CHAPTER

11

Introduction to Inference

When statistics are used to summarize the distribution of a given sample or population, we call this summarizing **descriptive statistics**. In descriptive statistics, we use statistical techniques to describe and to summarize data. **Inferential statistics** is the use of quantitative techniques to generalize from a sample to a population. In short, with inferential statistics, we hope to use a small subset or sample of data to infer what all the data or the population look like.

For example, the city fathers (and mothers) of Pittsburgh, Pennsylvania, would like to know how pleased Pittsburgh citizens are with their new bus service. The best way to do this would be to ask every citizen of Pittsburgh how he or she feels about the transit system. Because the population of Pittsburgh is about 350,000 people, this interviewing could take forever (not to mention that it would cost more to do that than to run the transit system). An alternative is to randomly select a subset of persons (say, 100) and ask them about the mass transit system. From this sample, we will infer what the people of Pittsburgh think.

In this chapter, we begin the presentation of inferential statistics by reviewing some basic definitions and describing some simple inferential techniques.

Some Definitions

A **population** is the total set of items that we are concerned about. In the preceding example, the population is all the people who live in Pittsburgh, Pennsylvania.

A measure that is used to summarize a population is called a **parameter**. For example, the mean education level in Pittsburgh is 12.9 years. This measure is a parameter. Thus far in this book, we have discussed a variety of parameters, including the mean, the median, and the standard deviation of the population.

A **sample** is a subset of a population. In this text, we will assume that all samples are selected randomly. A random sample is a sample in which every member of the population has an equal chance of being included. If a sample is not a random sample, then the rules of statistical inference introduced here do not necessarily hold.

A **statistic** is a measure that is used to summarize a sample. The mean, the standard deviation, and the median of a sample are all statistics.

Table 11.1 Symbols for Parameters and Statistics

Measure	Population Parameter	Sample Statistic
Mean	μ	$\overline{X}$
Standard deviation	σ	S
Number of cases	**N**	N

To create a bit more complexity in the interest of clarity, statisticians use different symbols for the mean and the standard deviation, depending on whether they are parameters or statistics. Table 11.1 illustrates the symbols.

The mean is always calculated the same way whether the data are taken from a sample or a population. The standard deviation, however, is calculated differently. Recall that the formula for calculating the standard deviation is

$$\sigma = \sqrt{\frac{\sum_{i=1}^{\mathbf{N}} (X_i - \mu)^2}{\mathbf{N}}}$$

The formula for the standard deviation of a sample is similar, with one slight twist:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - X)^2}{n - 1}}$$

The difference here is that the sum of the squared deviations from the mean is divided by $n - 1$ rather than by **N**. Later in this chapter, we will explain why this correction is made.

Estimating a Population Mean

The best estimate of the population mean μ is the mean of the sample $\overline{X}$. To illustrate why this is true, we will use the data in Table 11.2, which lists the number of arrests by all 10 Yukon, Oklahoma, police officers in 2011.

The mean number of arrests by Yukon police officers is 15.0. But in circumstances where the population parameter cannot be calculated, either because the population is too large or because the data are not available, the mean of a sample can be used to estimate the population mean.

Assume that we took a random sample of five Yukon police officers and calculated the mean for those five as follows (the sample was selected by using a random number table):

Officers in Sample	Arrests	Mean
1, 3, 2, 8, 4	14, 10, 16, 20, 18	15.6

Table 11.2 Number of Arrests by Police Officers in Yukon, Oklahoma, 2011

Police Officer	Number of Arrests, 2011
1	14
2	16
3	10
4	18
5	8
6	15
7	17
8	20
9	19
10	13

The sample mean is one estimate of the population mean. Note that although our estimate is close (15.6 compared with 15.0), it is not exact. This discrepancy occurs because of sampling error—that is, our sample is not perfectly representative of the population. We would expect, however, that if we took numerous samples of five, the average sample mean (i.e., the average of the sample means) would approach the population mean. Let's try it. The data and calculations are shown in Table 11.3.*

Notice how the average sample mean quickly approaches the population mean and fluctuates around it. Statisticians have worked out this general problem and have logically demonstrated that the average sample mean over the long run will equal the population mean. The best estimate of the population mean, therefore, is the sample mean.

Estimating a Population Standard Deviation

The best estimate of the population standard deviation σ is the sample standard deviation s. Remember that s has $n - 1$ as a denominator rather than **N**. Again, let's use the Yukon police arrests example to illustrate the estimating accuracy of the sample standard deviation. The population standard deviation is 3.7. (If you do not believe us, calculate it from the data in Table 11.2.)

Drawing a sample of five officers, we get the estimate of the standard deviation shown in Table 11.4 (note that we use the same sample in this illustration that we used before).

*An entire field of statistics has developed around the idea of taking repeated samples with replacement. See Mooney and Duval (1993).

Table 11.3 Calculating the Average Sample Mean from Samples of Five

Officers in Sample	Number of Arrests	Sample Mean	Average of Means
1, 3, 2, 8, 4	14, 10, 16, 20, 18	15.6	15.6
1, 6, 8, 3, 4	14, 15, 20, 10, 18	15.4	15.5
7, 4, 1, 5, 9	17, 18, 14, 8, 19	15.2	15.4
2, 10, 7, 4, 6	16, 13, 17, 18, 15	15.8	15.5
7, 10, 3, 6, 5	17, 13, 10, 15, 8	12.6	14.9
10, 7, 2, 1, 4	13, 17, 16, 14, 18	15.6	15.0
10, 3, 7, 4, 2	13, 10, 17, 18, 16	14.8	15.0
10, 3, 8, 9, 4	13, 10, 20, 19, 18	16.0	15.1
6, 4, 8, 9, 10	15, 18, 20, 19, 13	17.0	15.3
2, 8, 4, 1, 5	16, 20, 18, 14, 8	15.2	15.3
8, 3, 9, 10, 5	20, 10, 19, 13, 8	14.0	15.2
2, 3, 7, 5, 1	16, 10, 17, 8, 14	13.0	15.0
10, 8, 5, 6, 4	13, 20, 8, 15, 18	14.8	15.0
9, 3, 6, 2, 7	19, 10, 15, 16, 17	15.4	15.0

Table 11.4 Calculating s for a Sample of Five

Officer	Arrests	Arrests − Mean	Squared
1	14	−1.6	2.56
3	10	−5.6	31.36
2	16	.4	.16
8	20	4.4	19.36
4	18	2.4	5.76
	$s = \sqrt{59.2 \div 4} = 3.85$		59.20 sum of squares

Our estimate of the population standard deviation is 3.85, which is close to the population value of 3.7. Note that had we divided by n rather than by $n - 1$, the standard deviation estimate would have been 3.4. Dividing by n gives us a consistently low estimate of the population standard deviation. For this reason, the estimate *always* is made with a denominator of $n - 1$. If you want to know why this is true, consult an advanced statistics text (William, 1994).

Most statistical programs calculate the sample standard deviation, not the population standard deviation. When you use a statistical package, you need to be aware of this fact.

The Standard Error

If you reported that the mean arrests per police officer in Yukon was 15.6, your supervisor might ask you if the mean was exactly 15.6. Your answer would be that you do not know but that your best estimate of the mean is 15.6. Your superior might then ask how good an estimate 15.6 is. Your superior really wants to know the range of values that the mean must fall within—that is, how much error can the mean estimate contain?

One way to answer this question is to take numerous samples, calculate a mean for each sample, and show the supervisor the range of mean estimates. If you wanted to be more sophisticated, you might calculate a standard deviation for all the mean estimates. The standard deviation for mean estimates has a special name; it is called the **standard error of the mean**.

Fortunately for the sanity of most management analysts, one does not need to take numerous samples, calculate a mean for each sample, and then calculate a standard deviation for the mean estimates to find the standard error of the mean. Statisticians have demonstrated that a good estimate of the standard error of the mean can be made with the following formula:

$$\text{s.e.} = \frac{\sigma}{\sqrt{n}}$$

where s.e. is the standard error of the mean, σ is the standard deviation of the population, and n is the sample size. Because we rarely know the population standard deviation, the estimated standard deviation can be used in computations:

$$\text{s.e.} = \frac{s}{\sqrt{n}}$$

Sometimes the symbol $S_{\overline{Y}}$ is used for the standard error of the mean to indicate that it is the standard deviation of mean estimates.

In the present example, we have a mean estimate of 15.6, a standard deviation estimate of 3.85, and a sample size of 5. Substituting these values into the equation for the standard error, we get

$$\text{s.e.} = \frac{3.85}{\sqrt{5}} = \frac{3.85}{2.236} = 1.7$$

How Sample Size Affects the Standard Error

The Yukon arrest data represent an ideal scenario as far as sampling is concerned. Because there were only 10 officers in the entire population, the chances of obtaining an accurate estimate of the population mean from a sample of 5 officers were very good. In the real world, we rarely have the time or resources to gather samples that are literally half the size of the population in question. For example, polling organizations typically use random samples in the range of 1,200 to 1,500 registered voters to make inferences about how tens of millions of voters are likely to vote in presidential elections. Because the goal of inference is to make a statement about the population based on one sample, what impact can sample size have on our results?

Statisticians have determined that larger samples generally provide better estimates of the population mean than smaller samples (provided the sampling method is not arbitrary or biased). This is true because the size of the standard error for larger samples is typically less than it is for smaller samples.

An example with real data will help illustrate this point. One of the authors has data on grades for a population of 283 public administration students. The data are normally distributed, with a mean of 79.3 and a standard deviation of 11.4. Using these data, we generated six random samples: three samples where $n = 5$ (for each sample) and three samples where $n = 35$ (for each sample). Table 11.5 presents the means, standard deviations, and standard errors.

Note that the means and standard deviations for the three larger samples are much closer to the true population values than the means and standard deviations for the smaller samples. The standard errors for the smaller samples are consistently larger than those for the larger samples. The results illustrate that larger samples typically do a better job at capturing the characteristics present in the population than do smaller samples.

Although there is no guarantee that we will always obtain a smaller standard error when using larger samples, well-known statistical principles called the **central limit theorem** and the **law of large numbers** form the basis for the idea that larger samples are generally more reflective of population characteristics than smaller samples.

If using larger samples is more desirable, how does an analyst determine the appropriate sample size? Statisticians have found that beyond a certain point,

Table 11.5 The Impact of Sample Size on the Standard Error

Three Samples (**N** = 5)	Three Samples (**N** = 35) each)
Sample 1: $\bar{X}$ = 72.4	Sample 1: $\bar{X}$ = 77.9
s = 5.89	s = 10.88
s.e. = 2.64	s.e. = 1.84
Sample 2: $\bar{X}$ = 87.6	Sample 2: $\bar{X}$ = 78.43
s = 9.20	s = 11.55
s.e. = 4.11	s.e. = 1.95
Sample 3: $\bar{X}$ = 78.8	Sample 3: $\bar{X}$ = 75.37
s = 9.34	s = 11.22
s.e. = 4.18	s.e. = 1.89

adding additional cases to a sufficiently large sample will have a limited effect on the quality and accuracy of the inferences we make about a population. How do we know when we have a sufficiently large sample? We will elaborate the formulas and steps used to determine appropriate sample size in Chapters 12 and 13.

The *t* Distribution

Sample estimates of a population mean fit a probability distribution called **Student's *t* distribution** or simply the ***t* distribution**. The *t* distribution is a sampling distribution. In other words, if we repeatedly took samples and calculated means for each sample and plotted them on a graph, the sample means would eventually take on a shape similar to that of the normal distribution.

The *t* distribution allows us to determine the probability of drawing a random sample with a particular mean and standard deviation, given a known or hypothesized population mean. In simpler terms, when we draw a random sample and calculate the mean, the *t* distribution enables us to evaluate which of the following statements is more likely to be correct:

> The sample mean is statistically indistinguishable from the known (or hypothesized) population mean. As a result, there is a high probability that we could draw a sample with this mean from the given population.
>
> The sample mean is statistically different from the known (or hypothesized) population mean. As a result, there is a low probability that we could draw a sample with this mean from the given population.

The interpretation of *t* scores is similar to that of *z* scores, or "standard normal scores" (see Chapter 8). Recall that as *z* scores become larger, the data points associated with the scores fall increasingly away from the mean. As *t* scores become larger, the values for the sample means in question fall increasingly away from the population mean to which they are being compared. In other words, as the distance between the sample mean and the population mean grows, it becomes less likely that the sample mean could have come from a population with that population mean. When we interpret *t* scores, we use this information to make judgments about whether a sample mean is similar to or different from a population mean.

When a sample has 30 cases or more, the normal distribution can be used in place of the *t* distribution. The *t* distribution resembles the normal distribution but has a flatter shape. Unlike the normal distribution, the *t* distribution differs for each sample size. To use the *t* distribution, the public or nonprofit manager needs to know the *degrees of freedom* (df), which corresponds to sample size and identifies the appropriate *t* values. For a sample mean, the degrees of freedom are $n - 1$. In our earlier example, we have four degrees of freedom. The *t* distribution is found in Table 3 in the Appendix of Statistical Tables. Note that rather than printing a large table for every sample size, only key values of the *t* distribution are printed.

To illustrate, if we wish to know between what two numbers 95% of the mean estimates will fall, we calculate the following estimates:

$$\overline{X} \pm \text{s.e.} \times t$$

This formula is simply the mean estimate plus or minus the standard error times the appropriate t score. If we want the t score for a 95% **confidence limit**, we know that there must be 2.5% in each tail of the curve (to total 5%), so we look up the t score for .025 with four degrees of freedom. This figure is 2.78. The upper 95% confidence limit is $15.6 + (1.7 \times 2.78)$, or 20.3. The lower limit is $15.6 - (1.7 \times 2.78)$, or 10.9. We can be 95% sure (95% "confident") that the average number of arrests by Yukon police officers is between 10.9 and 20.3.

An Example

Several years ago, the Wiese school system was criticized because the average Wiese High School (WHS) graduate could read at only the ninth-grade level. After this report was made public, a special reading program was established to upgrade skills. The WHS principal wants to know whether the reading program has improved reading scores. If it has, she will request that an accreditation team visit the school. If the program has not improved the reading level to at least 10.0 (sophomore level), the administrator would like to avoid the embarrassment of a poor review.

To determine the average reading level of the senior class, we select 10 seniors at random. We assume that reading scores are normally distributed. The reading scores for these 10 seniors are shown in Table 11.6. We proceed as follows:

Step 1: Estimate the population mean. Calculate the sample mean reading score in the space provided next to Table 11.6. The answer should be 10.6.

Table 11.6 Reading Scores

Senior	Reading Score
1	13.4
2	12.1
3	11.4
4	10.6
5	10.3
6	10.2
7	9.8
8	9.7
9	9.4
10	8.6

Step 2: Estimate the population standard deviation. In this case, the estimated standard deviation is 1.4. Calculate in the space next to the table.

Step 3: Calculate the standard error of the mean.

$$\text{s.e.} = \frac{s}{\sqrt{n}} = \frac{1.4}{\sqrt{10}} = \frac{1.4}{3.16} = .44$$

Step 4: Provide an **interval estimate** of the mean. An interval estimate is an interval such that the probability that the mean falls within the interval is acceptably high. The most common interval is the 95% confidence interval (we are 95% sure that the mean falls within the interval). The *t* score (with 9 degrees of freedom) for 95% confidence limits is 2.262 (.025 in each tail; see Table 3 in the Appendix). Using this *t* score, we find that the 95% confidence interval is equal to

$$10.6 \pm 2.262 \times .44$$
$$10.6 \pm 1.0$$
$$9.6 \text{ to } 11.6$$

We are 95% sure ("confident") that the mean reading level falls within the range of 9.6 to 11.6 years of school.

This answer bothers the principal because she wants to be certain that the mean is above 10.0. The principal wants us to calculate the probability that the population mean is 10.0 or less. The statistically correct way to ask this question is, If μ is 10.0 or less, what is the probability of drawing a sample with a mean of 10.6? This probability can be determined easily because we have both a mean and a standard error. We convert 10.6 into a *t* score:

$$t = \frac{\bar{X} - \mu}{\text{s.e.}} = \frac{10.6 - 10.00}{.44} = \frac{.6}{.44} = 1.36$$

Looking up a *t* score of 1.36 in Table 3, we find a probability greater than .10. In other words, the probability is greater than .1 that with a mean of less than 10.0 we could get a sample mean estimate of 10.6. To say this another way, the chance of a sample mean of 10.6 if the true (population) mean is only 10.0 is greater than 10%. Computer programs can calculate and report the exact probability for a *t* score. A *t* score of 1.36 with 9 degrees of freedom has a probability of.103. To use the *t* table, one needs to know the degrees of freedom (in this case, 9) and compare the *t* scores with the listed values. A *t* score of 1.9, for example, is larger than 1.833 (the value for .05) and smaller than 2.262 (the value for .025), so the probability of a *t* score of 1.9 with 9 degrees of freedom is less than .05 (but more than .025).

The principal decides that she would like to be more certain. Greater certainty can be achieved if the standard error of the mean can be reduced. The formula for the standard error of the mean,

$$\text{s.e.} = \frac{s}{\sqrt{n}}$$

shows that the standard error can be reduced if the standard deviation can be reduced (which is not likely because this is a function of the population) or if the sample size can be increased. Let's increase the sample size to 100. By a quirk of fate, our sample of 100 has a mean of 10.6 and a standard deviation of 1.4. In the space provided, calculate the new standard error.

If you found the standard error to be .14, congratulations.

In the space provided, calculate the new 95% confidence limits for the Wiese High School senior class reading scores. (*Hint*: With 99 degrees of freedom, you may use the normal curve to approximate the *t* distribution.)

From this new information, what can you tell the principal?

Chapter Summary

Whenever the public or nonprofit manager wants to say something about a population of items or people based on a subset or sample of those items or people, the manager must engage in statistical inference. Inferential statistics use quantitative techniques to generalize from a sample (with summary measures called *statistics*) to a population (with summary measures called *parameters*). We normally use samples for our data and estimations because it is too difficult or too expensive to access or use the entire population. This chapter illustrated how to make and use estimates of the mean and standard deviation of a population based on sample data. The standard error of the mean tells us how much error is contained in the estimate of the mean—that is, how good the estimate is. Because larger samples tend to have smaller standard error values than smaller samples, larger samples generally allow more accurate inferences.

Problems

11.1 George Fastrack, head of the Bureau of Obfuscation's United Way drive, wants to know the average United Way pledge among Obfuscation Bureau employees (pledges are normally distributed). George takes a sample of 10 and gets the results shown in the accompanying table. What is George's best estimate of the mean donation? What is George's best estimate of the standard deviation of the donations? Between what two values can George be 95% sure the mean value lies?

Person	Pledge
1	$ 25
2	0
3	35
4	100
5	0
6	0
7	15
8	50
9	25
10	50

11.2 Captain E. Garth Beaver has been warned by Colonel Sy Verleaf that if the mean efficiency rating for the 150 platoons under Verleaf's command falls below 80, Captain Beaver will be transferred to Minot Air Force Base (a fate worse than Diego Garcias). Beaver wants to know in advance what his fate will be, so he will know whether to send change-of-address cards to his magazine subscriptions. Beaver takes a sample of 20 platoons and finds the following:

$$\overline{X} = 85 \quad s = 13.5$$

Would you advise Beaver to send change-of-address cards? (Assume that efficiency ratings are normally distributed.)

11.3 Last year, sanitation engineer crews in Buffalo, New York, collected 124 tons of trash per day. This year, larger, more efficient trucks were purchased. A sample of 100 truck-days shows that a mean of 130 tons of trash were collected, with a standard deviation of 30 tons. What is the probability that a sample with this mean could be drawn if the new trucks are no improvement (i.e., the population mean $=$ 124)?

11.4 Current Tinderbox Park water pumps can pump 2,000 gallons of water per minute. The park tests 10 new Fastwater brand pumps and finds a mean of 2,200 and a standard deviation of 500. What is the probability that the Fastwater pumps were selected from a population with a mean no better than that of the present pumps?

11.5 If the absenteeism rates for a school district rise above 10%, the state reduces its aid to the school district. Stermerville Independent School District takes a sample of five schools within the district and finds the following absenteeism rates: 5.4%, 8.6%, 4.1%, 8.9%, and 7.8%. What is your best estimate of the absenteeism rate in Stermerville? Is it likely that the absenteeism rate is greater than 10%?

11.6 The Department of Animal Husbandry at State University believes that adding cement to cattle feed will increase the cattle's weight gain. The average weekly gain for State U cattle last year was 12.7 pounds. A sample of 30 cattle are fed cement-fortified feed with the following results:

$$\overline{X} = 14.1 \text{ pounds} \quad s = 5.0$$

What can you tell the department?

11.7 Last year, there were 512 burglaries in Groton, Georgia. The police chief wants to know the average economic loss associated with burglaries in Groton and wants to know it this afternoon. There isn't time to analyze all 512 burglaries, so the department's research analyst selects 10 burglaries at random, which show the following losses:

\$1,550	\$1,874
\$1,675	\$2,595
\$1,324	\$1,835
\$1,487	\$1,910
\$2,246	\$1,612

What is the best estimate of the average loss on a burglary? Place 80% confidence limits around this estimate.

11.8 Last year, the Department of Vocational Rehabilitation was able to place people in jobs with an average salary of \$30,600. This year, placement is handled by a private agency that charges \$200 per placement. Using the following placement figures for this year (sample of 100), what can you tell the department?

$$\overline{X} = \$30{,}600 \quad s = 2{,}000$$

11.9 The General Accounting Office is auditing Branflake International Airways, a company that flies numerous charters for the government. The contract with Branflake specifies that the average flight can be no more than 15 minutes late. A sample of 20 flights reveals the results in the accompanying table. Write a brief memo interpreting this information.

Flight Number	Status
217	On time
167	20 minutes late
133	17 minutes late
207	64 minutes late
219	On time
457	96 minutes late
371	30 minutes late
612	On time
319	6 minutes late
423	12 minutes late
684	11 minutes late

Flight Number	Status
661	61 minutes late
511	On time
536	On time
493	17 minutes late
382	12 minutes late
115	6 minutes late
107	3 minutes late
19	26 minutes late
123	19 minutes late

11.10 The Secretary of Welfare hypothesizes that the average district office has 5% or fewer fraudulent or ineligible recipients. A sample of 10 offices reveals a mean of 4.7% with a standard deviation of 1.2%. What can be said about the secretary's hypothesis?

11.11 The Department of Health and Human Services wants to know the average income of general assistance recipients. A sample of 60 recipients shows a sample mean of $17,400 with a standard deviation of $3,150. (a) Place a 90% confidence limit around your best estimate of the average income of general assistance recipients. (b) What is the probability that the average income could be $18,500 or more? (c) What is the probability that the average income might be as low as $17,000?

11.12 As an analyst for the Overseas Private Investment Corporation, you are required to report to Congress about guaranteed loans to companies doing business in Central American countries. You do not have time to find all the loans, so you take a sample of six loans. The loans have the following values, in millions of dollars:

223 247 187 17 215 275

Use this information to calculate a mean, and put 90% confidence limits around your estimate.

11.13 Complaints about how long it takes the city of Shorewood to pay its bills have reached the city manager. City policy requires that bills be paid within 30 days. A sample of 100 bills shows a mean of 34 days with a standard deviation of 15. Is it possible that a sample mean of 34 days could be generated from a population with a true mean of 30?

11.14 Last year, the Texas State Penitentiary averaged 14.1 violent incidents per day in its prisons. At the end of last year, the federal courts held that inmates could not supervise other inmates. Warden John Law thinks that this ruling will generate more violent incidents because, in the past, inmates used the supervision hierarchy to maintain a pecking order inside the prison. A sample of 40 days of records reveals a mean of 17.5 and a standard deviation of 2.0. What can you tell Warden Law?

11.15 The Metro City Bus system is concerned about the number of people who complain about the service. Metro City Bus managers suspect that many people do not know how to complain. Last year, complaints averaged 47.3 per day. This year, bus systems manager Ralph Kramden has posted a sign in all buses listing a number to call with complaints. Ralph would like to know whether this effort has generated any additional complaints. He takes a sample of 50 days and finds a mean of 54.7 and a standard deviation of 25.4. What can you tell Ralph?

11.16 The director of the Birkfield, California, Women's Shelter is planning the budget for next year and wants to know whether it's reasonable to assume an average stay of 21 days per client. The director asks her assistant to take a random sample of 35 clients who recently stayed at either of the shelter's two locations. The sample shows an average stay of 23 days with a standard deviation of 9.3 days. What should the assistant tell the director?

11.17 Last year, the average weight of babies born to clients of the Houston Women's Health Clinic was 7.5 pounds. The clinic's new director recently implemented an aggressive new educational program on the importance of proper nutrition during pregnancy. A random sample of 50 clients who took part in this program reveals an average birth weight of 8.1 pounds with a standard deviation of 1.1 pounds. What can the director conclude from these data?

11.18 Janice Johnson, director of the Montana Bureau of Prisons, is concerned about the number of violent incidents among prisoners. Last year, the system averaged 1.5 violent incidents per day. In an attempt to lower this number, Director Johnson implemented an aggressive new training protocol to help prison staff better understand the conditions leading to the occurrence of such incidents. The director's assistant takes a random sample of 100 daily incident reports from the state's prisons to get a quick read of the situation. She finds an average incident rate of 1.41 per day with a standard deviation of .12. What should the assistant tell the director? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.19 The Indiana Superintendent of Public Education is concerned about whether the state's schools will meet this year's achievement targets for performance on standardized tests. Specifically, this year's goal is an average pass rate of 81% at the state's 1,500 high schools. Because the state has not yet received all of the data from each school, the superintendent takes a random sample of 150 schools from the available reports to see whether this goal is within reach. The sample mean is 80.6 with a standard deviation of 7.6. What should the superintendent conclude from these data? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.20 The National Council on Nonprofit Monitoring (NCNM) has issued guidelines stating that nonprofit organizations should spend at least 70% of all revenues collected from donations on direct program expenses. As part of an ongoing monitoring program, the Minnesota Lakes and Rivers Appreciation

Foundation samples monthly expenditure reports from its offices throughout the state to see whether the local chapters are below the NCNM minimum. After taking a sample of 100 monthly expenditure reports from local chapters, the foundation's chief data analyst finds a mean of 77.4 with a standard deviation of 7.8. What can the foundation conclude about its conformance with the NCNM standard? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.21 The mayor of Chud, Wisconsin, has faced criticism over the number of water quality advisories issued last year for the city's water supply. Specifically, Chud's average water contaminant score last year was .29 (a rating of "marginally acceptable" according to the Wisconsin Department of Natural Resources). The mayor and city council took a number of aggressive steps to improve the quality of the city's water supply this year. Ideally the mayor's target for this year is an average water contaminant score of no greater than .25 (an "acceptable" rating). The mayor hires an independent consultant to analyze 80 days' worth of water samples for contaminants. What can the consultant tell the mayor after analyzing these scores? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.22 The director of the North Carolina Local Government Insurance Pool (NCLGIP) has implemented mandatory annual risk assessments for all members of the pool in hopes of reducing both the number of claims made by members and the average size of these claims. The average payout per claim last year was $34,000. To see whether the mandatory risk assessments are working, the director takes a random sample of 100 claims filed through the first 4 months of this year. What can the director conclude about the impact of this program? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.23 Two years ago, the Harris County Nonprofit Council (HCNC) did a comprehensive study on the financial position of nonprofits in the Harris County area. The study revealed that budgetary operating reserves in the population of Harris County nonprofit organizations averaged 5.8 months. Officials at HCTNC have updated these data by surveying a random sample of 400 nonprofits in Harris County (150 organizations actually responded to the survey). Based on these data, can the staff at HCTNC conclude that the mean for budgetary operating reserves in the population is still at least 5.8 months? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.24 In 2009, the average household income for micro-enterprise operators in Vermont was $36,752. Officials at the Vermont Foundation for Social Enterprise believe that household incomes among micro-entrepreneurs have gone up over time, but they do not have the time to contact all of the individuals in their micro-enterprise database to get an updated estimate. Accordingly, officials contact a randomly selected sample of 125 micro-entrepreneurs from their listings. Based on the data collected, can officials confidently conclude that the household incomes of the state's micro-entrepreneurs have gone up? *(Note: The data set for this problem is available on the book's Companion Website.)*

11.25 The Metro City Rescue Mission (MCRM) depends heavily on food donations from individuals. Last year, food donations at the MCRM's curbside collection program averaged 9.1 pounds per donation. The director of MCRM would like to know whether making assumptions based on a 9.1 pound average donation is still realistic. To find out, she asks two interns to randomly select and weigh 50 unprocessed donations from the curbside collection program currently in the MCRM's warehouse. Upon analyzing the data, what should the interns tell the director? *(Note: The data set for this problem is available on the book's Companion Website.)*