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

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

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

$$Y = a_1X_1 + a_2X_2 \sim N\left(\mu_Y = a_1\mu_1 + a_2\mu_2, \sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_{12}\right) \quad \text{(P.30)}$$

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.

P.7 Exercises

Answers to exercises marked * appear on the web page www.wiley.com/college/hill.

P.1* You are organizing an outdoor concert for next week and believe attendance will depend on the weather. You consider the following possibilities are appropriate:

Weather	Probability $= f(x)$	Attendance $= X$
Terrible weather	0.2	500
Mediocre weather	0.6	1000
Great weather	0.2	2000

(a) Let X denote the attendance. Why is X a random variable?
(b) What is the expected attendance?
(c) Suppose that each ticket costs \$5 and that the total cost of giving the concert is a fixed \$2,000. Let Y = profit = total sales revenue – total cost $= 5X - 2000$. What is the expected profit?
(d) If the variance of attendance is $\sigma_X^2 = 240{,}000$, find the variance of profit Y.

P.2 As you walk into your econometrics exam, a friend bets you \$10 that she will outscore you on the exam. Let X be a random variable denoting your winnings. X can take the values 10, 0 if there is a tie, or −10. 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.

		Y		
	$f(x,y)$	0	1	$f(x)$
	−10	0.18	?	?
X	0	0	?	0.3
	10	?	0.45	?
	$f(y)$	?	0.75	

(a) Fill in the missing elements 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.

P.3* A firm's marketing manager believes that total sales X can be modeled using a normal distribution with mean $\mu = \$2.5$ million and standard deviation $\sigma = \$300{,}000$. What is the probability that the firm's sales will exceed \$3 million? Draw a sketch to illustrate your calculation.

P.4 In the U.S. the North and South are quite different. Below is the joint probability distribution of political affiliation (R = Republican, I = Independent and D = Democrat) for a Northern city and a Southern city.

	Political Affiliation (*PA*)		
	R	*I*	*D*
Southern	0.24	0.04	0.12
Northern	0.18	0.12	0.30

(a) What is the probability of selecting a Republican given that we sample from the Northern city? Show your calculation.
(b) Are political affiliation and region of residence statistically independent random variables? Explain.
(c) Assign the values $R = 0, I = 2$ and $D = 5$ to political affiliation (PA). That is, if a citizen is selected at random, the variable PA can take the values 0, 2 and 5. Find the mathematical expectation of the random variable PA.
(d) Find the expected value of $X = 2PA + 2PA^2$, where PA is the random variable political affiliation.

P.5* Before the 2009 Super Bowl there was a coin flip to determine who kicked off and who received. The NFC (National Football Conference) had won 11 prior coin flips.
(a) Given that the NFC had won 11 straight flips, what is the probability that they would win the 12th flip? Explain.
(b) Before the 2010 Super Bowl (won by the New Orleans Saints) the NFC won the coin toss for the 13th consecutive time. What is the probability that the NFC will win the next two consecutive tosses?

P.6 At supermarkets in a Midwestern city the sales of canned tuna varies 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, $SALES = 40710 - 430PRICE$ where $SALES$ are cans sold 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 $\mu = 75$ cents and standard deviation $\sigma = 5$ cents. That is $PRICE \sim N(75, 25)$.
(a) What is the numerical expected value of $SALES$? Show your work.
(b) What is the numerical value of the variance of $SALES$? Show your work.
(c) Find the probability that more than 6,300 cans are sold in a week. Draw a sketch illustrating the calculation.

P.7* "Charley Chicken" and "Bradley Bee" are brands of canned tuna. During a week a certain amount of advertising appears for these products. There may be no advertising, one form of advertising (newspaper coupon), or two forms (coupon and a special store display). Let C denote the level of advertising for Charley Chicken. It can take the values $c = 0, 1$ or 2. Let B denote the level of advertising of Bradley Bee; B can take the values $b = 0$, 1 or 2. Suppose the following table represents the joint probability distribution of the advertising levels for these two brands of canned tuna.

		B		
		0	1	2
	0	0.05	0.05	0.05
C	1	0.05	0.20	0.15
	2	0.05	0.25	0.15

(a) What is the marginal probability distribution of Charley Chicken advertising, C?
(b) What is the expected value of C? Show your work.
(c) What is the variance of C? Show your work.
(d) Are the two companies' advertising strategies statistically independent? Explain.
(e) Bradley Bee pays its advertising firm \$5,000 per week plus \$1,000 for each level of advertising B. What is the probability distribution of Bradley Bee's advertising outlay, A?
(f) What is the correlation between Bradley Bee's advertising level (B) and its advertising expenditure (A)? Explain.

P.8 Let X be a discrete random variable that is the value shown on a single roll of a fair die.
(a) Represent the probability density function $f(x)$ in tabular form.
(b) What is the probability that $X = 4$? That $X = 4$ or $X = 5$?
(c) What is the expected value of X? Explain the meaning of $E(X)$ in this case.
(d) Find the expected value of X^2.
(e) Find the variance of X.
(f) Obtain a die. Roll it 20 times and record the values obtained. What is the average of the first 5 values? The first 10? What is the average of the 20 rolls?

P.9 Let X be a continuous random variable whose probability density function is

$$f(x) = \begin{cases} \frac{2}{3} - \frac{2}{9}x & 0 \leq x \leq 3 \\ 0 & otherwise \end{cases}$$

(a) Sketch the probability density function $f(x)$. Is the area under the curve equal to one?
(b) Geometrically calculate the probability that X falls between 0 and ½.
(c) Geometrically calculate the probability that X falls between ¼ and 3/4.

P.10 Suppose that X and Y are random variables with expected values $\mu_X = \mu_Y = \mu$ and variances $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Let $Z = (X + Y)/2$.
(a) Find $E(Z)$.
(b) Find var(Z) assuming that X and Y are statistically independent.
(c) Find var(Z) assuming that $\text{cov}(X, Y) = 0.5\sigma^2$.

P.11* The length of life (in years) of a personal computer is approximately normally distributed with mean 3.4 years and variance 1.6 years.

(a) What fraction of computers will fail in the first year?
(b) What fraction of computers will last 4 years or more?
(c) What fraction of computers will last at least 2 years?
(d) What fraction of computers will last more than 2.5 years but less than 4 years?
(e) If the manufacturer adopts a warranty policy in which only 5% of the computers have to be replaced, what will be the length of the warranty period?

P.12 Based on many years of experience, an instructor in econometrics has determined that the probability distribution of X, the number of students absent on Mondays, is as follows:

x	0	1	2	3	4	5	6	7
$f(x)$	0.02	0.03	0.26	0.34	0.22	0.08	0.04	0.01

(a) Sketch the probability function of X.
(b) Find the probability that on a given Monday either 2, or 3 or 4 students will be absent.
(c) Find the probability that on a given Monday more than 3 students are absent.
(d) Compute the expected value of the random variable X. Interpret this expected value.
(e) Compute the variance and standard deviation of the random variable X.
(f) Compute the expected value and variance of $Y = 7X + 3$.

P.13* Suppose a certain mutual fund has an annual rate of return that is approximately normally distributed with mean (expected value) 5% and standard deviation 4%. Use Table 1, the table of cumulative probabilities for the standard normal distribution, for parts (a)–(c).

(a) Find the probability that your 1-year return will be negative.
(b) Find the probability that your 1-year return will exceed 15%.
(c) If the mutual fund managers modify the composition of its portfolio, they can raise its mean annual return to 7%, but will also raise the standard deviation of returns to 7%. Answer parts (a) and (b) in light of these decisions. Would you advise the fund managers to make this portfolio change?
(d) Verify your computations in (a)–(c) using your computer software.

P.14 An investor holding a portfolio consisting of two stocks invests 25% of assets in Stock A and 75% into Stock B. The return R_A from Stock A has a mean of 4% and a standard deviation of $\sigma_A = 8\%$. Stock B has an expected return $E(R_B) = 8\%$ with a standard deviation of $\sigma_B = 12\%$. The portfolio return is $P = 0.25R_A + 0.75R_B$.

(a) Compute the expected return on the portfolio.
(b) Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns are perfectly positively correlated.
(c) Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns have a correlation of 0.5.
(d) Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns are uncorrelated.

P.15* Let $x_1 = 7$, $x_2 = 2$, $x_3 = 4$, $x_4 = -7$, $y_1 = 5$, $y_2 = 2$, $y_3 = 3$, $y_4 = 12$. Calculate the following:

(a) $\sum_{i=1}^{2} x_i$

(b) $\bar{x} = \sum_{i=1}^{4} x_i \Big/ 4$ [Note: $\bar{x}$ is called the arithmetic average or arithmetic mean.]

(c) $\sum_{i=1}^{4} (x_i - \bar{x})$

(d) $\sum_{i=1}^{4} (x_i - \bar{x})^2$

(e) $\sum_{i=1}^{4} (x_i - \bar{x})(y_i - \bar{y})$ where $\bar{y} = \sum_{i=1}^{4} y_i \Big/ 4$

(f) $\dfrac{\left(\sum_{i=1}^{4} x_i y_i\right) - 4 \times \bar{x} \times \bar{y}}{\left(\sum_{i=1}^{4} x_i^2\right) - 4 \times \bar{x}^2}$

P.16 Express each of the following sums in summation notation:

(a) $x_1 + x_2 + x_3 + x_4$
(b) $x_2 + x_3$
(c) $x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4$
(d) $x_1 y_3 + x_2 y_4 + x_3 y_5 + x_4 y_6$
(e) $x_3 y_3^2 + x_4 y_4^2$
(f) $(x_1 - y_1) + (x_2 - y_2) + (x_3 - y_3)$

P.17* Write out each of the following sums and compute where possible.

(a) $\sum_{i=1}^{4} (a + b x_i)$

(b) $\sum_{i=1}^{3} i^2$

(c) $\sum_{x=0}^{3} (x^2 + 2x + 2)$

(d) $\sum_{x=2}^{4} f(x + 2)$

(e) $\sum_{x=0}^{2} f(x, y)$

(f) $\sum_{x=2}^{4} \sum_{y=1}^{2} (x + 2y)$

P.18 Let X take 4 values $x_1 = 1,\ x_2 = 3,\ x_3 = 5,\ x_4 = 3.$

(a) Calculate the arithmetic average $\bar{x} = \sum_{i=1}^{4} x_i \Big/ 4$

(b) Calculate $\sum_{i=1}^{4} (x_i - \bar{x})$

(c) Calculate $\sum_{i=1}^{4} (x_i - \bar{x})^2$

(d) Calculate $\left(\sum_{i=1}^{4} x_i^2\right) - 4\bar{x}^2$

(e) Show algebraically that $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2$

P.19 Show that $\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\,\bar{y}$

There are two tricks with summations that now come in handy. First, if c is a constant, then $\sum_{i=1}^{n} c = nc$, so $\sum_{i=1}^{n} 1 = 1 + 1 + \cdots + 1 = n$. Second, $\sum_{i=1}^{k} i = k(k+1)/2$, so

$$\sum_{i=1}^{n}(i-1) = 0 + 1 + \cdots + (n-1) = \sum_{i=1}^{n-1} i = \frac{(n-1)(n)}{2}$$

Using these two expressions, we can simplify the second line of (A.16) as

$$\begin{aligned} S_n &= 2a\frac{(b-a)}{n}n + \frac{2(b-a)^2}{n^2} \cdot \frac{(n-1)(n)}{2} \\ &= 2a(b-a) + (b-a)^2 \cdot \frac{n-1}{n} \end{aligned} \tag{A.17}$$

This sum, S_n, is an approximation to the area under the line $f(x) = 2x$ between the points a and b. The approximation becomes better when more rectangles are used—that is, when n, the number of divisions between a and b, is larger. In fact, the exact area under the graph can be obtained by evaluating the limit of S_n as $n \to \infty$. The only place in (A.17) where n appears is in the last term. The limit of this term is

$$\lim_{n\to\infty} \frac{n-1}{n} = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right) = 1 - \lim_{n\to\infty}\frac{1}{n} = 1 \tag{A.18}$$

Using (A.18) we can take the limit of (A.17) to obtain

$$\begin{aligned} Area &= \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x = 2a(b-a) + (b-a)^2 \\ &= (b-a)(b+a) = b^2 - a^2 \end{aligned}$$

This solution is identical to the result in (A.13) using the geometry of triangles and the fundamental theorem of calculus.

A.5 Exercises

Answers to exercises marked * can be found at www.wiley.com/college/hill.

A.1* Let $Q^s = -3 + 1.5P$, where Q^s is the quantity supplied of a good and P is the market price.
(a) State the interpretation of the slope in economic terms.
(b) Calculate the elasticity at $x = 10$ and at $x = 50$, and state their interpretations.

A.2 Suppose the rate of inflation INF, the annual percentage increase in the general price level, is related to the annual unemployment rate $UNEMP$ by the equation $INF = -2 + 6 \times (1/UNEMP)$.
(a) Sketch the curve for values of $UNEMP$ between 1 and 10.
(b) Where is the impact of a change in the unemployment rate the largest?
(c) If the unemployment rate is 5%, what is the marginal effect of an increase in the unemployment rate on the inflation rate?

A.3* Simplify the following expressions:

(a) $x^{1/2}x^{1/6}$
(b) $x^{2/3} \div x^{7/8}$
(c) $(x^4y^3)^{-1/2}$

A.4 (a) The velocity of light is 186,000 miles per second. Write the velocity of light in scientific notation.
(b) Find the number of seconds in a year and write in scientific notation.
(c) Express the distance light travels in one year in scientific notation.

A.5* Technology affects agricultural production by increasing yield over time. Let $WHEAT_t$ = average wheat production (tonnes per hectare) for the period 1950–2000 ($t = 1, \ldots, 51$) in the Western Australia shire of Chapman Valley.
(a) Suppose production is defined by $WHEAT_t = 0.5 + 0.20\ln(t)$. Plot this curve. Find the slope and elasticity at the point $t = 49$ (1998).
(b) Suppose production is defined by $WHEAT_t = 0.80 + 0.0004\,t^2$. Plot this curve. Find the slope and elasticity at the point $t = 49$ (1998).

A.6 Forensic scientists can deduce the amount of arsenic in drinking water from concentrations (in parts per million) in toenails. Let y = toenail concentration of arsenic and x = drinking water concentration of arsenic. The following three equations describe the relationship:

$$\ln(y) = 0.8 + 0.4\ln(x)$$
$$y = 1.5 + 0.2\ln(x)$$
$$\ln(y) = -1.75 + 20x$$

(a) Plot each of the functions for $x = 0$ to $x = 0.15$
(b) Calculate the slope of each function at $x = 0.10$. State the interpretation of the slope.
(c) Calculate the elasticity of each function at $x = 0.10$ and give its interpretation.

A.7* Consider the numbers $x = 4573239$ and $y = 59757.11$.
(a) Write each number in scientific notation.
(b) Use scientific notation to obtain the product xy.
(c) Use scientific notation to obtain the quotient x/y.
(d) Use scientific notation to obtain the sum $x + y$. [Hint: write each number as a numeric part times 10^6.]

A.8 Consider the function $y = f(x) = 3 + 2x + 3x^2$.
(a) Sketch the curve for values of x between $x = 0$ and $x = 4$.
(b) Find the derivative dy/dx and evaluate it at $x = 2$. Sketch the tangent to the curve at this point.
(c) Compute $y_1 = f(1.99)$ and $y_2 = f(2.01)$. Locate these values (approximately) on your sketch.
(d) Evaluate $m = [f(2.01) - f(1.99)]/.02$. Compare this value to the value of the derivative computed in (b). Explain, geometrically, why the values should be close. The value m is a "numerical derivative," which is useful for approximating derivatives.

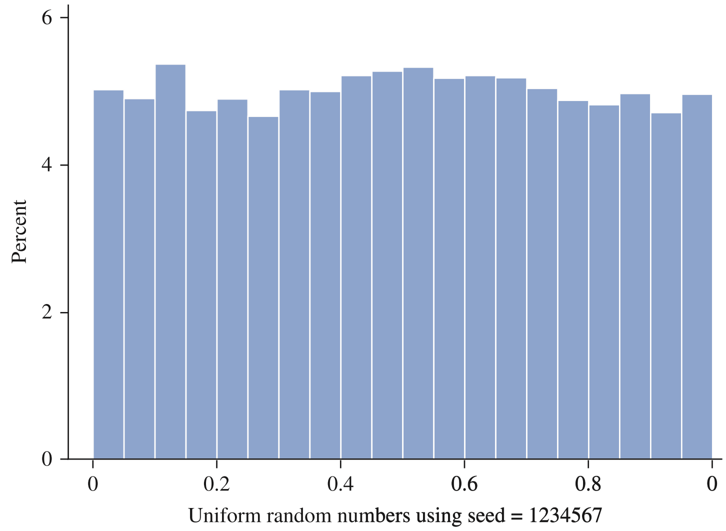

FIGURE **B.15** Histogram of 10,000 generated random values.

B.5 Exercises

Answers to exercise marked * can be found at www.wiley.com/college/hill.

B.1* Let X_1, X_2, …, X_n be independent random variables which all have the same probability distribution, with mean μ and variance σ^2. Let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(a) Use the properties of expected values to show that $E(\overline{X}) = \mu$.
(b) Use the properties of variance to show that $\text{var}(\overline{X}) = \sigma^2/n$. How have you used the assumption of independence?

B.2 Suppose that Y_1, Y_2, Y_3 is a sample of observations from a $N(\mu, \sigma^2)$ population but that $Y_1, Y_2,$ and Y_3 are *not* independent. In fact, suppose that

$$\text{cov}(Y_1, Y_2) = \text{cov}(Y_2, Y_3) = \text{cov}(Y_1, Y_3) = \frac{\sigma^2}{2}$$

Let $\overline{Y} = (Y_1 + Y_2 + Y_3)/3$.
(a) Find $E(\overline{Y})$.
(b) Find $\text{var}(\overline{Y})$.

B.3* Let X be a continuous random variable with probability density function given by

$$f(x) = -\frac{1}{2}x + 1, \quad 0 \le x \le 2$$

(a) Graph the density function $f(x)$.
(b) Find the total area beneath $f(x)$ for $0 \le x \le 2$.
(c) Find $P(X \ge 1)$ using both geometry and integration.
(d) Find $P\left(X \le \frac{1}{2}\right)$.
(e) Find $P\left(X = 1\frac{1}{2}\right)$.

(f) Find the expected value and variance of X.
(g) Find the cumulative distribution function of X.

B.4 Let X be a uniform random variable on the interval (a,b).
(a) Use integration techniques to find the mean and variance of X.
(b) Find the cumulative distribution function of X.

B.5* Use the recursive relationship in (B.52) with $X_0 = 79$, $m = 100$, $a = 263$, and $c = 71$ to generate 40 values $X_1, X_2, \ldots, X_{40}$. Do the resulting numbers appear random? Is this a good random number generator, or not?

B.6 Let X have a normal distribution with mean μ and variance σ^2. Use the change of variable technique to find the probability density function of $Y = aX + b$.

B.7* Show that if $E(Y|X) = E(Y)$, then $\text{cov}(Y, g(X)) = 0$ for any function $g(X)$.

B.8 Normal random numbers are useful for Monte Carlo simulations. One way to generate them is using the Box-Muller transformation. The Box-Muller transformation creates two new random variables, *Z1* and *Z2*, that have independent $N(0,1)$ distributions, using

$$Z1 = \sqrt{-2\ln(U1)}\ \cos(2\pi U2), \quad Z2 = \sqrt{-2\ln(U1)}\ \sin(2\pi U2)$$

(a) Construct a histogram of *Z1* and *Z2* obtained by using the 1,000 uniform random values *U1* and *U2* in *uniform1.dat* (or the 10,000 values in *uniform2.dat*). Is the distribution of values "bell-shaped"?
(b) Calculate the summary statistics for *Z1* and *Z2*. Are the sample mean and variance close to zero and one, respectively?
(c) Construct a scatter diagram for *Z1* and *Z2*. That is, plot *Z1* (vertical axis) and *Z2* (horizontal axis) in the *x-y* plane. Is there any evidence of positive or negative correlation?

B.9* Let X be a continuous random variable with *pdf* $f(x) = 3x^2/8$ for $0 < x < 2$. Compute
(a) $P\left(0 < X < \frac{1}{2}\right)$
(b) $P(1 < X < 2)$

B.10 A continuous random variable X is said to have an exponential distribution if its *pdf* is $f(x) = e^{-x},\ x \geq 0$.
(a) Plot this density function for $0 \leq x \leq 10$.
(b) The cumulative distribution function for X is $F(x) = 1 - e^{-x}$. Plot this function over the interval $0 \leq x \leq 10$. Is it strictly increasing or decreasing, or are you unsure?
(c) Use the inverse transformation method to draw random values *X1* from this distribution. Use the 1,000 values for *U1* in *uniform1.dat* or the 10,000 values for *U1* in *uniform2.dat*. Construct a histogram of the values you have created. Does it resemble the plot in (a)?
(d) The true mean and variance of X are $\mu = 1$ and $\sigma^2 = 1$. How close are the sample mean and the sample variance to the true values?

B.11 Use the recursive relationship in (B.52) with $X_0 = 1234567$, $m = 2^{32}$, $a = 1103515245$, and $c = 12345$ to generate 1,000 random values called *U1*. Do the resulting numbers appear random? Is this a good random number generator, or not? Choose another seed value and generate another 1,000 values called *U2*. Find the summary statistics and sample correlation for *U1* and *U2*. Do the values behave as you expect them to, or not?

B.12* Suppose that the joint *pdf* of the continuous random variables X and Y is $f(x, y) = 6x^2y$ for $0 \leq x \leq 1,\ 0 \leq y \leq 1$.
(a) Does this function satisfy the conditions for a valid *pdf*?
(b) Find the marginal *pdf* of X, as well as its mean and variance.
(c) Find the marginal *pdf* of Y.
(d) Find the conditional *pdf* of X given $Y = \frac{1}{2}$.
(e) Find the conditional mean and variance of X given $Y = \frac{1}{2}$.
(f) Are X and Y independent? Explain.

B.13 Suppose that X and Y are continuous random variables with joint *pdf* $f(x, y) = \frac{1}{2}$ for $0 \leq x \leq y \leq 2$ and $f(x, y) = 0$ otherwise. Note that the values of X are less than or equal to the values of Y.
(a) Verify that the volume under the joint *pdf* is 1.
(b) Find the marginal *pdf*s of X and Y.
(c) Find $P(X < \frac{1}{2})$.
(d) Find the *cdf* of Y.
(e) Find the conditional probability $P\left(X < \frac{1}{2} \middle| Y = 1.5\right)$. Are X and Y independent?
(f) Find the expected value and variance of Y.
(g) Use the law of iterated expectations to find $E(X)$.

and Ullah, A., *Nonparametric Econometrics*, Cambridge University Press, 1999; and Li, Q and Racine, J. S. *Nonparametric Econometrics: Theory and Practice,* Princeton University Press, 2007.

C.11 Exercises

Answers to exercises marked * can be found at www.wiley.com/college/hill.

C.1 Suppose $Y_1, Y_2, \ldots, Y_N$ is a random sample from a population with mean μ and variance σ^2. Rather than using all N observations, consider an easy estimator of μ that uses only the first two observations

$$Y^* = \frac{Y_1 + Y_2}{2}$$

(a) Show that Y^* is a linear estimator.
(b) Show that Y^* is an unbiased estimator.
(c) Find the variance of Y^*.
(d) Explain why the sample mean of all N observations is a better estimator than Y^*.

C.2 Suppose that Y_1, Y_2, Y_3 is a random sample from a $N(\mu, \sigma^2)$ population. To estimate μ, consider the weighted estimator

$$\tilde{Y} = \frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3$$

(a) Show that $\tilde{Y}$ is a linear estimator.
(b) Show that $\tilde{Y}$ is an unbiased estimator.
(c) Find the variance of $\tilde{Y}$ and compare it to the variance of the sample mean $\overline{Y}$.
(d) Is $\tilde{Y}$ as good an estimator as $\overline{Y}$?
(e) If $\sigma^2 = 9$, calculate the probability that each estimator is within one unit on either side of μ.

C.3* The hourly sales of fried chicken at Louisiana Fried Chicken are normally distributed with mean 2,000 pieces and standard deviation 500 pieces. What is the probability that in a nine-hour day more than 20,000 pieces will be sold?

C.4 Starting salaries for economics majors have a mean of $47,000 and a standard deviation of $8,000. What is the probability that a random sample of 40 economics majors will have an average salary of more than $50,000?

C.5* A store manager designs a new accounting system that will be cost-effective if the mean monthly charge account balance is more than $170. A sample of 400 accounts is randomly selected. The sample mean balance is $178 and the sample standard deviation is $65. Can the manager conclude that the new system will be cost effective?
(a) Carry out a hypothesis test to answer this question. Use the $\alpha = 0.05$ level of significance.
(b) Compute the p-value of the test.

C.6 An econometric professor's rule of thumb is that students should expect to spend two hours outside of class on coursework for each hour in class. For a three-hour-per-week class, this means that students are expected to do six hours of work outside class. The professor randomly selects eight students from a class, and asks how many hours they studied econometrics during the past week. The sample values are 1, 3, 4, 4, 6, 6, 8, 12.

(a) Assuming that the population is normally distributed, can the professor conclude at the 0.05 level of significance that the students are studying on average more than six hours per week?

(b) Construct a 90% confidence interval for the population mean number of hours studied per week.

C.7 Modern labor practices attempt to keep labor costs low by hiring and laying off workers to meet demand. Newly hired workers are not as productive as experienced ones. Assume that assembly line workers with experience handle 500 pieces per day. A manager concludes that it is cost-effective to maintain the current practice if new hires, with a week of training, can process more than 450 pieces per day. A random sample of $N = 50$ trainees is observed. Let Y_i denote the number of pieces each handles on a randomly selected day. The sample mean is $\bar{y} = 460$, and the estimated sample standard deviation is $\hat{\sigma} = 38$.

(a) Carry out a test of whether or not there is evidence to support the conjecture that current hiring procedures are effective, at the 5% level of significance. Pay careful attention when formulating the null and alternative hypotheses.

(b) What exactly would a Type I error be in this example? Would it be a costly one to make?

(c) Compute the *p*-value for this test.

C.8* To evaluate alternative retirement benefit packages for its employees, a large corporation must determine the mean age of its workforce. Assume that the age of its employees is normally distributed. Since the corporation has thousands of workers, a sample is to be taken. If the standard deviation of ages is known to be $\sigma = 21$ years, how large should the sample be to ensure that a 95% interval estimate of mean age is no more than four years wide?

C.9 Consider the discrete random variable Y that takes the values $y = 1, 2, 3$, and 4 with probabilities 0.1, 0.2, 0.3, and 0.4, respectively.

(a) Sketch this *pdf*.

(b) Find the expected value of Y.

(c) Find the variance of Y.

(d) If we take a random sample of size $N = 3$ from this distribution, what are the mean and variance of the sample mean, $\overline{Y} = (Y_1 + Y_2 + Y_3)/3$?

C.10 This exercise is a low-tech simulation experiment related to Exercise C.9. It can be a group or class exercise if desired. Have each group member create a set of 10 numbered, identical, slips of paper like the following table.

1	2	2	3	3
3	4	4	4	4

(a) Draw a slip of paper at random and record its value, preferably entering each number into a data file for use with your computer software. Draw a total of 10 times, each time replacing the slip into the pile and stirring them well. Compare the average of these values to the expected value in Exercise C.9(b). Draw 10 more values with replacement. What is the average of all 20 values?

(b) Calculate the sample variance of the 20 values obtained in part (a). Compare this value to the true variance in Exercise C.9(c).

(c) Draw three slips of paper at random, with replacement. Calculate the average of the numbers on these $N = 3$ slips of paper, $\overline{Y} = (Y_1 + Y_2 + Y_3)/3$. Repeat this process at least $NSAM = 20$ times, obtaining *NSAM* average values, $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_{NSAM}$. Calculate the sample average and sample variance of these *NSAM* values. Compare these to the expected value and variance of the sample mean obtained in Exercise C.9(d).

(d) Enter the *NSAM* values $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_{NSAM}$ into a data file. Standardize these values by subtracting the true mean and dividing by the true standard deviation of the mean, from Exercise C9(d). Use your computer software to create a histogram. Discuss the central limit theorem and how it relates to the figure you have created.

(e) Repeat parts (c) and (d) using *NSAM* samples of more than $N = 3$ slips of paper, perhaps five or seven. How do the histograms compare to the one in part (d)?

(f) Discuss the terms ''sampling variation'' and ''sampling distribution'' in the context of the experiments you have performed.

C.11 At the famous Fulton Fish Market in New York City, sales of whiting (a type of fish) vary from day to day. Over a period of several months, daily quantities sold (in pounds) were observed. These data are in the file *fultonfish.dat*.

(a) Using the data for Monday sales, test the null hypothesis that the mean quantity sold is greater than or equal to 10,000 pounds per day against the alternative that the mean quantity sold is less than 10,000 pounds. Use the $\alpha = 0.05$ level of significance. Be sure to (i) state the null and alternative hypotheses, (ii) give the test statistic and its distribution, (iii) indicate the rejection region, including a sketch, (iv) state your conclusion, and (v) calculate the *p*-value for the test. Include a sketch showing the *p*-value.

(b) Assume that daily sales on Tuesday (X_2) and Wednesday (X_3) are normally distributed with means μ_2 and μ_3, and variances σ_2^2 and σ_3^2, respectively. Assume that sales on Tuesday and Wednesday are independent of each other. Test the hypothesis that the variances σ_2^2 and σ_3^2 are equal against the alternative that the variance on Tuesday is larger. Use the $\alpha = 0.05$ level of significance. Be sure to (i) state the null and alternative hypotheses, (ii) give the test statistic and its distribution, (iii) indicate the rejection region, including a sketch, (iv) state your conclusion, and (v) calculate the *p*-value for the test. Include a sketch showing the *p*-value.

(c) We wish to test the hypothesis that mean daily sales on Tuesday and Wednesday are equal against the alternative that they are not equal. Using the result in part (b) as a guide to the appropriate version of the test (Section C.7), carry out this hypothesis test using the 5% level of significance.

(d) Let the daily sales for Monday, Tuesday, Wednesday, Thursday, and Friday be denoted as X_1, X_2, X_3, X_4, and X_5, respectively. Assume that $X_i \sim N(\mu_i, \sigma_i^2)$, and that sales from day to day are independent. Define total weekly sales as $W = X_1 + X_2 + X_3 + X_4 + X_5$. Derive the expected value and variance of *W*. Be sure to show your work and justify your answer.

(e)◆ Referring to part (d), let $E(W) = \mu$. Assume that we estimate μ using

$$\hat{\mu} = \overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \overline{X}_4 + \overline{X}_5$$

where $\overline{X}_i$ is the sample mean for the *i*thday. Derive the probability distribution of $\hat{\mu}$ and construct an approximate (valid in large samples) 95% interval estimate for μ. Justify the validity of your interval estimator.