AN INTRODUCTION TO STATISTICS

WITH

DATA ANALYSIS

by

SHELLEY RASMUSSEN

Department of Mathematical Sciences
Olney 428T
University of Massachusetts/Lowell
Lowell, MA 01854

Originally published by Brooks/Cole Publishing Company,
Division of Wadsworth, Inc.

ISBN 0-534-13578-1

Shelley_Rasmussen@uml.edu

Minitab is a statistical package, a computer program that performs many statistical procedures. The versions of Minitab now available for use on personal computers are menu-driven and much easier to use than the main-frame version originally discussed in this text. Those sections are not included in this online edition of the text. At this time, the most recent version is Minitab 15, available at very reasonable prices for purchase or rental from:

www.e-academy.com/minitab

System Requirements

Processor:	PC with a 1 GHz 32- or 64-bit processor
Memory:	512 MB or more of available RAM
Disk Space:	125 MB free space available
Operating System:	Microsoft Windows 2000, XP, or Vista.
Display:	A display capable of 1024 X 768 or higher resolution
Software:	Adobe Acrobat Reader 5.0 or higher for Meet Minitab

CHAPTER 10

Inferences About a Measure of Central Tendency

IN THIS CHAPTER

Large-sample inferences about a mean
Large-sample inferences about a proportion
t test and confidence intervals based on a *t* distribution
Wilcoxon signed rank test and associated confidence intervals
Sign test and associated confidence intervals

Is the average birth weight of babies born to exercised goats different from the average birth weight of babies born to unexercised goats? Do bomb bases meet height specifications, on average? Is the median lifetime of a new, cheaper type of electric cord equal to the median lifetime of the older, more expensive type? These questions concern *measures of central tendency,* also called *location parameters.*

> A **measure of central tendency,** or **location parameter,** describes the location or center of a distribution or set of numbers.

We would like to use information in a sample to make inferences about central tendency in a population. For instance, we exercise a sample of pregnant goats. When the baby goats are born, we compare their average weight with the average weight of babies born to unexercised goats. We use the results to draw inferences about the effect of exercise on birth weights.

In an industrial setting, we measure a sample of bomb bases. We use the heights of bomb bases in the sample to make inferences about the average height of bomb bases produced. Or, we subject a sample of a new type of electric cord to accelerated life testing. We use the lifetimes in the sample to draw inferences about the median lifetime of this new, cheaper type of cord.

How do we decide on a formal procedure for making inferences about a measure of central tendency? The choice depends on sample size and the assumptions we are willing to make about the observations.

The most stringent assumptions, that we have independent observations from a Gaussian distribution, form the basis for classical statistical analysis. Assuming a specific form of distribution for the observations gives us a *parametric model.* This model leads to the t test and interval estimates for a mean based on a *t distribution.* We will discuss small-sample inferences about a mean based on a t distribution in Section 10-3.

Suppose we drop the assumption that the observations come from a Gaussian distribution. However, we continue to assume that the observations are independent and from a continuous, symmetric distribution. Then we have a *nonparametric model* with inferences based on a *Wilcoxon signed rank distribution.* Small-sample inference about a mean based on a Wilcoxon signed rank distribution is the subject of Section 10-4.

If we assume only that our observations are independent and come from a continuous distribution, then we have a simpler *nonparametric model.* For small samples, we base our inferences on a *binomial distribution,* as we will see in Section 10-5.

Section 10-1 discusses *large-sample inferences* about a population mean. We assume that the observations form a random sample from a population. We also assume that the sample size is large enough to ensure that the sample mean has approximately a Gaussian distribution (see the Central Limit Theorem results in Section 8-3). Then we base tests of hypotheses and interval estimates on the *standard Gaussian distribution.* Large-sample inference about a proportion is a special case, as we see in Section 10-2.

10-1

Large-Sample Inference About a Population Mean Based on the Standard Gaussian Distribution

Recall that in Example 9-3 we considered the thickness of gold applied to computer parts during an electroplating process. Emily was in charge of the process, and she was concerned that too much gold was being used. She took a large random sample from a production lot and used the parts in the sample to make inferences about the average gold thickness on parts in the entire lot. In particular, Emily compared a null hypothesis (mean gold thickness on parts in the lot equals a target value) with an alternative hypothesis (the mean exceeds the target value). Also, she calculated a confidence interval to provide an interval estimate or range of reasonable values for the mean gold thickness on parts in the production lot. Emily used large-sample techniques to make these inferences about the mean gold thickness in the lot.

In this section, we will outline the general approach to making inferences about a population mean based on large samples. Then we will apply this approach to the following example.

EXAMPLE 10-1

When you pick up your 510-gram box of corn flakes or puffed wheat do you ever wonder whether it really weighs 510 grams? You might be surprised if every 510-gram box of cereal weighed exactly 510 grams. After all, slight variations in conditions during processing could affect the amount of cereal deposited in the boxes.

Perhaps the 510-gram label printed on the box refers to an average over many boxes rather than a guaranteed weight in each box. Of course, if the box contains much less than 510 grams, then you, the consumer, feel cheated; and the cereal company does not want to put much more than 510 grams in a box.

A company packaging 510-gram boxes of cereal has an interest in keeping the weight close to 510 grams for each box. Suppose James is in charge of quality control at a company packaging 510-gram boxes of cereal. Out of a large production lot, he randomly selects 81 boxes and weighs the cereal in each box. A dot plot of the 81 weights is shown in Figure 10-1.

Figure 10-1 clearly shows that the weights in the sample do not all equal 510 grams. The weights range from 509.2 grams to 511.2 grams. The distribution appears to be fairly symmetrical about a value somewhat larger than 510 grams.

There are many questions James can ask about the boxes of cereal in the production lot. If government standards specify a range of acceptable weights for 510-gram boxes of cereal, he can ask what proportion of boxes are within this acceptable range (he might address this question with the techniques of Section 10-2). He may want to estimate the variation about the target weight of 510 grams (perhaps using techniques described in Section 14-1). He can also ask if the average weight of cereal packaged equals 510 grams per box.

Since we are considering inferences about a population mean, we will concentrate here on questions about the average weight of cereal packaged

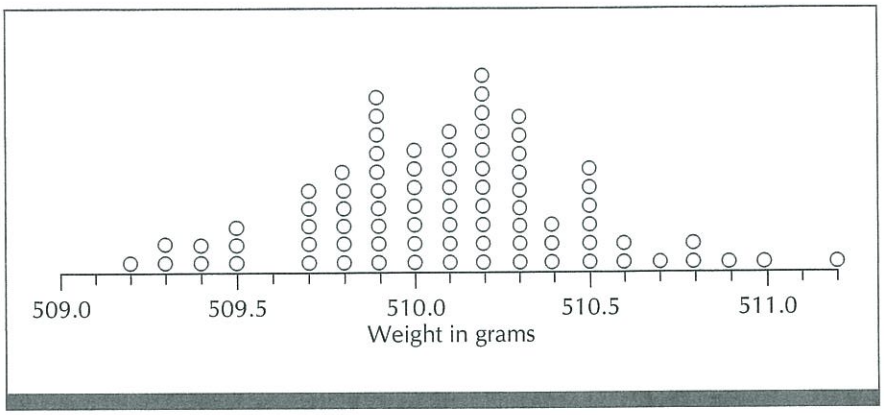


FIGURE 10-1 Dot plot of the weight (in grams) of cereal in 81 boxes, Example 10-1

per box. Let μ denote the mean weight of cereal packaged in boxes in the entire production lot. James wants to compare the hypotheses

$$H_0: \ \mu = 510 \text{ grams} \qquad \text{and} \qquad H_a: \ \mu \neq 510 \text{ grams}$$

He also wants to estimate the mean weight of cereal per box packaged in the entire production lot. Let's discuss the general approach to large-sample inferences about a population mean, and then see how James might make inferences about the lot of packaged cereal.

Testing Hypotheses About a Population Mean Based on Large Samples

Suppose we have a large random sample from a population with mean μ, and we want to use the sample to make inferences about μ. Let n represent the sample size, $\bar{X}$ the sample mean, s the sample standard deviation, and $\text{SE} = s/\sqrt{n}$ the standard error of the sample mean. To make large-sample inferences about μ, we assume that the sample size n is large enough that $(\bar{X} - \mu)/\text{SE}$ has approximately the standard Gaussian distribution (see the Central Limit Theorem results in Section 8-3). We now outline the significance level approach to testing hypotheses about a mean based on large samples. Then we will discuss how to find confidence intervals for a population mean.

The significance level approach to testing hypotheses about a population mean μ based on large samples

1. State null and alternative hypotheses, $H_0: \mu = \mu_0$ and $H_a: \mu \neq \mu_0$, for a specific number μ_0.
2. The test statistic is

$$\frac{\bar{X} - \mu_0}{\text{SE}}$$

3. Assume that we have independent observations from a distribution with mean μ. Also assume that the sample size is large enough that $\bar{X}$ has approximately a Gaussian distribution. Then under the null hypothesis, the test statistic has approximately the standard Gaussian distribution.
4. Select a significance level α.
5. The acceptance region is the interval $(-c, c)$ and the rejection region includes the intervals $(-\infty, -c]$ and $[c, \infty)$. The number c is chosen so that $P(Z \leq c) = 1 - \alpha/2$, where Z has the standard Gaussian distribution.
6. The decision rule is:

   If $-c <$ test statistic $< c$, say the results are consistent with the null hypothesis.

   If test statistic $\leq -c$ or test statistic $\geq c$, say the results are inconsistent with the null hypothesis.
7. Collect a large random sample from the population of interest. Calculate the test statistic based on the sample. Use the decision rule in step 6 to decide whether the observations are consistent with the null hypothesis. Draw conclusions based on what we see in the sample.

If we have the one-sided alternative $H_a: \mu > \mu_0$, then in step 5 the acceptance region is the interval $(-\infty, c)$. The rejection region is the interval $[c, \infty)$. The number c is chosen so that $P(Z \leq c) = 1 - \alpha$, where Z has the standard Gaussian distribution. The decision rule in step 6 is:

If test statistic $< c$, say the results are consistent with the null hypothesis.
If test statistic $\geq c$, say the results are inconsistent with the null hypothesis.

If our one-sided alternative is $H_a: \mu < \mu_0$, then in step 5 the acceptance region is the interval $(-c, \infty)$. The rejection region is the interval $(-\infty, -c]$. The number c is chosen so that $P(Z \leq -c) = \alpha$. The decision rule in step 6 is:

If test statistic $> -c$, say the results are consistent with the null hypothesis.
If test statistic $\leq -c$, say the results are inconsistent with the null hypothesis.

Large-Sample Confidence Intervals for a Population Mean

Large-sample confidence intervals for μ are of the form $(\bar{X} - c\text{SE}, \bar{X} + c\text{SE})$. We find the number c from the standard Gaussian distribution to give the desired confidence level. If the area from $-c$ to c under the standard Gaussian curve equals A, then

$$A \doteq P\left(-c \leq \frac{\bar{X} - \mu}{\text{SE}} \leq c\right) = P(\bar{X} - c\text{SE} \leq \mu \leq \bar{X} + c\text{SE})$$

We find c from Table B such that $P(Z \leq c) = (1 + A)/2$, where Z has the standard Gaussian distribution. We say the interval from $\bar{X} - c\text{SE}$ to $\bar{X} + c\text{SE}$ is an approximate $100A\%$ confidence interval for μ. For instance, $(\bar{X} - 1.96\text{SE}, \bar{X} + 1.96\text{SE})$ is an approximate 95% confidence interval for μ and $(\bar{X} - 1.65\text{SE}, \bar{X} + 1.65\text{SE})$ is an approximate 90% confidence interval for μ.

The correct interpretation of a confidence interval is this:

> If the sampling process were repeated over and over, and if a $100A\%$ confidence interval for μ were calculated each time, about $100A\%$ of these confidence intervals would contain μ and about $100(1 - A)\%$ would not. Once a specific confidence interval has been calculated, it either contains μ or it does not. The **confidence level** refers to what we would expect if the sampling process were repeated many times.

Confidence intervals and tests of hypotheses are related. If the value μ_0 specified in the null hypothesis is in the confidence interval, we say the results are consistent with the null hypothesis. If μ_0 is not in the confidence interval, we say the results are inconsistent with the null hypothesis.

EXAMPLE 10-1 *(continued)*

In Example 10-1, James wants to use his sample to make inferences about the average weight of cereal packaged per box in the production lot. He has a large random sample from a much larger population, so he feels justified in using large-sample inference based on the standard Gaussian distribution. One thing he wants to do is to compare the hypotheses H_0: $\mu = 510$ grams and H_a: $\mu \neq 510$ grams, where μ is the mean weight of cereal packaged in the production lot. He chooses significance level $\alpha = .05$. Since $P(Z \leq 1.96) = .975$, where Z has the standard Gaussian distribution, James chooses the following decision rule:

If $-1.96 <$ test statistic < 1.96, say the results are consistent with the null hypothesis that the mean weight per box in the lot equals 510 grams.

If test statistic ≤ -1.96 or test statistic ≥ 1.96, say the results are inconsistent with the null hypothesis, suggesting that the mean weight per box in the lot does not equal 510 grams.

James calculates the sample mean, standard deviation, and standard error for his observations: $\bar{X} = 510.10$ grams, $s = .39$ gram, $\text{SE} = s/9 = .043$ gram. For a test statistic, he calculates

$$\frac{\bar{X} - 510}{\text{SE}} = \frac{510.10 - 510}{.043} = 2.3$$

Since the observed value of his test statistic is 2.3, James decides the results are inconsistent with his null hypothesis that the mean weight per box in the lot equals 510 grams. The approximate p-value is .02, the probability that a standard Gaussian random variable is greater than or equal to 2.3 in absolute value. If the average weight of cereal packaged in the production lot were really 510 grams, there would be about a 2 in 100 chance of seeing a test statistic at least as extreme as the one observed. Since the sample mean is 510.10 grams, these results suggest that the average weight per box in the lot is greater than 510 grams.

James decides to estimate the average weight μ of cereal packaged in the production lot. His point estimate is $\bar{X} = 510.10$ grams. As an interval estimate

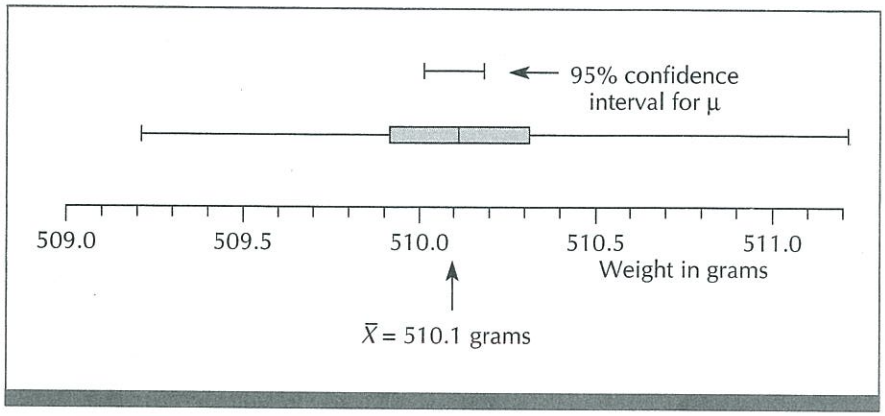


FIGURE 10-2 Box plot of the 81 sample weights in Example 10-1. Also shown is a 95% confidence interval for the population mean weight μ in the production lot.

of μ, he decides to use a 95% confidence interval. He knows that for a standard Gaussian random variable Z, $P(-1.96 \leq Z \leq 1.96) = .95$, so he calculates the confidence interval

$$(\bar{X} - 1.96\text{SE}, \bar{X} + 1.96\text{SE})$$
$$= (510.10 - 1.96 \times .043, 510.10 + 1.96 \times .043) = (510.02, 510.18)$$

James estimates that the mean weight of cereal per box in the production lot is from 510.02 grams to 510.18 grams. This confidence interval is illustrated in Figure 10-2, along with a box plot of the observations. Because the sample size is large, the confidence interval for the mean is narrow relative to the range of the observed weights. The null hypothesis mean, $\mu_0 = 510$ grams, is not in the confidence interval, in agreement with what we found using the test of hypotheses.

The sample mean of the 81 weights is 510.1 grams. (The sample median is also 510.1 grams.) From the 95% confidence interval, we estimate that the average packaged weight of cereal is from .02 to .18 gram above the labeled value of 510 grams. These results suggest that the average weight packaged in the cereal boxes during production is greater than the labeled weight of 510 grams. However, such inferences about the mean are not completely satisfactory for this quality control problem. Even though the sample average weight exceeds 510 grams, 29 (almost 36%) of the 81 sample boxes contain less than 510 grams of cereal. Eight (about 10%) of the sample boxes contain 509.5 grams or less. A fairly large proportion of consumers might feel cheated, if they expect at least 510 grams of cereal in a box. On the other hand, the smallest packaged weight in the sample, 509.2 grams, is less than 1 gram short of the labeled weight. Whether this sample represents an acceptable packaging process depends on tolerances or variations in packaged weight that are acceptable to consumers, the government, and the cereal company.

Large-sample inference about a population mean based on the standard Gaussian distribution is *nonparametric* because we assume no specific type of probability distribution for the observations. We do not even assume that the distribution is continuous. We will take advantage of this simplicity in Section 10-2 when we discuss large-sample inference about a proportion.

10-2 Large-Sample Inference About a Proportion

Before we discuss large-sample inference about a proportion in general, let's consider an example.

EXAMPLE 10-2

Tubal infertility refers to infertility resulting from damage to the Fallopian tubes, often caused by infection. Investigators wanted to study the relationship between tubal infertility and method of contraception (Cramer et al., 1987). They interviewed 283 women with tubal infertility and no children. Of these 283 women, 131 had at some time used barrier methods of contraception (for example, use of a diaphragm by the woman or condoms by her partner).

In a large control group of women with children, 51.5% had used barrier methods of contraception at some time. The women in the control group were similar to the women with tubal infertility with respect to age, race, and income status.

The investigators wanted to make inferences about p, the proportion of women with tubal infertility who had sometime used barrier methods of contraception. In particular, they wanted to ask how p compared with .515, the proportion for the control group.

Why do you think these investigators were interested in methods of contraception used by women with tubal infertility? What relationship might method of contraception (in particular, barrier methods versus other methods) have with infection and possible damage to the Fallopian tubes?

We will continue with this example after discussing the general approach to large-sample inference about a proportion.

Large-sample inference about a proportion is really just a special case of large-sample inference about a mean. Let's see why this is true.

Suppose p is the proportion we are interested in. Imagine n independent repetitions of a two-outcome experiment, where p is the probability of success and $1 - p$ is the probability of failure for each repetition. Let the random variable X_i equal 1 if the ith repetition results in success. Let X_i equal 0 if the ith repetition results in failure. Then X_1, X_2, through X_n represent a random sample from a Binomial$(1, p)$ distribution.

A reasonable estimate of the probability of success p is $\hat{p}$, where $\hat{p}$ denotes the proportion of successes observed in the sample. But $\hat{p}$ is the same as $\bar{X}$. Therefore, for large samples, we can base inferences about p on the standard Gaussian distribution.

Suppose we want to test the hypotheses $H_0: p = p_0$ and $H_a: p \neq p_0$ for a specified number p_0. The test statistic has the form

$$\frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$$

where $\sqrt{p_0(1 - p_0)/n}$ is the standard deviation of $\hat{p}$ when $p = p_0$. We carry out the test of hypotheses as outlined for the large-sample case in Section 10-1.

A confidence interval for p has the form

$$\hat{p} \pm c\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Here, $\sqrt{\hat{p}(1 - \hat{p})/n}$ is the standard error of $\hat{p}$. We find the number c from the standard Gaussian distribution, as in Section 10-1.

EXAMPLE 10-2 *(continued)*

In Example 10-2, p is the proportion of women with tubal infertility who had sometime used barrier methods of contraception. The investigators wanted to compare p with .515, the proportion of women in the control group who had sometime used barrier methods of contraception. In particular, they wanted to test the hypotheses $H_0: p = .515$ and $H_a: p \neq .515$.

For large-sample inference about p to be appropriate, we must assume that the sample represents 283 independent observations from a much larger population of women with tubal infertility. How likely is it that the sample is a random sample from this population? If these 283 women were all patients at a fertility clinic, how likely is it that the sample is representative of all women with tubal infertility? And, as long as we are posing questions, how reliable would you consider the information on contraception history for the women in the sample and the women in the control group?

A point estimate for p is $\hat{p}$, the proportion of women in the sample who had used barrier methods of contraception: $\hat{p} = 131/283 = .463$. The large-sample test statistic is

$$\text{Test statistic} = \frac{.463 - .515}{\sqrt{\dfrac{(.515)(.485)}{283}}} = -1.75$$

If Z has the standard Gaussian distribution, then $P(Z \geq 1.75) = .0401$. Therefore, the p-value is $P(\text{test statistic} \leq -1.75 \text{ or test statistic} \geq 1.75 \text{ when } H_0 \text{ is true}) \doteq .0802$. If the null hypothesis were true, there would be about an 8% chance of seeing a test statistic at least as far from .515 as the one observed. We might call this a borderline result. We cannot say the sample is strongly consistent with or strongly inconsistent with the null hypothesis.

An approximate 90% confidence interval for p is

$$.463 \pm 1.65\sqrt{\frac{(.463)(.537)}{283}} \quad \text{or} \quad (.414, .512)$$

An approximate 95% confidence interval for p is

$$.463 \pm 1.96\sqrt{\frac{(.463)(.537)}{283}} \quad \text{or} \quad (.405, .521)$$

The null hypothesis value .515 is in the 95% confidence interval, but not in the 90% confidence interval. This agrees with the borderline significance of the test of hypotheses. There is some evidence that the proportion of women with tubal infertility who had sometime used barrier methods of contraception is smaller than the corresponding proportion for the control group. But the evidence is not overwhelming.

In Section 10-3, we consider the classical approach to inferences about a population mean when we have a small sample.

10-3

Inferences About a Population Mean (or Median) Based on a *t* Distribution

We want to use a small sample to make inferences about the center of a population. In the classical approach, we must assume that the sample represents independent observations from a Gaussian distribution. Then our analysis is based on a t distribution. The corresponding test of hypotheses is called the *t test* or *one-sample t test.* In this section, we will discuss inferences about a population mean based on a t distribution, as applied to the following example.

EXAMPLE 10-3

Does exercise affect birth weights in Pygmy goats? In a study designed to answer this question, investigators trained pregnant Pygmy goats to walk on a motor-driven treadmill (Dhindsa, Metcalfe, and Hummels, 1978). Six of these goats subsequently gave birth to twins. The average weights in grams of the six pairs of twins were:

745.5 1,175.0 1,290.0 1,364.5 1,397.5 1,660.0

In a large control group of similar pregnant Pygmy goats who did not exercise, the average weight of twins was 1,592.0 grams.

A dot plot of the six average weights of twins born to exercised goats is shown in Figure 10-3. We see that five of these six values are less than the average weight (1,592.0 grams) of twins born to unexercised goats. The distribution of the six plotted values is concentrated around an interval from about 1,200 to 1,400 grams. What does the plot suggest about the relative weights of twins born to exercised and unexercised goats?

Let μ denote the mean weight of twins born to exercised goats. We would like to test whether μ equals 1,592.0, the mean weight of twins born to unexercised goats. We would also like to calculate an interval estimate, or range of

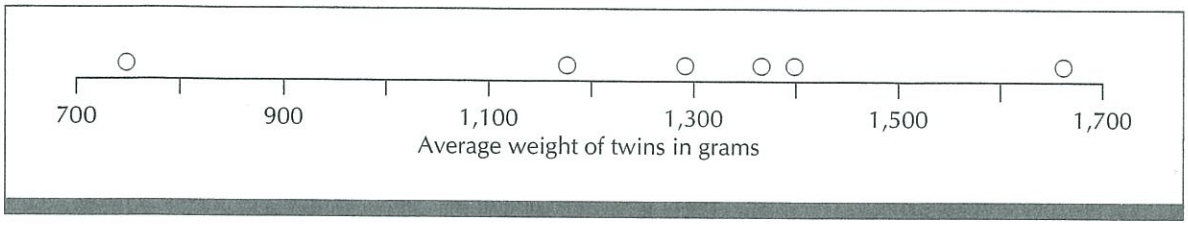


FIGURE 10-3 Dot plot of the average weight of twins born to six exercised goats in Example 10-3

reasonable values, for μ. Each of these analyses will be based on a t distribution. Let's discuss the general approach and then apply it to this example.

First consider the assumptions we must make for analysis based on a t distribution. Suppose we have a sample of size n. We assume that these n observations are independent, with the same Gaussian probability distribution. We want to make inferences about the mean μ of this distribution. Because a Gaussian distribution is symmetric about its mean, μ is also the median of the distribution.

Let $\bar{X}$ denote the sample mean, s the sample standard deviation, and $\text{SE} = s/\sqrt{n}$ the standard error of the mean. Under the assumptions just stated, $(\bar{X} - \mu)/\text{SE}$ has a probability distribution known as the t distribution or Student t distribution with $n - 1$ degrees of freedom.

A t distribution is characterized by a parameter called its *degrees of freedom*. Recall that when we calculate the sample variance, we add up the squared deviations of the observations from the sample mean and then divide by the sample size minus 1. This denominator, $n - 1$, is the degrees of freedom for the t distribution of $(\bar{X} - \mu)/\text{SE}$.

Suppose X_1, X_2, through X_n are independent observations from a Gaussian distribution with mean μ. Let $\bar{X}$ denote the sample mean and SE the standard error of the mean for these observations. Then

$$\frac{\bar{X} - \mu}{\text{SE}}$$

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

A t distribution is a continuous probability distribution that is symmetrical about 0, as illustrated in Figure 10-4a. A graph of a t distribution looks similar to that of the standard Gaussian distribution. However, a random variable with a t distribution is more likely to be far from zero than is a standard Gaussian random variable. We say a t distribution has fatter tails than the standard Gaussian distribution, illustrated in Figure 10-4b. The exact shape of a t distribution depends on its degrees of freedom. As the degrees of freedom (or equivalently, the sample size) increase, the differences between a t distribution and the standard Gaussian distribution diminish. For large sample sizes, probabilities determined from the corresponding t distribution are practically the same as those obtained from the standard Gaussian distribution.

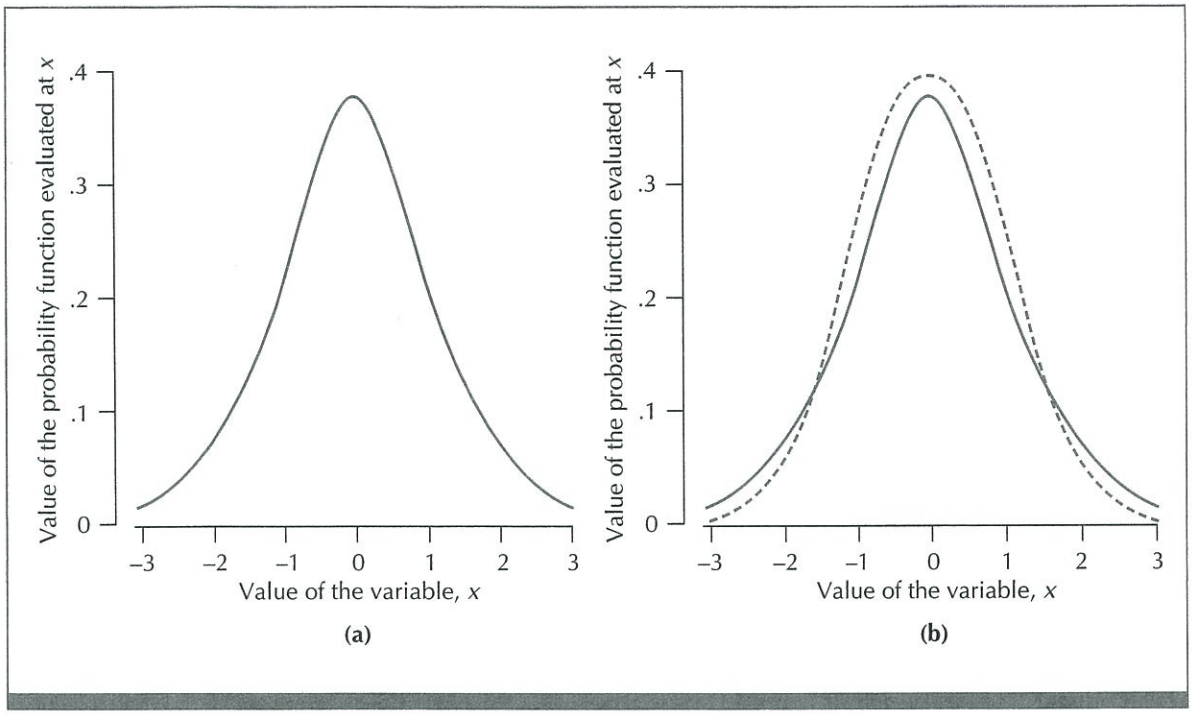


FIGURE 10-4 **a.** Graph of the probability function for the t distribution with 5 degrees of freedom. **b.** Comparison of the probability function for the t distribution with 5 degrees of freedom (solid line) with the probability function for the standard Gaussian distribution (dashed line). A t distribution has fatter tails than does the standard Gaussian distribution.
Adapted from *Mathematical Statistics and Data Analysis*, by J. A. Rice. (Pacific Grove, CA: Brooks/Cole Publishing Co., 1988, p. 170.)

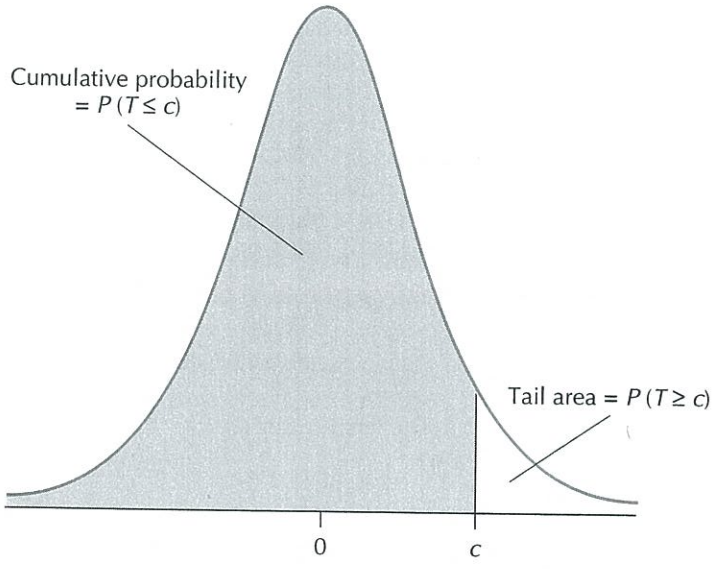
Let T denote a random variable having the t distribution with d degrees of freedom. We are interested in finding probabilities such as $P(T \geq c)$ or $P(-c \leq T \leq c)$, where c is a positive number. These probabilities correspond to areas under the graph of the appropriate probability function. For instance, the area corresponding to the cumulative probability $P(T \leq c)$ is shaded in Figure 10-5. The tail area $P(T \geq c)$ is the unshaded region under the curve in Figure 10-5.

Table C at the back of the book lists numbers c and cumulative probabilities $P(T \leq c)$ for several degrees of freedom d. All the numbers c in Table C are greater than 0. Because a t distribution is symmetrical about 0, we know that

$$P(T \leq -c) = P(T \geq c) = 1 - P(T \leq c)$$

for any positive number c.

Let's see how to use Table C. The line in the table for 5 degrees of freedom is reproduced in Figure 10-5. We see that for a cumulative probability of



Degrees of freedom	Cumulative probability							
5	**.60**	**.70**	**.80**	**.90**	**.95**	**.975**	**.99**	**.995**
	.267	.559	.920	1.476	2.015	2.571	3.365	4.032

FIGURE 10-5 The line in Table C corresponding to the t distribution with 5 degrees of freedom is reproduced here, along with an illustration.

.80, the tabled value of c is .920. This means that if the random variable T has the t distribution with 5 degrees of freedom, then

$$P(T \leq .920) = P(T \geq -.920) = .80$$

From this information, we know that the tail probabilities $P(T > .920)$ and $P(T < -.920)$ both equal .20, and we can find the probability $P(-.920 \leq T \leq .920)$ this way:

$$P(-.920 \leq T \leq .920) = 1 - P(T > .920) - P(T < -.920) = .60$$

Similarly, we can see that

$$P(T \leq 2.571) = P(T \geq -2.571) = .975 \quad \text{and}$$
$$P(-2.571 \leq T \leq 2.571) = .95$$

Table C gives probabilities for t distributions with degrees of freedom up to 30, and three values of d larger than 30. The last line of the table, for $d = \infty$,

corresponds to the standard Gaussian distribution. For degrees of freedom greater than 30, we will use the standard Gaussian table (Table B) since $(\bar{X} - \mu)/\text{SE}$ has approximately the standard Gaussian distribution in that case.

A t distribution is a continuous probability distribution, so if a random variable T has a t distribution, then $P(T = c) = 0$ for any number c. Therefore, we know that $P(T \geq c)$ equals $P(T > c)$, for instance, and $P(-c \leq T \leq c) = P(-c < T < c)$. (The argument is the same as for a Gaussian random variable. Recall Figure 8-3d and the discussion in Section 8-1.)

Let's outline the significance level approach to testing hypotheses about a population mean μ based on a t distribution.

The significance level approach to testing hypotheses about a population mean μ based on a t distribution

1. State null and alternative hypotheses, H_0: $\mu = \mu_0$ and H_a: $\mu \neq \mu_0$, for some number μ_0.
2. The test statistic is
$$\frac{\bar{X} - \mu_0}{\text{SE}}$$
3. Assume that we have independent observations from a Gaussian distribution with mean μ. Then the test statistic has the t distribution with $n - 1$ degrees of freedom under the null hypothesis.
4. Specify the significance level α.
5. Find the number c in Table C such that $P(T \leq c) = 1 - \alpha/2$, where T is a random variable having the t distribution with $n - 1$ degrees of freedom. The acceptance region is the interval $(-c, c)$. The rejection region includes $(-\infty, -c]$ and $[c, \infty)$.
6. The decision rule is:
   If $-c <$ test statistic $< c$, say the results are consistent with the null hypothesis.
   If test statistic $\leq -c$ or test statistic $\geq c$, say the results are inconsistent with the null hypothesis.
7. Collect a random sample satisfying the stated assumptions. Calculate the test statistic. Use the decision rule in step 6 to decide whether the observations are consistent or inconsistent with the null hypothesis. Draw conclusions and discuss the experimental results.

If in step 1 we specify a one-sided alternative, then we must alter steps 5 and 6 appropriately. We find the number c in Table C such that $P(T \leq c) = 1 - \alpha$. If we consider the one-sided alternative H_a: $\mu > \mu_0$, then the acceptance region is the interval $(-\infty, c)$ and the rejection region is the interval $[c, \infty)$. We say values of the test statistic less than c are consistent with the null hypothesis, while values greater than or equal to c are inconsistent with the null hypothesis.

If we specify the one-sided alternative H_a: $\mu < \mu_0$, then the acceptance region is the interval $(-c, \infty)$ and the rejection region is the interval $(-\infty, -c]$.

We say values of the test statistic greater than $-c$ are consistent with the null hypothesis, while values less than or equal to $-c$ are inconsistent with the null hypothesis.

EXAMPLE 10-3 *(continued)*

Let's test the hypotheses of interest in Example 10-3. The population is a hypothetical population of Pygmy goats pregnant with twins who might have been subjected to the exercise regimen. If we let μ denote the mean weight of twins born to such goats, then the hypotheses are H_0: $\mu = 1{,}592.0$ grams and H_a: $\mu \neq 1{,}592.0$ grams.

The test statistic is $(\bar{X} - 1{,}592.0)/\text{SE}$, where $\bar{X}$ is the sample mean and SE the standard error of the six observations in the sample. Assume that the observations are independent, from a Gaussian distribution with mean μ. Then the test statistic has the t distribution with 5 degrees of freedom under the null hypothesis.

Independence means that the results for one pregnant goat did not affect in any way the results for another pregnant goat. We cannot judge the appropriateness of this assumption without more details about the experiment.

The dot plot in Figure 10-3 gives us no reason to doubt the assumption that the observations come from a Gaussian distribution. With so few observations, we cannot say we have strong evidence supporting this assumption, either. The t test, however, tends to be *robust* to deviations from the Gaussian assumption. This means that even when the observations do not follow exactly a Gaussian distribution, significance levels (and confidence levels) tend to be close to the values we select.

> We say a statistical procedure is **robust** if the actual significance level (or confidence level) is close to the level we select, even under deviations from assumptions.

Letting the significance level α equal .10, we find $\alpha/2 = .05$. Referring to Table C or Figure 10-5, we see that if T is a random variable having the t distribution with 5 degrees of freedom, then $P(T \leq 2.015) = .95$. The acceptance region is the interval $(-2.015, 2.015)$. The rejection region includes the two intervals $(-\infty, -2.015]$ and $[2.015, \infty)$. The decision rule is:

If $-2.015 <$ test statistic < 2.015, say the results are consistent with the null hypothesis.

If test statistic ≤ -2.015 or test statistic ≥ 2.015, say the results are inconsistent with the null hypothesis.

Using the six sample observations, we find $\bar{X} = 1{,}272.08$ grams and SE $= 124.07$ grams. The test statistic is

$$\frac{\bar{X} - 1{,}592.0}{\text{SE}} = \frac{1{,}272.08 - 1{,}592.0}{124.07} = -2.6$$

Since -2.6 is less than -2.015, we say the experimental results are inconsistent with the null hypothesis, at the $\alpha = .10$ significance level.

The p-value is the probability of seeing a test statistic as extreme as or more extreme than the one observed, if the null hypothesis were true. Since the test statistic equals -2.6, the p-value equals

$$p\text{-value} = P(T \leq -2.6) + P(T \geq 2.6)$$

where T has the t distribution with 5 degrees of freedom. From Table C (or Figure 10-5), we see the p-value is about .05. If the null hypothesis were true, there would be about a 5% chance of seeing results as extreme as or more extreme (in the direction of the alternative) than those observed.

The experimental results suggest that the population mean μ does not equal 1,592.0 grams. If we want to estimate μ, a reasonable point estimate is $\bar{X} = 1{,}272.08$, or 1,272.1 grams. We know that a different sample would probably result in a different sample mean, so we would like an interval estimate or range of reasonable values for μ. The general procedure for finding a confidence interval for a population mean μ based on a t distribution is outlined below.

Confidence Intervals for a Population Mean Based on a t Distribution

A confidence interval for the population mean μ, based on a t distribution, has the form $(\bar{X} - c\text{SE}, \bar{X} + c\text{SE})$, where c is the number that gives the confidence level we want. Suppose A is a number between 0 and 1. If we want a $100A\%$ confidence interval for μ, then we find c such that

$$A = P\left(-c \leq \frac{\bar{X} - \mu}{\text{SE}} \leq c\right) = P(\bar{X} - c\text{SE} \leq \mu \leq \bar{X} + c\text{SE})$$

This is the number c from Table C that satisfies $P(T \leq c) = (1 + A)/2$, where T has the t distribution with $n - 1$ degrees of freedom.

As with large-sample confidence intervals, the correct interpretation is this:

> If the sampling process were repeated over and over, and if a $100A\%$ confidence interval for μ were calculated each time, about $100A\%$ of these confidence intervals would contain μ and about $100(1 - A)\%$ would not. A specific confidence interval either contains μ or it does not; the **confidence level** refers to what we would expect if the sampling process were repeated many times.

Suppose we want a 90% confidence interval for the population mean in Example 10-3. We can think of μ as the mean weight of twins born to (a hypothetical population of) exercised Pygmy goats. Then $A = .90$ and $(1 + A)/2 = .95$. If T has a t distribution with 5 degrees of freedom, then $P(T \leq 2.015) = .95$. Therefore, a 90% confidence interval for μ is

$$(\bar{X} - 2.015\text{SE}, \bar{X} + 2.015\text{SE}) = (1{,}022.1, 1{,}522.1)$$

This confidence interval is shown graphically in Figure 10-6, along with a dot plot of the six observations. Note that because the sample size is small, the

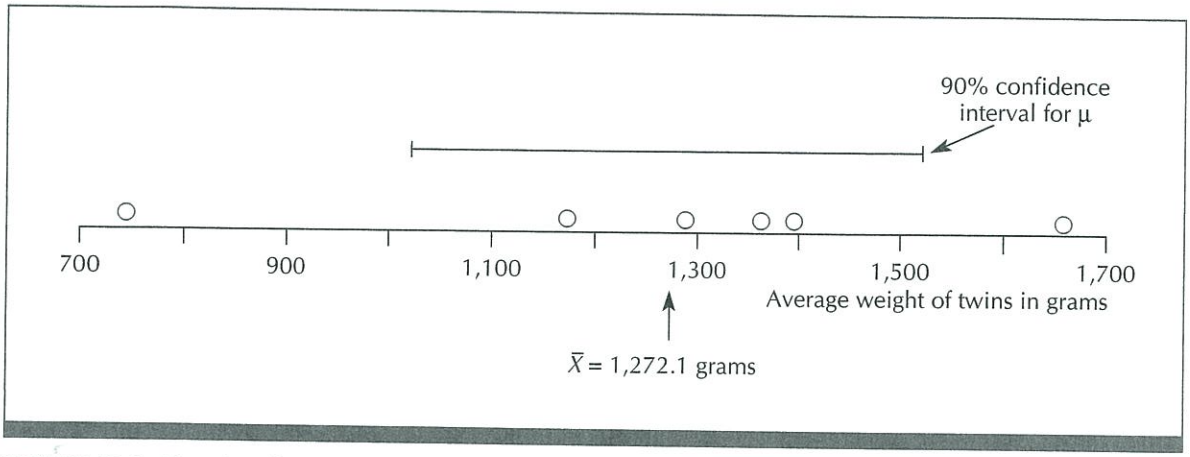


FIGURE 10-6 Dot plot of average weight of twins born to six exercised Pygmy goats in Example 10-3., Also shown is a 90% confidence interval for the mean weight μ of twins born to a hypothetical population of exercised Pygmy goats.

confidence interval is wide compared with the range of data values. We say the interval from 1,022.1 grams to 1,522.1 grams is a range of reasonable values for μ. The null hypothesis value, $\mu_0 = 1{,}592.0$ grams, is not in this interval, in agreement with the test of hypotheses.

What do the results of our analysis suggest about the effect of exercise on average birth weights? If you were asked to develop a theory for why the offspring of exercised goats weighed on average less than the offspring of unexercised goats, what might it be? Would you be willing to extrapolate the results of this experiment on Pygmy goats and hazard a guess about the effect of maternal exercise on birth weights in humans?

Some time after this experiment, the investigators decided a possible extraneous factor might have affected the outcome. A trainer induced a goat to exercise by applying a mild electric shock to a rear leg whenever she stopped exercising! In a subsequent experiment, investigators trained the goats to exercise with positive reinforcement such as petting or food. In this second study, birth weights were on average no different in the exercise and no-exercise groups (Hohimer et al., 1984). How does this new information affect any conclusions we might have drawn based on the first experiment?

In Section 10-4, we consider small-sample inferences about a population mean when we are not willing to assume the observations come from a Gaussian distribution.

10-4 Inferences About a Population Mean (or Median) Based on a Wilcoxon Signed Rank Distribution

Again, we want to use a sample to make inferences about the center of a population. We assume that we have a random sample of observations from a con-

Making Large-Sample Inferences About a Mean

For large-sample inference about a population mean, we can use the TTEST and TINTERVAL commands, because for large degrees of freedom, a t distribution is close to the standard Gaussian distribution. There is no direct way in Minitab to carry out a large-sample test about a proportion.

Exercises for Chapter 10

In each exercise, describe the population sampled (whether real or hypothetical). Graph the data in any way that seems helpful. For each statistical procedure, state the assumptions that make the analysis appropriate. Do these assumptions seem reasonable? Discuss the results of your analysis.

EXERCISE 10-1

Investigators measured plasma citrate concentrations at 8 A.M., before breakfast, for 10 volunteers (from a contribution by E. B. Jensen to a collection of problems in Andrews and Herzberg, 1985, page 237; from Andersen, Jensen, and Schou, 1981). The measurements are shown below (in μmol per liter).

93 116 125 144 105 109 89 116 151 137

a. Plot the observations.

b. Calculate a confidence interval for the mean plasma citrate concentration, using a t distribution.

c. Calculate a confidence interval for the mean plasma citrate concentration, using a Wilcoxon signed rank distribution.

d. Calculate a confidence interval for the median plasma citrate concentration, using a binomial distribution.

e. Discuss and compare your answers to parts (b), (c), and (d).

EXERCISE 10-2

As part of a patent application for a new cake mix, applicants compared two types of cake mix (Box, Hunter, and Hunter, 1978, page 160; from U. S. Patent 3,505,079, April 7, 1970). They prepared five recipes with each of two types of cake mix, the new mix and an old mix. The difference in volume (units not given) between the two mixes is shown below for each recipe:

Recipe	Difference in volume new − old
1	18
2	8
3	6
4	18
5	8

a. Plot these observations.

b. The patent claims that the new cake mix results in significantly greater volume than the old mix. Is this claim justified by these observations? State and test appropriate hypotheses.

EXERCISE 10-3

In a large set of measurements on nonpregnant women, the average fasting blood glucose level was about 80 milligrams/100 milliliters of blood. Researchers determined fasting blood sugar levels for 52 women during their third trimester of pregnancy (contributed by C. M. Mahan to a collection of problems in Andrews and Herzberg, 1985, pages 211–214; from O'Sullivan and Mahan, 1966). The results (in mg/100 ml) are shown here (sample size = 52, mean = 70.12, sample standard deviation = 9.68).

60	56	80	55	62	74	64	73	68	69	60	70
66	83	68	78	103	77	66	70	75	91	66	75
74	76	74	74	67	78	64	67	78	64	71	63
90	60	48	66	74	60	63	66	77	70	73	78
73	72	65	52								

a. Plot these measurements.

b. How do these blood sugar levels for pregnant women compare with the average level of 80 mg/100 ml for nonpregnant women? State and test appropriate hypotheses.

c. Calculate an interval estimate for the mean fasting blood sugar level of women in their third trimester of pregnancy.

EXERCISE 10-4

Does a heart defect known as patent foramen ovale contribute to the bends (decompression sickness) experienced by some scuba divers? Researchers used echocardiography to examine 30 divers with a history of decompression sickness (*Science News,* March 25, 1989, volume 135, page 188). Eleven of the 30 divers showed evidence of the heart defect. About 5% of the general population has this heart defect.

a. Does the proportion of divers with a history of decompression sickness who have this heart defect seem to differ from that of the general population? State and test appropriate hypotheses.

b. Calculate a confidence interval for the proportion of divers with a history of decompression sickness who have this heart defect. Discuss your findings.

EXERCISE 10-5

An engineer subjected 35 motors to a life test. The engineer set up a machine for recording time to failure, 16 hours after the beginning of the life test. Four motors failed before the recording machine was set up at 16 hours. Ten motors were still working at 30 hours, when the engineer ended the test. The times to failure (in hours) for the other 21 motors are shown below (Shapiro, 1986, page 22; from Brain and Shapiro, 1983).

16.0	16.3	16.7	16.9	17.0	17.1	17.3	17.8	17.9
18.3	18.4	18.6	19.1	19.5	20.6	21.4	22.9	23.0
24.6	25.9	28.6						

a. Plot these times to failure.

b. Test the null hypothesis that the median time to failure is 20 hours versus the alternative that it is not 20 hours.

c. Calculate a confidence interval for the median time to failure for this type of motor.

EXERCISE 10-6 Some health care workers estimate that one-third of people with coronary artery disease may be clinically depressed. In one study, researchers found 9 cases of major depression among 52 patients with newly diagnosed coronary disease (*Science News,* January 7, 1989, volume 135, page 13).

a. Does the health care workers' estimate seem reasonable based on this sample? State and test appropriate hypotheses.

b. Calculate a confidence interval for the proportion of patients with newly diagnosed coronary artery disease suffering from major depression. Discuss your findings.

EXERCISE 10-7 Scientists developed a new method of determining serum iron concentrations. To check the accuracy of the method, they made 20 analyses of control sera, with a concentration of 105 μg serum iron per 100 milliliters (Hollander and Wolfe, 1973, pages 85–86; a portion of the data of Jung and Parekh, 1970). The determinations of serum iron concentration (μg/100 ml) are shown below.

96	98	99	100	103	103	104	104	105	105
106	106	107	108	108	108	110	113	114	114

a. Plot these observations.

b. Test the null hypothesis that the average serum iron determination using the new method is 105 μg/100 ml.

c. Calculate a confidence interval for the average serum iron determination using the new method. Discuss your findings.

EXERCISE 10-8 A major problem in liver transplantation is the short time (at most 10 hours) that donor livers can be preserved. In a large study, researchers treated 185 donor livers with an experimental solution and 180 livers with the traditional solution for organ preservation. Eighty-one of the livers treated with the experimental solution lasted longer than 9.5 hours, while none of the livers treated with the traditional solution lasted that long (numbers calculated from percentages reported in *Science News,* February 4, 1989, volume 135, page 70).

a. Calculate a confidence interval for the proportion of livers treated with the new solution that are still viable after 9.5 hours.

b. Discuss the experimental results.

EXERCISE 10-9 In a tomato processing factory, the drained weight after filling cans of tomatoes in puree averages 21.8 ounces during the morning. One afternoon, a quality control worker selects five cans of tomatoes filled that afternoon and weighs

the drained contents in ounces (based on Duncan, 1974, page 569; Grant and Leavenworth, 1972, page 41):

19.0 19.5 19.5 20.5 21.5

a. Plot the observations.

b. Use a t distribution to test the null hypothesis that the average drained weight in cans filled that afternoon equaled 21.8 ounces. Calculate a confidence interval for the average afternoon drained weight.

c. Repeat part (b) using a Wilcoxon signed rank distribution.

d. Use a binomial distribution to test the null hypothesis that the median drained weight in cans filled that afternoon equaled 21.8 ounces. Calculate a confidence interval for the median afternoon drained weight.

e. Compare your results in parts (b), (c), and (d). Discuss your findings.

EXERCISE 10-10 Researchers have worked to identify serum markers for the genetically transmitted disease Duchenne muscular dystrophy. The average serum level of the enzyme creatine kinase in a large group of women who were not carriers of the disease was 39.8 (units not given). Measurements of the serum level of creatine kinase are shown below for 38 women who were genetic carriers of the disease (data contributed by M. Percy of Mount Sinai Hospital in Toronto to a collection of problems in Andrews and Herzberg, 1985, pages 223–228). (Sample size = 38, mean = 175.9, sample standard deviation = 192.8.)

167	104	30	65	440	58	129	265	285	124
53	657	168	286	73	19	113	57	78	69
48	109	925	59	363	37	101	99	560	85
197	154	80	28	57	326	100	115		

a. Plot these observations.

b. Test the null hypothesis that the mean serum creatine kinase level among female carriers equals 39.8.

c. Calculate a confidence interval for the mean serum creatine kinase level among female carriers.

d. Discuss your findings.

EXERCISE 10-11 Refer to the exercise experiment on pregnant Pygmy goats in Example 10-3.

a. Use a Wilcoxon signed rank distribution to test the null hypothesis that the mean weight of twins born to exercised goats equals 1,592.0 grams. Calculate a confidence interval for the mean weight of twins born to exercised goats.

b. Use a binomial distribution to test the null hypothesis that the median weight of twins born to exercised goats equals 1,592.0 grams. Calculate a confidence interval for the median weight of twins born to exercised goats.

c. Compare your results in parts (a) and (b) with our results using a t distribution in Example 10-3.

EXERCISE 10-12 Refer to the quality control problem involving heights of bomb bases in Example 10-4.

a. Use a t distribution to test the null hypothesis that the mean height of bomb bases in the production lot equaled .830 inch. Calculate a confidence interval for the mean height of bomb bases in the production lot.

b. Use a binomial distribution to test the null hypothesis that the median height of bomb bases in the production lot equaled .830 inch. Calculate a confidence interval for the median height of bomb bases in the production lot.

c. Compare your results in parts (a) and (b) with our results using a Wilcoxon signed rank distribution in Example 10-4.

EXERCISE 10-13 In Example 10-5, we asked if the median life of a new type of electric cord equaled 114 hours. Use the large-sample version of the sign test to test this null hypothesis. Compare your results with our results using a binomial distribution in Example 10-5.

EXERCISE 10-14 In the appendix on the Wilcoxon signed rank distributions at the end of the book, we find the Wilcoxon signed rank distribution of $T+$ for sample size 3. Show that $T-$ has the same probability distribution.

EXERCISE 10-15 Find the Wilcoxon signed rank distribution for a sample of size 4.

EXERCISE 10-16 Twelve pairs of siblings were involved in a study to investigate whether children can recognize siblings by the sense of smell (Porter and Moore, 1981). In each pair, the children were full siblings living together with their parents. The younger child was from 36 to 49 months of age and the older child was from 62 to 95 months of age.

Each child was given one of 24 identical new T-shirts. Parents were asked to have each child wear his or her T-shirt to bed three nights in a row. T-shirts were stored during the day in individual sealed plastic bags.

On the morning following the third night of the experiment, each child sniffed each of two T-shirts: one worn during the experiment by his or her sibling and the other worn by another child of about the same age as the sibling. The child was asked to identify the T-shirt worn by his or her sibling. Nineteen of the 24 children made correct selections.

a. Why were T-shirts stored during the day in individual sealed plastic bags?

b. Test the null hypothesis that the performance of the children was no different from what we would expect from chance guessing.

c. What does this experiment suggest about a child's ability to recognize a sibling by the sense of smell?

EXERCISE 10-17 Ten pairs of siblings and 18 of the 20 parents participated in a study to see if parents could distinguish between their two children by sense of smell. (This was a part of the experiment discussed in Exercise 10-16.) Each child wore a

T-shirt for three consecutive nights. Each parent was then asked to sniff each of the T-shirts worn by his or her two children and identify the T-shirt worn by each child. Sixteen of the 18 parents correctly distinguished between the T-shirts worn by their two children (Porter and Moore, 1981).

a. State and test appropriate hypotheses.

b. What does this experiment suggest about a parent's ability to distinguish between his or her children by the sense of smell?

EXERCISE 10-18 Can a growth hormone gene be transferred from one type of fish to another? If so, faster growing fish might be developed, shortening the time for fish farmers to raise full-grown fish. To investigate this idea, experimenters injected a growth hormone gene from rainbow trout into thousands of carp eggs. Of 400 fish that grew from those eggs, 20 incorporated the gene into their DNA (*Science News,* June 11, 1988, volume 133, page 374). Calculate a 95% confidence interval for the proportion of carp that would incorporate this growth hormone gene under the same experimental conditions.

EXERCISE 10-19 The conventional treatment for patients with severe ulcerative colitis, an inflammatory disease of the colon, is surgical removal of the colon. In one study, 11 patients who were candidates for surgery chose to be treated with the experimental drug cyclosporin (known for its ability to prevent organ rejection in organ transplant surgery). At the end of six months, 5 of the 11 patients were in complete remission. A sixth patient had responded to treatment, but still needed drug therapy to control colitis symptoms (*Science News,* May 20, 1989, volume 135, page 310). Calculate a 90% confidence interval for the proportion of patients with severe ulcerative colitis who would respond to drug therapy.

EXERCISE 10-20 Discuss the sampling situations in which the t test, the Wilcoxon signed rank test, and the sign test are appropriate. Which test is preferred in each of these situations?