part of the formula is zero at the x-intercept of the demand curve (when P = 0), and undefined at the y-intercept (because we are dividing by zero.) As we move along demand towards the y-intercept, the fraction grows without limit, so at the y-intercept it is best defined as infinite. In between, it takes every possible value between zero and infinity. Relaxing our assumption of a downward sloping demand curve, there are other types of dividing by zero for vertical and horizontal demand curves, but it makes sense to say that the vertical demand curve is perfectly inelastic and the horizontal one is perfectly elastic, as we did in Chapter 9.

Is there a demand curve that has the same elasticity at every point on the curve and is neither a vertical or horizontal line? Looking at the formula, we see that the P/Q ratio gets smaller and smaller as we move from left to right, which we could counter if the derivative becomes bigger and bigger. One can show that the following demand curve does the trick:

$$Q = AP^{\varepsilon}$$

where: A is a positive constant and
ε is the constant price elasticity of demand

This formula is very useful for estimating demand curves, as we shall see below in the section on regression, where the elasticity of demand may be found from a particular specification of a regression equation.

3 Statistics

In the modern world, managers are expected to be "data driven" that is, to use data to make intelligent decisions about how they should manage their organizations. Doing this often involves employing techniques from statistics. To learn statistics, it is vital to first understand the concept of *probability*. Probability undergirds the analysis technique of "hypothesis testing," which is central to statistical analysis.

Much of what we now call statistics grew out of the world of gambling. If you draw a card at random from a regular deck of cards, what is the chance that card will be red (1/2), what is the chance that card will be a queen (1/13), what is the chance that card will be a red heart (1/52)? These chances can be characterized as probabilities, and have particular properties, as described below.

Probabilities

A *probability* is a number assigned to some outcome of an experiment, such as drawing a card from a standard deck of cards, which must conform to two rules:

1. Each probability must be a value between zero and one; $0 \leq p(x) \leq 1$ where x is the outcome in question.
2. The sum of the probabilities associated with mutually exclusive and collectively exhaustive outcomes must sum to one; $\Sigma p(x) = 1$ for all outcomes x.

Indeed, any assignment of probability values that follows these rules can be called a "probability distribution" even if these values are assigned subjectively and even if they do not reflect reality. So, if your home team in football has a woefully poor record, despite that fact, you can still assign optimistic probabilities to the chance for it to win the Super Bowl, for example:

$p(\text{win}) = 0.95$
$p(\text{lose}) = 0.05.$

Such a set of numbers is called a *subjective probability distribution* since it is based on an individual's or group's judgment. Even if those probabilities do not seem to make sense to anyone else given your team's history, since it follows the rules specified above it is still a legitimate probability distribution.

More commonly, probabilities are specified by looking at outcomes either in terms of relative frequency, as in determining the probability of drawing a red card from a standard deck of cards, or in the long term, as in the proportion of heads one would see if a coin was flipped a large number of times.

Assigning probability values to possible outcomes of experiments allows us to talk about whether an outcome is a rare or a common event. The probability of a common event is relatively large; for example, the probability that a randomly selected dog will have four legs is close to one. By contrast, the probability of a rare event is small, close to zero. For example, the probability of a randomly selected person earning a perfect score on the GRE test is very small.

Hypothesis testing

This perspective about large and small probabilities associated with common and rare events, allows us to discuss a technique used in statistics and econometrics called "hypothesis testing." Hypothesis testing is not a mysterious endeavor, but something that people do every day in some form or another. It simply involves assigning probabilities to the way the world is, analyzing some information that you have recently gathered, and asking whether you should change your view of the way that world works as a result of that analysis.

The classic example of hypothesis testing is found in the criminal justice system. In the United States, defendants are assumed to be innocent until they are proven, beyond a reasonable doubt, to be guilty. Other situations that involve hypothesis testing are also found in ordinary life, whenever a decision is made based on data available.

For example, imagine you are looking for something to cook for dinner, and there is a package of frozen hamburger meat in the freezer. Should you use that meat in cooking dinner tonight? Assuming that you can grill up the meat and make some tacos or burgers, what does it take for you to decide not to do that? Suppose you notice a very old "sell by" date on the package, or that it has freezer burn, or that the meat is brown or damaged. After assessing this information, you may decide not to use it for dinner. You have just performed a "hypothesis test" by assigning a probability to the statement that the meat is good to eat. You may have just spared your body a difficult evening.

In formal statistics, an "observation" is a set of variable values for an "observational unit." For example, if our observational unit is a person, an observation for Fred might include Fred's donations, income, gender, and shoe size. If the observational unit is an organization, an observation for Economists

Anonymous might include EA's total revenues, total expenditures, expenditures on coffee, and donations received. A set of observations is called a "data set." The ideal data set is the *population*, including an observation for every observational unit in existence. The population Fred belongs to could include everyone alive today, everyone alive today in the USA, or everybody who has ever lived or might be born in the future, depending on what you want to study. A *sample* is a more limited set of observations, for example, a representative group of people who answered a survey asking about their donations, income, gender, and shoe size. A *statistic* is a number calculated from a data set, such as the average shoe size.

A statistical hypothesis is a statement about a statistic or a pair of statistics that could be computed from the population data set if we only had data on everyone in the population. For example, we might want to know the average donations made by every household in America. We might hypothesize that the true population value for this statistic is $2,000, but we only have data for about 6000 households who answered the Philanthropy Panel Study survey in a particular year. Or we might want to know whether population-average giving by unmarried females equals the population-average giving by unmarried males. Here we are comparing two sample statistics.

Usually we care about whether a hypothesis (called the "null hypothesis" and symbolized H_0) is true, but since we cannot directly prove that it is true, we set up an alternative hypothesis (H_1) that will not be true given our null hypothesis. If we reject the null hypothesis, we do so because it makes a prediction inconsistent with the evidence from a sample. If we don't reject the null hypothesis, we say that we have "retained" the null hypothesis, although we cannot say that we have "accepted" it. In the first example, the hypothesis that there is no difference between the population average and $2000 is the one we call null, so we have:

H_0: Population-average household giving is equal to $2000

H_1: Population-average household giving is not equal to $2000 (it is either greater than or less than $2000).

In the second example, the hypothesis that there is no difference between female and male giving is the one we call null, so we have:

H_0: Population-average giving by single females is equal to population-average giving by males.

H_1: Population-average giving by females is not equal to population-average giving by males.

We then perform the hypothesis test, using sample statistics, to see if the evidence supports our alternative hypothesis. Suppose the average donation in our sample is $1950. This evidence gives us little reason to reject the null hypothesis because it is relatively common to have a sample mean only 2.5 percent different from the population mean. Alternatively, suppose another sample from the same population yielded an average donation of $6 million, which could happen if the second sample included the Bill and Melinda Gates household while the first sample did not. The second sample gives us reason to reject the null because it would be very unusual to have a sample mean that is so vastly different from the population mean.

In the classic example of hypothesis testing, in the criminal justice example, the null hypothesis is that the defendant is innocent, while the alternative is that the defendant is guilty. It is up to the prosecutor to present enough evidence to reject the null hypothesis in favor of the alternative. Some such evidence might be videos of a bank robbery, stolen money on the robber's person, or witnesses that observed the crime. The job of the defense attorney is to hold the prosecutor to this burden, and to make sure that no laws are broken in that process.

Suppose the manager wants to know how to recruit volunteers, and her choice depends on the volunteer's opportunity cost of volunteering, which is the after-tax wage rate (at least to a first approximation – see Chapter 6). The manager believes that the average after-tax wage is about $10.00 per hour and wants to know whether this is a reasonable estimate. She could do so by surveying a sample of volunteers and calculating several things in order to determine whether the evidence does not reject the hypothesis that the average after-tax wage is $10.00 (in which case this is a "reasonable" estimate) or does reject the hypothesis (suggesting this estimate should not be used).

First, we need to know the hypothesized value of the population mean we are testing for. This is often designated as "μ," the Greek letter mu that corresponds with our m (as in m̲ean). We also need to know the value of the *variance* which is symbolized by "σ^2" (sigma squared). The population variance of the variable X is the average squared deviation of Xs from the population mean, where the average is computed across all values that X takes in the population. When there are a finite number of possible values for X, the formula is:

$$\sigma_X^2 = \sum_{i=1}^{I} \frac{(X_i - \mu)^2}{I}$$

where: i is an index for observations in the population. If there are I observations, i takes each value in the set {1, 2, 3, . . ., I}.

The variance of X is measured in units that are the square of the units used for X, but it is often convenient to work with the positive square root of variance because it is measured the same way as X. This square root is called the *standard deviation* of the variable used, symbolized by the Greek letter sigma "σ" that corresponds with our s (as in standard deviation). The standard deviation measures how spread out the values of our variable are. If all observations in the population take the same value, there will be no spread between the observations of X and the mean of X and so σ will equal 0. The larger the value of σ, the more spread out is the data.

When σ is known for the population, hypothesis testing uses the standard normal distribution, which is a bell curve with $\mu = 0$ and $\sigma = 1$. Most hypothesis tests are about the population mean of a variable. Suppose the null hypothesis is that $\mu = k$, where k is some constant number. For example, if we wanted to know whether variable X is associated with a change in variable Y, the variable we want to test is the slope of the graph of X against Y. A null hypothesis of no association is equivalent to the hypothesis that the average slope (k) in repeated sampling would equal zero. We test this hypothesis using a sample average, denoted as $\overline{X}$, and a sample standard deviation, denoted as "s." The formulas for these sample statistics are:

$$\overline{X} = \sum_{n=1}^{N} \frac{X_n}{N} \text{ and}$$

$$s = \left[\sum_{n=1}^{N} \frac{(X_n - \overline{X})^2}{N - 1} \right]^{1/2}$$

where: n indexes observations in the sample, rather than those in the population.[3]

One more sample statistic must be defined before we conduct any hypothesis tests, the *standard error*. This measures the spread-outness of the distribution of sample means, rather than the spread-outness of the individual observations in the sample, but the two measures are related by this formula:

$$SE = \frac{s}{\sqrt{N}}$$

These sample statistics are random variables because they vary across samples. This allows us to calculate a "Z-statistic," another random variable, as:

$$Z = \frac{\bar{X} - k}{SE}$$

Using a remarkable theorem called the Central Limit Theorem, we know that the Z statistic approaches the standard normal distribution as sample size increases, provided X can take values from negative to positive infinity and has finite variance (the details of the theorem are far beyond this brief introduction). For this distribution, 95 percent of sample means will lie between a Z statistic of -1.96 and +1.96. Rounding slightly, this means that if your sample average is two standard deviations or more from the null hypothesis k, you can reject the null hypothesis with 95 percent confidence. The approximation is less good when sample size is less than 30, so for small samples we calculate a more exact "t-statistic" in a similar fashion and compare it with appropriate probability tables.

In rejecting a null hypothesis, there is always the possibility of making a mistake, and such errors can occur in two directions. We are concerned with the possibilities of either rejecting a true hypothesis, or of not rejecting a false hypothesis. However, we are most often particularly concerned with the chance that we incorrectly reject a true hypothesis, as would be the case if an innocent person is sent to prison in a system where people are assumed to be innocent until proven guilty. This probability characterizes a "Type I error" and is often designated as "alpha." If the probability of finding a particular value is smaller than the value of alpha (say 5 percent) that we are willing to live with, then we can reasonably reject the null hypothesis.

Although the term "reject" sounds very negative, it is actually what we want to do. No one is interested in affirming that the world looks exactly as we thought. Rather, we want to set up the null and alternative hypotheses so that a rejection of the null hypothesis will lead us to re-define a standard view of the world. For example, in 1492, most people thought that the world was flat, and that one could sail off the edge of it if one sailed too far. However, Christopher Columbus, suspecting that the world was round, sailed for quite a while and seemed to have sailed around the entire planet. This led people to rethink their view of the Earth as a disk suspended in space. The original view of the flat world, the null hypothesis, was rejected, in favor of the alternative hypothesis that proposed the spherical world that we now take for a fact. However, it is important to remember that a failure to reject the null hypothesis does not mean the null hypothesis is true. It just means that the evidence is insufficient, and the null hypothesis might turn out to be false if more evidence (a larger data set or one with more variation in the variable in question) is introduced.

Hypothesis testing with computers

In numerical situations, it is often useful to perform hypothesis tests using a statistical computer program, such as EXCEL, SPSS, Stata, or SAS. In performing such tests, a value is proposed for a hypothesized mean value of a variable. The computer uses this information, the test data provided, and the mean, standard deviation, and standard error calculated from that data to compute a Z-statistic or a t-statistic in much the same way we did above. The programs look up the probabilities and report the results of hypothesis tests in terms of the confidence level or "p-value," which is 1– the confidence level. Rejection of the null with 95 percent confidence corresponds to a p-value of 5 percent or 0.05, meaning that the sample mean statistic would be this far from the hypothesized population mean for less than 5 percent of possible samples when the null is true.

Computer programs that allow one to test null hypotheses on a single set of data also allow us to compare two or more subsets of data. Doing so sometimes requires assuming that the data subsets are somehow naturally related to each other. If this is the case, a "paired samples test" is most appropriate – for example, comparing test scores for incoming students to those of the same students after they complete a training program. Alternatively, it is possible that the data points are not naturally related to each other, as when comparing outcomes from an educational program that runs in several different distinct neighborhoods in a city. If there is no natural relationship between sets of observations, a hypothesis test on "independent samples" is appropriate. For independent samples, whether or not the variation (standard deviations) within the different groups are assumed to be equal will play a role in the calculation of the test statistic and therefore the p-value obtained from the test. How this is done varies among statistical programs, so details should be consulted in the documentation accompanying any such program.

Analysis of variance (ANOVA)

One type of hypothesis test that may be particularly useful for nonprofit managers and social entrepreneurs is "Analysis of Variance" generally known as *ANOVA*.[4] This test examines several groups of data and asks whether the mean of a particular variable is the same for the different groups. ANOVA compares variance between different groups to the variance within those groups. If the variance between groups is greater than the variance within the groups, then one concludes that the means of the groups are different. ANOVAs are usually used to compare sample means among three or more groups, testing the null hypothesis that all of those means are equal. Again, this is best done with software, where the analyst specifies a discrete grouping variable that may influence a specified dependent variable. For example, a nonprofit manager may want to know if students from different neighborhoods show the same level of improvement in study skills after attending an after-school program. The dependent variable will be the level of improvement in study skills, while the grouping variable will be the different neighborhoods in which the program is offered. As before, if the p-value obtained from this hypothesis test is smaller than some pre-determined level (typically 0.05, 0.01, or 0.001 depending on what convinces you that the null should be treated as false) the null hypothesis (that the programs work equally well for the average student) is rejected in favor of the alternative (that not all the average levels of improvement are equal[5]). This implies that at least one group showed a significantly different increase in study skills than did the others.

Nonparametric tests

The statistical tests outlined above require the analyst to make assumptions about the distribution of variables in the data. For good reason, it is common to assume that statistics calculated from a sample are normally-distributed across samples regardless of the distribution of the raw data used to construct those averages. This is the central limit theorem, but it relies on limiting distributions that may not be accurate for small or moderate sample sizes, and the central limit theorem does not apply when the statistic cannot take values outside a certain range. The tests we have described so far are parametric, meaning that they depend on the two parameters that fully describe the normal distribution (mean and variance). Parametric tests are most powerful, meaning that they are highly likely to reject the null hypothesis when the null hypothesis is false for the population, but only when the normality assumption is true for the calculated statistics. Nonparametric tests are appropriate when you are not sure whether the normality assumption is accurate, particularly in small samples. The Mann–Whitney U test (also known as the Wilcoxon Rank Sum test) is a nonparametric extension of the parametric t-tests and Z-tests. The null hypothesis for this test is that two samples are drawn from the same population against the alternative that they are drawn from different populations. Two nonparametric generalizations of the paired samples t-test are the sign test and the Wilcoxon Signed Rank Test. The Kruskal–Wallis test is used to compare medians among multiple comparison groups. Some details of such tests, which may be very useful for managers wishing to learn from data sets that are not necessarily normally distributed may be found in documentation accompanying the statistical programs that perform these tests, but see, for example Sullivan (n.d.) for a more complete description of the tests given herein or Gibbons and Chakraborti (2011) for a more advanced treatment.

Correlation

Two variables are said to be correlated when they move systematically together in the same direction or opposite to each other. If they move in the same direction, the "correlation coefficient" is a positive value between 0 and 1; when they move opposite each other, the correlation coefficient has a negative value between 0 and -1. Again, while it is possible to compute these values by hand, it is usually best to use a statistical computer package.

Note that two values can be correlated but not derived from a cause and effect relationship. For example, it is possible for an organization to use both more volunteers and more paid labor when demand for its services increases. While both paid and volunteer labor may increase at the same time (leading to a positive correlation coefficient between the time series of each variable), one is not causing the other. Rather, these changes come about because of an increase in demand for the organization's services. To analyze causation, one needs a model derived from theory and the use of more sophisticated statistical techniques (see Angrist and Pischke, 2014).

Regression

The hypothesis tests discussed above analyze information about a single variable or the relationship between two different variables. Similar analysis can be extended to discern how several different variables interact. For example, how are the outcomes of an after-school program influenced by the educational backgrounds of a child and the child's parents? In *regression* analysis the relationship between variables is studied while holding the influence of other variables constant. As noted below multivariate regression analysis is an important analytical tool that economists use to estimate economic relationships from empirical data.

4 Econometrics

Managers of nonprofit organizations and social enterprises require information such as knowledge about the demand curve for their services. However, getting this information may be difficult. Some obvious approaches may not yield reliable estimates. Asking people how much they are willing to pay for a service may lead to a situation of "moral hazard," where it is in the respondent's own interest to underestimate what they are willing to pay, knowing that their response will have a direct effect on their wallet. Two other approaches are to analyze natural data using econometrics (this section), the science of statistics applied to economics, or to analyze data from a controlled experiment (the last section of this Appendix).

Simple and multiple regression

Regression analysis is a set of statistical techniques used to quantify the relationship between one dependent variable and one or more explanatory variables and is the basic workhorse of econometrics. The dependent variable is something you would like to explain or predict, and the explanatory variables do just that. Some synonyms, or near synonyms, for explanatory variables are *regressors, covariates,* independent variables, or right-hand variables. Different formulas are used to compute this relationship under different circumstances; an *estimator* is a particular choice of formula. The most common model is the linear one, which, with one explanatory variable, looks like this:

$$D_i = \alpha + \beta X_i + \varepsilon_i$$

where: D is the dependent variable, for example, donations, plotted on the vertical y-axis

X is the independent variable, for example, income, plotted on the horizontal x-axis;

α is an unknown constant representing the y-axis intercept for the regression line (average donations for those with zero income);

β is an unknown constant representing the slope of the regression line (the increase in average donations when income increases by one unit);

ε is a variable known as the *error term* and represents the difference between the amount of donations predicted by the model and the actual amount of donations for each observation in the data;

i is an index for observations. If there are I observations, i takes each value from the set $\{1, 2, 3, \ldots, I\}$.

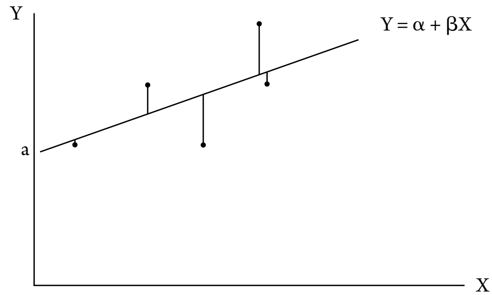

Figure A.2 Ordinary least squares regression

Standard econometrics software requires that you supply a data set containing the variables D and X and specify that you want D to be the dependent variable and X to be an independent variable. The program will calculate estimated values for the two parameters α and β by calculating the line that best fits the scatter-plot of the data. By far, the most common estimator is *Ordinary Least Squares* (OLS), which finds the values of α and β that minimize the sum of the squared deviations between the observed and predicted values of the dependent variable (see Figure A.2).

The α and β estimates produced by a regression are typically symbolized by placing a caret (hat) above the corresponding Greek letter. So $\hat{\alpha}$ (alpha hat) is the estimated intercept of the regression line, and when certain assumptions about the data are met, is the best approximation you can obtain to the true population intercept α, and $\hat{\beta}$ (beta hat) is the best approximation to the population slope β. Because the estimates are sample-dependent, $\hat{\alpha}$ and $\hat{\beta}$ are random variables, and we can base hypothesis tests on the values they take. We might want to test whether there is any linear relationship between income and donations, which amounts to a test of H_0: $\hat{\beta} = 0$. The computer software will calculate the standard error and run a t-test automatically, reporting a p-value or a significance level.

In the example above, donations are a function of a single explanatory variable, but there are many other variables that also explain donations. Regression with one explanatory variable is called *simple regression*, so next we consider *multiple regression* with two explanatory variables. Suppose we also have data on the age of the household head. We could then estimate:

$$D_i = \alpha + \beta_1 X_i + \beta_2 A_i + \varepsilon_i$$

where: A is the age of the household head;

β_1 is an unknown constant representing the slope of the relationship between income and donations while age is held constant;

β_2 is an unknown constant representing the slope of the relationship between age and donations while income is held constant.

all other variables are as defined previously.

When this equation is estimated, the software will compute a plane of best fit in a three-dimensional scatter-plot of the data, with income and age on the x- and y-axes and donations on the z-axis. $\widehat{\beta_1}$ is the estimated slope of a slice through the plane that holds age constant, interpreted as the marginal impact of income on donations, while $\widehat{\beta_2}$ is the estimated slope of a slice through the plane that holds income constant, interpreted as the marginal impact of age on donations.

Comparing our simple and multiple regression examples, we get a more accurate estimate of the independent effect of income on donations in the latter. This is because income and age are somewhat related, as average wages generally rise with age. When we don't control for age, some of the effect of age on giving is wrongfully attributed to income. This is an example of *omitted variable bias*, which occurs whenever you omit a relevant explanatory variable from your estimation. Here, a relevant omitted variable is one that (a) is correlated with the dependent variable, and (b) is correlated with one or more of the included explanatory variables. In contrast, our multiple regression calculates the independent effect of income after controlling for age using the independent variation of each explanatory variable.

There is no particular reason to suppose the relationship between income and donations is best summarized by a straight line, and the pattern of error terms may suggest a particular curved function better fits the data. Consider the following example:

$$D_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

In this equation, we are finding the parabola of best fit between income and donations. The alphas and betas will be estimated by your software for your particular data set but suppose that it just so happened that $\hat{\alpha} = 1$, $\widehat{\beta_1} = 6$, and $\widehat{\beta_2} = -1$. Then our estimated relationship between income and donations would be the parabola already illustrated in Figure A.1. That parabola starts with a steep slope, which would be interpreted here as a large marginal effect of income on donations for low levels of income, and that flattens and turns downward sloping, so the marginal effect decreases then turns negative as income increases. The scale of this parabola is such that this would be a very silly result to obtain – donations are not maximal when income = $3, and a person earning $5 would not give less than a person with income of $3. But let's change the scale a bit. Suppose we are measuring income (on the horizontal axis) in dollars per second. There are 3600 seconds in an hour, so with this scale, donations are maximal when income is $10,800 per hour, a well-paying job indeed, but the parabolic shape is now a whole lot less silly. This parabola still has the problem of predicting something we don't believe for people earning more than $10,800 per hour, but the parabola may fit the data reasonably well for everyone in our sample, which is unlikely to contain any super-high earners, and so this form is often used.

Another curve used instead of a straight line is this one:

$$\ln Q_i^D = \alpha + \beta \ln P_i = \varepsilon_i$$

This natural log/natural log specification is perfect for estimating demand curves that have constant price elasticity, and as we argued in Chapter 9, $\hat{\beta}$ is an estimate of the price elasticity.

When and why should ordinary least squares be used?

Recall that an estimator is the formula used to fit the curve to the data, and that OLS, which minimizes the sum of the squared error terms, is the most commonly used estimator. It is time to unpack that statement. First, some background on desirable properties of estimators:

- Bias: An estimator is biased if the expected value of the parameter is different than the population parameter. This means if we compute alphas and betas from an unlimited number of independent samples, the average values of these coefficients should be the same as the intercepts and slopes of a regression line fitted to the population as a whole. All else equal, we want an unbiased estimator, although sometimes in more advanced econometrics, we are willing to trade a little bias against more precision.
- Efficiency: An estimator is efficient if it has the smallest possible standard errors for the parameter estimates. This means that the "plus/minus" range for the slope and intercepts is as small as possible. We note that this is a vastly simplified definition designed to get the main idea across, and standard textbooks will give you a more nuanced definition.

- Consistency: An estimator is consistent if, as sample size increases, parameter estimates converge on their population values.

OLS estimates can be shown to be unbiased, consistent, and efficient (within the class of linear unbiased estimators) when the data satisfy certain assumptions:[6]

- Sample size: You cannot pin down the slope and intercept if you only have a single observation. Any line drawn through that point would have perfect fit. While it is always better to have a large sample, the minimum number of observations before you can use any estimator is equal to the number of regressors plus one.
- Multicollinearity: when two variables always move together, they are said to be perfectly collinear. When the way in which one variable moves is always determined by a linear combination of changes in other variables, they are said to be perfectly collinear. For example, suppose our data consisted of couples with no children, and we wanted to find the impact of income from the husband, impact of income from the wife, and income of the household on giving. Our explanatory variable household income is perfectly collinear with the other two variables because household income is defined as income of the husband plus income of the wife. In order to estimate coefficients for each of our betas, our data cannot be perfectly collinear. This assumption has a less obvious implication – in order to avoid perfect multicollinearity, each of our regressors has to vary independently within the sample. Such variation is a good thing to have, because we obviously could not estimate the effect of income on donations from a sample where everyone had the exact same income.

These first two assumptions are needed for any estimator. The next two assumptions are necessary and (together with the first two) sufficient to establish that OLS is unbiased:

- Random Sample: Your sample is obtained in a way that makes it representative of the population you are trying to model. This is an issue, for example, in telephone surveys. It is cheap and easy to randomly select land-line phone numbers, but costlier and more problematic to randomly select from all phone numbers, including cellular. If you run a cheap survey selecting from land-line phone numbers, your results will not necessary be typical for the population.
- Zero Conditional Mean: The error term represents the effect of all the determinants of the dependent variable that are not included as

explanatory variables in your regression. The zero conditional mean assumption, intuitively, is that none of the excluded variables (hidden in the error term) covary with the included ones, and also that all relevant variables are included so that no bias results. More precisely, this assumption states that the expected average error conditional on the included variables in the regression is the same as the expected average unconditional errors, and that both are equal to zero.

Our simple regression example above fails the zero conditional mean assumption because it leaves out age, which is correlated with both income and donations. The multiple regression example does better here, although there are still many excluded variables that covary with both donations and at least one of the variables income and age.

In order to satisfy the zero conditional mean assumption, the curve you estimate has to have the correct shape. Suppose that the true relationship between income and donations is a parabola, but you mistakenly fit a straight line to the data, omitting the income squared term. Income squared is correlated with income and is also a determinant of donations, thus correlated with the error term, so the assumption fails. The worst case here would be if your data contains an equal number of observations where the income level is on the upward-sloping part of the parabola as on the downward-sloping part. Then you might get a horizontal straight line and your test on the slope would be unable to reject the null hypothesis. You would wrongly conclude that there is no evidence of a relationship between income and donations.

One additional assumption insures that OLS is also efficient:

- Homoskedasticity: When the expected variance of the error term is the same for all possible observations, the error process is called *homoskedastic*. When there is a systematic difference in the variance of the error term associated with any of your regressors, the error process is heteroskedastic.

When all these assumptions hold: sufficient sample size, absence of perfect multicollinearity, zero conditional mean, and homoskedasticity, OLS produces unbiased estimates that are efficient within the class of linear unbiased estimators. This is the Gauss–Markov theorem (useful to know if you want to show off) and is often acronymically summarized: OLS is BLUE (Best Linear Unbiased Estimator).

If the first two assumptions are violated, then the only way to fix this is to get a new sample. But there are other known estimators that have good properties when the last two assumptions are violated. For example, heteroskedasticity can be remedied using EGLS (Estimated Generalized Least Squares) or the White/Huber Heteroskedasticity Robust estimator.

One additional assumption is often suggested – that the error term is normally distributed. This assumption is not necessary for OLS to be unbiased and efficient but helps to specify the hypothesis tests. The normality assumption is critical when the zero conditional mean assumption is violated.

An example

To do these calculations, we use a statistical package to create the resulting equation. There are many such packages, including Excel, Minitab, Stata, SAS, and SPSS. As Excel is the statistical package that is almost always available to managers, the following illustration uses EXCEL. The interpretation is very similar for other statistical packages, but one should consult the instructions and documentation that accompany any such package for details.

To run a regression, first enter data into an EXCEL spreadsheet. To perform linear estimation, make sure that the program includes the LINEST function in EXCEL. If it is not immediately available, add it to the package.[7]

Under "Data Analysis," choose "regression" and highlight the appropriate columns to use as dependent (y) variables and independent (x) variable(s). At least one independent variable must take on all possible values on the number line, and not be limited to values such as 0, 1 or a collection of a limited number of values, as with a Likert scale. Once dependent and independent variables are entered, the computer will give you output that describes the relationship between the dependent variable and the independent variable(s).

An example of the results from a regression performed on EXCEL asking about the relationship between age and education is shown in Table A.1. This example uses data from the 1993 "General Social Survey" or "GSS" The regression estimates a relationship between age and educational level of respondents to this survey. This survey is collected by the government and is most commonly used by social scientists. While this example uses only one independent variable, the concepts presented here are similar to the results found from multiple regressions.

Table A.1 EXCEL regression – relationship between age and education

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.132246
R Square	0.017489
Adjusted R Square	0.014192
Standard Error	2.977084
Observations	300.00
S	5.30

ANOVA TABLE

	df	SS	MS	F	Significance F
Regression	1	47.01401	47.01401	5.304508	0.021959
Residual	298	2641.183	8.863029		
Total	299	2688.197			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Upper 95%	Lower 95%
Intercept	15.050	0.526	28.56	2.48E-87	14.01	16.08	14.01	16.086
Variable (Age)	−0.0244	0.0106	−2.303	0.02195	−0.045	−0.003	−0.045	−0.0035

The output from this regression is a line representing the relationship between age and education level. Putting the coefficients together gives the estimated line (linear equation) as:

Education = 15.050 – 0.0244 (**Age of respondent**)

Note that this equation indicates that for each additional year of life, the level of education actually falls in this data set, which may seem odd. However, think about who probably answers this survey, conducted over the phone to a sample of people in the U.S. It is likely that it is mainly older people, and that higher levels of education are less likely to appear in the responses of this population. (The proliferation of college degrees and advanced degrees is only a recent phenomenon, and these degrees are less likely to be found in people in the later years of their lives, the very people who may be most likely to answer such a survey.) When looked at from this perspective, the negative coefficient on Age is not surprising. Note also that the constant, 15.050, reflects all the information not determined by the variable Age.

Several other pieces of information in this output require comment. "R square" is a measure of the percentage of the variation in the dependent variable that is explained by the independent variable(s). Here, R^2 is equal to 0.017489, so less than two percent of the variation in education has been explained. However, this just means that there are probably other variables besides Age that help determine the value of a respondent's education.

Next, note the ANOVA part of the printout. While an ANOVA test can be done directly on two variables, the ANOVA test as part of a regression tests the hypothesis that the entire regression is useless. Of particular interest in this output are the values presented as "p-values" (or, in some programs, "sig" values). These numbers ask what the probability is that we would find such results, if there was, indeed, no relationship between the variables being studied. In econometrics, we usually look for p-values to be at most 0.05, or even less. That means that there is still a 5 percent chance that there is no relationship between the variables being studied, given the statistical results that emerged from the regression. Here, the p-value for this test is reported as 0.021959. This means that there is only about a 2 percent chance that the resulting output would be found if there is actually no relationship among these variables. One can thus conclude that the hypothesis that there is no relationship between the variables must be rejected.

Finally, note the p-value reported for the independent variable. The p-value of 0.021959 on Age leads to a rejection of the hypothesis that this variable has no effect on the dependent variable, Education. We therefore conclude that Age is a statistically significant variable for the determination of a respondent's level of education.

Other estimators

Special kinds of data require special estimators. Here is a catalog of the most important ones:

1. If your dependent variable can only take the values 0 and 1. There are three estimators that are commonly used to estimate slopes and intercepts when the dependent variable has two categories. For example, we might want to test hypotheses about who will make a donation. For this, or any other yes/no question, we would code the dependent variable as a 0 if the person did not make a donation and a 1 if the person did make a donation. For such data, you would select between the linear probability model (which is what OLS is called for this kind of data), logit, or probit. Any of these three can estimate the marginal effect of a variable on the probability that the dependent variable will take the value of 1. For example, if we explain the yes/no donation decision using household income as an explanatory variable, we can use the estimates to compute the effect of a $1 increase in household income on the probability that a household will make a donation.

2. If your dependent variable is censored. A variable is said to be *censored* when it is only reported for a subset of the range of values taken by a statistical distribution. When a variable is normally distributed, there will be a nonzero probability that it takes any value, positive or negative, but variables like donations cannot take negative values. One approach to dealing with donations regressions is to assume that there is a latent (hidden) variable that can take positive or negative values, and reported donations are the greater of $0 or the value of the latent variable. When "donations" in the latent variable are negative, this information is censored and a zero is reported. Another type of truncation occurs when a variable is "top-coded." For example, survey respondents may be asked to report their actual donations if less than $1 million or check a box for "$1 million or more" for larger donations. When the data is censored, regressions will violate the zero-conditional-mean assumption, particularly for observations around the censoring point or points, and OLS estimates are said to suffer from "censorship bias."

There are many estimators designed to eliminate censorship bias, and it is beyond the scope of this book to discuss the advantages and disadvantages of each estimator. At this time, the predominant estimator in the literature is Tobit, which uses the latent variable artifice. Another latent variable approach is non-parametric, CLAD (Censored Least Absolute Deviation). Others abandon the normal distribution assumption and the fiction of hidden negative donations, such as the GLM-Gamma estimator (Generalized Linear Model with a gamma link function). Some are specially tailored for duration/survival analysis, with more complex censorship issues. For example, if we want to study how many donors continue to give every year for one year, two years, three years, and so on, and our data collection ends in year five, duration of donating is censored for anyone giving continuously for five or more years. The basic estimator for this is the Cox Proportional Hazards model, and many more complicated models are available.

3. If your dependent variable is a set of categories. Suppose you wanted to explain whether a household will make a donation of money only, a donation of time only (volunteering), donations of both, or donations of neither. We have four categories, and there is no natural ordering between them that enables us to assign numbers like 1, 2, 3, and 4 for the dependent variable. In this case, the common estimators are multinomial logit or multinomial probit.

Regression results can be translated into marginal probability statements, as in regular logit or probit, but now there are three independent marginal effects to calculate. For example, if income was an explanatory variable, we could estimate the marginal impact of a $1 increase in income on the probability the household will give money only, the probability they will give time only, and the probability they will give both. The three independent probabilities give us the fourth probability (that the household will give neither) for free, since the sum of the probabilities must equal 100 percent.

A related problem occurs when the dependent variable is a set of categories that can be put into a natural order. For example, our survey may have asked households to answer a multiple-choice question like "I gave: (a) nothing; (b) between $0 and $100; (c) between $101 and $1000; (d) more than $1000." The common estimators here are called Ordered Logit and Ordered Probit.

Conclusion

You can learn a lot from the statistical analysis of naturally-generated data. The main problem with such data is that everything is changing at once, but regression methods allow the analyst to sort out the independent effects of each explanatory variable for prediction and hypothesis testing.

5 Techniques in experimental economics

Experiments using human subjects provide an alternative to econometrics, with advantages and disadvantages. In this section, we provide some details on how to conduct experiments that can help nonprofit and social enterprise managers with their work. First, we review the main ideas of an experiment, then we discuss recruiting and using subjects, and highlight some issues of experimental design. We follow with sections on analysis of results, and conclusions. As with the other sections of this Appendix, a full course on the topics is highly recommended.

Main ideas

An experiment starts with two or more groups of subjects. The groups differ from each other on the basis of a single characteristic under the control of the experimenter. Most commonly, the groups are called the treatment group and the control group, the former distinguished by having a characteristic called the treatment. For example, one experiment used people collecting money for the Salvation Army during the holiday season (Andreoni, Rao, and Trachtman, 2017). In the control group, the experimenters used the current Salvation Army protocol, with someone dressed in a Santa suit ringing a bell and waiting for donations. The treatment group used the same method but asked the bell ringers to specifically ask shoppers to make a donation. The authors found that asking had a dramatic effect on the amount given when asked, but also on the tendency of shoppers to avoid being asked.

Each group has the same measurable outcome of interest, which may be a continuous or categorical variable. The main goal of the experiment is to estimate *treatment effects,* which are the differences in outcomes between each pair of groups. Experimenters assign subjects to each group randomly, and this provides the key advantage of experiments over econometric analysis of naturally-generated data. *Random assignment* greatly reduces the possibility of *self-selection bias* that infects many econometric analyses.

In natural data, people typically choose whether to experience the treatment, and they do so for reasons that might correlate with the outcomes. For example, we might use natural data to see whether commercial nonprofits that also ask for donations are better-off financially than commercial nonprofits that do not ask for donations. Asking for donations could be a treatment in an experiment but it is also a binary variable that you would want to use in a regression explaining the financial status of commercial nonprofits. The trouble in the latter case is that asking for donations is a random variable that is correlated with the error term, violating the assumption that the conditional mean of the error term is zero. Intuitively, commercial nonprofits may choose to fund-raise because they are in a great financial environment, where people are wealthier and supportive of the cause. This great financial environment results in a spurious correlation between financial success and use of fund-raising.

Experiments eliminate self-selection bias by using random assignment. For example, the experimenter could select subjects from a pool consisting of commercial nonprofits that did not fund-raise, then randomly assign a treatment that caused the treatment group to begin fund-raising. Now, use of fund-raising is not correlated with any of the determinants of financial success, and the problem would seem to be eliminated. The only caution is that sometimes attrition – where some organizations stop participating before the experiment is complete – would be correlated with the error term (the prospects for financial success). Participation payments of sufficient size to subjects in the experiment would reduce or eliminate attrition and the possibility of attrition bias.

To isolate the treatment effect, the groups should be treated as identically as possible, differing only in the treatment each receives. The experimenter can control many things, but not all determinants of the outcome. For example, it is known that many households concentrate their giving in the month of December. When running an experiment with a treatment designed to increase donations, the experimenter can avoid statistical bias by simply having both the treatment and control sessions in December, or by having both in months other than December. Other differences between the two groups are harder to deal with because random selection does not result in the same average level of subject characteristics in the treatment and control groups. For example, men and women respond differently to incentives to give (e.g., Croson and Gneezy, 2009), and when there is random assignment, the proportion of females in the control group is unlikely to be exactly equal to the proportion of females in the treatment group, confounding the estimation of the treatment effect. This problem can be fixed several ways – by conducting separate experiments for men and for women, or by random assignment with gender quotas (which simply means that if the number of males assigned to a group is equal to the experimenter-determined limit, only females will be assigned to that group thereafter, and similarly if the limit for females is reached). Other factors are not measured, like individual-subject beliefs, attitudes, emotional states, and experience, so you cannot conduct separate experiments or use random assignment with quotas to solve the problem. Although the average level of each of these characteristics is unlikely to be exactly equal across groups, the experimenter deals with these problems by placing a large number of subjects in each group, which makes average differences small.

In sum, a true experiment has the following characteristics: two or more groups, random assignment of subjects to groups, one or more treatments applied to some groups but not others, the treatments are under the control of the experimenter, and the only difference between groups, among those differences that the experimenter can control, is in treatments. Variations on these themes distinguish several types of experiments, and we shall consider three of them: lab experiments, field experiments, and natural experiments.

Lab experiments are most like true experiments. Economic decisions take place in a tightly-controlled environment, where the experimental design insures minimal differences between groups other than the treatment. Subjects in a lab experiment are aware that they are participating in an experiment and have given informed consent to do so. The classroom experiment on free riding detailed in Chapter 12 is an example of a lab experiment.

Field experiments use subjects who are unaware of their participation in an experiment. Subjects make their decisions in a natural environment as they go about their lives. For example, Landry et al. (2006) approached 5000 households in a door-to-door fund-raising campaign, with each household randomly assigned to one of four treatments.

A natural experiment occurs when natural data is generated by a process approximating random assignment. The treatment is not under the control of the experimenter, so this is least like a true experiment. To approximate random assignment, the experimenter must persuade the reader that there are no variables that both determine whether the treatment is imposed and also what the outcome will be. For example, Zhang and Zhu (2011) looked at whether the reduction in group size when mainland China blocked access to the Chinese Wikipedia affected contributions (volunteering to write and revise content) by comparing residents in China with residents of Taiwan, Hong Kong, Singapore, and other regions of the world where people aware of the Chinese Wikipedia continued to have access.

Subjects

1. How many subjects are needed? It is always better to have more subjects than less, but larger studies cost more. The number of subjects in each group determines the statistical distribution of group-average outcomes. In many cases, group-average outcomes will be well-approximated by the normal distribution as long as there are 30 or more subjects in each group. But the number of subjects also determines the statistical power of tests on the size of the treatment effect – that is, the ability of the test to reject the null hypothesis (typically, that the treatment effect is zero) when the null hypothesis is false. Statistical power is inversely proportional to the square of the number of subjects in each group, so in order to cut the standard error of effect size in half, you need four times as many subjects in each group. Many statistics packages include software that tells you the sample size necessary to obtain a given level of statistical power. More generally, group sizes should be related to the ratios of expected variances in the different groups and/or the relative cost of collecting data from a group (List, Sandoff, and Wagner, 2011). These all account for sampling error, but there is an additional effect of group sizes on the homogeneity of groups. Recall that in order to control for unmeasured differences in the characteristics of groups, we need a large number of subjects. This effect also follows an inverse square law, but it is worth commenting on because sometimes treatment effects are different across individuals because there is an interaction between the treatment and an unmeasured variable. Accounting for all of this, when the budget is very limited, you need at least 100 subjects in each group. That number is arbitrary but informed by experience.

2. Who should be in your subject pool? Designing an experiment requires careful thought about who participates in the experiment, and how their participation may affect the outcome of the experiment. Many laboratory studies use students as a convenient subject pool, but students are hardly representative of the population of greatest interest in most studies. For example, bonuses might affect students differently than they affect non-profit workers or major corporate executives. Ideally, you want subjects that are randomly drawn from the population you are studying – line workers if you want to study the effect of bonuses on line workers, major executives if you want to study the behavior of this group. But it is cheaper to work with students and certainly more convenient, and many researchers make that trade-off, reducing the persuasiveness of results in order to keep things affordable.

In many circumstances, variations on pure random sampling may be necessary or preferred. It may be desirable to "oversample" certain categories of subjects so that the subsamples are large enough to analyze separately. We discussed random assignment with quotas above, but the problem goes beyond the random assignment causing mismatches in unmeasured characteristics. For example, if you want to know whether donations by Jews or by Protestants in the U.S. respond the same way to matching incentives, a sample of 200 subjects would be expected to contain only 2.8 Jews, not nearly enough for proper statistical analysis. For such a study, you would employ *blocking*, which, in experiments, means the same thing as stratified sampling in survey methodology. In this case, you would want a Jewish block and a Protestant block of comparable sizes, then you would use random assignment to treatments within each block.

3. Ethical considerations. Experimental research using human subjects requires preapproval in many countries to ensure that subjects will not be harmed. In the U.S. (and many countries have followed suit), this generally takes the form of approval by IRBs (Institutional Review Boards) set up at college campuses and other institutions that conduct federally funded research. Under the law, the requirement governs research funded by the federal government, but in practice it has come to be applied to any research conducted by college and university staff, and, regardless of employment status, to any research that the researcher would like to publish in an academic journal. IRBs require that risks be minimized and reasonable in light of anticipated benefits to subjects, that vulnerable populations receive special review, that informed consent be obtained from each subject, and that data be maintained in a way that protects the privacy and confidentiality of subjects. Generally, subjects need to receive a payment for the use of their time, else they would suffer unnecessary harm in the form of opportunity costs.

Experimental design

Experimental design concerns planning the details of an experiment to minimize challenges to the validity of findings and to weigh cost versus quality issues. We have discussed some elements of experimental design above; in this section we will talk about the method of applying treatments, steps to include in your experiment, and the order of these steps.

When designing an experiment, the researcher must choose whether to use a *between-subjects design* or a *within-subjects design*. In a between-subjects design, different people (or organizations) are in the treatment group and any other groups. Thus, we must rely on the law of large numbers (differences in average characteristics of groups shrink as the number of subjects in each group increases) to claim that the only factor that explains differences in outcomes is the presence or absence of a treatment.

In a within-subjects design, the same people (or organizations) serve in both groups. We calculate a treatment effect for each subject by looking at the difference in outcomes when that subject receives the treatment versus experiencing control conditions. The overall treatment effect is then the average of individual-specific treatment effects. This design does not need the law of large numbers to isolate treatment effects from most confounding factors, because most of the unmeasured variables remain constant over time for the subject and therefore cannot explain any difference in outcomes. But the order of the two conditions matters. Perhaps subjects respond more strongly to whichever condition comes first, because they are more interested at the beginning of the experimental session or for any other reason. If so, the order itself could be a confounding variable. Therefore, when using a within-subjects design, we randomly assign the order of conditions (control then treatment; treatment then control) rather than randomly assigning a group.

Sometimes you want to study interactions between treatments. For example, Goering et al. (2011) wanted to know the effect of three factors on donations in response to direct mail solicitation – the persuasion strategy (which can be an appeal to facts, an appeal to emotions, or an appeal based on the authority of the letter-writer), the visual design (with or without bullet points), and the complexity of the letter (at two grade levels for the language employed).

They employed a 3 x 2 x 2 *factorial design* (the numbers indicate the number of treatments in the persuasion, visual, and complexity factors respectively) that included twelve groups – one for each combination of the three factors. For example, one group received a letter that appealed to facts, used bullets, and used simple language; another group received a letter that appealed to emotions, used bullets, and used advanced language. They could have analyzed each factor separately, with three experiments, but chose the factorial design because of the possibility of interaction effects between the factors. For example, the effect of simple language might be different when an appeal is made to the facts than when an appeal is made to emotions; the factorial design allowed them to detect all such interactions. The factorial design is complex and requires a large number of subjects but is often cheaper than one-factor-at-a-time experiments.

Economists traditionally ignored context effects because they were inconsistent with the simplified theory of choice that they had developed, but with the development of behavioral economics that is changing. Increasingly, economics experiments borrow techniques developed in the other behavioral sciences. One such useful if controversial technique is called *priming*. Priming occurs when exposure to one stimulus influences the response to another stimulus or treatment, and the theory of priming includes "how cues that activate . . . the recall of specific social contexts and events alter current preferences and choices" (Molden, 2014, p. 4). A classic priming experiment was conducted by Shih, Pittinsky, and Ambady (1999) who found that Asian-American women score higher on a math test when primed to identify themselves as Asian than when primed to identify as women. Some of the many experimental studies that use priming to better understand donations include Small, Loewenstein, and Slovic (2007), Ekström (2012), and Eckel, Grossman, and Milano (2007).

Sometimes, we want to understand why the priming works the way it does, that is, the mechanism behind the treatment effect. For example, if we want to test whether an altruistic frame of mind affects giving, we might show subjects a video of parents caring for young children. The experiment can show whether watching the video affects giving, but unless the experimentalist can show that the prime actually induced an altruistic state of mind, we have not properly tested the theorized mechanism. In such cases, it is best to include a *manipulation check*, a way of measuring state of mind to see whether the prime is doing what we would like it to do. Psychometric scales, constructed from survey responses, are commonly used for this purpose.

Inexperienced experimentalists often use a pretest-posttest design, which means that the outcome variable of interest is measured both before and after the experimental condition (treatment or control) is administered. That is unnecessary with random assignment and large groups, because we can safely assume that the treatment and control groups started with the same level of the outcome. Experiments only need to measure the effect by comparing the average outcome in each group after the conditions are administered.

A good logical order for the steps of the experiment is:

- obtain informed consent,
- provide instructions for the experiment and/or any necessary priming,
- conduct manipulation checks to see that the instructions are properly understood, and that the primes have their desired effect on perceived context and state of mind,
- administer the treatment and control conditions,
- measure outcomes,
- pay the subjects in a way that preserves their anonymity, and
- if desirable, administer a post-experiment survey or interview.

But the experimenter should carefully consider whether this ordering is the best way to proceed for the experiment that will be conducted. The order of the steps might bias the results, affecting validity and interpretation. For example, suppose (as above) the manipulation check occurs before the experimental conditions are administered. Then there is the possibility that the manipulation *check*, rather than the manipulation itself, explains the results. If, instead, the experimental conditions come first, then perhaps the success or failure of the manipulation check is determined by experience of these conditions rather than the initial priming step.

Because experiments are expensive and experimental design is a complicated subject, you should always run a pre-test of the design on a small, independent, sample of subjects. Each subject should be carefully debriefed after the pretest to see how well each instruction was understood and to see whether the subject's explanation of their behavior suggests that the experiment is testing what you want to test.

Analyzing results

A first step to analyzing experimental data is to create tables that display the mean outcome for each treatment or condition, along with standard errors. This may be displayed as a bar chart with standard error bars (lines centered atop the results bars that display the height of the bar plus or minus a standard error) or confidence interval bars (like standard error bars, but extend roughly two standard errors above and below the bar at the 95% confidence level). The extent of overlap between the two confidence bars can show whether the treatment effect is statistically significant, but to be sure, you should use the formal statistical test, particularly in a within-subjects experimental design (Belia et al., 2005). Another useful set of graphs is a histogram (a chart illustrating the frequency that a variable takes different values) of individual-subject outcomes within each group. This can show whether there are subtler patterns in the two groups that do not show up in comparisons of averages. For example, if half of the people in the control group donate $0 and half donate $100, average control group donations are $50. If everyone gives exactly $50 in the treatment group, the average treatment effect is $0. But without the histograms, the analysis would miss the fact that the treatment effect is +$50 for half the sample and -$50 for the other half. The within group histograms can also reveal whether an outlier is responsible for an apparent treatment effect. If Bill Gates were in the treatment group in a between-subjects design, then it will always look like there is a positive treatment effect on donations (one billion-dollar gift really messes up the average) whether a treatment effect exists in the population or not. Tests for differences in median outcomes, rather than mean outcomes, would be desirable in this case.

The central limit theorem, discussed in the Appendix section on statistics above, does not always apply to group averages generated in experiments, particularly for truncated outcomes where individual outcomes are concentrated around the truncation point. The histograms reveal such non-normality and suggest that nonparametric tests of the null hypothesis are more appropriate.

If you want to know the effect of measured variables that are not under experimental control (like subject income or religious denomination) you would use regression analysis to calculate these effects. Controlling for variables that are only approximately equal across groups also improves the efficiency of the estimated treatment effect, although without covariates, treatment effects are unbiased. Finally, some disciplines emphasize ANOVA and related procedures rather than regression results, for testing null hypotheses.

Conclusion

No single study is fully persuasive, and each study faces challenges to internal and external validity. Although there are certainly exceptions, econometric analysis of natural data has better external validity because it includes a range of conditions and characteristics, whereas analysis of experimental data has better internal validity because interdependent variables are not varying and covarying simultaneously; instead there are one-at-a-time changes known as treatments. Both kinds of studies have advantages and should be conducted when time and money are not obstacles for the researcher.

We have certainly not exhausted the subject of experiments. We haven't covered the choice between conducting a lab or a field experiment, and the two have different sets of validity challenges. Nor have we covered the choice between within-subject and between-subject designs. We haven't catalogued all the challenges to validity and discussed ways to deal with each challenge. Finally, we haven't discussed quasi-experimental methods, which attempt to statistically adjust naturally-generated data, so it can be analyzed like an experiment. Instead, we have tried to give the social entrepreneur or nonprofit manager enough background to understand new and useful results in behavioral economics, and the ability to design (perhaps with help) experimental studies particular to the management challenges he faces.

SELECTED REFERENCES AND CITATIONS

Allen, W.B., Doherty, N.A., Weigelt, K., and Mansfield, E. (2013). *Managerial Economics, Theory, Applications and Cases* (8th ed.). New York: W.W. Norton & Company.

Andreoni, J., Rao, J., and Trachtman, H. (2017). Avoiding the Ask. *Journal of Political Economy, 121(3)*, 625–653.

Angrist, J.D. and Pischke, J-S. (2014) *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press.

Ariely, D. (2008). *Predictably Irrational*. New York: Harper Collins Publishing.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological methods, 10*(4), 389.

Bengtsson, N. and Engstrom, P. (2013). Replacing Trust with Control: A Field Test of Motivation Crowd Out Theory. *Economic Journal, 124 (577)*, 833–858.

Bloom, H. (Ed.) (2005). *Learning More from Social Experiments, Evolving Analytic Approaches*. New York: Russel Sage Foundation.

Blundell, R. and Costa Dias, M. (2002). Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources, 44(3)*, 640–654.

Chiang, A.C. and Wainwright, K. (2004). *Fundamental Methods of Mathematical Economics* (4th ed.). Boston: McGraw-Hill/Irwin.

Croson, R., and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature, 47(2)*, 1–27.

Eckel, C., Grossman, P.J., and Milano, A. (2007). Is More Information Always Better? An Experimental Study of Charitable Giving and Hurricane Katrina. *Southern Economic Journal*, 388–411.

Ekström, M. (2012). Do Watching Eyes Affect Charitable Giving? Evidence from a Field Experiment. *Experimental Economics, 15(3)*, 530–546.

Friedman, D. and Sunder, S. (1994). *Experimental Methods, a Primer for Economists*. Cambridge, UK and New York: The Press Syndicate of the University of Cambridge.

Gibbons, J.D. and Chakraborti, S. (2011). *Nonparametric Statistical Inference* (5th ed.). Boca Raton: Chapman and Hall/CRC.

Goering, E., Connor, U.M., Nagelhout, E., and Steinberg, R. (2011). Persuasion in Fundraising Letters: An Interdisciplinary Study. *Nonprofit and Voluntary Sector Quarterly, 40(2)*, 228–246.

Gujarati, D. (2010). *Essentials of Econometrics* (4th ed.). New York: McGraw-Hill Publishing.

Harrison, G. and List, J. (2004). Field Experiments. *Journal of Economic Literature, 42(4)*, 1009–1055.

James, E. (1998). Commercialism Among Nonprofits: Objectives, Opportunities and Constraints. In Weisbrod, B.A. (Ed.), *To Profit or Not to Profit* (pp. 271–285). New York: Cambridge University Press.

Johnston, J. (1984). *Econometric Methods* (3rd ed.). New York: McGraw-Hill Book Company.

Kagel, J. and Roth, R. (1995). *Handbook of Experimental Economics.* Princeton, NJ and West Sussex, UK: Princeton University Press.

Kephart, C. (n.d.). Interpret Regression Coefficient Estimates. Retrieved from http://www.cazaar.com/ta/econ113/interpreting-beta (accessed December 23, 2017).

Landry, C.E., Lange, A., List, J.A., Price, M.K., and Rupp, N.G. (2006). Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. *The Quarterly Journal of Economics, 121(2)*, 747–782.

Levitt, S. (2006). White Color Crime Writ Small: A Case Study of Bagels, Donuts, and the Honor System. *American Economic Review, 96(2)*, 290–294.

Levitt, S. and List, J. (2009). Field Experiments in Economics: The Past, the Present, and the Future. *European Economic Review, 53(1)*, 1–18.

List, J.A. (2011). Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives, 25(3)*, 3–16.

List, J.A., Sandoff, S., and Wagner, M. (2011). So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design. *Experimental Economics, 14*, 439–457.

Loomes, G. (2005). Modelling the Stochastic Component of Behavior in Experiments: Some Issues for the Interpretation of Data. *Experimental Economics, 8(4)*, 301–323.

Molden, D.C. (2014). Understanding Priming Effects in Social Psychology: What is “Social Priming” and How does it Occur? *Social Cognition, 32(special issue)*, 1–11.

Niskanen, W.A. (1971). *Bureaucracy & Representative Government.* Chicago and New York: Aldine-Atherton.

Shih, M., Pittinsky, T.L., and Ambady, N. (1999). Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance. *Psychological Science, 10(1)*, 80–83.

Small, D.A., Loewenstein, G., and Slovic, P. (2007). Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims. *Organizational Behavior and Human Decision Processes, 102(2)*, 143–153.

Stein, R. (2017). Logs Transformation in a Regression Equation. Retrieved from http://www-stat.wharton.upenn.edu/~stine/stat621/handouts/LogsInRegression.pdf (accessed December 23, 2017).

Steinberg, R. (2006). Economic Theories of Nonprofit Organization. In Powell, W.W. and Steinberg, R. (Eds.), *The Nonprofit Sector: A Research Handbook* (2nd ed.) (pp. 117–139). New Haven: Yale University Press.

Steinberg, W. (2011). *Statistics Alive! Volume 2*. Thousand Oaks, CA: Sage Publishing.

Stewart, J. (2016). *Calculus* (8th ed.). Belmont, CA: Brooks/Cole.

Sullivan, L. (n.d.). Nonparametric Statistics. Retrieved from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/BS704_Nonparametric_print.html (accessed July 21, 2018).

Tuckman, H.P. and Chang, C.F. (2006). Commercial Activity, Technological Change, and Nonprofit Mission. In Powell, W.W. and Steinberg, R. (Eds.), *The Nonprofit Sector: A Research Handbook* (2nd ed.) (pp. 629–644). New Haven: Yale University Press.
Witte, R. and Witte, J. (2009). *Statistics* (10th ed.). Hoboken, NJ: Wiley Publishers.
Wooldridge, J.M. (2016). *Introductory Econometrics: A Modern Approach* (6th ed.). Toronto: Nelson Education.
Young, D.R., Searing, E.A.M., and Brewer, C.V. (2016). *The Social Enterprise Zoo*. Cheltenham, UK and Northampton, MA: Edward Elgar Publishing.
Zhang, X.M. and Zhu, F. (2011). Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review, 101(4)*, 1601–1615.

REVIEW CONCEPTS

Analysis of Variance (ANOVA): An important technique for analyzing the effect of categorical factors on a response. An ANOVA decomposes the variability in the response variable amongst the different factors. Depending upon the type of analysis, it may be important to determine: (a) which factors have a significant effect on the response, and/or (b) how much of the variability in the response variable is attributable to each factor.

Between-Subjects Design: When subjects in an experiment are randomly assigned to groups and some groups experience treatments and others don't.

Block Design: When you want to estimate treatment effects for separate subgroups of the sample, each subgroup is called a block. You recruit a sufficient number of subjects in each block and then randomly assign subjects to treatments within blocks.

Censored Data: Data where the dependent variable is not reported when it falls beyond a range of values. If we imagine a hidden "donation" variable that could take negative values for certain levels of income and family size, the reported measure of donations available to the researcher is zero for all "donations" below zero.

Correlation: The way in which two variables move together. If they move in lockstep, so that every time an observation of one variable is above its average value, the corresponding value for the other variable will be above its average, we have perfect positive correlation and the correlation coefficient takes its maximum value (+1). If they move by contraries, so that in any observation where the first variable takes an above average value, the second variable has a below-average value, we have perfect negative correlation and the correlation coefficient takes its minimum value (-1). When the correlation coefficient takes a value close to zero, there is poor correlation between the variables and there is no correlation at all if the correlation coefficient is zero.

Covariate: A variable used to explain the dependent variable. Synonymous with "explanatory variable" or "regressor."

Derivative: The slope of a function at a point. The limit of the slope of a line segment connecting two points on a function as the points move closer together.

Econometrics: Statistical analysis that is informed by economic theory and incorporates that theory into estimates. Used to test hypotheses, predict the future, and generate "what if" predictions.

Error term: A variable in regression analysis that incorporates the combined effect of all the unmeasured and absent explanatory variables on the dependent variable. The vertical distance between the regression line and each observation.

Estimator: A formula for calculating a statistic used to infer unknown parameters in a statistical model. Applied to regression, the way in which a curve is fit to the data.

Exogenous: An explanatory variable is exogenous if its value is determined by forces that do not also determine the value of the dependent variable.

Experimental Economics: The application of experimental methods to study economic questions.

Factorial Design: Used to study interactions between treatments. Treatments are classified into "factors" with levels. For example, an experiment might want income as a factor, which takes two levels: high and low. They might also want knowledge about the organization as a factor, with four levels: None; Read their form 990 data; Read their annual report; Serve as a Volunteer. The factorial design in this case would set up eight treatments, one for each combination of income and knowledge.

Homoskedastic: A property of the error term needed to ensure that OLS produces the Best Linear Unbiased Estimates. The error term is homoskedastic if the variance of the error is the same for all levels of covariates in the sample.

Multiple Regression: A way of fitting a curved surface to data with one dependent and more than one independent variables.

Nonparametric Statistics: Parametric statistics are random variables with known distributions that can be summarized by a small and definite number of parameters, like the normal distribution that is completely specified by two parameters – the mean and variance. Nonparametric statistics either lack a distribution or have distributions where the number and nature of parameters is flexible and not fixed in advance.

Omitted Variable Bias: Bias in the estimated coefficients of a regression equation caused by the omission of some explanatory variables from the calculation due to lack of appropriate data. Important when the omitted variable is correlated with other explanatory variables and is also a statistically significant determinant of the dependent variable.

Optimization: Mathematical techniques to detect variable values that maximize or minimize an objective function, perhaps subject to constraints.

Ordinary Least Squares (OLS): A regression estimator that assigns parameter values by minimizing the sum of the squared deviations of dependent-variable observations from that predicted by the regression curve.

Population: The complete set of possible observations, which may be infinite in number.

Probability: A number between 0 and 1 (or 0 percent and 100 percent) that describes the likelihood that an event will occur. Probabilities may be subjective or based on observed or theoretical frequencies.

Random Assignment: A process of placing subjects into groups (treatment and control) where each subject has an equal probability of assignment to each group.

Regression Analysis: A set of statistical processes for estimating the relationships among variables. It is a technique used for predicting the unknown value of a variable known as the (dependent variable) from the known value of two or more variables (independent variables) also called the predictors. It includes many techniques for modeling and analyzing several variables and for testing hypotheses about the effect of the independent variables on the dependent variable.

Regressor: A variable used to explain the dependent variable. Synonymous with "explanatory variable" or "covariate."

Sample: A subset of the population for which you have data and use that data in the analysis.

Self-Selection Bias: Bias in the estimated coefficients of a regression equation caused by the fact that the observational unit chooses the value of some covariates using factors that also determine the value of the dependent variable but are omitted from the list of covariates used. This causes a degree of spurious correlation that shows up as a violation of the zero conditional mean assumption.

Simple Regression: A technique for fitting a function to the relationship between a single independent variable and a single dependent variable.

Standard Deviation: A measure of the spread-outness of a variable in a sample or population, calculating as the square root of the sum of squared deviations from that variable's mean divided by the number of observations.

Standard Error: A measure of the spread-outness of a statistic calculated from a sample across samples. The standard error of the mean of a variable is generally equal to the standard deviation of that variable divided by the square root of the number of observations.

Standard Normal Distribution: A normal distribution with a zero mean and a unit standard deviation. The "normal distribution" is a bell-shaped curve indicating the likelihood that certain variables will take values along an infinite number line.

Statistic: A number calculated from a sample, such as the sample mean and variance.

Subjective Probability Distribution: A set of beliefs about the likelihood that a variable will take each possible value. These beliefs are updated by observations of the variable according to the rules of probability.

Treatment Effect: The difference in the average outcome between the treated and control groups. Interpreted as causal for well-designed experiments.

Variance: A measure of the spread-outness of a variable in a sample or population equal to the square of the standard deviation.

Within-Subjects Design: When subjects in an experiment experience all the treatment and control conditions, with random assignment of the order in which the conditions are experienced.

NOTES

1 Technically, the demand function specifies the quantity consumers wish to buy at any Price, that is $Q^D = f(P)$. For this problem, we work with the inverse demand function, that is, we solve for P as a function of Q.

2 See sections on second order conditions from Stewart (2016). This text also provides additional information about maximization in more than one dimension, as when there are several choice variables.

3 The alert reader may notice that we divide the sum of squared deviations by N-1 instead of N when calculating the sample standard deviation s. This is because we use the sample statistic to estimate the population statistic, and need to correct for the fact that the numerator uses an estimate – the sample mean – instead of the true population mean.

4 Not to be confused with "ARNOVA," the Association for Research on Nonprofit Organizations and Voluntary Action, a group worth joining for those working in or studying the nonprofit sector.

5 Note that the alternative is not worded as "all the groups are not equal," as it is possible for two or more of the groups to be equal, with at least one definitely different.

6 This version of the assumptions is based on Wooldridge (2016).

7 To do this, include it as an "add in" through "options" under "file."