CHAPTER 1

An Introduction to Econometrics

1.1 Why Study Econometrics?

Econometrics is fundamental for economic measurement. However, its importance extends far beyond the discipline of economics. Econometrics is a set of research tools also employed in the business disciplines of accounting, finance, marketing, and management. It is used by social scientists, specifically researchers in history, political science, and sociology. Econometrics plays an important role in such diverse fields as forestry and agricultural economics. This breadth of interest in econometrics arises in part because economics is the foundation of business analysis and is the core social science. Thus, research methods employed by economists, which includes the field of econometrics, are useful to a broad spectrum of individuals.

Econometrics plays a special role in the training of economists. As a student of economics, you are learning to "think like an economist." You are learning economic concepts such as opportunity cost, scarcity, and comparative advantage. You are working with economic models of supply and demand, macroeconomic behavior, and international trade. Through this training you become a person who better understands the world in which we live; you become someone who understands how markets work, and the way in which government policies affect the marketplace.

If economics is your major or minor field of study, a wide range of opportunities is open to you upon graduation. If you wish to enter the business world, your employer will want to know the answer to the question, "What can you do for me?" Students taking a traditional economics curriculum answer, "I can think like an economist." While we may view such a response to be powerful, it is not very specific and may not be very satisfying to an employer who does not understand economics.

The problem is that a gap exists between what you have learned as an economics student and what economists actually do. Very few economists make their livings by studying economic theory alone, and those who do are usually employed by universities. Most economists, whether they work in the business world or for the government, or teach in universities, engage in economic analysis that is in part "empirical." By this we mean that they use economic data to estimate economic relationships, test economic hypotheses, and predict economic outcomes.

Studying econometrics fills the gap between being "a student of economics" and being "a practicing economist." With the econometric skills you will learn from this book, including how to work with econometric software, you will be able to elaborate on your answer to the employer's question above by saying "I can predict the sales of your product." "I can estimate the effect on your sales if your competition lowers its price by $1 per unit." "I can test whether your new ad campaign is actually increasing your sales." These answers are music to an employer's ears, because they reflect your ability to think like an economist and to analyze economic data.

Such pieces of information are keys to good business decisions. Being able to provide your employer with useful information will make you a valuable employee and increase your odds of getting a desirable job.

On the other hand, if you plan to continue your education by enrolling in graduate school or law school, you will find that this introduction to econometrics is invaluable. If your goal is to earn a master's or Ph.D. degree in economics, finance, data analytics, data science, accounting, marketing, agricultural economics, sociology, political science, or forestry, you will encounter more econometrics in your future. The graduate courses tend to be quite technical and mathematical, and the forest often gets lost in studying the trees. By taking this introduction to econometrics you will gain an overview of what econometrics is about and develop some "intuition" about how things work before entering a technically oriented course.

1.2 What Is Econometrics About?

At this point we need to describe the nature of econometrics. It all begins with a theory from your field of study—whether it is accounting, sociology, or economics—about how important variables are related to one another. In economics we express our ideas about relationships between economic variables using the mathematical concept of a function. For example, to express a relationship between income and consumption, we may write

$$CONSUMPTION = f(INCOME)$$

which says that the level of consumption is *some* function, $f(\bullet)$, of income.

The demand for an individual commodity—say, the Honda Accord—might be expressed as

$$Q^d = f(P, P^s, P^c, INC)$$

which says that the quantity of Honda Accords demanded, Q^d, is a function $f(P, P^s, P^c, INC)$ of the price of Honda Accords P, the price of cars that are substitutes P^s, the price of items that are complements P^c (like gasoline), and the level of income INC.

The supply of an agricultural commodity such as beef might be written as

$$Q^s = f\left(P, P^c, P^f\right)$$

where Q^s is the quantity supplied, P is the price of beef, P^c is the price of competitive products in production (e.g., the price of hogs), and P^f is the price of factors or inputs (e.g., the price of corn) used in the production process.

Each of the above equations is a general economic model that describes how we visualize the way in which economic variables are interrelated. Economic models of this type *guide our economic analysis*.

Econometrics allows us to go further than knowing that certain economic variables are interrelated, or even the direction of a relationship. Econometrics allows us to assign magnitudes to questions about the interrelationships between variables. One aspect of econometrics is **prediction** or **forecasting**. If we know the value of *INCOME*, what will be the magnitude of *CONSUMPTION*? If we have values for the prices of Honda Accords, their substitutes and complements, and income, how many Honda Accords will be sold? Similarly, we could ask how much beef would be supplied given values of the variables on which its supply depends.

A second contribution of econometrics is to enable us to say **how much** a change in one variable affects another. If the price for Honda Accords is increased, by how much will quantity demanded decline? If the price of beef goes up, by how much will quantity supplied increase? Finally, econometrics contributes to our understanding of the interrelationships between variables by giving us the ability to **test** the validity of hypothesized relationships.

> **Econometrics** is about how we can use theory and data from economics, business, and the social sciences, along with tools from statistics, to predict outcomes, answer "how much" type questions, and test hypotheses.

1.2.1 Some Examples

Consider the problem faced by decision makers in a central bank. In the United States, the Federal Reserve System and, in particular, the Chair of the Board of Governors of the FRB must make decisions about interest rates. When prices are observed to rise, suggesting an increase in the inflation rate, the FRB must make a decision about whether to dampen the rate of growth of the economy. It can do so by raising the interest rate it charges its member banks when they borrow money (the discount rate) or the rate on overnight loans between banks (the federal funds rate). Increasing these rates sends a ripple effect through the economy, causing increases in other interest rates, such as those faced by would-be investors, who may be firms seeking funds for capital expansion or individuals who wish to buy consumer durables like automobiles and refrigerators. This has the economic effect of increasing costs, and consumers react by reducing the quantity of the durable goods demanded. Overall, aggregate demand falls, which slows the rate of inflation. These relationships are suggested by economic theory.

The real question facing the Chair is "*How much* should we increase the discount rate to slow inflation and yet maintain a stable and growing economy?" The answer will depend on the responsiveness of firms and individuals to increases in the interest rates and to the effects of reduced investment on gross national product (GNP). The key elasticities and multipliers are called **parameters**. The values of economic parameters are unknown and must be estimated using a sample of economic data when formulating economic policies.

Econometrics is about how to best estimate economic parameters given the data we have. "Good" econometrics is important since errors in the estimates used by policymakers such as the FRB may lead to interest rate corrections that are too large or too small, which has consequences for all of us.

Every day, decision-makers face "how much" questions similar to those facing the FRB Chair:

- A city council ponders the question of how much violent crime will be reduced if an additional million dollars is spent putting uniformed police on the street.
- The owner of a local Pizza Hut must decide how much advertising space to purchase in the local newspaper and thus must estimate the relationship between advertising and sales.
- Louisiana State University must estimate how much enrollment will fall if tuition is raised by $300 per semester and thus whether its revenue from tuition will rise or fall.
- The CEO of Proctor & Gamble must predict how much demand there will be in 10 years for the detergent Tide and how much to invest in new plant and equipment.
- A real estate developer must predict by how much population and income will increase to the south of Baton Rouge, Louisiana, over the next few years and whether it will be profitable to begin construction of a gambling casino and golf course.
- You must decide how much of your savings will go into a stock fund and how much into the money market. This requires you to make predictions of the level of economic activity, the rate of inflation, and interest rates over your planning horizon.
- A public transportation council in Melbourne, Australia, must decide how an increase in fares for public transportation (trams, trains, and buses) will affect the number of travelers who switch to car or bike and the effect of this switch on revenue going to public transportation.

To answer these questions of "how much," decision-makers rely on information provided by empirical economic research. In such research, an economist uses economic theory and reasoning to construct relationships between the variables in question. Data on these variables are collected and econometric methods are used to estimate the key underlying parameters and to make predictions. The decision-makers in the above examples obtain their "estimates" and "predictions" in different ways. The FRB has a large staff of economists to carry out econometric analyses. The CEO of Proctor & Gamble may hire econometric consultants to provide the firm with projections of sales. You may get advice about investing from a stock broker, who in turn is provided with econometric projections made by economists working for the parent company. Whatever the source of your information about "how much" questions, it is a good bet that there is an economist involved who is using econometric methods to analyze data that yield the answers.

In the next section, we show how to introduce parameters into an economic model and how to convert an economic model into an econometric model.

1.3 The Econometric Model

What is an econometric model, and where does it come from? We will give you a general overview, and we may use terms that are unfamiliar to you. Be assured that before you are too far into this book, all the terminology will be clearly defined. In an econometric model we must first realize that economic relations are not exact. Economic theory does not claim to be able to predict the specific behavior of any individual or firm, but rather describes the *average* or *systematic* behavior of *many* individuals or firms. When studying car sales we recognize that the *actual* number of Hondas sold is the sum of this systematic part and a random and unpredictable component e that we will call a **random error**. Thus, an **econometric model** representing the sales of Honda Accords is

$$Q^d = f(P, P^s, P^c, INC) + e$$

The random error e accounts for the many factors that affect sales that we have omitted from this simple model, and it also reflects the intrinsic uncertainty in economic activity.

To complete the specification of the econometric model, we must also say something about the form of the algebraic relationship among our economic variables. For example, in your first economics courses quantity demanded was depicted as a *linear* function of price. We extend that assumption to the other variables as well, making the systematic part of the demand relation

$$f(P, P^s, P^c, INC) = \beta_1 + \beta_2 P + \beta_3 P^s + \beta_4 P^c + \beta_5 INC$$

The corresponding econometric model is

$$Q^d = \beta_1 + \beta_2 P + \beta_3 P^s + \beta_4 P^c + \beta_5 INC + e$$

The coefficients $\beta_1, \beta_2, \ldots, \beta_5$ are unknown **parameters** of the model that we estimate using economic data and an econometric technique. The functional form represents a hypothesis about the relationship between the variables. In any particular problem, one challenge is to determine a functional form that is compatible with economic theory and the data.

In every econometric model, whether it is a demand equation, a supply equation, or a production function, there is a systematic portion and an unobservable random component. The systematic portion is the part we obtain from economic theory, and includes an assumption about the functional form. The random component represents a "noise" component, which obscures our understanding of the relationship among variables, and which we represent using the random variable e.

We use the econometric model as a basis for **statistical inference**. Using the econometric model and a sample of data, we make inferences concerning the real world, learning something in the process. The ways in which statistical inference are carried out include the following:

- **Estimating** economic parameters, such as elasticities, using econometric methods

- **Predicting** economic outcomes, such as the enrollment in two-year colleges in the United States for the next 10 years
- **Testing** economic hypotheses, such as the question of whether newspaper advertising is better than store displays for increasing sales

Econometrics includes all of these aspects of statistical inference. As we proceed through this book, you will learn how to properly estimate, predict, and test, given the characteristics of the data at hand.

1.3.1 Causality and Prediction

A question that often arises when specifying an econometric model is whether a relationship can be viewed as both causal and predictive or only predictive. To appreciate the difference, consider an equation where a student's grade in Econometrics *GRADE* is related to the proportion of class lectures that are skipped *SKIP*.

$$GRADE = \beta_1 + \beta_2 SKIP + e$$

We would expect β_2 to be negative: the greater the proportion of lectures that are skipped, the lower the grade. But, can we say that skipping lectures **causes** grades to be lower? If lectures are captured by video, they could be viewed at another time. Perhaps a student is skipping lectures because he or she has a demanding job, and the demanding job does not leave enough time for study, and this is the underlying cause of a poor grade. Or, it might be that skipping lectures comes from a general lack of commitment or motivation, and this is the cause of a poor grade. Under these circumstances, what can we say about the equation that relates *GRADE* to *SKIP*? We can still call it a predictive equation. *GRADE* and *SKIP* are (negatively) correlated and so information about *SKIP* can be used to help predict *GRADE*. However, we cannot call it a causal relationship. Skipping lectures does not cause a low grade. The parameter β_2 does not convey the direct causal effect of skipping lectures on grade. It also includes the effect of other variables that are omitted from the equation and correlated with *SKIP*, such as hours spent studying or student motivation.

Economists are frequently interested in parameters that can be interpreted as causal. Honda would like to know the direct effect of a price change on the sales of their Accords. When there is technological improvement in the beef industry, the price elasticities of demand and supply have important implications for changes in consumer and producer welfare. One of our tasks will be to see what assumptions are necessary for an econometric model to be interpreted as causal and to assess whether those assumptions hold.

An area where predictive relationships are important is in the use of "big data." Advances in computer technology have led to storage of massive amounts of information. Travel sites on the Internet keep track of destinations you have been looking at. Google targets you with advertisements based on sites that you have been surfing. Through their loyalty cards, supermarkets keep data on your purchases and identify sale items relevant for you. Data analysts use big data to identify predictive relationships that help predict our behavior.

In general, the type of data we have impacts on the specification of an econometric model and the assumptions that we make about it. We turn now to a discussion of different types of data and where they can be found.

1.4 How Are Data Generated?

In order to carry out statistical inference we must have data. Where do data come from? What type of real processes generate data? Economists and other social scientists work in a complex world in which data on variables are "observed" and rarely obtained from a controlled experiment. This makes the task of learning about economic parameters all the more difficult. Procedures for using such data to answer questions of economic importance are the subject matter of this book.

1.4.1 Experimental Data

One way to acquire information about the unknown parameters of economic relationships is to conduct or observe the outcome of an experiment. In the physical sciences and agriculture, it is easy to imagine controlled experiments. Scientists specify the values of key control variables and then observe the outcome. We might plant similar plots of land with a particular variety of wheat, and then vary the amounts of fertilizer and pesticide applied to each plot, observing at the end of the growing season the bushels of wheat produced on each plot. Repeating the experiment on N plots of land creates a sample of N observations. Such controlled experiments are rare in business and the social sciences. A key aspect of experimental data is that the values of the explanatory variables can be fixed at specific values in repeated trials of the experiment.

One business example comes from marketing research. Suppose we are interested in the weekly sales of a particular item at a supermarket. As an item is sold it is passed over a scanning unit to record the price and the amount that will appear on your grocery bill. But at the same time, a data record is created, and at every point in time the price of the item and the prices of all its competitors are known, as well as current store displays and coupon usage. The prices and shopping environment are controlled by store management, so this "experiment" can be repeated a number of days or weeks using the same values of the "control" variables.

There are some examples of planned experiments in the social sciences, but they are rare because of the difficulties in organizing and funding them. A notable example of a planned experiment is Tennessee's Project Star.[1] This experiment followed a single cohort of elementary school children from kindergarten through the third grade, beginning in 1985 and ending in 1989. In the experiment children and teachers were randomly assigned within schools into three types of classes: small classes with 13–17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. The objective was to determine the effect of small classes on student learning, as measured by student scores on achievement tests. We will analyze the data in Chapter 7 and show that small classes significantly increase performance. This finding will influence public policy toward education for years to come.

1.4.2 Quasi-Experimental Data

It is useful to distinguish between "pure" experimental data and "quasi"-experimental data. A pure experiment is characterized by random assignment. In the example where varying amounts of fertilizer and pesticides are applied to plots of land for growing wheat, the different applications of fertilizer and pesticides are randomly assigned to different plots. In Tennessee's Project Star, students and teachers are randomly assigned to different sized classes with and without a teacher's aide. In general, if we have a control group and a treatment group, and we want to examine the effect of a policy intervention or treatment, pure experimental data are such that individuals are randomly assigned to the control and treatment groups.

Random assignment is not always possible however, particularly when dealing with human subjects. With quasi-experimental data, allocation to the control and treatment groups is not random but based on another criterion. An example is a study by Card and Krueger that is studied in more detail in Chapter 7. They examined the effect of an increase in New Jersey's minimum wage in 1992 on the number of people employed in fast-food restaurants. The treatment group was fast-food restaurants in New Jersey. The control group was fast-food restaurants in eastern Pennsylvania where there was no change in the minimum wage. Another example is the effect on spending habits of a change in the income tax rate for individuals above a threshold income. The treatment group is the group with incomes above the threshold. The control group is those with incomes below the threshold. When dealing with quasi-experimental data, one must be aware that the effect of the treatment may be confounded with the effect of the criterion for assignment.

[1] See https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/10766 for program description, public use data, and extensive literature.

1.4.3 Nonexperimental Data

An example of nonexperimental data is survey data. The Public Policy Research Lab at Louisiana State University (www.survey.lsu.edu) conducts telephone and mail surveys for clients. In a telephone survey, numbers are selected randomly and called. Responses to questions are recorded and analyzed. In such an environment, data on all variables are collected simultaneously, and the values are neither fixed nor repeatable. These are nonexperimental data.

Such surveys are carried out on a massive scale by national governments. For example, the Current Population Survey (CPS)[2] is a monthly survey of about 50,000 households conducted by the U.S. Bureau of the Census. The survey has been conducted for more than 50 years. The CPS website says "CPS data are used by government policymakers and legislators as important indicators of our nation's economic situation and for planning and evaluating many government programs. They are also used by the press, students, academics, and the general public." In Section 1.8 we describe some similar data sources.

1.5 Economic Data Types

Economic data comes in a variety of "flavors." In this section we describe and give an example of each. In each example, be aware of the different data characteristics, such as the following:

1. Data may be collected at various levels of aggregation:
   - *micro*—data collected on individual economic decision-making units such as individuals, households, and firms.
   - *macro*—data resulting from a pooling or aggregating over individuals, households, or firms at the local, state, or national levels.
2. Data may also represent a flow or a stock:
   - *flow*—outcome measures over a period of time, such as the consumption of gasoline during the last quarter of 2018.
   - *stock*—outcome measured at a particular point in time, such as the quantity of crude oil held by ExxonMobil in its U.S. storage tanks on November 1, 2018, or the asset value of the Wells Fargo Bank on July 1, 2018.
3. Data may be quantitative or qualitative:
   - *quantitative*—outcomes such as prices or income that may be expressed as numbers or some transformation of them, such as real prices or per capita income.
   - *qualitative*—outcomes that are of an "either-or" situation. For example, a consumer either did or did not make a purchase of a particular good, or a person either is or is not married.

1.5.1 Time-Series Data

A **time-series** is data collected over discrete intervals of time. Examples include the annual price of wheat in the United States and the daily price of General Electric stock shares. Macroeconomic data are usually reported in monthly, quarterly, or annual terms. Financial data, such as stock prices, can be recorded daily, or at even higher frequencies. The key feature of time-series data is that the same economic quantity is recorded at a regular time interval.

For example, the annual real gross domestic product (GDP) for the United States is depicted in Figure 1.1. A few values are given in Table 1.1. For each year, we have the recorded value. The data are annual, or yearly, and have been "deflated" by the Bureau of Economic Analysis to billions of real 2009 dollars.

[2] www.census.gov/cps/

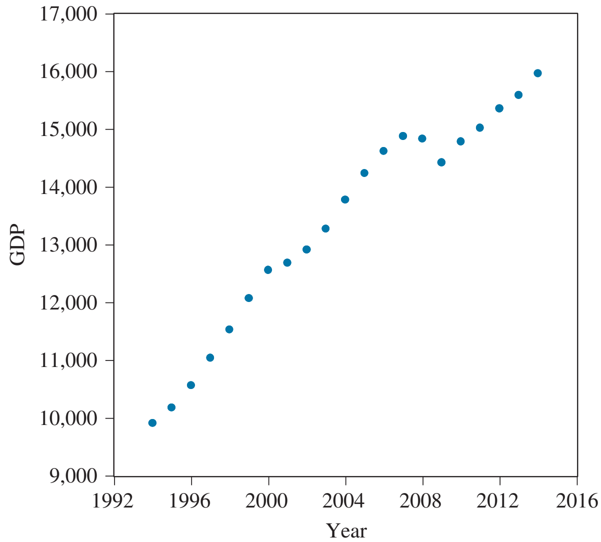

FIGURE 1.1 **Real U.S. GDP, 1994–2014.**[3]

TABLE 1.1 **U.S. Annual GDP (Billions of Real 2009 Dollars)**

Year	GDP
2006	14,613.8
2007	14,873.7
2008	14,830.4
2009	14,418.7
2010	14,783.8
2011	15,020.6
2012	15,354.6
2013	15,583.3
2014	15,961.7

1.5.2 Cross-Section Data

A cross-section of data is collected across sample units in a particular time period. Examples are income by counties in California during 2016 or high school graduation rates by state in 2015. The "sample units" are individual entities and may be firms, persons, households, states, or countries. For example, the CPS reports results of personal interviews on a monthly basis, covering items such as employment, unemployment, earnings, educational attainment, and income. In Table 1.2, we report a few observations from the March 2013 survey on the variables *RACE, EDUCATION, SEX*, and *WAGE* (hourly wage rate).[4] There are many detailed questions asked of the respondents.

[3] Source: www.bea.gov/national/index.htm

[4] In the actual raw data, the variable descriptions are coded differently to the names in Table 1.2. We have used shortened versions for convenience.

TABLE 1.2 Cross-Section Data: CPS, March 2013

	Variables			
Individual	*RACE*	*EDUCATION*	*SEX*	*WAGE*
1	White	Assoc Degree	Male	10.00
2	White	Master's Degree	Male	60.83
3	Black	Bachelor's Degree	Male	17.80
4	White	High School Graduate	Female	30.38
5	White	Master's Degree	Male	12.50
6	White	Master's Degree	Female	49.50
7	White	Master's Degree	Female	23.08
8	Black	Assoc Degree	Female	28.95
9	White	Some College, No Degree	Female	9.20

1.5.3 Panel or Longitudinal Data

A "panel" of data, also known as "longitudinal" data, has observations on individual micro-units that are followed over time. For example, the Panel Study of Income Dynamics (PSID)[5] describes itself as "a nationally representative longitudinal study of nearly 9000 U.S. families. Following the same families and individuals since 1969, the PSID collects data on economic, health, and social behavior." Other national panels exist, and many are described at "Resources for Economists," at www.rfe.org.

To illustrate, data from two rice farms[6] are given in Table 1.3. The data are annual observations on rice farms (or firms) over the period 1990–1997.

The key aspect of panel data is that we observe each micro-unit, here a farm, for a number of time periods. Here we have amount of rice produced, area planted, labor input, and fertilizer use. If we have the same number of time period observations for each micro-unit, which is the case here, we have a **balanced panel**. Usually the number of time-series observations is small relative to the number of micro-units, but not always. The Penn World Table[7] provides purchasing power parity and national income accounts converted to international prices for 182 countries for some or all of the years 1950–2014.

1.6 The Research Process

Econometrics is ultimately a research tool. Students of econometrics plan to do research or they plan to read and evaluate the research of others, or both. This section provides a frame of reference and guide for future work. In particular, we show you the role of econometrics in research.

Research is a process, and like many such activities, it flows according to an orderly pattern. Research is an adventure, and can be *fun!* Searching for an answer to your question, seeking new knowledge, is very addictive—for the more you seek, the more new questions you will find.

A research project is an opportunity to investigate a topic that is important to you. Choosing a good research topic is essential if you are to complete a project successfully. A starting point is the question "What are my interests?" Interest in a particular topic will add pleasure to the

[5] http://psidonline.isr.umich.edu

[6] These data were used by O'Donnell, C.J. and W.E. Griffiths (2006), Estimating State-Contingent Production Frontiers, *American Journal of Agricultural Economics*, 88(1), 249–266.

[7] www.rug.nl/ggdc/productivity/pwt

TABLE 1.3 Panel Data from Two Rice Farms

FARM	*YEAR*	*PROD*	*AREA*	*LABOR*	*FERT*
1	1990	7.87	2.50	160	207.5
1	1991	7.18	2.50	138	295.5
1	1992	8.92	2.50	140	362.5
1	1993	7.31	2.50	127	338.0
1	1994	7.54	2.50	145	337.5
1	1995	4.51	2.50	123	207.2
1	1996	4.37	2.25	123	345.0
1	1997	7.27	2.15	87	222.8
2	1990	10.35	3.80	184	303.5
2	1991	10.21	3.80	151	206.0
2	1992	13.29	3.80	185	374.5
2	1993	18.58	3.80	262	421.0
2	1994	17.07	3.80	174	595.7
2	1995	16.61	4.25	244	234.8
2	1996	12.28	4.25	159	479.0
2	1997	14.20	3.75	133	170.0

research effort. Also, if you begin working on a topic, other questions will usually occur to you. These new questions may put another light on the original topic or may represent new paths to follow that are even more interesting to you. The idea may come after lengthy study of all that has been written on a particular topic. You will find that "inspiration is 99% perspiration." That means that after you dig at a topic long enough, a new and interesting question will occur to you. Alternatively, you may be led by your natural curiosity to an interesting question. Professor Hal Varian[8] suggests that you look for ideas outside academic journals—in newspapers, magazines, etc. He relates a story about a research project that developed from his shopping for a new TV set.

By the time you have completed several semesters of economics classes, you will find yourself enjoying some areas more than others. For each of us, specialized areas such as health economics, economic development, industrial organization, public finance, resource economics, monetary economics, environmental economics, and international trade hold a different appeal. If you find an area or topic in which you are interested, consult the *Journal of Economic Literature* (*JEL*) for a list of related journal articles. The *JEL* has a classification scheme that makes isolating particular areas of study an easy task. Alternatively, type a few descriptive words into your favorite search engine and see what pops up.

Once you have focused on a particular idea, begin the research process, which generally follows steps like these:

1. Economic theory gives us a way of thinking about the problem. Which economic variables are involved, and what is the possible direction of the relationship(s)? Every research project, given the initial question, begins by building an economic model and listing the questions (hypotheses) of interest. More questions will arise during the research project, but it is good to list those that motivate you at the project's beginning.
2. The working economic model leads to an econometric model. We must choose a functional form and make some assumptions about the nature of the error term.

[8] Varian, H. How to Build an Economic Model in Your Spare Time, *The American Economist*, 41(2), Fall 1997, pp. 3–10.

3. Sample data are obtained and a desirable method of statistical analysis chosen, based on initial assumptions and an understanding of how the data were collected.
4. Estimates of the unknown parameters are obtained with the help of a statistical software package, predictions are made, and hypothesis tests are performed.
5. Model diagnostics are performed to check the validity of assumptions. For example, were all of the right-hand side explanatory variables relevant? Was an adequate functional form used?
6. The economic consequences and the implications of the empirical results are analyzed and evaluated. What economic resource allocation and distribution results are implied, and what are their policy-choice implications? What remaining questions might be answered with further study or with new and better data?

These steps provide some direction for what must be done. However, research always includes some surprises that may send you back to an earlier point in your research plan or that may even cause you to revise it completely. Research requires a sense of urgency, which keeps the project moving forward, the patience not to rush beyond careful analysis, and the willingness to explore new ideas.

1.7 Writing an Empirical Research Paper

Research rewards you with new knowledge, but it is incomplete until a research paper or report is written. The process of writing forces the distillation of ideas. In no other way will your depth of understanding be so clearly revealed. When you have difficulty explaining a concept or thought, it may mean that your understanding is incomplete. Thus, writing is an integral part of research. We provide this section as a building block for future writing assignments. Consult it as needed. You will find other tips on writing economics papers on the book website, www.principlesofeconometrics.com.

1.7.1 Writing a Research Proposal

After you have selected a specific topic, it is a good idea to write up a brief project summary, or proposal. Writing it will help to focus your thoughts about what you really want to do. Show it to your colleagues or instructor for preliminary comments. The summary should be short, usually no longer than 500 words, and should include the following:

1. A concise statement of the problem
2. Comments on the information that is available, with one or two key references
3. A description of the research design that includes
   a. the economic model
   b. the econometric estimation and inference methods
   c. data sources
   d. estimation, hypothesis testing, and prediction procedures, including the econometric software and version used
4. The potential contribution of the research

1.7.2 A Format for Writing a Research Report

Economic research reports have a standard format in which the various steps of the research project are discussed and the results interpreted. The following outline is typical.

1. *Statement of the Problem* The place to start your report is with a summary of the questions you wish to investigate as well as why they are important and who should be interested in the results. This introductory section should be nontechnical and should motivate the reader to continue reading the paper. It is also useful to map out the contents of the following sections of the report. This is the first section to work on and also the last. In today's busy world, the reader's attention must be garnered very quickly. A clear, concise, well-written introduction is a must and is arguably the most important part of the paper.
2. *Review of the Literature* Briefly summarize the relevant literature in the research area you have chosen and clarify how your work extends our knowledge. By all means, cite the works of others who have motivated your research, but keep it brief. You do not have to survey everything that has been written on the topic.
3. *The Economic Model* Specify the economic model that you used and define the economic variables. State the model's assumptions and identify hypotheses that you wish to test. Economic models can get complicated. Your task is to explain the model clearly, but as briefly and simply as possible. Don't use unnecessary technical jargon. Use simple terms instead of complicated ones when possible. Your objective is to display the quality of your thinking, not the extent of your vocabulary.
4. *The Econometric Model* Discuss the econometric model that corresponds to the economic model. Make sure you include a discussion of the variables in the model, the functional form, the error assumptions, and any other assumptions that you make. Use notation that is as simple as possible, and do not clutter the body of the paper with long proofs or derivations; these can go into a technical appendix.
5. *The Data* Describe the data you used, as well as the source of the data and any reservations you have about their appropriateness.
6. *The Estimation and Inference Procedures* Describe the estimation methods you used and why they were chosen. Explain hypothesis testing procedures and their usage. Indicate the software used and the version, such as Stata 15 or EViews 10.
7. *The Empirical Results and Conclusions* Report the parameter estimates, their interpretation, and the values of test statistics. Comment on their statistical significance, their relation to previous estimates, and their economic implications.
8. *Possible Extensions and Limitations of the Study* Your research will raise questions about the economic model, data, and estimation techniques. What future research is suggested by your findings, and how might you go about performing it?
9. *Acknowledgments* It is appropriate to recognize those who have commented on and contributed to your research. This may include your instructor, a librarian who helped you find data, or a fellow student who read and commented on your paper.
10. *References* An alphabetical list of the literature you cite in your study, as well as references to the data sources you used.

Once you've written the first draft, use your computer's spell-check software to check for spelling errors. Have a friend read the paper, make suggestions for clarifying the prose, and check your logic and conclusions. Before you submit the paper, you should eliminate as many errors as possible. Your work should look good. Use a word processor, and be consistent with font sizes, section headings, style of footnotes, references, and so on. Often software developers provide templates for term papers and theses. A little searching for a good paper layout before beginning is a good idea. Typos, missing references, and incorrect formulas can spell doom for an otherwise excellent paper. Some do's and don'ts are summarized nicely, and with good humor, by Deidre N. McClosky in *Economical Writing*, 2nd edition (Prospect Heights, IL: Waveland Press, Inc., 1999).

While it is not a pleasant topic to discuss, you should be aware of the rules of **plagiarism**. You must not use someone else's words as if they were your own. If you are unclear about what you can and cannot use, check with the style manuals listed in the next paragraph, or consult your instructor. Your university may provide a plagiarism-checking software, such as Turnitin or iThenticate, that will compare your paper to millions of online sources and look for problem areas. There are some free online versions as well. The paper should have clearly defined sections and subsections. The pages, equations, tables, and figures should be numbered. References and footnotes should be formatted in an acceptable fashion. A style guide is a good investment. Two classics are the following:

- *The Chicago Manual of Style*, 16th edition, is available online and in other formats.
- *A Manual for Writers of Research Papers, Theses, and Dissertations: Chicago Style for Students and Researchers*, 8th edition, by Kate L. Turabian; revised by Wayne C. Booth, Gregory G. Colomb, and Joseph M Williams (2013, University of Chicago Press).

1.8 Sources of Economic Data

Economic data are much easier to obtain since the development of the World Wide Web. In this section we direct you to some places on the Internet where economic data are accessible. During your study of econometrics, browse some of the sources listed to gain some familiarity with data availability.

1.8.1 Links to Economic Data on the Internet

There are a number of fantastic sites on the World Wide Web for obtaining economic data.

Resources for Economists (RFE) www.rfe.org is a primary gateway to resources on the Internet for economists. This excellent site is the work of Bill Goffe. Here you will find links to sites for economic data and sites of general interest to economists. The **Data** link has these broad data categories:

- *U.S. Macro and Regional Data* Here you will find links to various data sources such as the Bureau of Economic Analysis, Bureau of Labor Statistics, *Economic Reports of the President*, and the Federal Reserve Banks.
- *Other U.S. Data* Here you will find links to the U.S. Census Bureau, as well as links to many panel and survey data sources. The gateway to U.S. government agencies is FedStats (fedstats.sites.usa.gov). Once there, click on *Agencies* to see a complete list of U.S. government agencies and links to their homepages.
- *World and Non-U.S. Data* Here there are links to world data, such as at the CIA World Factbook and the Penn World Tables, as well as international organizations such as the Asian Development Bank, the International Monetary Fund, the World Bank, and so on. There are also links to sites with data on specific countries and sectors of the world.
- *Finance and Financial Markets* Here are links to sources of U.S. and world financial data on variables such as exchange rates, interest rates, and share prices.
- *Journal Data and Program Archives* Some economic journals post data used in articles. Links to these journals are provided here. (Many of the articles in these journals will be beyond the scope of undergraduate economics majors.)

National Bureau of Economic Research (NBER) www.nber.org/data provides access to a great amount of data. There are headings for

- Macro Data
- Industry Productivity and Digitalization Data
- International Trade Data

- Individual Data
- Healthcare Data—Hospitals, Providers, Drugs, and Devices
- Demographic and Vital Statistics
- Patent and Scientific Papers Data
- Other Data

Economagic Some websites make extracting data relatively easy. For example, Economagic (www.economagic.com) is an excellent and easy-to-use source of macro time series (some 100,000 series available). The data series are easily viewed in a copy and paste format, or graphed.

1.8.2 Interpreting Economic Data

In many cases it is easier to obtain economic data than it is to understand the meaning of the data. It is essential when using macroeconomic or financial data that you understand the definitions of the variables. Just what is the index of leading economic indicators? What is included in personal consumption expenditures? You may find the answers to some questions like these in your textbooks. Another resource you might find useful is *A Guide to Everyday Economic Statistics*, 7th edition, by Gary E. Clayton and Martin Gerhard Giesbrecht, (Boston: Irwin/McGraw-Hill, 2009). This slender volume examines how economic statistics are constructed, and how they can be used.

1.8.3 Obtaining the Data

Finding a data source is not the same as obtaining the data. Although there are a great many easy-to-use websites, "easy-to-use" is a relative term. The data will come packaged in a variety of formats. It is also true that there are many, many variables at each of these websites. A primary challenge is identifying the specific variables that you want, and what exactly they measure. The following examples are illustrative.

The Federal Reserve Bank of St. Louis [9] has a system called **FRED** (Federal Reserve Economic Data). Under "Categories" there are links to financial variables, population and labor variables, national accounts, and many others. Data on these variables can be downloaded in a number of formats. For reading the data, you may need specific knowledge of your statistical software. Accompanying *Principles of Econometrics, 5e*, are computer manuals for Excel, EViews, Stata, SAS, R, and Gretl to aid this process. See the publisher website www.wiley.com/college/hill, or the book website at www.principlesofeconometrics.com for a description of these aids.

The CPS (www.census.gov/cps) has a tool called **DataFerrett**. This tool will help you find and download data series that are of particular interest to you. There are tutorials that guide you through the process. Variable descriptions, as well as the specific survey questions, are provided to aid in your selection. It is somewhat like an Internet shopping site. Desired series are "ticked" and added to a "Shopping Basket." Once you have filled your basket, you download the data to use with specific software. Other Web-based data sources operate in this same manner. One example is the PSID.[10] The Penn World Tables[11] offer data downloads in both Excel and Stata formats.

You can expect to find massive amounts of readily available data at the various sites we have mentioned, but there is a learning curve. You should not expect to find, download, and process the data without considerable work effort. Being skilled with Excel and statistical software is a must if you plan to regularly use these data sources.

[9] https://fred.stlouisfed.org

[10] http://psidonline.isr.umich.edu

[11] www.rug.nl/ggdc/productivity/pwt

Probability Primer

LEARNING OBJECTIVES

Remark

Learning Objectives and *Keywords* sections will appear at the beginning of each chapter. We urge you to think about, and possibly write out answers to the questions, and make sure you recognize and can define the keywords. If you are unsure about the questions or answers, consult your instructor. When examples are requested in *Learning Objectives* sections, you should think of examples *not* in the book.

Based on the material in this primer, you should be able to

1. Explain the difference between a random variable and its values, and give an example.
2. Explain the difference between discrete and continuous random variables, and give examples of each.
3. State the characteristics of a probability density function (*pdf*) for a discrete random variable, and give an example.
4. Compute probabilities of events, given a discrete probability function.
5. Explain the meaning of the following statement: "The probability that the discrete random variable takes the value 2 is 0.3."
6. Explain how the *pdf* of a continuous random variable is different from the *pdf* of a discrete random variable.
7. Show, geometrically, how to compute probabilities given a *pdf* for a continuous random variable.
8. Explain, intuitively, the concept of the mean, or expected value, of a random variable.
9. Use the definition of expected value for a discrete random variable to compute expectations, given a *pdf* $f(x)$ and a function $g(X)$ of X.
10. Define the variance of a discrete random variable, and explain in what sense the values of a random variable are more spread out if the variance is larger.
11. Use a joint *pdf* (table) for two discrete random variables to compute probabilities of joint events and to find the (marginal) *pdf* of each individual random variable.
12. Find the conditional *pdf* for one discrete random variable given the value of another and their joint *pdf*.
13. Work with single and double summation notation.
14. Give an intuitive explanation of statistical independence of two random variables, and state the conditions that must hold to prove statistical independence. Give examples of two independent random variables and two dependent random variables.

15. Define the covariance and correlation between two random variables, and compute these values given a joint probability function of two discrete random variables.
16. Find the mean and variance of a sum of random variables.
17. Use Statistical Table 1, Cumulative Probabilities for the Standard Normal Distribution, and your computer software to compute probabilities involving normal random variables.
18. Use the Law of Iterated Expectations to find the expected value of a random variable.

KEYWORDS

conditional expectation
conditional *pdf*
conditional probability
continuous random variable
correlation
covariance
cumulative distribution function
discrete random variable
expected value
experiment
indicator variable
iterated expectation
joint probability density function
marginal distribution
mean
normal distribution
population
probability
probability density function
random variable
standard deviation
standard normal distribution
statistical independence
summation operations
variance

We assume that you have had a basic probability and statistics course. In this primer, we review some essential probability concepts. Section P.1 defines discrete and continuous random variables. Probability distributions are discussed in Section P.2. Section P.3 introduces joint probability distributions, defines conditional probability and **statistical independence**. In Section P.4, we digress and discuss operations with summations. In Section P.5, we review the properties of probability distributions, paying particular attention to expected values and variances. In Section P.6, we discuss the important concept of **conditioning**, and how knowing the value of one variable might provide information about, or help us predict, another variable. Section P.7 summarizes important facts about the normal probability distribution. In Appendix B, "Probability Concepts," are enhancements and additions to this material.

P.1 Random Variables

Benjamin Franklin is credited with the saying "The only things certain in life are death and taxes." While not the original intent, this bit of wisdom points out that almost everything we encounter in life is uncertain. We do not know how many games our football team will win next season. You do not know what score you will make on the next exam. We don't know what the stock market index will be tomorrow. These events, or outcomes, are uncertain, or random. **Probability** gives us a way to talk about possible outcomes.

A **random variable** is a variable whose value is unknown until it is observed; in other words, it is a variable that is not perfectly predictable. Each random variable has a set of possible values it can take. If W is the number of games our football team wins next year, then W can take the values 0, 1, 2, …, 13, if there are a maximum of 13 games. This is a **discrete random variable** since it can take only a limited, or **countable**, number of values. Other examples of discrete random variables are the number of computers owned by a randomly selected household, and the number of times you will visit your physician next year. A special case occurs when a random variable can only be one of two possible values—for example, in a phone survey, if you are asked if you are a college graduate or not, your answer can only be "yes" or "no." Outcomes like this can be characterized by an **indicator variable** taking the values one if yes or zero if no. Indicator variables are discrete and are used to represent qualitative characteristics such as sex (male or female) or race (white or nonwhite).

The U.S. GDP is yet another example of a random variable, because its value is unknown until it is observed. In the third quarter of 2014 it was calculated to be 16,164.1 billion dollars. What the value will be in the second quarter of 2025 is unknown, and it cannot be predicted perfectly. GDP is measured in dollars and it *can* be counted in whole dollars, but the value is so large that counting individual dollars serves no purpose. For practical purposes, GDP can take any value in the interval zero to infinity, and it is treated as a **continuous random variable**. Other common macroeconomic variables, such as interest rates, investment, and consumption, are also treated as continuous random variables. In finance, stock market indices, like the Dow Jones Industrial Index, are also treated as continuous. The key attribute of these variables that makes them continuous is that they can take any value in an interval.

P.2 Probability Distributions

Probability is usually defined in terms of **experiments**. Let us illustrate this in the context of a simple experiment. Consider the objects in Table P.1 to be a population of interest. In statistics and econometrics, the term **population** is an important one. A population is a group of objects, such as people, farms, or business firms, having something in common. The population is a complete set and is the focus of an analysis. In this case the population is the set of ten objects shown in Table P.1. Using this population, we will discuss some probability concepts. In an **empirical analysis,** a sample of observations is collected from the population of interest, and using the sample observations we make inferences about the population.

If we were to select one cell from the table at random (imagine cutting the table into 10 equally sized pieces of paper, stirring them up, and drawing one of the slips without looking), that would constitute a **random experiment**. Based on this random experiment, we can define several random variables. For example, let the random variable X be the numerical value showing on a slip that we draw. (We use uppercase letters like X to represent random variables in this primer). The term **random variable** is a bit odd, as it is actually a rule for assigning numerical values to experimental outcomes. In the context of Table P.1, the rule says, "Perform the experiment (stir the slips, and draw one) and for the slip that you obtain assign X to be the number showing." The values that X can take are denoted by corresponding lowercase letters, x, and in this case the values of X are $x = 1, 2, 3$, or 4.

For the experiment using the population in Table P.1,[1] we can create a number of random variables. Let Y be a discrete random variable designating the color of the slip, with $Y = 1$ denoting

TABLE P.1 **The Seussian Slips: A Population**

1	2	3	4	4
2	3	3	4	4

[1]A table suitable for classroom experiments can be obtained at www.principlesofeconometrics.com/poe5/extras/table_p1. We thank Veronica Deschner McGregor for the suggestion of "One slip, two slip, white slip, blue slip" for this experiment, inspired by Dr. Seuss's "One Fish Two Fish Red Fish Blue Fish (I Can Read It All by Myself)," Random House Books for Young Readers (1960).

a shaded slip and $Y = 0$ denoting a slip with no shading. The numerical values that Y can take are $y = 0, 1$.

Consider X, the numerical value on the slip. If the slips are equally likely to be chosen after shuffling, then in a large number of experiments (i.e., shuffling and drawing one of the ten slips), 10% of the time we would observe $X = 1$, 20% of the time $X = 2$, 30% of the time $X = 3$, and 40% of the time $X = 4$. These are probabilities that the specific values will occur. We would say, for example, $P(X = 3) = 0.3$. This interpretation is tied to the **relative frequency** of a particular outcome's occurring in a **large** number of experiment replications.

We summarize the probabilities of possible outcomes using a **probability density function** (*pdf*). The *pdf* for a discrete random variable indicates the probability of each possible value occurring. For a discrete random variable X the value of the *pdf* $f(x)$ is the probability that the random variable X takes the value x, $f(x) = P(X = x)$. Because $f(x)$ is a probability, it must be true that $0 \le f(x) \le 1$ and, if X takes n possible values $x_1, \ldots, x_n$, then the sum of their probabilities must be one

$$f\left(x_1\right) + f\left(x_2\right) + \cdots + f\left(x_n\right) = 1 \tag{P.1}$$

For discrete random variables, the *pdf* might be presented as a table, such as in Table P.2.

As shown in Figure P.1, the *pdf* may also be represented as a bar graph, with the height of the bar representing the probability with which the corresponding value occurs.

The **cumulative distribution function** (*cdf*) is an alternative way to represent probabilities. The *cdf* of the random variable X, denoted $F(x)$, gives the probability that X is less than or equal to a specific value x. That is,

$$F(x) = P(X \le x) \tag{P.2}$$

TABLE P.2 Probability Density Function of X

x	$f(x)$
1	0.1
2	0.2
3	0.3
4	0.4

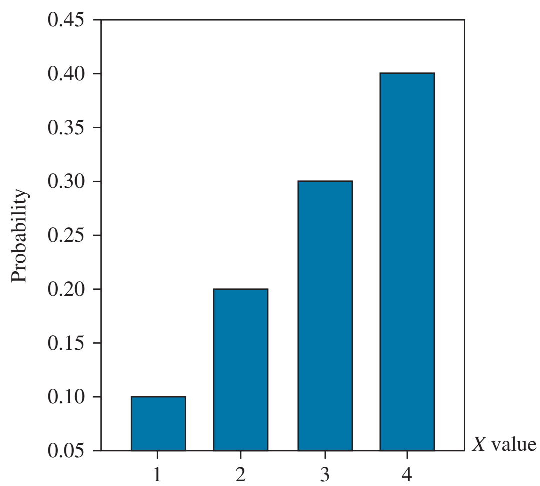

FIGURE P.1 Probability density function for X.

EXAMPLE P.1 | Using a *cdf*

Using the probabilities in Table P.2, we find that $F(1) = P(X \le 1) = 0.1$, $F(2) = P(X \le 2) = 0.3$, $F(3) = P(X \le 3) = 0.6$, and $F(4) = P(X \le 4) = 1$. For example, using the *pdf* $f(x)$ we compute the probability that X is less than or equal to 2 as

$$F(2) = P(X \le 2) = P(X = 1) + P(X = 2) = 0.1 + 0.2 = 0.3$$

Since the sum of the probabilities $P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1$, we can compute the probability that X is greater than 2 as

$$P(X > 2) = 1 - P(X \le 2) = 1 - F(2) = 1 - 0.3 = 0.7$$

An important difference between the *pdf* and *cdf* for X is revealed by the question "Using the probability distribution in Table P.2, what is the probability that $X = 2.5$?" This probability is zero because X cannot take this value. The question "What is the probability that X is less than or equal to 2.5?" does have an answer.

$$\begin{aligned} F(2.5) = P(X \le 2.5) &= P(X = 1) + P(X = 2) \\ &= 0.1 + 0.2 = 0.3 \end{aligned}$$

The cumulative probability can be calculated for any x between $-\infty$ and $+\infty$.

Continuous random variables can take any value in an interval and have an uncountable number of values. Consequently, the probability of any specific value is zero. For continuous random variables, we talk about outcomes being in a certain range. Figure P.2 illustrates the *pdf* $f(x)$ of a continuous random variable X that takes values of x from 0 to infinity. The shape is representative of the distribution for an economic variable such as an individual's income or wage. Areas under the curve represent probabilities that X falls in an interval. The *cdf* $F(x)$ is defined as in (P.2). For this distribution,

$$\begin{aligned} P(10 < X < 20) = F(20) - F(10) &= 0.52236 - 0.17512 \\ &= 0.34724 \end{aligned} \tag{P.3}$$

How are these areas obtained? The integral from calculus gives the area under a curve. We will not compute many integrals in this book.[2] Instead, we will use the computer and compute *cdf* values and probabilities using software commands.

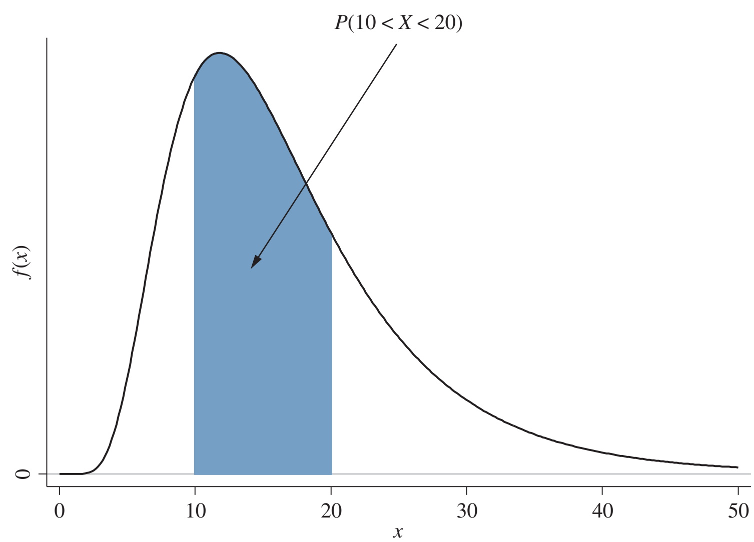

FIGURE P.2 **Probability density function for a continuous random variable.**

[2]See Appendix A.4 for a brief explanation of integrals, and illustrations using integrals to compute probabilities in Appendix B.2.1. The calculations in (P.3) are explained in Appendix B.3.9.

P.3 Joint, Marginal, and Conditional Probabilities

Working with more than one random variable requires a **joint probability density function**. For the population in Table P.1 we defined two random variables, X the numeric value of a randomly drawn slip and the indicator variable Y that equals 1 if the selected slip is shaded, and 0 if it is not shaded.

Using the joint *pdf* for X and Y we can say "The probability of selecting a shaded 2 is 0.10." This is a joint probability because we are talking about the probability of two events occurring simultaneously; the selection takes the value $X = 2$ **and** the slip is shaded so that $Y = 1$. We can write this as

$$P(X = 2 \text{ and } Y = 1) = P(X = 2, Y = 1) = f(x = 2, y = 1) = 0.1$$

The entries in Table P.3 are probabilities $f(x, y) = P(X = x, Y = y)$ of joint outcomes. Like the *pdf* of a single random variable, the sum of the joint probabilities is 1.

P.3.1 Marginal Distributions

Given a joint *pdf*, we can obtain the probability distributions of individual random variables, which are also known as **marginal distributions**. In Table P.3, we see that a shaded slip, $Y = 1$, can be obtained with the values $x = 1, 2, 3,$ and 4. The probability that we select a shaded slip is the sum of the probabilities that we obtain a shaded 1, a shaded 2, a shaded 3, and a shaded 4. The probability that $Y = 1$ is

$$P(Y = 1) = f_Y(1) = 0.1 + 0.1 + 0.1 + 0.1 = 0.4$$

This is the sum of the probabilities across the second row of the table. Similarly the probability of drawing a white slip is the sum of the probabilities across the first row of the table, and $P(Y = 0) = f_Y(0) = 0 + 0.1 + 0.2 + 0.3 = 0.6$, where $f_Y(y)$ denotes the *pdf* of the random variable Y. The probabilities $P(X = x)$ are computed similarly by summing down across the values of Y. The joint and marginal distributions are often reported as in Table P.4.[3]

TABLE P.3 Joint Probability Density Function for X and Y

	x			
y	1	2	3	4
0	0	0.1	0.2	0.3
1	0.1	0.1	0.1	0.1

TABLE P.4 Joint and Marginal Probabilities

y/x	1	2	3	4	$f(y)$
0	0	0.1	0.2	0.3	0.6
1	0.1	0.1	0.1	0.1	0.4
$f(x)$	0.1	0.2	0.3	0.4	1.0

[3] Similar calculations for continuous random variables use integration. See Appendix B.2.3 for an illustration.

P.3.2 Conditional Probability

What is the probability that a randomly chosen slip will take the value 2 **given that** it is shaded? This question is about the **conditional probability** of the outcome $X = 2$ *given* that the outcome $Y = 1$ has occurred. The effect of the conditioning is to reduce the set of possible outcomes. Conditional on $Y = 1$ we only consider the four possible slips that are shaded. One of them is a 2, so the **conditional probability** of the outcome $X = 2$ *given* that $Y = 1$ is 0.25. There is a one in four chance of selecting a 2 given only the shaded slips. Conditioning reduces the size of the population under consideration, and conditional probabilities characterize the reduced population. For discrete random variables the probability that the random variable X takes the value x *given* that $Y = y$ is written $P(X = x|Y = y)$. This conditional probability is given by the **conditional** ***pdf*** $f(x|y)$

$$f(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{f_Y(y)} \tag{P.4}$$

where $f_Y(y)$ is the marginal *pdf* of Y.

EXAMPLE P.2 | Calculating a Conditional Probability

Using the marginal probability $P(Y=1) = 0.4$, the conditional *pdf* of X given $Y = 1$ is obtained by using (P.4) for each value of X. For example,

$$\begin{aligned} f(x = 2|y = 1) &= P(X = 2|Y = 1) \\ &= \frac{P(X = 2, Y = 1)}{P(Y = 1)} = \frac{f(x = 2, y = 1)}{f_Y(1)} \\ &= \frac{0.1}{0.4} = 0.25 \end{aligned}$$

A key point to remember is that by conditioning we are considering only the subset of a population for which the condition holds. Probability calculations are then based on the "new" population. We can repeat this process for each value of X to obtain the complete conditional *pdf* given in Table P.5.

P.3.3 Statistical Independence

When selecting a shaded slip from Table P.1, the probability of selecting each possible outcome, $x = 1, 2, 3$, and 4 is 0.25. In the population of shaded slips the numeric values are **equally likely**. The probability of randomly selecting $X = 2$ from the entire population, from the marginal *pdf*, is $P(X = 2) = f_X(2) = 0.2$. This is different from the conditional probability. Knowing that the slip is shaded tells us something about the probability of obtaining $X = 2$. Such random variables are **dependent** in a statistical sense. Two random variables are **statistically independent**, or simply **independent**, if the conditional probability that $X = x$ given that $Y = y$ is the same as the unconditional probability that $X = x$. This means, if X and Y are independent random variables, then

$$P(X = x|Y = y) = P(X = x) \tag{P.5}$$

TABLE P.5 Conditional Probability of X Given $Y = 1$

x	1	2	3	4
$f(x\|y = 1)$	0.25	0.25	0.25	0.25

Equivalently, if X and Y are independent, then the conditional *pdf* of X given $Y = y$ is the same as the unconditional, or marginal, *pdf* of X alone.

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = f_X(x) \tag{P.6}$$

Solving (P.6) for the joint *pdf*, we can also say that X and Y are statistically independent if their joint *pdf* factors into the product of their marginal *pdf*s

$$P(X = x, Y = y) = f(x, y) = f_X(x)\, f_Y(y) = P(X = x) \times P(Y = y) \tag{P.7}$$

If (P.5) or (P.7) is true for each and every pair of values x and y, then X and Y are statistically independent. This result extends to more than two random variables. The rule allows us to check the independence of random variables X and Y in Table P.4. If (P.7) is violated for any pair of values, then X and Y are not statistically independent. Consider the pair of values $X = 1$ and $Y = 1$.

$$P(X = 1, Y = 1) = f(1, 1) = 0.1 \neq f_X(1)\, f_Y(1) = P(X = 1) \times P(Y = 1) = 0.1 \times 0.4 = 0.04$$

The joint probability is 0.1 and the product of the individual probabilities is 0.04. Since these are not equal, we can conclude that X and Y are not statistically independent.

P.4 A Digression: Summation Notation

Throughout this book we will use a **summation sign**, denoted by the symbol $\sum$, to shorten algebraic expressions. Suppose the random variable X takes the values $x_1, x_2, \ldots, x_{15}$. The sum of these values is $x_1 + x_2 + \cdots + x_{15}$. Rather than write this sum out each time we will represent it as $\sum_{i=1}^{15} x_i$, so that $\sum_{i=1}^{15} x_i = x_1 + x_2 + \cdots + x_{15}$. If we sum n terms, a general number, then the summation will be $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$. In this notation

- The symbol $\sum$ is the capital Greek letter sigma and means "the sum of."
- The letter i is called the **index of summation**. This letter is arbitrary and may also appear as t, j, or k.
- The expression $\sum_{i=1}^{n} x_i$ is read "the sum of the terms x_i, from i equals 1 to n."
- The numbers 1 and n are the **lower limit** and **upper limit** of summation.

The following rules apply to the **summation operation**.

Sum 1. The sum of n values $x_1, \ldots, x_n$ is

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

Sum 2. If a is a constant, then

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

Sum 3. If a is a constant, then

$$\sum_{i=1}^{n} a = a + a + \cdots + a = na$$

Sum 4. If X and Y are two variables, then

$$\sum_{i=1}^{n}\left(x_i + y_i\right) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Sum 5. If X and Y are two variables, then

$$\sum_{i=1}^{n}\left(ax_i + by_i\right) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$$

Sum 6. The arithmetic mean (average) of n values of X is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Sum 7. A property of the arithmetic mean (average) is that

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

Sum 8. We often use an abbreviated form of the summation notation. For example, if $f(x)$ is a function of the values of X,

$$\begin{aligned}\sum_{i=1}^{n} f\left(x_i\right) &= f\left(x_1\right) + f\left(x_2\right) + \cdots + f\left(x_n\right)\\ &= \sum_{i} f\left(x_i\right) \text{ ("Sum over all values of the index } i\text{")}\\ &= \sum_{x} f(x) \text{ ("Sum over all possible values of } X\text{")}\end{aligned}$$

Sum 9. Several summation signs can be used in one expression. Suppose the variable Y takes n values and X takes m values, and let $f(x, y) = x + y$. Then the **double summation** of this function is

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f\left(x_i, y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_i + y_j\right)$$

To evaluate such expressions work from the innermost sum outward. First set $i = 1$ and sum over all values of j, and so on. That is,

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f\left(x_i, y_j\right) = \sum_{i=1}^{m}\left[f\left(x_i, y_1\right) + f\left(x_i, y_2\right) + \cdots + f\left(x_i, y_n\right)\right]$$

The *order* of summation does not matter, so

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f\left(x_i, y_j\right) = \sum_{j=1}^{n}\sum_{i=1}^{m} f\left(x_i, y_j\right)$$

P.5 Properties of Probability Distributions

Figures P.1 and P.2 give us a picture of how frequently values of the random variables will occur. Two key features of a probability distribution are its center (location) and width (dispersion). A key measure of the center is the **mean**, or **expected value**. Measures of dispersion are **variance**, and its square root, the **standard deviation**.

P.5.1 Expected Value of a Random Variable

The **mean** of a random variable is given by its **mathematical expectation**. If X is a discrete random variable taking the values $x_1, \ldots, x_n$, then the mathematical expectation, or **expected value**, of X is

$$E(X) = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_n P(X = x_n) \tag{P.8}$$

The expected value, or mean, of X is a weighted average of its values, the weights being the probabilities that the values occur. **The uppercase letter "*E*" represents the expected value operation**. $E(X)$ is read as "the expected value of X." The expected value of X is also called the **mean of *X***. The mean is often symbolized by μ or μ_X. It is the average value of the random variable in an infinite number of repetitions of the underlying experiment. The mean of a random variable is the **population mean**. We use Greek letters for **population parameters** because later on we will use data to estimate these real world unknowns. In particular, keep separate the population mean μ and the arithmetic (or sample) mean $\bar{x}$ that we introduced in Section P.4 as Sum 6. This can be particularly confusing when a conversation includes the term "mean" without the qualifying term "population" or "arithmetic." Pay attention to the usage context.

EXAMPLE P.3 | Calculating an Expected Value

For the population in Table P.1, the expected value of X is

$$\begin{aligned} E(X) &= 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3) + 4 \times P(X = 4) \\ &= (1 \times 0.1) + (2 \times 0.2) + (3 \times 0.3) + (4 \times 0.4) = 3 \end{aligned}$$

For a discrete random variable the probability that X takes the value x is given by its *pdf* $f(x)$, $P(X = x) = f(x)$. The expected value in (P.8) can be written equivalently as

$$\begin{aligned} \mu_X = E(X) &= x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) \\ &= \sum_{i=1}^{n} x_i f(x_i) = \sum_{x} x f(x) \end{aligned} \tag{P.9}$$

Using (P.9), the expected value of X, the numeric value on a randomly drawn slip from Table P.1 is

$$\mu_X = E(X) = \sum_{x=1}^{4} x f(x) = (1 \times 0.1) + (2 \times 0.2) + (3 \times 0.3) + (4 \times 0.4) = 3$$

What does this mean? Draw one "slip" at random from Table P.1, and observe its numerical value X. This constitutes an experiment. If we repeat this experiment many times, the values $x = 1, 2, 3$, and 4 will appear 10%, 20%, 30%, and 40% of the time, respectively. The arithmetic average of all the numerical values will approach $\mu_X = 3$, as the number of experiments becomes large. The key point is that **the expected value of the random variable is the average value that occurs in many repeated trials of an experiment**.

For continuous random variables, the interpretation of the expected value of X is unchanged—it is the average value of X if many values are obtained by repeatedly performing the underlying random experiment.[4]

[4]Since there are now an uncountable number of values to sum, mathematically we must replace the "summation over all possible values" in (P.9) by the "integral over all possible values." See Appendix B.2.2 for a brief discussion.

P.5.2 Conditional Expectation

Many economic questions are formulated in terms of **conditional expectation**, or the **conditional mean**. One example is "What is the mean (expected value) wage of a person who has 16 years of education?" In expected value notation, what is $E(WAGE|EDUCATION = 16)$? For a discrete random variable, the calculation of conditional expected value uses (P.9) with the conditional *pdf* $f(x|y)$ replacing $f(x)$, so that

$$\mu_{X|Y} = E(X|Y = y) = \sum_x x f(x|y)$$

EXAMPLE P.4 | Calculating a Conditional Expectation

Using the population in Table P.1, what is the expected numerical value of X given that $Y = 1$, the slip is shaded? The conditional probabilities $f(x|y = 1)$ are given in Table P.5. The conditional expectation of X is

$$\begin{aligned} E(X|Y = 1) &= \sum_{x=1}^{4} x f(x|1) = 1 \times f(1|1) + 2 \times f(2|1) \\ &\quad + 3 \times f(3|1) + 4 \times f(4|1) \\ &= 1(0.25) + 2(0.25) + 3(0.25) + 4(0.25) = 2.5 \end{aligned}$$

The average value of X in many repeated trials of the experiment of drawing from the shaded slips is 2.5. This example makes a good point about expected values in general, namely that the expected value of X does not have to be a value that X can take. The expected value of X is **not** the value that you expect to occur in any single experiment.

What is the conditional expectation of X given that $Y = y$ if the random variables are statistically independent? If X and Y are statistically independent the conditional *pdf* $f(x|y)$ equals the *pdf* of X alone, $f(x)$, as shown in (P.6). The conditional expectation is then

$$E(X|Y = y) = \sum_x x f(x|y) = \sum_x x f(x) = E(X)$$

If X and Y are statistically independent, conditioning does not affect the expected value.

P.5.3 Rules for Expected Values

Functions of random variables are also random. If $g(X)$ is a function of the random variable X, such as $g(X) = X^2$, then $g(X)$ is also random. If X is a discrete random variable, then the expected value of $g(X)$ is obtained using calculations similar to those in (P.9).

$$E\left[g(X)\right] = \sum_x g(x) f(x) \tag{P.10}$$

For example, if a is a constant, then $g(X) = aX$ is a function of X, and

$$\begin{aligned} E(aX) = E\left[g(X)\right] &= \sum_x g(x) f(x) \\ &= \sum_x ax f(x) = a \sum_x x f(x) \\ &= aE(X) \end{aligned}$$

Similarly, if a and b are constants, then we can show that

$$E(aX + b) = aE(X) + b \tag{P.11}$$

If $g_1(X)$ and $g_2(X)$ are functions of X, then

$$E\left[g_1(X) + g_2(X)\right] = E\left[g_1(X)\right] + E\left[g_2(X)\right] \tag{P.12}$$

This rule extends to any number of functions. Remember the phrase "**the expected value of a sum is the sum of the expected values.**"

P.5.4 Variance of a Random Variable

The **variance** of a discrete or continuous random variable X is the expected value of

$$g(X) = [X - E(X)]^2$$

The variance of a random variable is important in characterizing the scale of measurement and the spread of the probability distribution. We give it the symbol σ^2, or σ_X^2, read "sigma squared." The variance σ^2 has a Greek symbol because it is a population parameter. Algebraically, letting $E(X) = \mu$, using the rules of expected values and the fact that $E(X) = \mu$ is not random, we have

$$\begin{aligned}\text{var}(X) = \sigma_X^2 &= E(X-\mu)^2 \\ &= E\left(X^2 - 2\mu X + \mu^2\right) = E\left(X^2\right) - 2\mu E(X) + \mu^2 \\ &= E\left(X^2\right) - \mu^2 \end{aligned} \tag{P.13}$$

We use the letters "var" to represent variance, and var(X) is read as "**the variance of** X," where X is a random variable. The calculation $\text{var}(X) = E\left(X^2\right) - \mu^2$ is usually simpler than $\text{var}(X) = E(X - \mu)^2$, but the solution is the same.

EXAMPLE P.5 | Calculating a Variance

For the population in Table P.1, we have shown that $E(X) = \mu = 3$. Using (P.10), the expectation of the random variable $g(X) = X^2$ is

$$\begin{aligned}E\left(X^2\right) &= \sum_{x=1}^{4} g(x)\,f(x) = \sum_{x=1}^{4} x^2 f(x) \\ &= \left[1^2 \times 0.1\right] + \left[2^2 \times 0.2\right] + \left[3^2 \times 0.3\right] + \left[4^2 \times 0.4\right] = 10\end{aligned}$$

Then, the variance of the random variable X is

$$\text{var}(X) = \sigma_X^2 = E\left(X^2\right) - \mu^2 = 10 - 3^2 = 1$$

The square root of the variance is called the **standard deviation**; it is denoted by σ or sometimes as σ_X if more than one random variable is being discussed. It also measures the spread or dispersion of a probability distribution and has the advantage of being in the same units of measure as the random variable.

A useful property of variances is the following. Let a and b be constants, then

$$\text{var}(aX + b) = a^2\text{var}(X) \tag{P.14}$$

An additive constant like b changes the mean (expected value) of a random variable, but it does not affect its dispersion (variance). A multiplicative constant like a affects the mean, and it affects the variance by the **square** of the constant.

To see this, let $Y = aX + b$. Using (P.11)

$$E(Y) = \mu_Y = aE(X) + b = a\mu_X + b$$

Then

$$\begin{aligned}\text{var}(aX + b) = \text{var}(Y) = E\left[\left(Y - \mu_Y\right)^2\right] &= E\left[\left(aX + b - \left(a\mu_X + b\right)\right)^2\right] \\ &= E\left[\left(aX - a\mu_X\right)^2\right] = E\left[a^2\left(X - \mu_X\right)^2\right] \\ &= a^2 E\left[\left(X - \mu_X\right)^2\right] = a^2\text{var}(X)\end{aligned}$$

The variance of a random variable is the *average* squared difference between the random variable X and its mean value μ_X. The larger the variance of a random variable, the more "spread out" the

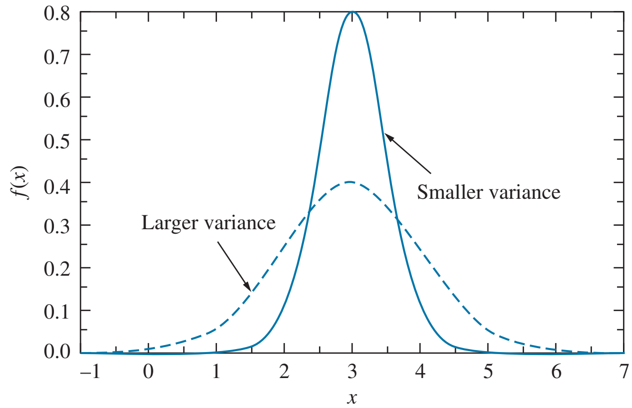

FIGURE P.3 **Distributions with different variances.**

values of the random variable are. Figure P.3 shows two *pdf*s for a continuous random variable, both with mean $\mu = 3$. The distribution with the smaller variance (the solid curve) is less spread out about its mean.

P.5.5 Expected Values of Several Random Variables

Let X and Y be random variables. The rule "the expected value of the sum is the sum of the expected values" applies. Then[5]

$$E(X + Y) = E(X) + E(Y) \tag{P.15}$$

Similarly

$$E(aX + bY + c) = aE(X) + bE(Y) + c \tag{P.16}$$

The product of random variables is not as easy. $E(XY) = E(X)E(Y)$ **if** X and Y are independent. These rules can be extended to more random variables.

P.5.6 Covariance Between Two Random Variables

The **covariance** between X and Y is a measure of linear association between them. Think about two continuous variables, such as height and weight of children. We expect that there is an association between height and weight, with taller than average children tending to weigh more than the average. The product of X minus its mean times Y minus its mean is

$$\left(X - \mu_X\right)\left(Y - \mu_Y\right) \tag{P.17}$$

In Figure P.4, we plot values (x and y) of X and Y that have been constructed so that $E(X) = E(Y) = 0$.

The x and y values of X and Y fall predominately in quadrants I and III, so that the arithmetic average of the values $(x - \mu_X)(y - \mu_Y)$ is positive. We define the covariance between two random variables as the expected (population average) value of the product in (P.17).

$$\text{cov}(X, Y) = \sigma_{XY} = E\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right] = E(XY) - \mu_X\mu_Y \tag{P.18}$$

We use the letters "cov" to represent covariance, and cov(X, Y) is read as "**the covariance between X and Y**," where X and Y are random variables. The covariance σ_{XY} of the random variables underlying Figure P.4 is positive, which tells us that when the values x are greater

[5]These results are proven in Appendix B.1.4.

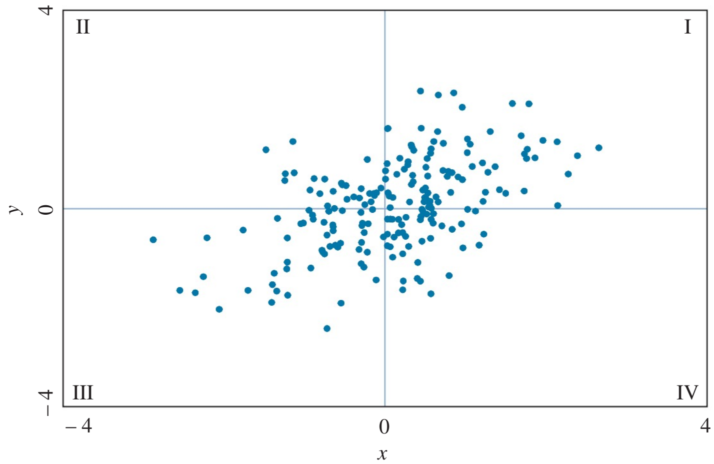

FIGURE P.4 **Correlated data.**

than μ_X, then the values y also tend to be greater than μ_Y; and when the values x are below μ_X, then the values y also tend to be less than μ_Y. If the random variables values tend primarily to fall in quadrants II and IV, then $(x - \mu_X)(y - \mu_Y)$ will tend to be negative and σ_{XY} will be negative. If the random variables values are spread evenly across the four quadrants, and show neither positive nor negative association, then the covariance is zero. The sign of σ_{XY} tells us whether the two random variables X and Y are positively associated or negatively associated.

Interpreting the actual value of σ_{XY} is difficult because X and Y may have different units of measurement. Scaling the covariance by the standard deviations of the variables eliminates the units of measurement, and defines the **correlation** between X and Y

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{P.19}$$

As with the covariance, the correlation ρ between two random variables measures the degree of *linear* association between them. However, unlike the covariance, the correlation must lie between –1 and 1. Thus, the correlation between X and Y is 1 or –1 if X is a perfect positive or negative linear function of Y. If there is *no linear* association between X and Y, then $\text{cov}(X, Y) = 0$ and $\rho = 0$. For other values of correlation the magnitude of the absolute value $|\rho|$ indicates the "strength" of the linear association between the values of the random variables. In Figure P.4, the correlation between X and Y is $\rho = 0.5$.

EXAMPLE P.6 | Calculating a Correlation

To illustrate the calculation, reconsider the population in Table P.1 with joint *pdf* given in Table P.4. The expected value of XY is

$$\begin{aligned} E(XY) &= \sum_{y=0}^{1} \sum_{x=1}^{4} xy f(x, y) \\ &= (1 \times 0 \times 0) + (2 \times 0 \times 0.1) + (3 \times 0 \times 0.2) \\ &\quad + (4 \times 0 \times 0.3) + (1 \times 1 \times 0.1) \\ &\quad + (2 \times 1 \times 0.1) + (3 \times 1 \times 0.1) + (4 \times 1 \times 0.1) \\ &= 0.1 + 0.2 + 0.3 + 0.4 \\ &= 1 \end{aligned}$$

The random variable X has expected value $E(X) = \mu_X = 3$ and the random variable Y has expected value $E(Y) = \mu_Y = 0.4$. Then the covariance between X and Y is

$$\text{cov}(X, Y) = \sigma_{XY} = E(XY) - \mu_X \mu_Y = 1 - 3 \times (0.4) = -0.2$$

The correlation between X and Y is

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{-0.2}{\sqrt{1} \times \sqrt{0.24}} = -0.4082$$

If X and Y are independent random variables, then their covariance and correlation are zero. The converse of this relationship is **not** true. Independent random variables X and Y have zero covariance, indicating that there is no linear association between them. However, just because the covariance or correlation between two random variables is zero **does not** mean that they are necessarily independent. There may be more complicated nonlinear associations such as $X^2 + Y^2 = 1$.

In (P.15) we obtain the expected value of a sum of random variables. There are similar rules for variances. If a and b are constants, then

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\,\text{cov}(X, Y) \tag{P.20}$$

A significant point to note is that the variance of a sum is **not** just the sum of the variances. There is a covariance term present. Two special cases of (P.20) are

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \tag{P.21}$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y) \tag{P.22}$$

To show that (P.22) is true, let $Z = X - Y$. Using the rules of expected value

$$E(Z) = \mu_Z = E(X) - E(Y) = \mu_X - \mu_Y$$

The variance of $Z = X - Y$ is obtained using the basic definition of variance, with some substitution,

$$\begin{aligned}
\text{var}(X - Y) = \text{var}(Z) = E\left[(Z - \mu_Z)^2\right] &= E\left[\left(X - Y - (\mu_X - \mu_Y)\right)^2\right] \\
&= E\left\{\left[(X - \mu_X) - (Y - \mu_Y)\right]^2\right\} \\
&= E\left\{(X - \mu_X)^2 + (Y - \mu_Y)^2 - 2(X - \mu_X)(Y - \mu_Y)\right\} \\
&= E\left[(X - \mu_X)^2\right] + E\left[(Y - \mu_Y)^2\right] - 2E\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y)
\end{aligned}$$

If X and Y are independent, or if $\text{cov}(X, Y) = 0$, then

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) \tag{P.23}$$

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \tag{P.24}$$

These rules extend to more random variables.

P.6 Conditioning

In Table P.4, we summarized the joint and marginal probability functions for the random variables X and Y defined on the population in Table P.1. In Table P.6 we make two modifications. First, the probabilities are expressed as fractions. The many calculations below are simpler using arithmetic

TABLE P.6 **Joint, Marginal, and Conditional Probabilities**

y/x	1	2	3	4	$f(y)$	$f(y\mid x=1)$	$f(y\mid x=2)$	$f(y\mid x=3)$	$f(y\mid x=4)$
0	0	1/10	2/10	3/10	6/10	0	1/2	2/3	3/4
1	1/10	1/10	1/10	1/10	4/10	1	1/2	1/3	1/4
$f(x)$	1/10	2/10	3/10	4/10					
$f(x\mid y=0)$	0	1/6	2/6	3/6					
$f(x\mid y=1)$	1/4	1/4	1/4	1/4					

with fractions. Second, we added the conditional probability functions (P.4) for Y given each of the values that X can take and the conditional probability functions for X given each of the values that Y can take. Now would be a good time for you to review Section P.3.2 on conditional probability. For example, what is the probability that $Y = 0$ given that $X = 2$? That is, if we only consider population members with $X = 2$, what is the probability that $Y = 0$? There are only two population elements with $X = 2$, one with $Y = 0$ and one with $Y = 1$. The probability of randomly selecting $Y = 0$ is one-half. For discrete random variables, the conditional probability is calculated as the joint probability divided by the probability of the conditioning event.

$$f(y = 0|x = 2) = P(Y = 0|X = 2) = \frac{P(X = 2, Y = 0)}{P(X = 2)} = \frac{1/10}{2/10} = \frac{1}{2}$$

In the following sections, we discuss the concepts of **conditional expectation** and **conditional variance**.

P.6.1 Conditional Expectation

Many economic questions are formulated in terms of a **conditional expectation**, or the **conditional mean**. One example is "What is the mean wage of a person who has 16 years of education?" In expected value notation, what is $E(WAGE|EDUC = 16)$? The effect of conditioning on the value of $EDUC$ is to reduce the population of interest to only individuals with 16 years of education. The mean, or expected value, of wage for these individuals may be quite different than the mean wage for all individuals regardless of years of education, $E(WAGE)$, which is the **unconditional expectation** or **unconditional mean**.

For discrete random variables,[6] the calculation of a conditional expected value uses equation (P.9) with the conditional *pdf* replacing the usual *pdf*, so that

$$\begin{aligned} E(X|Y = y) &= \textstyle\sum_x x f(x|y) = \sum_x xP(X = x|Y = y) \\ E(Y|X = x) &= \textstyle\sum_y y f(y|x) = \sum_y yP(Y = y|X = x) \end{aligned} \tag{P.25}$$

EXAMPLE P.7 | Conditional Expectation

Using the population in Table P.1, what is the expected numerical value of X given that $Y = 1$? The conditional probabilities $P(X = x|Y = 1) = f(x|y = 1) = f(x|1)$ are given in Table P.6. The conditional expectation of X is

$$\begin{aligned} E(X|Y = 1) &= \textstyle\sum_{x=1}^{4} x f(x|1) \\ &= [1 \times f(1|1)] + [2 \times f(2|1)] + [3 \times f(3|1)] \\ &\quad + [4 \times f(4|1)] \\ &= 1(1/4) + 2(1/4) + 3(1/4) + 4(1/4) = 10/4 \\ &= 5/2 \end{aligned}$$

The average value of X in many repeated trials of the experiment of drawing from the shaded slips ($Y = 1$) is 2.5. This example makes a good point about expected values in general, namely that the expected value of X does not have to be a value that X can take. The expected value of X is **not** the value that you expect to occur in any single experiment. It is the average value of X after many repetitions of the experiment.

What is the expected value of X given that we only consider values where $Y = 0$? Confirm that $E(X|Y = 0) = 10/3$. For comparison purposes recall from Section P.5.1 that the **unconditional expectation** of X is $E(X) = 3$.

Similarly, if we condition on the X values, the conditional expectations of Y are

$$E(Y|X = 1) = \textstyle\sum_y y f(y|1) = 0(0) + 1(1) = 1$$

$$E(Y|X = 2) = \textstyle\sum_y y f(y|2) = 0(1/2) + 1(1/2) = 1/2$$

$$E(Y|X = 3) = \textstyle\sum_y y f(y|3) = 0(2/3) + 1(1/3) = 1/3$$

$$E(Y|X = 4) = \textstyle\sum_y y f(y|4) = 0(3/4) + 1(1/4) = 1/4$$

Note that $E(Y|X)$ varies as X varies; it is a function of X. For comparison, the unconditional expectation of Y, $E(Y)$, is

$$E(Y) = \textstyle\sum_y y f(y) = 0(6/10) + 1(4/10) = 2/5$$

[6]For continuous random variables the sums are replaced by integrals. See Appendix B.2.

P.6.2 Conditional Variance

The **unconditional variance** of a discrete random variable X is

$$\operatorname{var}(X) = \sigma_X^2 = E\left[\left(X - \mu_X\right)^2\right] = \sum_x \left(x - \mu_X\right)^2 f(x) \tag{P.26}$$

It measures how much variation there is in X around the unconditional mean of X, μ_X. For example, the unconditional variance var(*WAGE*) measures the variation in *WAGE* around the unconditional mean $E(WAGE)$. In (P.13) we show that equivalently

$$\operatorname{var}(X) = \sigma_X^2 = E\left(X^2\right) - \mu_X^2 = \sum_x x^2 f(x) - \mu_X^2 \tag{P.27}$$

In Section P.6.1 we discussed how to answer the question "What is the mean wage of a person who has 16 years of education?" Now we ask "How much variation is there in wages for a person who has 16 years of education?" The answer to this question is given by the **conditional variance**, $\operatorname{var}(WAGE|EDUC = 16)$. The conditional variance measures the variation in *WAGE* around the conditional mean $E(WAGE|EDUC = 16)$ for individuals with 16 years of education. The conditional variance of *WAGE* for individuals with 16 years of education is the average squared difference in the population between *WAGE* and the conditional mean of *WAGE*,

$$\underbrace{\operatorname{var}(WAGE \mid EDUC = 16)}_{\text{conditional variance}} = E\left\{\left[WAGE - \underbrace{E(WAGE \mid EDUC = 16)}_{\text{conditional mean}}\right]^2 \middle| EDUC = 16\right\}$$

To obtain the conditional variance we modify the definitions of variance in equations (P.26) and (P.27); replace the unconditional mean $E(X) = \mu_X$ with the conditional mean $E(X|Y = y)$, and the unconditional *pdf* $f(x)$ with the conditional *pdf* $f(x|y)$. Then

$$\operatorname{var}(X|Y = y) = E\left\{\left[X - E(X|Y = y)\right]^2 \middle| Y = y\right\} = \sum_x \left(x - E(X|Y = y)\right)^2 f(x|y) \tag{P.28}$$

or

$$\operatorname{var}(X|Y = y) = E\left(X^2|Y = y\right) - \left[E(X|Y = y)\right]^2 = \sum_x x^2 f(x|y) - \left[E(X|Y = y)\right]^2 \tag{P.29}$$

EXAMPLE P.8 | Conditional Variance

For the population in Table P.1, the unconditional variance of X is $\operatorname{var}(X) = 1$. What is the variance of X given that $Y = 1$? To use (P.29) first compute

$$\begin{aligned} E\left(X^2|Y = 1\right) &= \sum_x x^2 f(x|Y = 1) \\ &= 1^2(1/4) + 2^2(1/4) + 3^2(1/4) + 4^2(1/4) = 15/2 \end{aligned}$$

Then

$$\begin{aligned} \operatorname{var}(X|Y = 1) &= E\left(X^2|Y = 1\right) - \left[E(X|Y = 1)\right]^2 \\ &= 15/2 - (5/2)^2 = 5/4 \end{aligned}$$

In this case, the conditional variance of X, given that $Y = 1$, is larger than the unconditional variance of X, $\operatorname{var}(X) = 1$.

To calculate the conditional variance of X given that $Y = 0$, we first obtain

$$\begin{aligned} E\left(X^2|Y = 0\right) &= \sum_x x^2 f(x|Y = 0) \\ &= 1^2(0) + 2^2(1/6) + 3^2(2/6) + 4^2(3/6) \\ &= 35/3 \end{aligned}$$

Then

$$\begin{aligned} \operatorname{var}(X|Y = 0) &= E\left(X^2|Y = 0\right) - \left[E(X|Y = 0)\right]^2 \\ &= 35/3 - (10/3)^2 = 5/9 \end{aligned}$$

In this case, the conditional variance of X, given that $Y = 0$, is smaller than the unconditional variance of X, $\operatorname{var}(X) = 1$. These examples have illustrated that in general the conditional variance can be larger or smaller than the unconditional variance. Try working out $\operatorname{var}(Y|X = 1)$, $\operatorname{var}(Y|X = 2)$, $\operatorname{var}(Y|X = 3)$, and $\operatorname{var}(Y|X = 4)$.

P.6.3 Iterated Expectations

The Law of **Iterated Expectations** says that we can find the expected value of Y in two steps. First, find the conditional expectation $E(Y|X)$. Second, find the expected value $E(Y|X)$ treating X as random.

$$\text{Law of Iterated Expectations: } E(Y) = E_X\left[E(Y|X)\right] = \sum_x E(Y|X=x)\, f_X(x) \tag{P.30}$$

In this expression we put an "X" subscript in the expectation $E_X[E(Y|X)]$ and the probability function $f_X(x)$ to emphasize that we are treating X as random. The Law of Iterated Expectations is true for both discrete and continuous random variables.

EXAMPLE P.9 | Iterated Expectation

Consider the conditional expectation $E(X|Y=y) = \sum_x xf(x|y)$. As we computed in Section P.6.1, $E(X|Y=0) = 10/3$ and $E(X|Y=1) = 5/2$. Similarly, the conditional expectation $E(Y|X=x) = \sum_y yf(y|x)$. For the population in Table P.1, these conditional expectations were calculated in Section P.6.1 to be $E(Y|X=1) = 1$, $E(Y|X=2) = 1/2$, $E(Y|X=3) = 1/3$ and $E(Y|X=4) = 1/4$. Note that $E(Y|X=x)$ changes when x changes. If X is allowed to vary randomly[7] then the conditional expectation varies randomly. The conditional expectation is a function of X, or $E(Y|X) = g(X)$, and is random when viewed this way. Using (P.10) we can find the expected value of $g(X)$.

$$\begin{aligned}
E_X\left[E(Y|X)\right] &= E_X\left[g(X)\right] = \sum_x g(x) f_X(x) = \sum_x E(Y|X=x)\, f_X(x)\\
&= \left[E(Y|X=1)\, f_X(1)\right] + \left[E(Y|X=2)\, f_X(2)\right]\\
&\quad + \left[E(Y|X=3)\, f_X(3)\right] + \left[E(Y|X=4)\, f_X(4)\right]\\
&= 1(1/10) + (1/2)(2/10) + (1/3)(3/10)\\
&\quad + (1/4)(4/10) = 2/5
\end{aligned}$$

If we draw many values x from the population in Table P.1, the average of $E(Y|X)$ is 2/5. For comparison the "unconditional" expectation of Y is $E(Y) = 2/5$. $E_X[E(Y|X)]$ and $E(Y)$ are the same.

Proof of the Law of Iterated Expectations To prove the Law of Iterated Expectations we make use of relationships between joint, marginal, and conditional *pdf*s that we introduced in Section P.3. In Section P.3.1 we discussed *marginal distributions*. Given a joint *pdf* $f(x, y)$ we can obtain the *marginal pdf* of y alone $f_Y(y)$ by summing, for each y, the joint *pdf* $f(x, y)$ across all values of the variable we wish to eliminate, in this case x. That is, for Y and X,

$$\begin{aligned}
f(y) &= f_Y(y) = \sum_x f(x,y)\\
f(x) &= f_X(x) = \sum_y f(x,y)
\end{aligned} \tag{P.31}$$

Because $f(\,)$ is used to represent *pdf*s in general, sometimes we will put a subscript, X or Y, to be very clear about which variable is random.

Using equation (P.4) we can define the conditional *pdf* of y given $X = x$ as

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

Rearrange this expression to obtain

$$f(x,y) = f(y|x)\, f_X(x) \tag{P.32}$$

A joint *pdf* is the product of the conditional *pdf* and the *pdf* of the conditioning variable.

---

[7] Imagine shuffling the population elements and randomly choosing one. This is an experiment and the resulting number showing is a value of X. By doing this repeatedly X varies randomly.

To show that the Law of Iterated Expectations is true[8] we begin with the definition of the expected value of Y, and operate with the summation.

$$\begin{aligned}
E(Y) &= \sum_y y f(y) = \sum_y y \left[\sum_x f(x,y)\right] && \left[\text{substitute for } f(y)\right] \\
&= \sum_y y \left[\sum_x f(y|x)\, f_X(x)\right] && \left[\text{substitute for } f(x,y)\right] \\
&= \sum_x \left[\sum_y y f(y|x)\right] f_X(x) && \left[\text{change order of summation}\right] \\
&= \sum_x E(Y|x) f_X(x) && \left[\text{recognize the conditional expectation}\right] \\
&= E_X\left[E(Y|X)\right]
\end{aligned}$$

While this result may seem an esoteric oddity it is very important and widely used in modern econometrics.

P.6.4 Variance Decomposition

Just as we can break up the expected value using the Law of Iterated Expectations we can decompose the variance of a random variable into two parts.

$$\text{Variance Decomposition: } \operatorname{var}(Y) = \operatorname{var}_X\left[E(Y|X)\right] + E_X\left[\operatorname{var}(Y|X)\right] \tag{P.33}$$

This "beautiful" result[9] says that the variance of the random variable Y equals the sum of the variance of the conditional mean of Y given X and the mean of the conditional variance of Y given X. In this section we will discuss this result.[10]

Suppose that we are interested in the wages of the population consisting of working adults. How much variation do wages display in the population? If $WAGE$ is the wage of a randomly drawn population member, then we are asking about the variance of $WAGE$, that is, $\operatorname{var}(WAGE)$. The variance decomposition says

$$\operatorname{var}(WAGE) = \operatorname{var}_{EDUC}\left[E(WAGE|EDUC)\right] + E_{EDUC}\left[\operatorname{var}(WAGE|EDUC)\right]$$

$E(WAGE|EDUC)$ is the expected value of $WAGE$ given a specific value of education, such as $EDUC = 12$ or $EDUC = 16$. $E(WAGE|EDUC = 12)$ is the average $WAGE$ in the population, given that we only consider workers who have 12 years of education. If $EDUC$ changes then the conditional mean $E(WAGE|EDUC)$ changes, so that $E(WAGE|EDUC = 16)$ is not the same as $E(WAGE|EDUC = 12)$, and in fact we expect $E(WAGE|EDUC = 16) > E(WAGE|EDUC = 12)$; more education means more "human capital" and thus the average wage should be higher. The first component in the variance decomposition $\operatorname{var}_{EDUC}\left[E(WAGE|EDUC)\right]$ measures the variation in $E(WAGE|EDUC)$ due to variation in education.

The second part of the variance decomposition is $E_{EDUC}\left[\operatorname{var}(WAGE|EDUC)\right]$. If we restrict our attention to population members who have 12 years of education, the mean wage is $E(WAGE|EDUC = 12)$. Within the group of workers who have 12 years of education we will observe wide ranges of wages. For example, using one sample of *CPS* data from 2013,[11] wages for those with 12 years of education varied from \$3.11/hour to \$100.00/hour; for those with 16 years of education wages varied from \$2.75/hour to \$221.10/hour. For workers with 12 and 16 years of education that variation is measured by $\operatorname{var}(WAGE|EDUC = 12)$ and

---

[8]The proof for continuous variables is in Appendix B.2.4.

[9]Tony O'Hagan, "A Thing of Beauty," *Significance Magazine*, Volume 9 Issue 3 (June 2012), 26–28.

[10]The proof of the variance decomposition is given in Appendix B.1.8 and Example B.1.

[11]The data file *cps5*.

$\text{var}(WAGE|EDUC = 16)$. The term $E_{EDUC}\left[\text{var}(WAGE|EDUC)\right]$ measures the average of $\text{var}(WAGE|EDUC)$ as education changes.

To summarize, the variation of $WAGE$ in the population can be attributed to two sources: variation in the conditional mean $E(WAGE|EDUC)$ and variation due to changes in education in the conditional variance of $WAGE$ given education.

P.6.5 Covariance Decomposition

Recall that the covariance between two random variables Y and X is $\text{cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$. For discrete random variables this is

$$\text{cov}(X,Y) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)\, f(x,y)$$

By using the relationships between marginal, conditional and joint *pdf*s we can show

$$\text{cov}(X,Y) = \sum_x (x-\mu_X)\, E(Y|X=x)\, f(x) \tag{P.34}$$

Recall that $E(Y|X) = g(X)$ so this result says that the covariance between X and Y can be calculated as the expected value of X, minus its mean, times a function of X, $\text{cov}(X,Y) = E_X[(X-\mu_X)E(Y|X)]$.

An important special case is important in later chapters. When the conditional expectation of Y given X is a constant, $E(Y|X=x) = c$, then

$$\text{cov}(X,Y) = \sum_x (x-\mu_X)\, E(Y|X=x)\, f(x) = c\sum_x (x-\mu_X)\, f(x) = 0$$

A special case is $E(Y|X=x) = 0$, which by direct substitution implies $\text{cov}(X,Y) = 0$.

EXAMPLE P.10 | Covariance Decomposition

To illustrate we compute $\text{cov}(X, Y)$ for the population in Table P.1 using the covariance decomposition. We have computed that $\text{cov}(X, Y) = -0.2$ in Section P.5.6. The ingredients are the values of the random variable X, its mean $\mu_X = 3$, the probabilities $P(X = x) = f(x)$ and conditional expectations

$$E(Y|X=1) = 1,\ E(Y|X=2) = 1/2,$$
$$E(Y|X=3) = 1/3 \text{ and } E(Y|X=4) = 1/4$$

Using the covariance decomposition we have

$$\begin{aligned}\text{cov}(X,Y) &= \sum_x (x-\mu_X)\, E(Y|X=x)\, f(x)\\ &= (1-3)(1)(1/10) + (2-3)(1/2)(2/10)\\ &\quad + (3-3)(1/3)(3/10) + (4-3)(1/4)(4/10)\\ &= -2/10 - 1/10 + 1/10 = -2/10 = -0.2\end{aligned}$$

We see that the covariance decomposition yields the correct result, and it is convenient in this example.

P.7 The Normal Distribution

In the previous sections we discussed random variables and their *pdf*s in a general way. In real economic contexts, some specific *pdf*s have been found to be very useful. The most important is the **normal distribution**. If X is a normally distributed random variable with mean μ and variance σ^2, it is symbolized as $X \sim N(\mu, \sigma^2)$. The *pdf* of X is given by the impressive formula

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty \tag{P.35}$$

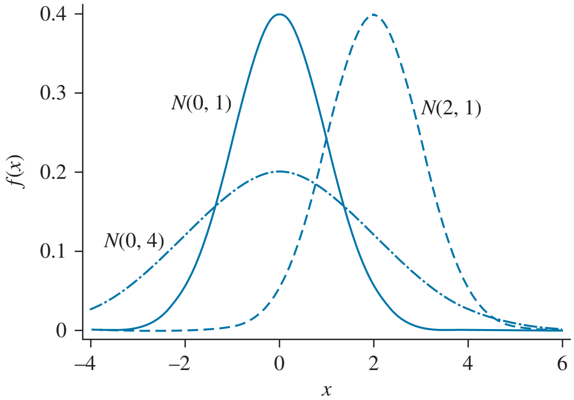

FIGURE P.5 **Normal probability density functions $N(\mu, \sigma^2)$.**

where exp(a) denotes the exponential[12] function e^a. The mean μ and variance σ^2 are the parameters of this distribution and determine its center and dispersion. The range of the continuous normal random variable is from minus infinity to plus infinity. Pictures of the normal *pdf*s are given in Figure P.5 for several values of the mean and variance. Note that the distribution is symmetric and centered at μ.

Like all continuous random variables, probabilities involving normal random variables are found as areas under the *pdf*. For calculating probabilities both computer software and statistical tables values make use of the relation between a normal random variable and its "standardized" equivalent. A **standard normal random variable** is one that has a normal *pdf* with mean 0 and variance 1. If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \tag{P.36}$$

The standard normal random variable Z is so widely used that its *pdf* and *cdf* are given their own special notation. The *cdf* is denoted $\Phi(z) = P(Z \le z)$. Computer programs, and Statistical Table 1 in Appendix D give values of $\Phi(z)$. The *pdf* for the standard normal random variable is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \quad -\infty < z < \infty$$

Values of the density function are given in Statistical Table 6 in Appendix D. To calculate normal probabilities, remember that the distribution is symmetric, so that $P(Z > a) = P(Z < -a)$, and $P(Z > a) = P(Z \ge a)$, since the probability of any one point is zero for a continuous random variable. If $X \sim N(\mu, \sigma^2)$ and a and b are constants, then

$$P(X \le a) = P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{P.37}$$

$$P(X > a) = P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = P\left(Z > \frac{a-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{P.38}$$

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{P.39}$$

[12]See Appendix A.1.2 for a review of exponents.

EXAMPLE P.11 | Normal Distribution Probability Calculation

For example, if $X \sim N(3, 9)$, then

$P(4 \leq X \leq 6) = P(0.33 \leq Z \leq 1) = \Phi(1) - \Phi(0.33) = 0.8413 - 0.6293 = 0.2120$

In addition to finding normal probabilities we sometimes must find a value z_α of a standard normal random variable such that $P(Z \leq z_\alpha) = \alpha$. The value z_α is called the **100α-percentile**. For example, $z_{0.975}$ is the value of Z such that $P(Z \leq z_{0.975}) = 0.975$. This particular percentile can be found using Statistical Table 1, Cumulative Probabilities for the Standard Normal Distribution. The cumulative probability associated with the value $z = 1.96$ is $P(Z \leq 1.96) = 0.975$, so that the 97.5 percentile is $z_{0.975} = 1.96$. Using Statistical Table 1 we can only roughly obtain other percentiles. Using the cumulative probabilities $P(Z \leq 1.64) = 0.9495$ and $P(Z \leq 1.65) = 0.9505$ we can say that the 95th percentile of the **standard normal distribution** is between 1.64 and 1.65, and is about 1.645.

Luckily computer software makes these approximations unnecessary. The **inverse normal** function finds percentiles z_α given α. Formally, if $P(Z \leq z_\alpha) = \Phi(z_\alpha) = \alpha$ then $z_\alpha = \Phi^{-1}(\alpha)$. Econometric software, even spreadsheets, have the inverse normal function built in. Some commonly used percentiles are shown in Table P.7. In the last column are the percentiles rounded to fewer decimals. It would be useful for you to remember the numbers 2.58, 1.96, and 1.645.

An interesting and useful fact about the normal distribution is that a weighted sum of normal random variables has a normal distribution. That is, if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ then

$$Y = a_1X_1 + a_2X_2 \sim N\left(\mu_Y = a_1\mu_1 + a_2\mu_2, \sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_{12}\right) \tag{P.40}$$

where $\sigma_{12} = \text{cov}(X_1, X_2)$. A number of important probability distributions are related to the normal distribution. The t-distribution, the chi-square distribution, and the F-distribution are discussed in Appendix B.

TABLE P.7 Standard Normal Percentiles

α	$z_\alpha = \Phi^{-1}(\alpha)$	Rounded
0.995	2.57583	2.58
0.990	2.32635	2.33
0.975	1.95996	1.96
0.950	1.64485	1.645
0.900	1.28155	1.28
0.100	−1.28155	−1.28
0.050	−1.64485	−1.645
0.025	−1.95996	−1.96
0.010	−2.32635	−2.33
0.005	−2.57583	−2.58

P.7.1 The Bivariate Normal Distribution

Two continuous random variables, X and Y, have a **joint normal**, or **bivariate normal**, distribution if their joint *pdf* takes the form

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] \Big/ 2\left(1-\rho^2\right)\right\}$$

where $-\infty < x < \infty$, $-\infty < y < \infty$. The parameters μ_X and μ_Y are the means of X and Y, σ_X^2 and σ_Y^2 are the variances of X and Y, so that σ_X and σ_Y are the standard deviations. The parameter ρ is the correlation between X and Y. If $\text{cov}(X, Y) = \sigma_{XY}$ then

$$\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

The complex equation for $f(x, y)$ defines a surface in three-dimensional space. In Figure P.6a[13] we depict the surface if $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and $\rho = 0.7$. The positive correlation means there is a positive linear association between the values of X and Y, as described in Figure P.4. Figure P.6b depicts the contours of the density, the result of slicing the density horizontally, at a given height. The contours are more "cigar-shaped" the larger the absolute value of the correlation ρ. In Figure P.7a the correlation is $\rho = 0$. In this case the joint density is symmetrical and the contours in Figure P.7b are circles. If X and Y are jointly normal then they are statistically independent if, and only if, $\rho = 0$.

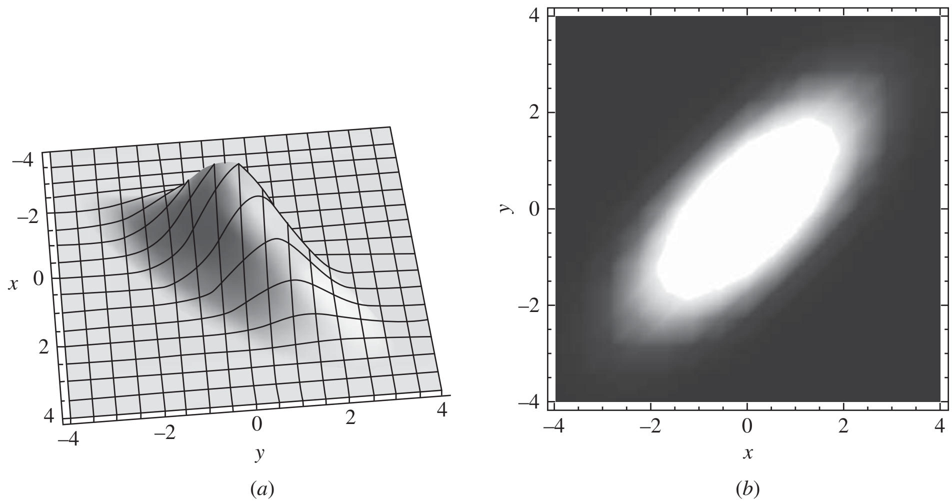

FIGURE P.6 **The bivariate normal distribution: $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and $\rho = 0.7$.**

[13]"The Bivariate Normal Distribution" from the Wolfram Demonstrations Project http://demonstrations.wolfram.com/. Figures P.6, P.7, and P.8 represent the interactive graphics on the site as static graphics for the primer. The site permits easy manipulation of distribution parameters. The joint density function figure can be rotated and viewed from different angles.

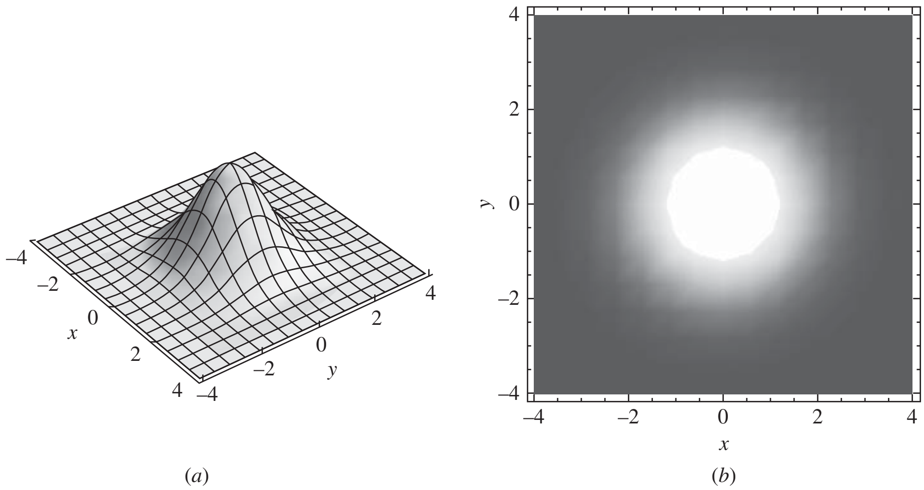

FIGURE P.7 **The bivariate normal distribution: $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and $\rho = 0$.**

There are several relations between the normal, bivariate normal, and the conditional distributions that are used in statistics and econometrics. First, if X and Y have a bivariate normal distribution then the marginal distributions of X and Y are normal distributions too, $X \sim N\left(\mu_X, \sigma_X^2\right)$ and $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$.

Second, the conditional distribution for Y given X is normal, with conditional mean $E(Y|X) = \alpha + \beta X$, where $\alpha = \mu_Y - \beta\mu_X$ and $\beta = \sigma_{XY}/\sigma_X^2$, and conditional variance $\text{var}(Y|X) = \sigma_Y^2\left(1 - \rho^2\right)$. Or $Y|X \sim N\left[\alpha + \beta X, \sigma_Y^2\left(1 - \rho^2\right)\right]$. Three noteworthy points about these results are (i) that the conditional mean is a linear function of X, and is called a **linear regression function**; (ii) the conditional variance is constant and does not vary with X; and (iii) the conditional variance is smaller than the unconditional variance if $\rho \neq 0$. In Figure P.8[14] we display a joint normal density with

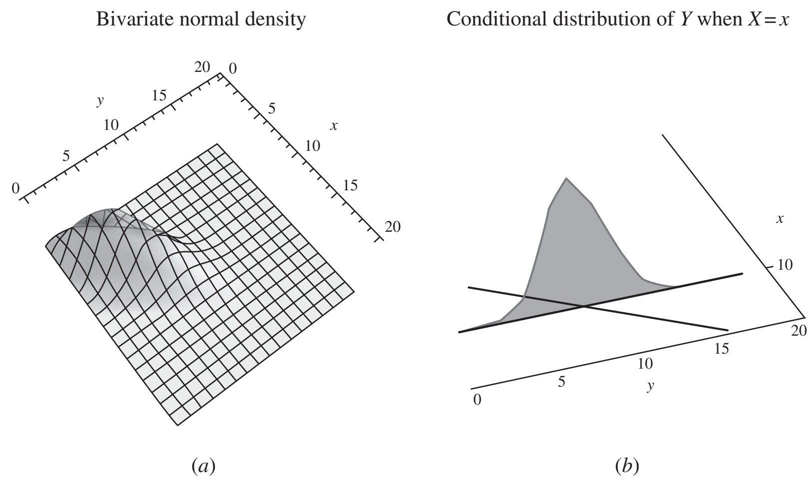

FIGURE P.8 **(*a*) Bivariate normal distribution with $\mu_X = \mu_Y = 5$, $\sigma_X = \sigma_Y = 3$, and $\rho = 0.7$; (*b*) conditional distribution of Y given $X = 10$.**

[14] "The Bivariate Normal and Conditional Distributions" from the Wolfram Demonstrations Project http://demonstrations.wolfram.com/TheBivariateNormalAndConditionalDistributions/. Both the bivariate distribution and conditional distributions can be rotated and viewed from different perspectives.

$\mu_X = \mu_Y = 5$, $\sigma_X = \sigma_Y = 3$, and $\rho = 0.7$. The covariance between X and Y is $\sigma_{XY} = \rho\sigma_X\sigma_Y = 0.7 \times 3 \times 3 = 6.3$ so that $\beta = \sigma_{XY}/\sigma_X^2 = 6.3/9 = 0.7$ and $\alpha = \mu_Y - \beta\mu_X = 5 - 0.7 \times 5 = 1.5$. The conditional mean of Y given $X = 10$ is $E(Y|X = 10) = \alpha + \beta X = 1.5 + 0.7X = 1.5 + 0.7 \times 10 = 8.5$. The conditional variance is $\text{var}(Y|X = 10) = \sigma_Y^2\left(1 - \rho^2\right) = 3^2\left(1 - 0.7^2\right) = 9(0.51) = 4.59$. That is, the conditional distribution is $(Y|X = 10) \sim N(8.5, 4.59)$.

P.8 Exercises

Answers to odd-numbered exercises are on the book website www.principlesofeconometrics.com/poe5.

P.1 Let $x_1 = 17, x_2 = 1, x_3 = 0; y_1 = 5, y_2 = 2, y_3 = 8$. Calculate the following:

a. $\sum_{i=1}^{2} x_i$
b. $\sum_{t=1}^{3} x_t y_t$
c. $\bar{x} = \left(\sum_{i=1}^{3} x_i\right)/3$ [Note: $\bar{x}$ is called the arithmetic average or arithmetic mean.]
d. $\sum_{i=1}^{3} (x_i - \bar{x})$
e. $\sum_{i=1}^{3} (x_i - \bar{x})^2$
f. $\left(\sum_{i=1}^{3} x_i^2\right) - 3\bar{x}^2$
g. $\sum_{i=1}^{3} (x_i - \bar{x})(y_i - \bar{y})$ where $\bar{y} = \left(\sum_{i=1}^{3} y_i\right)/3$
h. $\left(\sum_{j=1}^{3} x_j y_j\right) - 3\bar{x}\bar{y}$

P.2 Express each of the following sums in summation notation.

a. $(x_1/y_1) + (x_2/y_2) + (x_3/y_3) + (x_4/y_4)$
b. $y_2 + y_3 + y_4$
c. $x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$
d. $x_3y_5 + x_4y_6 + x_5y_7$
e. $(x_3/y_3^2) + (x_4/y_4^2)$
f. $(x_1 - y_1) + (x_2 - y_2) + (x_3 - y_3) + (x_4 - y_4)$

P.3 Write out each of the following sums and compute where possible.

a. $\sum_{i=1}^{3} (a - bx_i)$
b. $\sum_{t=1}^{4} t^2$
c. $\sum_{x=0}^{2} (2x^2 + 3x + 1)$
d. $\sum_{x=2}^{4} f(x + 3)$
e. $\sum_{x=1}^{3} f(x, y)$
f. $\sum_{x=3}^{4} \sum_{y=1}^{2} (x + 2y)$

P.4 Show algebraically that

a. $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2$
b. $\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}$
c. $\sum_{j=1}^{n} (x_j - \bar{x}) = 0$

P.5 Let *SALES* denote the monthly sales at a bookstore. Assume *SALES* are normally distributed with a mean of \$50,000 and a standard deviation of \$6000.

a. Compute the probability that the firm has a month with *SALES* greater than \$60,000. Show a sketch.
b. Compute the probability that the firm has a month with *SALES* between \$40,000 and \$55,000. Show a sketch.
c. Find the value of *SALES* that represents the 97th percentile of the distribution. That is, find the value $SALES_{0.97}$ such that $P(SALES > SALES_{0.97}) = 0.03$.
d. The bookstore knows their *PROFITS* are 30% of *SALES* minus fixed costs of \$12,000. Find the probability of having a month in which *PROFITS* were zero or negative. Show a sketch. [*Hint*: What is the distribution of *PROFITS*?]

P.6 A venture capital company feels that the rate of return (X) on a proposed investment is approximately normally distributed with a mean of 40% and a standard deviation of 10%.

a. Find the probability that the return X will exceed 55%.
b. The banking firm who will fund the venture sees the rate of return differently, claiming that venture capitalists are always too optimistic. They perceive that the distribution of returns is $V = 0.8X - 5\%$, where X is the rate of return expected by the venture capital company. If this is correct, find the probability that the return V will exceed 55%.

P.7 At supermarkets sales of "Chicken of the Sea" canned tuna vary from week to week. Marketing researchers have determined that there is a relationship between sales of canned tuna and the price of canned tuna. Specifically, $SALES = 50000 - 100\ PRICE$. $SALES$ is measured as the number of cans per week and $PRICE$ is measured in cents per can. Suppose $PRICE$ over the year can be considered (approximately) a normal random variable with mean $\mu = 248$ cents and standard deviation $\sigma = 10$ cents.

a. Find the expected value of $SALES$.
b. Find the variance of $SALES$.
c. Find the probability that more than 24,000 cans are sold in a week. Draw a sketch illustrating the calculation.
d. Find the $PRICE$ such that $SALES$ is at its 95th percentile value. That is, let $SALES_{0.95}$ be the 95th percentile of $SALES$. Find the value $PRICE_{0.95}$ such that $P\,(SALES > SALES_{0.95}) = 0.05$.

P.8 The Shoulder and Knee Clinic knows that their expected monthly revenue from patients depends on their level of advertising. They hire an econometric consultant who reports that their expected monthly revenue, measured in \$1000 units, is given by the following equation $E(REVENUE|ADVERT) = 100 + 20\,ADVERT$, where $ADVERT$ is advertising expenditure in \$1000 units. The econometric consultant also claims that $REVENUE$ is normally distributed with variance $\text{var}(REVENUE|ADVERT) = 900$.

a. Draw a sketch of the relationship between expected $REVENUE$ and $ADVERT$ as $ADVERT$ varies from 0 to 5.
b. Compute the probability that $REVENUE$ is greater than 110 if $ADVERT = 2$. Draw a sketch to illustrate your calculation.
c. Compute the probability that $REVENUE$ is greater than 110 if $ADVERT = 3$.
d. Find the 2.5 and 97.5 percentiles of the distribution of $REVENUE$ when $ADVERT = 2$. What is the probability that $REVENUE$ will fall in this range if $ADVERT = 2$?
e. Compute the level of $ADVERT$ required to ensure that the probability of $REVENUE$ being larger than 110 is 0.95.

P.9 Consider the U.S. population of registered voters, who may be Democrats, Republicans or independents. When surveyed about the war with ISIS, they were asked if they strongly supported war efforts, strongly opposed the war, or were neutral. Suppose that the proportion of voters in each category is given in Table P.8:

TABLE P.8 **Table for Exercise P.9**

		War Attitude		
		Against	Neutral	In Favor
	Republican	0.05	0.15	0.25
Political Party	Independent	0.05	0.05	0.05
	Democrat	0.35	0.05	0

a. Find the "marginal" probability distributions for war attitudes and political party affiliation.
b. What is the probability that a randomly selected person is a political independent given that they are in favor of the war?
c. Are the attitudes about war with ISIS and political party affiliation statistically independent or not? Why?

d. For the attitudes about the war assign the numerical values $AGAINST = 1$, $NEUTRAL = 2$, and $IN\ FAVOR = 3$. Call this variable WAR. Find the expected value and variance of WAR.

e. The Republican party has determined that monthly fundraising depends on the value of WAR from month to month. In particular the monthly contributions to the party are given by the relation (in millions of dollars) $CONTRIBUTIONS = 10 + 2 \times WAR$. Find the mean and standard deviation of $CONTRIBUTIONS$ using the rules of expectations and variance.

P.10 A firm wants to bid on a contract worth \$80,000. If it spends \$5000 on the proposal it has a 50–50 chance of getting the contract. If it spends \$10,000 on the proposal it has a 60% chance of winning the contract. Let X denote the net revenue from the contract when the \$5000 proposal is used and let Y denote the net revenue from the contract when the \$10,000 proposal is used.

X	$f(x)$
−5,000	0.5
75,000	0.5

y	$f(y)$
−10,000	0.4
70,000	0.6

a. If the firm bases its choice solely on expected value, how much should it spend on the proposal?

b. Compute the variance of X. [*Hint*: Using scientific notation simplifies calculations.]

c. Compute the variance of Y.

d. How might the variance of the net revenue affect which proposal the firm chooses?

P.11 Prior to presidential elections citizens of voting age are surveyed. In the population, two characteristics of voters are their registered party affiliation (republican, democrat, or independent) and for whom they voted in the previous presidential election (republican or democrat). Let us draw a citizen at random, defining these two variables.

$$PARTY = \begin{cases} -1 & \text{registered republican} \\ 0 & \text{independent or unregistered} \\ 1 & \text{registered democrat} \end{cases}$$

$$VOTE = \begin{cases} -1 & \text{voted republican in previous election} \\ 1 & \text{voted democratic in previous election} \end{cases}$$

a. Suppose that the probability of drawing a person who voted republication in the last election is 0.466, and the probability of drawing a person who is registered republican is 0.32, and the probability that a randomly selected person votes republican given that they are a registered republican is 0.97. Compute the joint probability $\text{Prob}[PARTY = -1, VOTE = -1]$. Show your work.

b. Are these random variables statistically independent? Explain.

P.12 Based on years of experience, an economics professor knows that on the first principles of economics exam of the semester 13% of students will receive an A, 22% will receive a B, 35% will receive a C, 20% will receive a D, and the remainder will earn an F. Assume a 4 point grading scale (A = 4, B = 3, C = 2, D = 1, and F = 0). Define the random variable $GRADE = 4, 3, 2, 1, 0$ to be the grade of a randomly chosen student.

a. What is the probability distribution $f(GRADE)$ for this random variable?

b. What is the expected value of $GRADE$? What is the variance of $GRADE$? Show your work.

c. The professor has 300 students in each class. Suppose that the grade of the ith student is $GRADE_i$ and that the probability distribution of grades $f(GRADE_i)$ is the same for all students. Define $CLASS\_AVG = \sum_{i=1}^{300} GRADE_i/300$. Find the expected value and variance of $CLASS\_AVG$.

d. The professor has estimated that the number of economics majors coming from the class is related to the grade on the first exam. He believes the relationship to be $MAJORS = 50 + 10CLASS\_AVG$. Find the expected value and variance of $MAJORS$. Show your work.

P.13 The LSU Tigers baseball team will play the Alabama baseball team in a weekend series of two games. Let $W = 0$, 1, or 2 equal the number of games LSU wins. Let the weekend's weather be designated as Cold or Not Cold. Let $C = 1$ if the weather is cold and $C = 0$ if the weather is not cold. The joint probability function of these two random variables is given in Table P.9, along with space for the marginal distributions.

TABLE P.9 Table for Exercise P.13

	$W = 0$	$W = 1$	$W = 2$	$f(c)$
$C = 1$	(i)	0.12	0.12	(ii)
$C = 0$	0.07	0.14	(iii)	(iv)
$f(w)$	(v)	(vi)	0.61	

a. Fill in the blanks, (i)–(vi).

b. Using the results of (a), find the conditional probability distribution of the number of wins, W, conditional on the weather being warm, $C = 0$. Based on a comparison of the conditional probability distribution $f(w|C = 0)$ and the marginal distribution $f(w)$, can you conclude that the number of games LSU wins W is statistically independent of the weather conditions C, or not? Explain.

c. Find the expected value of the number of LSU wins, W. Also find the conditional expectation $E(W|C = 0)$. Show your work. What kind of weather is more favorable for the LSU Tigers baseball team?

d. The revenue of vendors at the LSU Alex Box baseball stadium depends on the crowds, which in turn depends on the weather. Suppose that food sales $FOOD = \$10{,}000 - 3000C$. Use the rules for expected value and variance to find the expected value and standard deviation of food sales.

P.14 A clinic specializes in shoulder injuries. A patient is randomly selected from the population of all clinic clients. Let S be the number of doctor visits for shoulder problems in the past six months. Assume the values of S are $s = 1, 2, 3$, or 4. Patients at the shoulder clinic are also asked about knee injuries. Let K = the number of doctor visits for knee injuries during the past six months. Assume the values of K are $k = 0, 1$ or 2. The joint probability distribution of the numbers of shoulder and knee injuries is shown in Table P.10. Use the information in the joint probability distribution to answer the following questions. Show **brief** calculations for each

TABLE P.10 Table for Exercise P.14

		Knee = K			
		0	1	2	$f(s)$
	1	0.15	0.09	0.06	
Shoulder = S	2	0.06			
	3	0.02	0.10		0.2
	4	0.02	0.08	0.10	
	$f(k)$			0.33	

a. What is the probability that a randomly chosen patient will have two doctor visits for shoulder problems during the past six months?

b. What is the probability that a randomly chosen patient will have two doctor visits for shoulder problems during the past six months given that they have had one doctor visit for a knee injury in the past six months?

c. What is the probability that a randomly chosen patient will have had three doctor visits for shoulder problems and two doctor visits for knee problems in the past six months?

d. Are the number of doctor visits for knee and shoulder injuries statistically independent? Explain.

e. What is the expected value of the number of doctor visits for shoulder injuries from this population?

f. What is the variance of the number of doctor visits for shoulder injuries from this population?

P.15 As you walk into your econometrics exam, a friend bets you \$20 that she will outscore you on the exam. Let X be a random variable denoting your winnings. X can take the values 20, 0 [if there is a tie], or -20. You know that the probability distribution for $X, f(x)$, depends on whether she studied for the exam or not. Let $Y = 0$ if she studied and $Y = 1$ if she did not study. Consider the following joint distribution Table P.11.

TABLE P.11 **Joint *pdf* for Exercise P.15**

		Y		
		0	1	$f(x)$
	−20	(i)	0	(ii)
X	0	(iii)	0.15	0.25
	20	0.10	(iv)	(v)
	$f(y)$	(vi)	0.60	

a. Fill in the missing elements (i)–(vi) in the table.
b. Compute $E(X)$. Should you take the bet?
c. What is the probability distribution of your winnings if you know that she did not study?
d. Find your expected winnings given that she did not study.
e. Use the Law of Iterated Expectations to find $E(X)$.

P.16 Breast cancer prevalence in the United Kingdom can be summarized for the population (data are in 1000s) as in Table P.12.

TABLE P.12 **Table for Exercise P.16**

	Sex		
	Female	Male	Total
Suffers from Breast Cancer	550	3	553
Not Suffering from Breast Cancer	30,868	30,371	61,239
Total	31,418	30,374	61,792

a. Compute the probability that a randomly drawn person has breast cancer.
b. Compute the probability that a randomly drawn female has breast cancer.
c. Compute the probability that a person is female given that the person has breast cancer.
d. What is the conditional probability function for the prevalence of breast cancer given that the person is female?
e. What is the conditional probability function for the prevalence of breast cancer given that the person is male?

P.17 A continuous random variable Y has *pdf*

$$f(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & otherwise \end{cases}$$

a. Sketch the *pdf*.
b. Find the *cdf*, $F(y) = P(Y \le y)$ and sketch it. [*Hint*: Requires calculus.]
c. Use the *pdf* and a geometric argument to find the probability $P(Y \le 1/2)$.
d. Use the *cdf* from part (b) to compute $P(Y \le 1/2)$.
e. Using the *pdf* and a geometric argument find the probability $P(1/4 \le Y \le 3/4)$.
f. Use the *cdf* from part (b) to compute $P(1/4 \le Y \le 3/4)$.

P.18 Answer each of the following:

a. An internal revenue service auditor knows that 3% of all income tax forms contain errors. Returns are assigned randomly to auditors for review. What is the probability that an auditor will have to view four tax returns until the first error is observed? That is, what is the probability of observing three returns with no errors, and then observing an error in the fourth return?

b. Let Y be the number of independent trials of an experiment before a success is observed. That is, it is the number of failures before the first success. Assume each trial has a probability of success of p and a probability of failure of $1 - p$. Is this a discrete or continuous random variable? What is the set of possible values that Y can take? Can Y take the value zero? Can Y take the value 500?

c. Consider the *pdf* $f(y) = P(Y = y) = p(1 - p)^y$. Using this *pdf* compute the probability in (a). Argue that this probability function generally holds for the experiment described in (b).

d. Using the value $p = 0.5$, plot the *pdf* in (c) for $y = 0, 1, 2, 3, 4$.

e. Show that $\sum_{y=0}^{\infty} f(y) = \sum_{y=0}^{\infty} p(1 - p)^y = 1$. [*Hint*: If $|r| < 1$ then $1 + r + r^2 + r^3 + \cdots = 1/(1-r)$.]

f. Verify for $y = 0, 1, 2, 3, 4$ that the *cdf* $P(Y \leq y) = 1 - (1 - p)^{y+1}$ yields the correct values.

P.19 Let X and Y be random variables with expected values $\mu = \mu_X = \mu_Y$ and variances $\sigma^2 = \sigma_X^2 = \sigma_Y^2$. Let $Z = (2X + Y)/2$.

a. Find the expected value of Z.

b. Find the variance of Z assuming X and Y are statistically independent.

c. Find the variance of Z assuming that the correlation between X and Y is -0.5.

d. Let the correlation between X and Y be -0.5. Find the correlation between aX and bY, where a and b are any nonzero constants.

P.20 Suppose the *pdf* of the continuous random variable X is $f(x) = 1$, for $0 < x < 1$ and $f(x) = 0$ otherwise.

a. Draw a sketch of the *pdf*. Verify that the area under the *pdf* for $0 < x < 1$ is 1.

b. Find the *cdf* of X. [*Hint*: Requires the use of calculus.]

c. Compute the probability that X falls in each of the intervals [0, 0.1], [0.5, 0.6], and [0.79, 0.89]. Indicate the probabilities on the sketch drawn in (a).

d. Find the expected value of X.

e. Show that the variance of X is 1/12.

f. Let Y be a discrete random variable taking the values 1 and 0 with conditional probabilities $P(Y = 1|X = x) = x$ and $P(Y = 0|X = x) = 1 - x$. Use the Law of Iterated Expectations to find $E(Y)$.

g. Use the variance decomposition to find var(Y).

P.21 A fair die is rolled. Let Y be the face value showing, 1, 2, 3, 4, 5, or 6 with each having the probability 1/6 of occurring. Let X be another random variable that is given by

$$X = \begin{cases} Y & \text{if } Y \text{ is even} \\ 0 & \text{if } Y \text{ is odd} \end{cases}$$

a. Find $E(Y)$, $E\left(Y^2\right)$, and var(Y).

b. What is the probability distribution for X? Find $E(X)$, $E\left(X^2\right)$, and var(X).

c. Find the conditional probability distribution of Y given each X.

d. Find the conditional expected value of Y given each value of X, $E(Y|X)$.

e. Find the probability distribution of $Z = XY$. Show that $E(Z) = E(XY) = E\left(X^2\right)$.

f. Find cov(X, Y).

P.22 A large survey of married women asked "How many extramarital affairs did you have last year?" 77% said they had none, 5% said they had one, 2% said two, 3% said three, and the rest said more than three. Assume these women are representative of the entire population.

a. What is the probability that a randomly selected married woman will have had one affair in the past year?

b. What is the probability that a randomly selected married woman will have had more than one affair in the past year?

c. What is the probability that a randomly chosen married woman will have had less than three affairs in the past year?

d. What is the probability that a randomly chosen married woman will have had one or two affairs in the past year?

e. What is the probability that a randomly chosen married woman will have had one or two affairs in the past year, given that they had at least one?

P.23 Let *NKIDS* represent the number of children ever born to a woman. The possible values of *NKIDS* are $nkids = 0, 1, 2, 3, 4, \ldots$. Suppose the *pdf* is $f(nkids) = 2^{nkids} / (7.389 nkids!)$, where ! denotes the factorial operation.

- **a.** Is *NKIDS* a discrete or continuous random variable?
- **b.** Calculate the *pdf* for $nkids = 0, 1, 2, 3, 4$. Sketch it. [Note: It may be convenient to use a spreadsheet or other software to carry out tedious calculations.]
- **c.** Calculate the probabilities $P[NKIDS \le nkids]$ for $nkids = 0, 1, 2, 3, 4$. Sketch the cumulative distribuiton function.
- **d.** What is the probability that a woman will have more than one child.
- **e.** What is the probability that a woman will have two or fewer children?

P.24 Five baseballs are thrown to a batter who attempts to hit the ball 350 feet or more. Let H denote the number of successes, with the *pd* for having h successes being $f(h) = 120 \times 0.4^h \times 0.6^{5-h} / [h!(5-h)!]$, where ! denotes the factorial operation.

- **a.** Is H a discrete or continuous random variable? What values can it take?
- **b.** Calculate the probabilities that the number of successes $h = 0, 1, 2, 3, 4$, and 5. [Note: It may be convenient to use a spreadsheet or other software to carry out tedious calculations.] Sketch the *pdf*.
- **c.** What is the probability of two or fewer successes?
- **d.** Find the expected value of the random variable H. Show your work.
- **e.** The prizes are $1000 for the first success, $2000 for the second success, $3000 for the third success, and so on. What is the *pdf* for the random variable *PRIZE*, which is the total prize winnings?
- **f.** Find the expected value of total prize winnings, *PRIZE*.

P.25 An author knows that a certain number of typographical errors (0, 1, 2, 3, …) are on each book page. Define the random variable T equaling the number of errors per page. Suppose that T has a Poisson distribution [Appendix B.3.3], with *pdf*, $f(t) = \mu^t \exp(-\mu)/t!$, where ! denotes the factorial operation, and $\mu = E(T)$ is the mean number of typographical errors per page.

- **a.** If $\mu = 3$, what is the probability that a page has one error? What is the probability that a page has four errors?
- **b.** An editor independently checks each word of every page and catches 90% of the errors, but misses 10%. Let Y denote the number of errors caught on a page. The values of y must be less than or equal to the actual number t of errors on the page. Suppose that the number of errors caught on a page with t errors has a binomial distribution [Appendix B.3.2].

$$g(y|t, p = 0.9) = \frac{t!}{y!(t-y)!} 0.9^y 0.1^{t-y}, \; y = 0, 1, \ldots, t$$

  Compute the probability that the editor finds one error on a page given that the page actually has four errors.
- **c.** Find the **joint** probability $P[Y = 3, T = 4]$.
- **d.** It can be shown that the probability the editor will find Y errors on a page follows a Poisson distribution with mean $E(Y) = 0.9\mu$. Use this information to find the conditional probability that there are $T = 4$ errors on a page given that $Y = 3$ are found.