c. Calculate and interpret a point estimate of the population standard deviation σ. Which estimator did you use? [*Hint*: $\sum x_i^2 = 1860.94$.]

d. Calculate a point estimate of the proportion of all such beams whose flexural strength exceeds 10 MPa. [*Hint*: Think of an observation as a "success" if it exceeds 10.]

e. Calculate a point estimate of the population coefficient of variation σ/μ, and state which estimator you used.

2. A sample of 20 students who had recently taken elementary statistics yielded the following information on the brand of calculator owned (T = Texas Instruments, H = Hewlett Packard, C = Casio, S = Sharp):

T	T	H	T	C	T	T	S	C	H
S	S	T	H	C	T	T	T	H	T

a. Estimate the true proportion of all such students who own a Texas Instruments calculator.

b. Of the 10 students who owned a TI calculator, 4 had graphing calculators. Estimate the proportion of students who do not own a TI graphing calculator.

3. Consider the following sample of observations on coating thickness for low-viscosity paint ("Achieving a Target Value for a Manufacturing Process: A Case Study," *J. of Quality Technology*, 1992: 22–26):

.83	.88	.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

Assume that the distribution of coating thickness is normal (a normal probability plot strongly supports this assumption).

a. Calculate a point estimate of the mean value of coating thickness, and state which estimator you used.

b. Calculate a point estimate of the median of the coating thickness distribution, and state which estimator you used.

c. Calculate a point estimate of the value that separates the largest 10% of all values in the thickness distribution from the remaining 90%, and state which estimator you used. [*Hint*: Express what you are trying to estimate in terms of μ and σ.]

d. Estimate $P(X < 1.5)$, i.e., the proportion of all thickness values less than 1.5. [*Hint*: If you knew the values of μ and σ, you could calculate this probability. These values are not available, but they can be estimated.]

e. What is the estimated standard error of the estimator that you used in part (b)?

4. The article from which the data in Exercise 1 was extracted also gave the accompanying strength observations for cylinders:

6.1	5.8	7.8	7.1	7.2	9.2	6.6	8.3	7.0	8.3
7.8	8.1	7.4	8.5	8.9	9.8	9.7	14.1	12.6	11.2

Prior to obtaining data, denote the beam strengths by $X_1, \ldots, X_m$ and the cylinder strengths by $Y_1, \ldots, Y_n$. Suppose that the X_i's constitute a random sample from a distribution with mean μ_1 and standard deviation σ_1 and that the Y_i's form a random sample (independent of the X_i's) from another distribution with mean μ_2 and standard deviation σ_2.

a. Use rules of expected value to show that $\overline{X} - \overline{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$. Calculate the estimate for the given data.

b. Use rules of variance from Chapter 5 to obtain an expression for the variance and standard deviation (standard error) of the estimator in part (a), and then compute the estimated standard error.

c. Calculate a point estimate of the ratio σ_1/σ_2 of the two standard deviations.

d. Suppose a single beam and a single cylinder are randomly selected. Calculate a point estimate of the variance of the difference $X - Y$ between beam strength and cylinder strength.

5. As an example of a situation in which several different statistics could reasonably be used to calculate a point estimate, consider a population of N invoices. Associated with each invoice is its "book value," the recorded amount of that invoice. Let T denote the total book value, a known amount. Some of these book values are erroneous. An audit will be carried out by randomly selecting n invoices and determining the audited (correct) value for each one. Suppose that the sample gives the following results (in dollars).

	Invoice				
	1	2	3	4	5
Book value	300	720	526	200	127
Audited value	300	520	526	200	157
Error	0	200	0	0	−30

Let

$\overline{Y}$ = sample mean book value

$\overline{X}$ = sample mean audited value

$\overline{D}$ = sample mean error

Propose three different statistics for estimating the total audited (i.e., correct) value—one involving just N and $\overline{X}$, another involving T, N, and $\overline{D}$, and the last involving T and $\overline{X}/\overline{Y}$. If $N = 5000$ and $T = 1{,}761{,}300$, calculate the three corresponding point estimates. (The article "Statistical Models and Analysis in Auditing," *Statistical Science*, 1989: 2–33 discusses properties of these estimators.)

6. Consider the accompanying observations on stream flow (1000s of acre-feet) recorded at a station in Colorado for the period April 1–August 31 over a 31-year span (from an article in the 1974 volume of *Water Resources Research*).

127.96	210.07	203.24	108.91	178.21
285.37	100.85	89.59	185.36	126.94
200.19	66.24	247.11	299.87	109.64
125.86	114.79	109.11	330.33	85.54
117.64	302.74	280.55	145.11	95.36
204.91	311.13	150.58	262.09	477.08
94.33				

An appropriate probability plot supports the use of the lognormal distribution (see Section 4.5) as a reasonable model for stream flow.

a. Estimate the parameters of the distribution. [*Hint*: Remember that X has a lognormal distribution with parameters μ and σ^2 if $\ln(X)$ is normally distributed with mean μ and variance σ^2.]

b. Use the estimates of part (a) to calculate an estimate of the expected value of stream flow. [*Hint*: What is $E(X)$?]

7. a. A random sample of 10 houses in a particular area, each of which is heated with natural gas, is selected and the amount of gas (therms) used during the month of January is determined for each house. The resulting observations are 103, 156, 118, 89, 125, 147, 122, 109, 138, 99. Let μ denote the average gas usage during January by all houses in this area. Compute a point estimate of μ.

b. Suppose there are 10,000 houses in this area that use natural gas for heating. Let τ denote the total amount of gas used by all of these houses during January. Estimate τ using the data of part (a). What estimator did you use in computing your estimate?

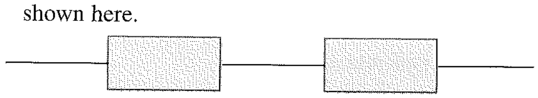
c. Use the data in part (a) to estimate p, the proportion of all houses that used at least 100 therms.

d. Give a point estimate of the population median usage (the middle value in the population of all houses) based on the sample of part (a). What estimator did you use?

8. In a random sample of 80 components of a certain type, 12 are found to be defective.

a. Give a point estimate of the proportion of all such components that are *not* defective.

b. A system is to be constructed by randomly selecting two of these components and connecting them in series, as shown here.



The series connection implies that the system will function if and only if neither component is defective (i.e., both components work properly). Estimate the proportion of all such systems that work properly. [*Hint*: If p denotes the probability that a component works properly, how can P(system works) be expressed in terms of p?]

9. Each of 150 newly manufactured items is examined and the number of scratches per item is recorded (the items are supposed to be free of scratches), yielding the following data:

Number of scratches per item	0	1	2	3	4	5	6	7
Observed frequency	18	37	42	30	13	7	2	1

Let X = the number of scratches on a randomly chosen item, and assume that X has a Poisson distribution with parameter μ.

a. Find an unbiased estimator of μ and compute the estimate for the data. [*Hint*: $E(X) = \mu$ for X Poisson, so $E(\overline{X}) = ?$]

b. What is the standard deviation (standard error) of your estimator? Compute the estimated standard error. [*Hint*: $\sigma_X^2 = \mu$ for X Poisson.]

10. Using a long rod that has length μ, you are going to lay out a square plot in which the length of each side is μ. Thus the area of the plot will be μ^2. However, you do not know the value of μ, so you decide to make n independent measurements $X_1, X_2, \ldots, X_n$ of the length. Assume that each X_i has mean μ (unbiased measurements) and variance σ^2.

a. Show that $\overline{X}^2$ is not an unbiased estimator for μ^2. [*Hint*: For any rv Y, $E(Y^2) = V(Y) + [E(Y)]^2$. Apply this with $Y = \overline{X}$.]

b. For what value of k is the estimator $\overline{X}^2 - kS^2$ unbiased for μ^2? [*Hint*: Compute $E(\overline{X}^2 - kS^2)$.]

11. Of n_1 randomly selected male smokers, X_1 smoked filter cigarettes, whereas of n_2 randomly selected female smokers, X_2 smoked filter cigarettes. Let p_1 and p_2 denote the probabilities that a randomly selected male and female, respectively, smoke filter cigarettes.

a. Show that $(X_1/n_1) - (X_2/n_2)$ is an unbiased estimator for $p_1 - p_2$. [*Hint*: $E(X_i) = n_i p_i$ for $i = 1, 2$.]

b. What is the standard error of the estimator in part (a)?

c. How would you use the observed values x_1 and x_2 to estimate the standard error of your estimator?

d. If $n_1 = n_2 = 200$, $x_1 = 127$, and $x_2 = 176$, use the estimator of part (a) to obtain an estimate of $p_1 - p_2$.

e. Use the result of part (c) and the data of part (d) to estimate the standard error of the estimator.

12. Suppose a certain type of fertilizer has an expected yield per acre of μ_1 with variance σ^2, whereas the expected yield for a second type of fertilizer is μ_2 with the same variance σ^2. Let S_1^2 and S_2^2 denote the sample variances of yields based on sample sizes n_1 and n_2, respectively, of the two fertilizers. Show that the pooled (combined) estimator

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

13. Consider a random sample $X_1, \ldots, X_n$ from the pdf

$$f(x; \theta) = .5(1 + \theta x) \qquad -1 \le x \le 1$$

where $-1 \le \theta \le 1$ (this distribution arises in particle physics). Show that $\hat{\theta} = 3\overline{X}$ is an unbiased estimator of θ. [*Hint*: First determine $\mu = E(X) = E(\overline{X})$.]

14. A sample of n captured Pandemonium jet fighters results in serial numbers $x_1, x_2, x_3, \ldots, x_n$. The CIA knows that the aircraft were numbered consecutively at the factory starting with α and ending with β, so that the total number of planes manufactured is $\beta - \alpha + 1$ (e.g., if $\alpha = 17$ and $\beta = 29$, then $29 - 17 + 1 = 13$ planes having serial numbers 17, 18, 19, . . . , 28, 29 were manufactured). However, the CIA does not know the values of α or β. A CIA statistician suggests using the