CHAPTER 4

Inferences on a Single Population

CONTENTS

■ Example 4.1: How Accurately Are Areas Perceived?

The data in Table 4.1 are from an experiment in perceptual psychology. A person asked to judge the relative areas of circles of varying sizes typically judges the areas on a perceptual scale that can be approximated by

$$\text{judged area} = a(\text{true area})^b.$$

For most people the exponent b is between 0.6 and 1. That is, a person with an exponent of 0.8 who sees two circles, one twice the area of the other, would judge the larger one to be only $2^{0.8} = 1.74$ as large. Note that if the exponent is less than 1 a person tends to underestimate the area; if larger than 1, he or she will overestimate the area. The data shown in Table 4.1 are the set of measured exponents for 24 people from one particular experiment (Cleveland *et al.*, 1982). A histogram of this data is given in Figure 4.1.

It may be of interest to estimate the mean value of b for the population from which this sample is drawn; however, because we do not know the value of the population standard deviation we cannot use the methods of

Statistical Methods, DOI: 10.1016/B978-0-12-374970-3.00004-4

Table 4.1 Measured Exponents

0.58	0.63	0.69	0.72	0.74	0.79
0.88	0.88	0.90	0.91	0.93	0.94
0.97	0.97	0.99	0.99	0.99	1.00
1.03	1.04	1.05	1.07	1.18	1.27

Note: Reprinted with permission from the American Statistical Association.

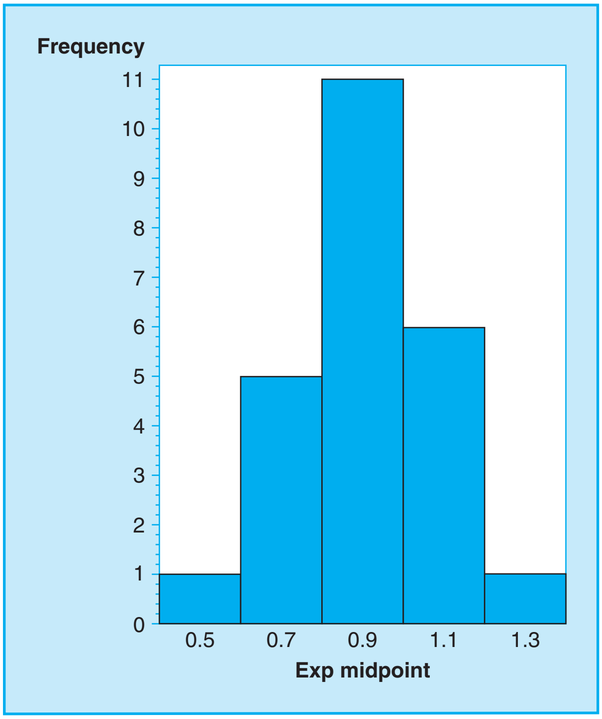

FIGURE 4.1
Histogram of Exponents in Example 4.1.

Chapter 3. Further, we might be interested in estimating the variance of these measurements as well. This chapter discusses methods for doing inferences on means when the population variance is unknown as well as inferences on the unknown population variance. The inferences for this example are presented in Sections 4.2 and 4.4. ■

4.1 INTRODUCTION

The examples used in Chapter 3 to introduce the concepts of statistical inference were not very practical, because they required outside knowledge of the population variance. This was intentional, as we wanted to avoid distractions from issues that were irrelevant to the principles we were introducing. We will now turn to examples that, although still quite simple, will have more useful applications. Specifically, we present procedures for

- making inferences on the mean of a normally distributed population where the variance is unknown,
- making inferences on the variance of a normally distributed population, and
- making inferences on the proportion of successes in a binomial population.

Increasing degrees of complexity are added in subsequent chapters. These begin in Chapter 5 with inferences for comparing two populations and in Chapter 6 with inferences on means from any number of populations. In Chapter 7 we present inference procedures for relationships between two variables through what we will refer to as the linear model, which is subsequently used as the common basis for many other statistical inference procedures. Additional chapters contain brief introductions to other statistical methods that cover different situations as well as methodology that may be used when underlying assumptions cannot be satisfied.

4.2 INFERENCES ON THE POPULATION MEAN

In Chapter 3 we used the sample mean $\bar{y}$ and its sampling distribution to make inferences on the population mean. For these inferences we used the fact that, for any approximately normally distributed population the statistic[1]

$$z = \frac{(\bar{y} - \mu)}{\sigma/\sqrt{n}}$$

has the standard normal distribution. This statistic has limited practical value because, if the population mean is unknown, it is also likely that the variance of the population is unknown.

In the discussion of the t distribution in Section 2.6 we noted that if, in the above equation, the known standard deviation is replaced by its estimate, s, the resulting statistic has a sampling distribution known as Student's t distribution. This distribution has a single parameter, called **degrees of freedom**, which is $(n - 1)$ for this case. Thus for statistical inferences on a mean from a normally distributed population, we can use the statistic

$$t = \frac{(\bar{y} - \mu)}{\sqrt{s^2/n}},$$

where $s^2 = \sum(y - \bar{y})^2/(n - 1)$.

[1]In Section 2.2 we adopted a convention that used capital letters to designate random variables and lowercase letters to represent realizations of those random variables. At that time we stated that the specificity of this designation would not be necessary after Chapter 3. Therefore, for this and subsequent chapters we will use lowercase letters exclusively.

It is very important to note that the degrees of freedom are based on the denominator of the formula used to calculate s^2, which reflects the general formula for computing s^2,

$$s^2 = \frac{\text{sum of squares}}{\text{degrees of freedom}} = \frac{\text{SS}}{\text{df}},$$

a form that will be used extensively in future chapters.

Inferences on μ follow the same pattern outlined in Chapter 3 with only the test statistic changed, that is, z and σ are replaced by t and s.

4.2.1 Hypothesis Test on μ

To test the hypothesis

$$H_0: \mu = \mu_0 \quad \textit{versus} \quad H_1: \mu \neq \mu_0$$

compute the test statistic

$$t = \frac{(\bar{y} - \mu_0)}{\sqrt{s^2/n}} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}.$$

The decision on the rejection of H_0 follows the rules specified in Chapter 3. That is, H_0 is rejected if the calculated value of t is in the rejection region, as defined by a specified α, found in the table of the t distribution, or if the calculated p value is smaller than a specified value of α. Since most tables of the t distribution have only limited numbers of probability levels available, the calculation of p values is usually provided only when the analysis is being performed on computers, which are not limited to using tables.[2]

Power curves for this test can be constructed; however, they require a rather more complex distribution. Charts do exist for determining the power for selected situations and are available in some texts (see, for example, Neter *et al.*, 1996).

■ Example 4.2

In Example 3.3 we presented a quality control problem in which we tested the hypothesis that the mean weight of peanuts being put in jars was the required 8 oz. We assumed that we knew the population standard deviation, possibly from

[2] We noted in Section 2.6 that when the degrees of freedom become large, the t distribution very closely approximates the normal. In such cases, the use of the tables of the normal distribution provides acceptable results even if σ^2 is not known. For this reason many textbooks treat such cases, usually specifying sample sizes in excess of 30, as large sample cases and specify the use of the z statistic for inferences on a mean. Although the results of such methodology are not incorrect, the large sample–small sample dichotomy does not extend to most other statistical methods. In addition, most computer programs correctly use the t distribution regardless of sample size.

experience. We now relax that assumption and estimate both mean and variance from the sample. Table 4.2 lists the data from a sample of 16 jars.

Table 4.2 Data for Peanuts Example (oz.)

8.08	7.71	7.89	7.72
8.00	7.90	7.77	7.81
8.33	7.67	7.79	7.79
7.94	7.84	8.17	7.87

Solution

We follow the five steps of a hypothesis test (Section 3.2).

1. The hypotheses are

$$H_0: \mu = 8,$$
$$H_1: \mu \neq 8.$$

2. Specify $\alpha = 0.05$. The table of the t distribution (Appendix Table A.2) provides the t value for the two-tailed rejection region for 15 degrees of freedom as $|t| > 2.1314$.
3. To obtain the appropriate test statistic, first calculate $\bar{y}$ and s^2:

$$\bar{y} = 126.28/16 = 7.8925,$$
$$s^2 = (997.141 - 996.6649)/15 = 0.03174.$$

The test statistic has the value

$$t = (7.8925 - 8)/\sqrt{(0.03174/16)} = (-0.1075)/0.04453 = -2.4136.$$

4. Since $|t|$ exceeds the critical value of 2.1314, reject the null hypothesis.
5. We will recommend that the machine be adjusted. Note that the chance that this decision is incorrect is at most 0.05, the chosen level of significance.

The actual p value of the test statistic cannot be obtained from Appendix Table A.2. The actual p value, obtained by a computer program, is 0.0290, and we may reject H_0 at any specified α greater than the observed value of 0.0290. ■

■ Example 4.3

One-sided alternative hypotheses frequently occur in regulatory situations. Suppose, for example, that the state environmental protection agency requires a paper mill to aerate its effluent so that the mean dissolved oxygen (DO) level is demonstrably above 6 mg/L. To monitor compliance, the state samples water specimens at 12 randomly selected dates. The data is given in Table 4.3. In view of the critical role of DO, the agency is requiring very strong evidence that mean DO is high. Has the paper mill demonstrated compliance, if α is set at 1%?

Table 4.3 Data for Example 4.3

5.85	6.28	6.50	6.21
5.94	6.12	6.65	6.14
6.34	6.19	6.29	6.40

Solution

Since dissolved oxygen is critical to aquatic life downstream of the plant, the state is placing the burden of proof on the company to show that its effluent has a high mean DO. This implies a one-tailed test.

1. Representing the true mean DO from the plant as μ, the hypotheses are:

$$H_0: \mu = 6 \qquad \textit{versus} \qquad H_1: \mu > 6.$$

2. The variance is estimated from the sample of 12, hence the t statistic has 11 degrees of freedom and we will reject H_0 if the calculated value of t exceeds 2.7181 (Appendix Table A.2).
3. From the sample, $\bar{y} = 6.2425$ and $s^2 = 0.04957$ and the test statistic is

$$t = (6.2425 - 6)\,/\sqrt{.04957/12} = 3.773.$$

4. The null hypothesis is rejected.
5. There is sufficient evidence that the mean (over all time periods) exceeds the state-required minimum.

If this problem was solved using a scientific calculator or computer software, the p value would be provided. Some calculators allow you to specify the one-tailed alternative, and therefore can give the appropriate p value of 0.0015. Many software packages default to the two-tailed alternative. If p_2 is the p value from the two-tailed test, then the one-tailed p value is $p_1 = p_2/2$ if the observed difference is in the direction specified by H_1. ■

■ Example 1.2: Revisited

Recall that in Example 1.2, John Mode had been offered a job in a midsized east Texas town. Obviously, the cost of housing in this city will be an important consideration in a decision to move. The Modes read an article in the paper from the town in which they presently live that claimed the "average" price of homes was $155,000. The Modes want to know whether the data collected in Example 1.2

indicate a difference between the two cities. They assumed that the "average" price referred to in the article was the mean, and the sample they collected from the new city represents a random sample of all home prices in that city.

For this purpose,

$$H_0\colon \mu = 155, \quad \text{and}$$
$$H_1\colon \mu \neq 155.$$

They computed the following results from Table 1.2:

$$\sum y = 9755.18, \qquad \sum y^2 = 1{,}876{,}762, \quad \text{and} \quad n = 69.$$

Thus,

$$\bar{y} = 141.4, \qquad \text{SS} = 497{,}580, \quad \text{and} \quad s^2 = 7317.4,$$

and then

$$t = \frac{141.4 - 155.0}{\sqrt{\frac{7317.4}{69}}} = -1.32,$$

which is insufficient evidence (at $\alpha = 0.05$) that the mean price is different. In other words, the mean price of housing appears not to be different from that of the city in which the Modes currently live. ■

4.2.2 Estimation of μ

Confidence intervals on μ are constructed in the same manner as those in Chapter 3 except that σ is replaced with s, and the table value of z for a specified confidence coefficient $(1 - \alpha)$ is replaced by the corresponding value from the table of the t distribution for the appropriate degrees of freedom. The general formula of the $(1 - \alpha)$ confidence interval on μ is

$$\bar{y} \pm t_{\alpha/2}\sqrt{\frac{s^2}{n}},$$

where $t_{\alpha/2}$ has $(n - 1)$ degrees of freedom.

A 0.95 confidence interval on the mean weight of peanuts in Example 4.2 (Table 4.2) is

$$7.8925 \pm 2.1314\,(0.04453) \text{ or,}$$

$$7.8925 \pm 0.0949,$$

or from 7.798 to 7.987. Remembering the equivalence of hypothesis tests and confidence intervals, we note that this interval does not contain the null hypothesis value of 8 used in Example 4.2, thus agreeing with the results obtained there.

Similarly, the one-sided lower 0.99 confidence interval for the mean DO level in Example 4.3 is

$$6.2425 - 2.7181\sqrt{.04957/12} \text{ or}$$
$$6.2425 - .1747 = 6.0678.$$

With confidence level 99%, the mean DO among all effluent from the mill is at least 6.0678. This is consistent with the results of the hypothesis test.

Solution to Example 4.1

We can now solve the problem in Example 4.1 by providing a confidence interval for the mean exponent. We first calculate the sample statistics: $\bar{y} = 0.9225$ and $s = 0.1652$. The t statistic is based on $24 - 1 = 23$ degrees of freedom, and since we want a 95% confidence interval we use $t_{0.05/2} = 2.069$ (rounded). The 0.95 confidence interval on μ is given by

$$0.9225 \pm (2.069)(0.165)/\sqrt{24} \text{ or}$$
$$0.9225 \pm 0.070, \text{ or from } 0.8527 \text{ to } 0.9923.$$

Thus we are 95% confident that the true mean exponent is between 0.85 and 0.99, rounded to two decimal places. This seems to imply that, on the average, people tend to underestimate the relative areas. ■

4.2.3 Sample Size

Sample size requirements for an estimation problem where σ is not known can be quite complicated. Obviously we cannot estimate a variance before we take the sample; hence the t statistic cannot be used directly to estimate sample size. Iterative methods that will furnish sample sizes for certain situations do exist, but they are beyond the scope of this text. Therefore most sample size calculations simply assume some known variance and proceed as discussed in Section 3.4.

4.2.4 Degrees of Freedom

For the examples in this section the degrees of freedom of the test statistic (the t statistic) have been $(n - 1)$, where n is the size of the sample. It is, however, important to remember that the degrees of freedom of the t statistic are always those used to estimate the variance used in constructing the test statistic. We will see that for many applications this is not $(n - 1)$.

For example, suppose that we need to estimate the average size of stones produced by a gravel crusher. A random sample of 100 stones is to be used. Unfortunately, we do not have time to weigh each stone individually. We can, however, weigh the entire 100 in one weighing, divide the total weight by 100 to obtain an estimate of

CASE STUDY 4.1

Kiefer and Sekaquaptewa (2007) studied the effects of women's degree of gender-math stereotyping and "stereotype threat level" on math proficiency. The authors measured the degree of gender-math stereotyping (a tendency to identify one gender as being better than the other at math) among 138 female undergraduates. The degree of stereotyping was assessed using an Implied Association Test (IAT). IATs attempt to measure the degree of association in concepts by taking the difference in reaction (or processing) times for a *concordant* and *discordant* task. For example, our difference in processing time for a task involving pairs like green/go and red/stop versus a task involving pairs like green/stop and red/go would measure the degree to which we associate these colors and actions. A value of 0 would indicate no association between the concepts. The researchers designed this IAT so that positive values denoted an association of men with math skills.

In the sample, the mean IAT score was 0.28 and the standard deviation was 0.45. A sensible question is whether there is any evidence that, on average, women undergraduates exhibit gender-math stereotyping. We check this, using $\alpha = 0.001$.

$\mu =$ mean IAT score if we could give the test to all female undergraduates at this college

H_0: $\mu = 0$ (on average, no association)

H_1: $\mu \neq 0$

$t = (0.28 - 0)/\sqrt{.45^2/138} = 7.31, df = 137$

Using Appendix Table A.2 with 120 df, we see this is far beyond the critical value that we would use with a two-tailed test and $\alpha = .001$. Hence, we can say the p value for this test is less than 0.001. The researchers conclude that there is significant evidence that women undergraduates do, on average, exhibit gender-math stereotyping.

Somewhat awkwardly, the authors give $t(137) = 6.62$, $p < 0.001$ in the article. The discrepancy seems somewhat too large to attribute to rounding, and these types of inconsistencies are distressingly common in research articles.

Note that the question of whether there is significant evidence of stereotyping is different from the question of whether the effect is large enough to be practically important. In large samples, a small sample mean may still be significantly different from 0. Whether a mean value in the vicinity of 0.28 represents a meaningful or important degree of stereotyping requires the expertise of the researchers. There is also the question of whether the inferences extend beyond the population that was actually sampled, which was female undergraduates at a particular university. The extent to which they are typical of other universities cannot be answered statistically.

μ, and call it $\bar{y}_{100}$. We then take a random subsample of 10 stones from the 100, which we weigh individually to compute an estimate of the variance,

$$s^2 = \frac{\sum(y - \bar{y}_{10})^2}{9},$$

where $\bar{y}_{10}$ is calculated from the subsample of 10 observations. The statistic

$$t = \frac{\bar{y}_{100} - \mu}{\sqrt{s^2/100}},$$

will have the t distribution with 9 (not 99) degrees of freedom.

Although situations such as this do not often arise in practice, it illustrates the fact that the degrees of freedom for the t statistic are associated with the calculation of s^2: it is always the denominator in the expression $s^2 = \text{SS}/\text{df}$. However, the variance of $\bar{y}_{100}$ is still estimated by $s^2/100$ because the variance of the sampling distribution of the mean is based on the sample size used to calculate that mean.

4.3 INFERENCES ON A PROPORTION

In a binomial population, the parameter of interest is p, the proportion of "successes." In Section 2.3 we described the nature of a binomial population and provided in Section 2.5 the normal approximation to the distribution of the proportion of successes in a sample of n from a binomial population. This distribution can be used to make statistical inferences about the parameter p, the proportion of successes in a population.

The estimate of p from a sample of size n is the sample proportion, $\hat{p} = y/n$, where y is the number of successes in the sample. Using the normal approximation, the appropriate statistic to perform inferences on p is

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}.$$

Under the conditions for binomial distributions stated in Section 2.3, this statistic has the standard normal distribution, assuming sufficient sample size for the approximation to be valid.

4.3.1 Hypothesis Test on *p*

The hypotheses are

$$H_0: p = p_0,$$
$$H_1: p \neq p_0.$$

The alternative hypothesis may, of course, be one-sided. To perform the test, compute the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}},$$

which is compared to the appropriate critical values from the normal distribution (Appendix Table A.1), or a p value is calculated from the normal distribution.

Note that we do not use the t distribution here because the variance is not estimated as a sum of squares divided by degrees of freedom. Of course, the use of the normal distribution is an approximation, and it is generally recommended to be used only if $np_0 \geq 5$ and $n(1-p_0) \geq 5$.

■ Example 4.4

An advertisement claims that more than 60% of doctors prefer a particular brand of painkiller. An agency established to monitor truth in advertising conducts a survey consisting of a random sample of 120 doctors. Of the 120 questioned, 82 indicated a preference for the particular brand. Is the advertisement justified?

Solution

The parameter of interest is p, the proportion of doctors in the population who prefer the particular brand. To answer the question, the following hypothesis test is performed:

$$H_0: p = 0.6,$$

$$H_1: p > 0.6.$$

Note that this is a one-tailed test and that rejection of the hypothesis supports the advertising claim. Is it likely that the manufacturer of the painkiller would use a slightly different set of hypotheses? A significance level of 0.05 is chosen. The test statistic is

$$\begin{aligned} z &= \frac{\frac{82}{120} - 0.6}{\sqrt{0.6(1-0.6)/120}} \\ &= \frac{0.083}{0.0447} \\ &= 1.86. \end{aligned}$$

The p value for this statistic (from Appendix Table A.1) is

$$p = P(z > 1.86) = 0.0314.$$

Since this p value is less than the specified 0.05, we reject H_0 and conclude that the proportion is in fact larger than 0.6. That is, the advertisement appears to be justified. ■

4.3.2 Estimation of p

A $(1 - \alpha)$ confidence interval on p based on a sample size of n with y successes is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Note that since there is no hypothesized value of p, the sample proportion $\hat{p}$ is substituted for p in the formula for the variance.

■ Example 4.5

A preelection poll using a random sample of 150 voters indicated that 84 favored candidate Smith, that is, $\hat{p} = 0.56$. We would like to construct a 0.99 confidence interval on the true proportion of voters favoring Smith.

Solution

To calculate the confidence interval, we use

$$0.56 \pm (2.576)\sqrt{\frac{(0.56)(1-0.56)}{150}} \text{ or}$$

$$0.56 \pm 0.104,$$

resulting in an interval from 0.456 to 0.664. Note that the interval does contain 50% (0.5) as well as values below 50%. This means that Smith cannot predict with 0.99 confidence that she will win the election. ■

An Alternate Approximation for the Confidence Interval

In Agresti and Coull (1998), it is pointed out that the method of obtaining a confidence interval on p presented above tends to result in an interval that does not actually provide the level of confidence specified. This is because the binomial is a discrete random variable and the confidence interval is constructed using the normal approximation to the binomial, which is continuous. Simulation studies reported in Agresti and Coull indicate that even with sample sizes as high as 100 and true proportion of 0.018, the actual number of confidence intervals containing the true p are closer to 84% than the nominal 95% specified.

The solution, as proposed in this article, is to add two successes and two failures and then use the standard formula to calculate the confidence interval. This adjustment results in much better performance of the confidence interval, even with relative small samples. Using this adjustment, the interval is based on a new estimate of p; $\tilde{p} = (y+2)/(n+4)$. For Example 4.5 the interval would be based on $\tilde{p} = (86)/154 = 0.558$. The resulting confidence interval would be

$$0.558 \pm (2.576)\sqrt{\frac{(0.558)(0.442)}{154}} \text{ or}$$

$$0.558 \pm 0.103,$$

resulting in an internal from 0.455 to 0.661. This interval is not much different from that constructed without the adjustment, mainly because the sample size is large and the estimate of p is close to 0.5. If the sample size were small, this approximation would result in a more reliable confidence interval.

4.3.3 Sample Size

Since estimation on p uses the standard normal sampling distribution, we are able to obtain the required sample sizes for a given degree of precision. In Section 3.4 we noted that for a $(1-\alpha)$ degree of confidence and a maximum error of estimation E, the required sample size is

$$n = (z_{\alpha/2}\sigma)^2/E^2.$$

This formula is adapted for a binomial population by substituting the quantity $p(1-p)$ for σ^2.

In most cases we may have an estimate (or guess) for p that can be used to calculate the required sample size. If no estimate is available, then 0.5 may be used for p, since this results in the largest possible value for the variance and, hence, also the largest n for a given E (and, of course, α). In other words, the use of 0.5 for the unknown p provides the most conservative estimate of sample size.

■ Example 4.6

In close elections between two candidates (p approximately 0.5), a preelection poll must give rather precise estimates to be useful. We would like to estimate the proportion of voters favoring the candidate with a maximum error of estimation of 1% (with confidence of 0.95). What sample size would be needed?

Solution

To satisfy the criteria specified would require a sample size of

$$n = (1.96)^2(0.5)(0.5)/(0.01)^2 = 9604.$$

This is certainly a rather large sample and is a natural consequence of the high degree of precision and confidence required. ■

4.4 INFERENCES ON THE VARIANCE OF ONE POPULATION

Inferences for the variance follow the same pattern as those for the mean in that the inference procedures use the sampling distribution of the point estimate. The point estimate for σ^2 is

$$s^2 = \sum \frac{(y-\bar{y})^2}{n-1},$$

or more generally SS/df. We also noted in Section 2.6 that the sample quantity

$$\frac{(n-1)s^2}{o^2} = \frac{\sum(y-\bar{y})^2}{\sigma^2} = \frac{\text{SS}}{\sigma^2}$$

has the χ^2 distribution with $(n-1)$ degrees of freedom, assuming a sample from a normally distributed population. As before, the point estimate and its sampling distribution provide the basis for hypothesis tests and confidence intervals.

4.4.1 Hypothesis Test on σ^2

To test the null hypothesis that the variance of a population is a prescribed value, say σ_0^2, the hypotheses are

$$H_0: \sigma^2 = \sigma_0^2,$$
$$H_1: \sigma^2 \neq \sigma_0^2,$$

with one-sided alternatives allowed. The statistic from Section 2.6 used to test the null hypothesis is

$$X^2 = \mathrm{SS}/\sigma_0^2,$$

where for this case $\mathrm{SS} = \sum(y - \bar{y})^2$. If the null hypothesis is true, this statistic has the χ^2 distribution with $(n-1)$ degrees of freedom.

If the null hypothesis is false, then the value of the quantity SS will tend to reflect the true value of σ^2. That is, if σ^2 is larger (smaller) than the null hypothesis value, then SS will tend to be relatively large (small), and the value of the test statistic will therefore tend to be larger (smaller) than those suggested by the χ^2 distribution. Hence the rejection region for the test will be two-tailed; however, the critical values will both be positive and we must find individual critical values for each tail. In other words, the rejection region is

$$\text{reject } H_0 \text{ if}: \left(\mathrm{SS}/\sigma_0^2\right) > \chi^2_{\alpha/2},$$
$$\text{or if}: \left(\mathrm{SS}/\sigma_0^2\right) < \chi^2_{(1-\alpha/2)}.$$

Like the t distribution, χ^2 is another distribution for which only limited tables are available. Thus it is difficult to calculate p values when performing hypothesis tests on the variance when such tables must be used.

Hypothesis tests on variances are often one-tailed because variability is used as a measure of consistency, and we usually want to maintain consistency, which is indicated by small variance. Thus, an alternative hypothesis of a larger variance implies an unstable or inconsistent process.

■ Example 4.2: Revisited

In filling the jar with peanuts, we not only want the average weight of the contents to be 8 oz., but we also want to maintain a degree of consistency in the amount of peanuts being put in jars. If one jar receives too many peanuts, it will overflow, and waste peanuts. If another jar gets too few peanuts, it will not be full and the consumer of that jar will feel cheated even though *on average* the jars have the specified amount of peanuts. Therefore, a test on the variance of weights of peanuts should also be part of the quality control process.

Suppose the weight of peanuts in at least 95% of the jars is required to be within 0.2 oz. of the mean. Assuming an approximately normal distribution we can use the empirical rule to state that the standard deviation should be at most $0.2/2 = 0.10$, or equivalently that the variance be at most 0.01.

Solution

We will use the sample data in Table 4.2 to test the hypothesis

$$H_0\colon \sigma^2 = 0.01 \quad \textit{versus} \quad H_1\colon \sigma^2 > 0.01,$$

using a significance level of $\alpha = 0.05$. If we reject the null hypothesis in favor of a larger variance we declare that the filling process is not in control. The rejection region is based on the statistic

$$X^2 = \mathrm{SS}/0.01,$$

which is compared to the χ^2 distribution with 15 degrees of freedom. From Appendix Table A.3 the rejection region for rejecting H_0 is for the calculated χ^2 value to exceed 25.00. From the sample, SS = 0.4761, and the test statistic has the value

$$X^2 = 0.4761/0.01 = 47.61.$$

Therefore the null hypothesis is rejected and we recommend the expense of modifying the filling process to ensure more consistency. That is, the machine must be adjusted or modified to reduce the variability. Naturally, after the modification, another series of tests would be conducted to ensure success in reducing variation. ■

■ Example 4.1: Revisited

Suppose in the study in perceptual psychology, the variability of subjects was of concern. In particular, suppose that the researchers wanted to know whether the variance of exponents differed from 0.02, corresponding to about 95% of the population lying within 0.28 of either side of the mean.

Solution

The hypotheses of interest would then be

$$H_0\colon \sigma^2 = 0.02,$$
$$H_1\colon \sigma^2 \neq 0.02.$$

Using a level of significance of 0.05, the critical region is

reject H_0 if SS/0.02 is larger than 38.08 (rounded)
or smaller than 11.69 (rounded).

The data in Table 4.1 produce SS $= 0.628$. Hence, the test statistic has a value of $0.628/0.02 = 31.4$, which is not in the critical region; thus, we cannot reject the null hypothesis that $\sigma^2 = 0.02$. The sample variance does not differ significantly from 0.02. ■

4.4.2 Estimation of σ^2

A confidence interval can be constructed for the value of the parameter σ^2 using the χ^2 distribution. Because the distribution is not symmetric, the confidence interval is not symmetric about s^2 and, as in the case of the two-sided hypothesis test, we need two individual values from the χ^2 distribution to calculate the confidence interval.

The lower limit of the confidence interval is

$$L = \text{SS}/\chi^2_{\alpha/2},$$

and the upper limit is

$$U = \text{SS}/\chi^2_{(1-\alpha/2)},$$

where the tail values come from the χ^2 distribution with $(n-1)$ degrees of freedom. Note that the upper tail value from the χ^2 distribution is used for the lower limit and vice versa.

For Example 4.2 we can calculate a 0.95 confidence interval on σ^2 based on the sample data given in Table 4.2. Since the hypothesis test for this example was one-tailed, we construct a corresponding one-sided confidence interval. In this case we would want the lower 95% limit, which would require the upper 0.05 tail of the χ^2 distribution with 15 degrees of freedom, which we have already seen to be 25.00. The lower confidence limit is $\text{SS}/\chi^2_\alpha = 0.4761/25.00 = 0.0190$. The lower 0.95 confidence limit for the standard deviation is simply the square root of the limit for the variance, resulting in the value 0.138. We are therefore 95% confident that the true standard deviation is at least 0.138. This value is larger than that specified by the null hypothesis and again the confidence interval agrees with the result of the hypothesis test.

4.5 ASSUMPTIONS

The mathematical elegance of statistical theory, coupled with the detailed output from statistical software, may give a false sense of security regarding statistical results. If the data are deficient, the results may be less reliable than indicated. How can data be deficient? There are two major sources:

- sloppy data gathering and recording, and
- failure of the distribution of the variable(s) to conform to the assumptions underlying the statistical inference procedure.

Avoiding errors in data gathering and recording is largely a matter of common sense. Double-checking of randomly selected records can find persistent sources of error. Graphical summaries of the data (e.g., box plots and scatterplots), coupled with simple frequency tables of qualitative variables, should be an integral part of an ongoing data quality process.

The failure to conform to assumptions is a subtler problem. In this section we briefly summarize the necessary assumptions, suggest a method for detecting violations, and suggest some remedial methods.

4.5.1 Required Assumptions and Sources of Violations

Two major assumptions are needed to assure correctness for statistical inferences:

- randomness of the sample observations, and
- the distribution of the variable(s) being studied.

We have already noted that randomness is a necessary requirement to define sampling distributions and the consequent use of probabilities associated with these distributions. Another aspect of randomness is that it helps to assure that the observations we obtain have the necessary independence. For example, a failure of the assumption of independence occurs when the sample is selected from the population in some ordered manner. This occurs in some types of economic data obtained on a regular basis at different time periods. These observations then become naturally ordered, and adjacent observations tend to be related, which is a violation of the independence assumption. This does not make the data useless; instead, the user must be aware of the trend and account for it in the analysis (see also Section 11.9).

The distributional assumptions arise from the fact that most of the sampling distributions we use are based on the normal distribution. We know that no "real" data are ever *exactly* normally distributed. However, we also know that the central limit theorem is quite robust so that the normality of the sampling distribution of the mean should not pose major problems except with small sample sizes and/or extremely nonnormal distributions. The χ^2 distribution used for the sampling distribution of the variance and consequently the t distribution are not quite as robust but again, larger sample sizes help.

Outliers or unusual observations are also a major source of nonnormality. If they arise from measurement errors or plain sloppiness, they can often be detected and corrected. However, sometimes they are "real," and no corrections can be made, and they certainly cannot simply be discarded and may therefore pose a problem.

4.5.2 Detection of Violations

The exploratory data analysis techniques presented in Chapter 1 should be used as a matter of routine throughout the data collection process and after the final data set is accumulated. These techniques not only help to reveal extreme recording errors, but

Table 4.4 Exponents from Example 4.1

```
N                                        24
MEAN                                 0.9225
STD  DEV                           0.165247
    50%  MED                          0.955

STEM  LEAF                                #    BOXPLOT
    12    7                               1       |
    10    034578                          6    +-----+
     8    88013477999                    11    *-----*
     6    39249                           5       |
     4    8                               1       |
         ---+---+---+---+
 MULTIPLY STEM.LEAF BY 10** -01
```

can also detect distributional problems. For example, a routine part of an analysis such as that done for Example 4.1 would be to produce a stem and leaf or box plot of the data, as shown in Table 4.4, showing no obvious problem with the normality assumption. This gives us confidence that the conclusions based on the t test and χ^2 test are valid.

The use of a **normal probability plot** allows a slightly more rigorous test of the normality assumption. A special plot, called a Q–Q plot (quantile–quantile), shows the observed value on one axis (usually the horizontal axis) and the value that is expected if the data are a sample from the normal distribution on the other axis. The points should cluster around a straight line for a normally distributed variable. If the data are skewed, the normal probability plot will have a very distinctive shape. Figures 4.2, 4.3, and 4.4 were constructed using the Q–Q graphics function in SPSS. Figure 4.2 shows a typical Q–Q plot for a distribution skewed negatively. Note how the points are all above the line for small values. Figure 4.3 shows a typical Q–Q plot for a distribution skewed positively. In this plot the larger points are all below the line. Figure 4.4 shows the Q–Q plot for the data in Example 4.1. Note that the points are reasonably close to the line, and there are no indications of systematic deviations from the line, thereby indicating that the distribution of the population is reasonably close to normal.

4.5.3 Tests for Normality

There are formal tests for the null hypothesis that a set of values is from a specified distribution, usually the normal. Such tests are known as **goodness-of-fit tests**. One such test is the χ^2 test discussed in Section 12.3. Another popular test is the Kolmogoroff-Smirnoff test, which compares the observed cumulative distribution with the cumulative distribution of the normal, measuring the maximum difference between the two. This is a tedious calculation to try by hand, but most statistical

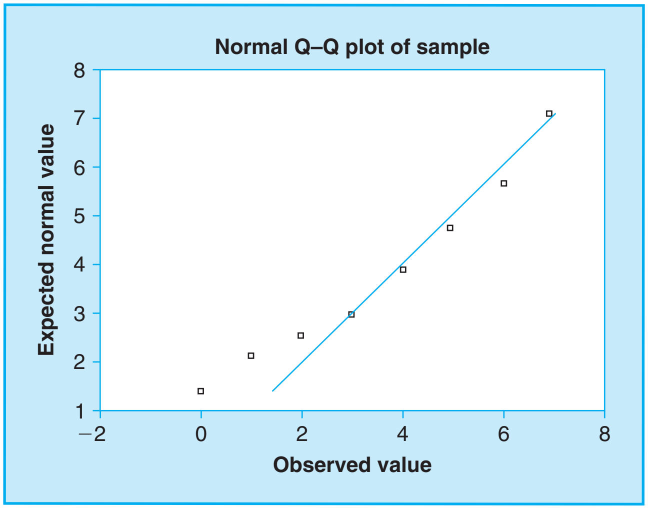

FIGURE 4.2
Normal Probability Plot for a Negatively Skewed Distribution.

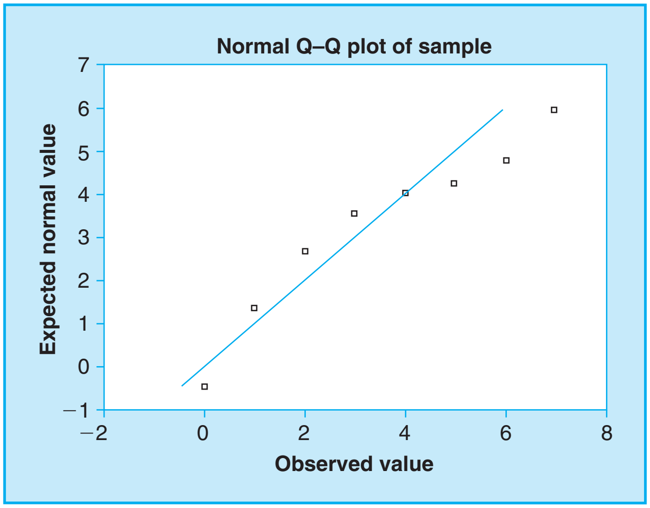

FIGURE 4.3
Normal Probability Plot for a Positively Skewed Distribution.

software contains an implementation of this and other goodness-of-fit tests. For example, using the tree data (Example 1.3), SAS' `Proc Univariate` gives p values for this test as $p > 0.14$ for `HT` and $p < 0.01$ for `HCRN`. Since the null hypothesis is "data is from a normal distribution" and the alternative is "data is not from a normal distribution," we interpret these results as saying that there is no significant evidence that `HT` is nonnormal, but there is strong evidence that `HCRN` is nonnormal. This test confirms what the histograms in Figs. 1.4 and 1.5 showed. Notice that we cannot "prove" that `HT` is normally distributed, we can say only that there is not strong evidence of nonnormality. It makes sense then, whenever the normality assumption is crucial, to run these checks at fairly high significance levels, such as $\alpha = 0.1$, so that

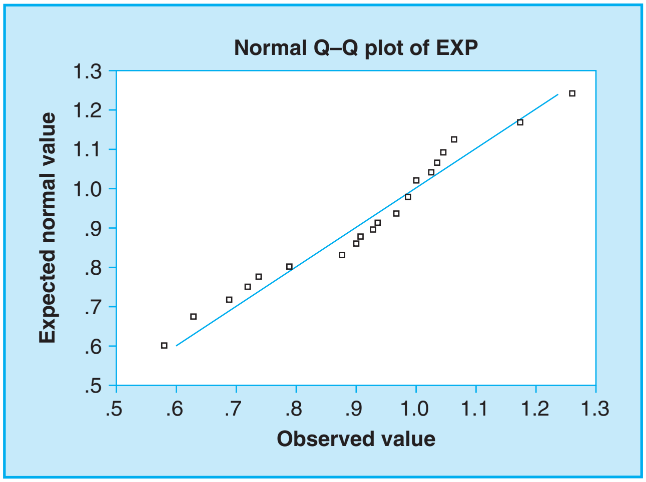

FIGURE 4.4
Normal Probability Plot for Example 4.1.

we raise an alarm if there is even moderate evidence of a violation of the normality assumption.

Goodness-of-fit tests must be treated with caution. In small samples, they have poor power to detect the kinds of violations that can undercut t tests and especially χ^2 tests. In large samples, they are overly sensitive, reporting minor violations that will not greatly disturb the tests. The best strategy is to rely strongly on graphical evidence, and whenever there is any doubt, to analyze data using a variety of statistical techniques. We will discuss one of these alternative techniques here, and more later in Chapter 14.

4.5.4 If Assumptions Fail

Now that we have scared you, we add a few words of comfort. Most statistical methods are reasonably **robust** with respect to the normality assumption. In statistics, we say a method is robust if it is not greatly affected by mild violations of a certain assumption. If the normality assumption appears approximately correct, then most statistical analyses can be used as advertised. If problems arise, all is not lost. The following example shows the effect of an extreme value (a severe violation of normality) on a t test and how an alternate analysis can be substituted.

■ Example 4.7

A supermarket chain is interested in locating a store in a neighborhood suspected of having families with relatively low incomes, a situation that may cause a store in that neighborhood to be unprofitable. The supermarket chain believes that if the average family income is more than \$13,000 the store will be profitable. To determine whether the suspicion is valid, income figures are obtained from a random sample of 20 families in that neighborhood. The data from the sample are given in Table 4.5. Assuming that the conditions for using the t test described in this chapter hold, what can be concluded about the average income in this neighborhood?

Table 4.5 Data on Household Income (Coded in Units of $1000)

No.	Income	No.	Income	No.	Income	No.	Income
1	17.1	6	12.3	11	15.7	16	16.2
2	12.7	7	13.2	12	93.4	17	13.6
3	16.5	8	13.3	13	14.9	18	12.8
4	14.0	9	17.9	14	13.0	19	13.4
5	14.2	10	12.5	15	13.8	20	16.6

Solution

The hypotheses

$$H_0\colon \mu = 13.0,$$
$$H_1\colon \mu > 13.0$$

are to be tested using a 0.05 significance level. The estimated mean and variance are

$$\bar{y} = 18.36,$$
$$s^2 = 314.9,$$

resulting in a t statistic of

$$t = (18.36 - 13.0)/\sqrt{314.9/20}$$
$$= 1.351.$$

We compare this with the 0.05 one-tailed t value of 1.729 and the conclusion is to fail to reject the null hypothesis. It appears that the store will not be built.

The developer involved in the proposed venture decides to take another look at the data and immediately notes an obvious anomaly. The observed income values are all less than $20,000 with one exception: One family reported its income as $93,400. Further investigation reveals that the observation is correct. This income belongs to a family of descendants of the original owner of the land on which the neighborhood is located and who are still living in the old family mansion.

The relevant question here is: What effect does this observation have on the conclusion reached by the hypothesis test? One would think that the large value of this observation would inflate the value of the sample mean and therefore tend to increase the probability of finding an adequate mean income in that area. However, the effect of the extreme value is not only on the mean, but also on the variance, and therefore the result is not quite so predictable. To illustrate, assume that the sampling procedure had picked a more typical family with an income of

16.4. This substitution does lower the sample mean from 18.36 to 14.51. However, it also reduces the variance from 314.86 to 3.05! The value of the test statistic now becomes 3.87, and the null hypothesis would be rejected. ■

4.5.5 Alternate Methodology

In the above example we were able to get a different result by replacing an extreme observation with one that seemed more reasonable. Such a procedure is definitely not recommended, because it could easily lead to abuse (data could be changed until the desired result was obtained). There are, however, more legitimate alternative procedures that can be used if the necessary assumptions appear to be unfulfilled. Such methods may be of two types:

1. The data are "adjusted" so that the assumptions fit.
2. Procedures that do not require as many assumptions are used.

Adjusting the data is accomplished by "transforming" the data. For example, the variable measured in an experiment may not have a normal distribution, but the natural logarithm of that variable may. Transformations take many forms, and are discussed in Section 6.4. More complete discussions are given in some texts (see, for example, Neter *et al.*, 1996).

Procedures of the second type are usually referred to as "nonparametric" or "distribution-free" methods since they do not depend on parameters of specified distributions describing the population. For illustration we apply a simple alternative procedure to the data of Example 4.7 that will illustrate the use of a nonparametric procedure for making the decision on the location of the store.

■ Example 4.7: Revisited

In Chapter 1 we observed that for a highly skewed distribution the median may be a more logical measure of central tendency. Remember that the specification for building the store said "average," a term that may be satisfied by the use of the median.

The median (see Section 1.5) is defined as the "middle" value of a set of population values. Therefore, in the population, half of the observations are above and half of the observations are below the median. In a random sample then, observations should be either higher or lower than the median with equal probability. Defining values above the median as successes, we have a sample from a binomial population with $p = 0.5$. We can then simply count how many of the sample values fall above the hypothesized median value and use the binomial distribution to conduct a hypothesis test.

Solution

The decision to locate a store in the neighborhood discussed in Example 4.7 is then based on testing the hypotheses

$$H_0\text{: the population median} = 13,$$
$$H_1\text{: the population median} > 13.$$

This is equivalent to testing the hypotheses

$$H_0\colon p = 0.5,$$

$$H_1\colon p > 0.5,$$

where p is the proportion of the population values exceeding 13.

This is an application of the use of inferences on a binomial parameter. In the sample shown in Table 4.5 we observe that 15 of the 20 values are strictly larger than 13. Thus $\hat{p}$, the sample proportion having incomes greater than 13, is 0.75. Using the normal approximation to the binomial, the value of the test statistic is

$$z = (0.75 - 0.5)/\sqrt{[(0.5)(0.5)/20]} = 2.23.$$

This value is compared with the 0.05 level of the standard normal distribution (1.645), or results in a p value of 0.012. The result is that the null hypothesis is rejected, leading to the conclusion that the store should be built. ■

■ Example 1.2: Revisited

After reviewing the housing data collected in Example 1.1, the Modes realized that the t test they performed might be affected by the small number of very-high-priced homes that appeared in Table 1.2. In fact, they determined that the median price of the data in Table 1.2 was \$119,000, which is quite a bit less than the sample mean of \$141,400 obtained from the data. Further, a re-reading of the article in the paper found that the "average" price of \$155,000 referred to was actually the median price. A quick check showed that 50 of the 69 (or 72.4%) of the housing prices given in Table 1.2 had values below 155. The test for the null hypothesis that the median is \$155,000 gives

$$z = \frac{0.724 - 0.500}{\sqrt{\frac{(0.5)(0.5)}{69}}} = 3.73,$$

which, when compared with the 0.05 level of the standard normal distribution ($z = 1.960$), provides significant evidence that the median price of homes is lower in their prospective new city than that of their current city of residence.

Despite its simplicity, the test based on the median should not be used if the assumptions necessary for the t test are fulfilled. The median does not use all of the information available in the observed values, since it is based on simply the count of sample observations larger than the hypothesized median. Hence, when the data does come from a normal distribution, the sample mean will lead to a more powerful test.

Other nonparametric methods exist for this particular example. Specifically, the Wilcoxon signed rank test (Chapter 14) may be considered appropriate here, but we defer presentation of all nonparametric methods to Chapter 14. ■

4.6 CHAPTER SUMMARY

This chapter provides the methodology for making inferences on the parameters of a single population. The specific inferences presented are

- inferences on the mean, which are based on the Student's t distribution,
- inferences on a proportion using the normal approximation to the binomial distribution, and
- inferences on the variance using the χ^2 distribution.

A final section discusses some of the assumptions necessary for ensuring the validity of these inference procedures and provides an example for which a violation has occurred and a possible alternative inference procedure for that situation.

4.7 CHAPTER EXERCISES

Concept Questions

Indicate true or false for the following statements. If false, specify what change will make the statement true.

1. ________ The t distribution is more dispersed than the normal.
2. ________ The χ^2 distribution is used for inferences on the mean when the variance is unknown.
3. ________ The mean of the t distribution is affected by the degrees of freedom.
4. ________ The quantity

$$\frac{(\bar{y} - \mu)}{\sqrt{\sigma^2/n}}$$

has the t distribution with $(n - 1)$ degrees of freedom.

5. ________ In the t test for a mean, the level of significance increases if the population standard deviation increases, holding the sample size constant.
6. ________ The χ^2 distribution is used for inferences on the variance.
7. ________ The mean of the t distribution is zero.
8. ________ When the test statistic is t and the number of degrees of freedom is >30, the critical value of t is very close to that of z (the standard normal).

9. ________ The χ^2 distribution is skewed and its mean is always 2.

10. ________ The variance of a binomial proportion is npq [or $np(1-p)$].

11. ________ The sampling distribution of a proportion is approximated by the χ^2 distribution.

12. ________ The t test can be applied with absolutely no assumptions about the distribution of the population.

13. ________ The degrees of freedom for the t test do not necessarily depend on the sample size used in computing the mean.

Practice Exercises

The following exercises are designed to give the reader practice in doing statistical inferences on a single population through simple examples with small data sets. The solutions are given in the back of the text.

1. Find the following upper one-tail values:
   - (a) $t_{0.05}(13)$
   - (b) $t_{0.01}(26)$
   - (c) $t_{0.10}(8)$
   - (d) $\chi^2_{0.01}(20)$
   - (e) $\chi^2_{0.10}(8)$
   - (f) $\chi^2_{0.975}(40)$
   - (g) $\chi^2_{0.99}(9)$

2. The following sample was taken from a normally distributed population:

   3, 4, 5, 5, 6, 6, 6, 7, 7, 9, 10, 11, 12, 12, 13, 13, 13, 14, 15.

   - (a) Compute the 0.95 confidence interval on the population mean μ.
   - (b) Compute the 0.90 confidence interval on the population standard deviation σ.

3. Using the data in Exercise 2, test the following hypotheses:
   - (a) H_0: $\mu = 13$,
     H_1: $\mu \neq 13$.
   - (b) H_0: $\sigma^2 = 10$,
     H_1: $\sigma^2 \neq 10$.

4. A local congressman indicated that he would support the building of a new dam on the Yahoo River if more than 60% of his constituents supported the dam. His legislative aide sampled 225 registered voters in his district and found 135 favored the dam. At the level of significance of 0.10 should the congressman support the building of the dam?

5. In Exercise 4, how many voters should the aide sample if the congressman wanted to estimate the true level of support to within 1%?

Exercises

1. Weight losses of 12 persons in an experimental one-week diet program are given below:

3.0	1.4	0.2	−1.2
5.3	1.7	3.7	5.9
0.2	3.6	3.7	2.0

Do these results indicate that a mean weight loss was achieved? (Use $\alpha = 0.05$).

2. In Exercise 1, determine whether a mean weight loss of more than 1 lb was achieved. (Use $\alpha = 0.01$.)

3. A manufacturer of watches has established that on the average his watches do not gain or lose. He also would like to claim that at least 95% of the watches are accurate to ±0.2 s per week. A random sample of 15 watches provided the following gains (+) or losses (−) in seconds in one week:

+0.17	−0.07	+0.13	−0.05	+0.23
+0.01	+0.06	+0.08	−0.14	−0.10
+0.08	+0.11	+0.05	−0.87	+0.05

Can the claim be made with a 5% chance of being wrong? (Assume that the inaccuracies of these watches are normally distributed.)

4. A sample of 20 insurance claims for automobile accidents (in $1000) gives the following values:

1.6	2.0	2.7	1.3	2.0
1.3	0.3	0.9	1.2	1.2
0.2	1.3	5.0	0.8	7.4
3.0	0.6	1.8	2.5	0.3

Construct a 0.95 confidence interval on the mean value of claims. Comment on the usefulness of this estimate (*Hint:* Look at the distribution.)

5. An advertisement for a headache remedy claims that 90% or more of headache sufferers get relief if they use the remedy. A truth in advertising agency is considering a suit for false advertising and obtains a sample of 100 individuals, which shows that 88 indicate that the remedy gave them relief.
   (a) Using $\alpha = 0.10$ can the suit be justified?
   (b) Comment on the implications of a type I or a type II error in this problem.
   (c) Suppose that the company manufacturing the remedy wants to conduct a promotional campaign that claims over 90% of the remedy users get relief from headaches. What would change in the hypotheses statements used in part (a)?
   (d) What about the implications discussed in part (b)?

6. Average systolic blood pressure of a normal male is supposed to be about 129. Measurements of systolic blood pressure on a sample of 12 adult males from a community whose dietary habits are suspected of causing high blood pressure are listed below:

115	134	131	143
130	154	119	137
155	130	110	138

Do the data justify the suspicions regarding the blood pressure of this community? (Use $\alpha = 0.01$.)

7. A public opinion poll shows that in a sample of 150 voters, 79 preferred candidate X. If X can be confident of winning, she can save campaign funds by reducing TV commercials. Given the results of the survey should X conclude that she has a majority of the votes? (Use $\alpha = 0.05$.)

8. Construct a 0.95 interval on the true proportion of voters preferring candidate X in Exercise 7.

9. It is said that the average weight of healthy 12-hour-old infants is supposed to be 7.5 lb. A sample of newborn babies from a low-income neighborhood yielded the following weights (in pounds) at 12 hours after birth:

6.0	8.2	6.4	4.8
8.6	8.0	6.0	
7.5	8.1	7.2	

At the 0.01 significance level, can we conclude that babies from this neighborhood are underweight?

10. Construct a 0.99 confidence interval on the mean weight of 12-hour-old babies in Exercise 9.

11. A truth in labeling regulation states that no more than 1% of units may vary by more than 2% from the weight stated on the label. The label of a product states that units weigh 10 oz. each. A sample of 20 units yielded the following:

10.01	9.92	9.82	10.04
10.04	10.06	9.97	9.94
9.97	9.86	10.02	10.14
9.97	9.97	9.97	10.05
10.19	10.10	9.95	10.00

At $\alpha = 0.05$ can we conclude that these units satisfy the regulation?

12. Construct a 0.95 confidence interval on the variance of weights given in Exercise 11.

13. A production line in a certain factory puts out washers with an average inside diameter of 0.10 in. A quality control procedure that requires the line to be shut down and adjusted when the standard deviation of inside diameters of washers exceeds 0.002 in. has been established. Discuss the quality control procedure

relative to the value of the significance level, type I and type II errors, sample size, and cost of the adjustment.

14. Suppose that a sample of size 25 from Exercise 13 yielded $s = 0.0037$. Should the machine be adjusted?

15. Using the data from Exercise 4, construct a stem and leaf plot and a box plot (Section 1.6). Do these graphs indicate that the assumptions discussed in Section 4.5 are valid? Discuss possible alternatives.

16. Using the data from Exercise 11, construct a stem and leaf plot and a box plot. Do these graphs indicate that the assumptions discussed in Section 4.5 are valid? Discuss possible alternatives.

17. In Exercise 13 of Chapter 1 the half-lives of aminoglycosides were listed for a sample of 43 patients, 22 of which were given the drug Amikacin. The data for the drug Amikacin are reproduced in Table 4.6. Use these data to determine a 95% confidence interval on the true mean half-life of this drug.

Table 4.6 Half-Life of Amikacin

2.50	1.20	2.60	1.44	1.87	2.48
2.20	1.60	1.00	1.26	2.31	2.80
1.60	2.20	1.50	1.98	1.40	0.69
1.30	2.20	3.15	1.98		

18. Using the data from Exercise 17, construct a 90% confidence interval on the variance of the half-life of Amikacin.

19. A certain soft drink bottler claims that less than 10% of its customers drink another brand of soft drink on a regular basis. A random sample of 100 customers yielded 18 who did in fact drink another brand of soft drink on a regular basis. Do these sample results support the bottler's claim? (Use a level of significance of 0.05.)

20. Draw a power curve for the test constructed in Exercise 19. (Refer to the discussion on power curves in Section 3.2 and plot $1 - \beta$ versus $p =$ proportion of customers drinking another brand.)

21. This experiment concerns the precision of four types of collecting tubes used for air sampling of hydrofluoric acid. Each type is tested three times at five different concentrations. The data shown in Table 4.7 give the differences between the three observed and true concentrations for each level of true concentration for each of the tubes.

    The differences are required to have a standard deviation of no more than 0.1. Do any of the tubes meet this criterion? (*Careful:* What is the most appropriate sum of squares for this test?)

Table 4.7 Data for Exercise 21

Type	Concentration	Differences		
1	1	−0.112	0.163	−0.151
1	2	−0.117	0.072	0.169
1	3	−0.006	−0.092	−0.268
1	4	0.119	0.118	0.051
1	5	−0.272	−0.302	0.343
2	1	−0.094	−0.137	0.308
2	2	−0.238	0.031	0.160
2	3	−0.385	−0.366	−0.173
2	4	−0.259	0.266	−0.303
2	5	−0.125	0.383	0.334
3	1	0.060	0.106	0.084
3	2	−0.016	−0.191	0.097
3	3	−0.024	−0.046	−0.178
3	4	0.040	0.028	0.619
3	5	0.062	0.293	−0.106
4	1	−0.034	0.116	0.055
4	2	−0.023	−0.099	−0.212
4	3	−0.256	−0.110	−0.272
4	4	−0.046	0.009	−0.134
4	5	−0.050	0.009	−0.034

22. The following data gives the average pH in rain/sleet/snow for the two-year period 2004–2005 at 20 rural sites on the U.S. West Coast. (Source: National Atmospheric Deposition Program)

(a) Box plot this data and identify any anomalous observations.

(b) Would the sample mean or the sample median be a better descriptor of typical pH values?

(c) Use the alternative method described in Section 4.7 to test the null hypothesis that the median pH is at least 5.40.

5.335	5.345	5.395	5.305	5.315
5.380	5.520	5.190	5.455	5.330
5.360	6.285	5.350	5.125	5.115
5.510	5.340	5.340	5.305	5.265

23. Warren and McFadyen (2010) interviewed 44 residents of Kintyre, Scotland. This rural area in southwest Scotland is home to a growing number of wind farms. Twenty-two of the interviewees rated the visual impact of the wind farms as Positive or Very Positive. Assuming this was a random sample, give a 90% confidence interval for the proportion of all Kintyre residents who would give one of these ratings.

24. Federal workplace safety standards for noise levels state that personal protective equipment is required if the time-weighted sound level at a work site exceeds 90 dBA. Suppose that this is interpreted as saying that the mean sound level at a site should not significantly exceed 90 dBA. At one location on its fabrication floor, a manufacturer records sound levels over 10 randomly selected time intervals. Should the company begin requiring ear protection, if:

(a) $\bar{x} = 81.2, \quad s = 10.4, \ \alpha = 0.1$?

(b) $\bar{x} = 97.2, \quad s = 10.4, \ \alpha = 0.1$?

(c) Why would $\alpha = 0.1$ be more reasonable than $\alpha = 0.01$ in this situation?

25. In Exercise 24, suppose that we interpret the standard as meaning that there should be only a small probability (no more than 10%) that the time-weighted sound level at a site will exceed 90 dBA. In 70 independent measurements of the sound level, you find 10 instances where the noise exceeds 90 dBA. Using $\alpha = 0.1$, should the company begin requiring ear protection?

26. The methods for proportions discussed in Section 4.3 assume that $np_0 \geq 5$ and $n(1 - p_0) \geq 5$. When sample sizes are smaller, then methods based directly on the binomial distribution can be used. Suppose a vendor claims that at most 5% of parts are defective. In a random sample of 20 parts shipped by this vendor, you find four that are defective.

(a) State the null and alternative hypotheses.

(b) Show that this sample size is too small for the z test for proportions.

(c) Calculate the p value for this test, using the binomial distribution with $p = 0.05$.

(d) What do you conclude, if you are using $\alpha = 0.05$?

27. In Example 4.3, the paper mill had to demonstrate that its effluent had mean DO greater than 6 mg/L. But to ensure against occasional very low DO, keeping the mean high is not enough, the mill must also keep the standard deviation low.

(a) Using the data in Table 4.3, is there evidence that the standard deviation is less than 0.5 mg/L? Use a significance level of 1%.

(b) The test used in (a) requires the data come from a normal distribution. Use a box plot and a normal probability plot to check this assumption.

28. McCluskey *et al.* (2008) conducted a survey of citizens' attitudes toward police in San Antonio, Texas. Before proceeding with their main analysis, they first compare the ethnic distribution in their sample to that of San Antonio as a whole. According to the 2000 Census, 59% of San Antonio residents are Hispanic. In their sample of 454, 36% were Hispanic. Does the proportion of Hispanics in the sample differ from the Census figure by more than can be attributed to chance? Use $\alpha = 0.01$. Note: The authors believe this difference is due to a tendency of poorer residents not to have telephones, and to have a greater tendency to refuse to answer surveys.

29. Each year, the Florida Department of Education grades K–12 schools on a scale of A to F. In spring 2009, 71 schools in Duval changed their grade compared to

the previous year, 46 improving, and 25 declining. In the nearby county of Putnam, which is much more rural, 9 schools changed their grade, 6 improving and 3 declining.

(a) Can Duval state that their observed proportion of improving schools differs significantly from what would be expected if improving and declining were equally likely events? (Use $\alpha = 0.05$).

(b) Can Putnam make the same statement? *Hint:* This data set is not large enough for the z approximation. You can calculate a p value using the binomial distribution.

Project

1. **Lake Data Set.** Total chlorophyll `(SMRCHLO)` levels for summer months are reported in the Florida Lakewatch data set, as described in Appendix C.1.

(a) Using box plots, show that the logarithm (natural or common) of this variable is more nearly normally distributed than the original variable.

(b) Assuming this data is randomly selected from among all lakes in North Central Florida, give a 95% confidence interval for the mean `ln(SMRCHLO)` in this population.

(c) Give a 95% confidence interval for the variance in `ln(SMRCHLO)` in this population.

This page intentionally left blank