9

Small Sample Inferences for Normal Populations

Collecting a Sample of Lengths of Anacondas

Jesus Rivas, a herpetologist, is currently doing the definitive research on green anacondas. These snakes, some of the largest in the world, can grow to 25 feet in length. They have been known to swallow live goats and even people. Jesus Rivas and fellow researchers walk barefoot in shallow water in the Llanos grasslands shared by Venezuela and Colombia during the dry season. When they feel a snake with their feet, they grab and hold it with the help of another person. After muzzling the snake with a sock and tape, they place a string along an imaginary centerline from head to tail. The measured length of string is the recorded length of the anaconda.

Females are typically larger than males. The lengths (feet) of 21 females are

10.2	11.4	13.6	17.1	16.5	11.8	15.6
11.8	11.3	11.9	9.6	14.4	13.2	13.5
12.4	12.1	11.6	8.6	13.0	16.3	14.4

If the captured snakes can be treated as a random sample, a 95% confidence interval for the mean length of a female anaconda in that area is given in the computer output

```
One-Sample T: Length(ft)

 Variable       N    Mean  StDev       95.0% CI
 Length(ft)   21  12.871  2.262  ( 11.842, 13.901)
```

Both capturing and measuring make data collection difficult.

1. INTRODUCTION

In Chapter 8, we discussed inferences about a population mean when a large sample is available. Those methods are deeply rooted in the central limit theorem, which guarantees that the distribution of $\overline{X}$ is approximately normal. By the versatility of the central limit theorem, we did not need to know the specific form of the population distribution.

Many investigations, especially those involving costly experiments, require statistical inferences to be drawn from small samples ($n \leq 30$, as a rule of thumb). Since the sample mean $\overline{X}$ will still be used for inferences about μ, we must address the question, "What is the sampling distribution of $\overline{X}$ when n is not large?" Unlike the large sample situation, here we do not have an unqualified answer. In fact, when n is small, the distribution of $\overline{X}$ does depend to a considerable extent on the form of the population distribution. With the central limit theorem no longer applicable, more information concerning the population is required for the development of statistical procedures. In other words, the appropriate methods of inference depend on the restrictions met by the population distribution.

In this chapter, we describe how to set confidence intervals and test hypotheses **when it is reasonable to assume that the population distribution is normal.**

We begin with inferences about the mean μ of a normal population. Guided by the development in Chapter 8, it is again natural to focus on the ratio

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

when σ is also unknown. The sampling distribution of this ratio, called Student's t distribution, is introduced next.

2. STUDENT'S *t* DISTRIBUTION

When $\overline{X}$ is based on a random sample of size n from a normal $N(\mu, \sigma)$ population, we know that $\overline{X}$ is exactly distributed as $N(\mu, \sigma/\sqrt{n})$. Consequently, the standardized variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

has the standard normal distribution.

Because σ is typically unknown, an intuitive approach is to estimate σ by the sample standard deviation S. Just as we did in the large sample situation, we consider the ratio

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Its probability density function is still symmetric about zero. Although estimating σ with S does not appreciably alter the distribution in large samples, it does make a substantial difference if the sample is small. The new notation T is required in order to distinguish it from the standard normal variable Z. In fact, this ratio is no longer standardized. Replacing σ by the sample quantity S introduces more variability in the ratio, making its standard deviation larger than 1.

The distribution of the ratio T is known in statistical literature as "Student's t distribution." This distribution was first studied by a British chemist W. S. Gosset, who published his work in 1908 under the pseudonym "Student." The brewery for which he worked apparently did not want the competition to know that it was using statistical techniques to better understand and improve its fermentation process.

Student's *t* Distribution

If $X_1, \ldots, X_n$ is a random sample from a normal population $N(\mu, \sigma)$ and

$$\overline{X} = \frac{1}{n}\sum X_i \qquad \text{and} \qquad S^2 = \frac{\sum (X_i - \overline{X})^2}{n - 1}$$

then the distribution of

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

is called **Student's t distribution with $n - 1$ degrees of freedom**

The qualification "with $n - 1$ degrees of freedom" is necessary, because with each different sample size or value of $n - 1$, there is a different t distribution. The choice $n - 1$ coincides with the divisor for the estimator S^2 that is based on $n - 1$ degrees of freedom.

The t distributions are all symmetric around 0 but have tails that are more spread out than the $N(0, 1)$ distribution. However, with increasing degrees of freedom, the t distributions tend to look more like the $N(0, 1)$ distribution. This agrees with our previous remark that for large n the ratio

$$\frac{\overline{X} - \mu}{S / \sqrt{n}}$$

is approximately standard normal. The density curves for t with 2 and 5 degrees of freedom are plotted in Figure 1 along with the $N(0, 1)$ curve.

Appendix B, Table 4, gives the upper α points t_α for some selected values of α and the degrees of freedom (abbreviated d.f.).

The curve is symmetric about zero so the lower α point is simply $-t_\alpha$. The entries in the last row marked "d.f. = infinity" in Appendix B, Table 4, are exactly the percentage points of the $N(0, 1)$ distribution.

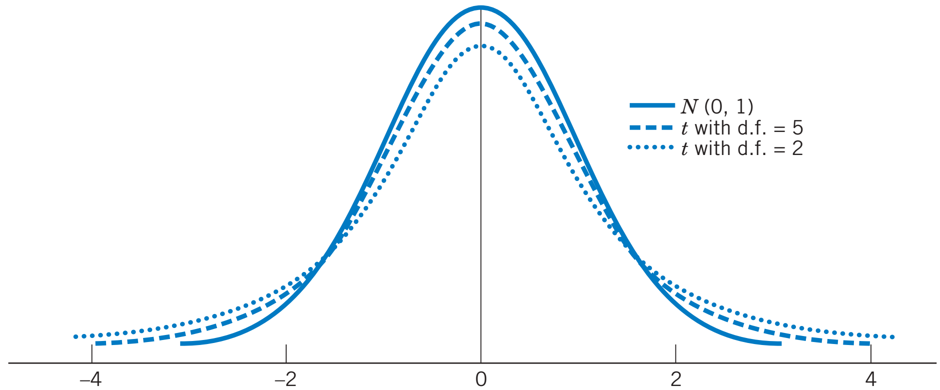


Figure 1 Comparison of $N(0, 1)$ and t density curves.

Example 1 Obtaining Percentage Points of t Distributions

Using Appendix B, Table 4, determine the upper .10 point of the t distribution with 5 degrees of freedom. Also find the lower .10 point.

SOLUTION With d.f. $= 5$, the upper .10 point of the t distribution is found from Appendix B, Table 4, to be $t_{.10} = 1.476$. Since the curve is symmetric about 0, the lower .10 point is simply $-t_{.10} = -1.476$. See Figure 2.

Percentage Points of t Distributions

d.f. \ α	$\cdots$	.10
.	.	.
.	.	.
.	.	.
5	$\cdots$	1.476

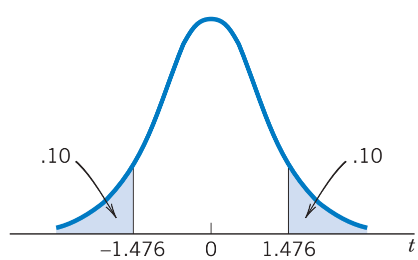


Figure 2 The upper and lower .10 points of the t distribution with d.f. $= 5$.

Example 2 Determining a Central Interval Having Probability .90

For the t distribution with d.f. $= 9$, find the number b such that $P[-b < T < b] = .90$.

SOLUTION In order for the probability in the interval $(-b, b)$ to be .90, we must have a probability of .05 to the right of b and, correspondingly, a probability of .05 to the left of $-b$ (see Figure 3). Thus, the number b is the upper $\alpha = .05$ point of the t distribution. Reading Appendix B, Table 4, at $\alpha = .05$ and d.f. $= 9$, we find $t_{.05} = 1.833$, so $b = 1.833$.

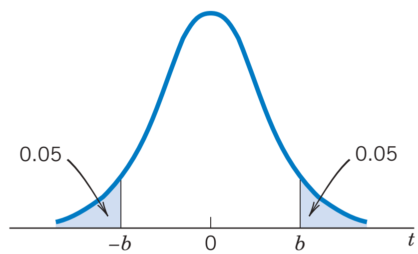


Figure 3 Finding the 5th percentile.

Exercises

9.1 Using the table for the t distributions, find:

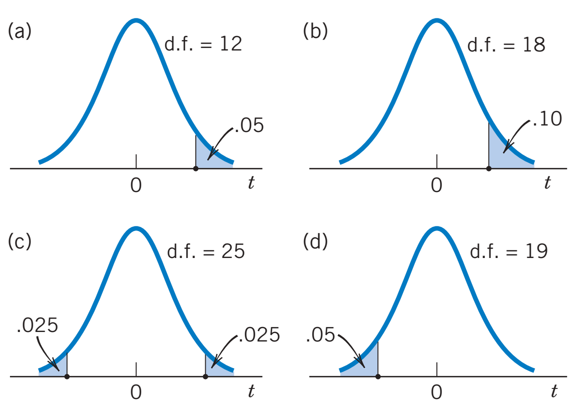
(a) The upper .05 point when d.f. $= 5$.

(b) The lower .025 point when d.f. $= 13$.

(c) The lower .01 point when d.f. $= 8$.

(d) The upper .10 point when d.f. $= 11$.

9.2 Name the t percentiles shown and find their values from Appendix B, Table 4.



9.3 Using the table for the t distributions find:

(a) The 90th percentile of the t distribution when d.f. $= 9$.

(b) The 99th percentile of the t distribution when d.f. $= 4$.

(c) The 5th percentile of the t distribution when d.f. $= 22$.

(d) The lower and upper quartiles of the t distribution when d.f. $= 18$.

9.4 Find the probability of

(a) $T < -1.761$ when d.f. $= 14$

(b) $|T| > 2.306$ when d.f. $= 8$

(c) $-1.734 < T < 1.734$ when d.f. $= 18$

(d) $-1.812 < T < 2.764$ when d.f. $= 10$

9.5 In each case, find the number b so that

(a) $P[T < b] = .95$ when d.f. $= 5$

(b) $P[-b < T < b] = .95$ when d.f. $= 16$

(c) $P[T > b] = .01$ when d.f. $= 7$

(d) $P[T > b] = .99$ when d.f. $= 12$

9.6 Record the $t_{.05}$ values for d.f. of 5, 10, 15, 20, and 29. Does this percentile increase or decrease with increasing degrees of freedom?

9.7 Using the table for the t distributions, make an assessment for the probability of the stated event. (The answer to part (a) is provided.)

(a) $T > 2.6$ when d.f. $= 7$ (Answer: $P[T > 2.6]$ is between .01 and .025 because 2.6 lies between $t_{.05} = 2.365$ and $t_{.01} = 2.998$.)

(b) $T > 1.9$ when d.f. $= 16$

(c) $T < -1.5$ when d.f. $= 11$

(d) $|T| > 1.9$ when d.f. $= 10$

(e) $|T| < 2.8$ when d.f. $= 17$

9.8 What can you say about the number c in each case? Justify your answer. (The answer to part (a) is provided.)

(a) $P[T > c] = .03$ when d.f. $= 5$

(Answer: c is between 2.015 and 2.571 because $t_{.05} = 2.015$ and $t_{.025} = 2.571$.)

(b) $P[T > c] = .016$ when d.f. $= 11$

(c) $P[T < -c] = .004$ when d.f. $= 13$

(d) $P[|T| > c] = .03$ when d.f. $= 6$

(e) $P[|T| < c] = .96$ when d.f. $= 27$

3. INFERENCES ABOUT μ—SMALL SAMPLE SIZE

3.1. CONFIDENCE INTERVAL FOR μ

The distribution of

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

provides the key for determining a **confidence interval for μ,** the mean of a normal population. For a $100(1 - \alpha)\%$ confidence interval, we consult the t table (Appendix B, Table 4) and find $t_{\alpha/2}$, the upper $\alpha/2$ point of the t distribution with $n - 1$ degrees of freedom (see Figure 4).

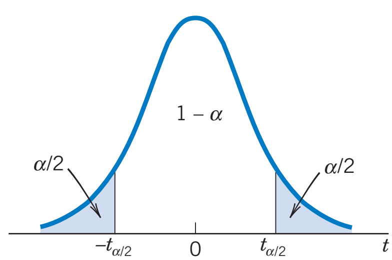


Figure 4 $t_{\alpha/2}$ and the probabilities.

Since $\dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ has the t distribution with d.f. $= n - 1$, we have

$$1 - \alpha = P\left[-t_{\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right] = P\left[-t_{\alpha/2}\frac{S}{\sqrt{n}} < \overline{X} - \mu < t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

In order to obtain a confidence interval, let us rearrange the terms inside the second set of brackets so that only the parameter μ remains in the center. The above probability statement then becomes

$$P\left[\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

which is precisely in the form required for a confidence statement about μ. The probability is $1 - \alpha$ that the random interval $\overline{X} - t_{\alpha/2}\,S/\sqrt{n}$ to $\overline{X} + t_{\alpha/2}\,S/\sqrt{n}$ will cover the true population mean μ. This argument is virtually the same as in Section 3 of Chapter 8. Only now the unknown σ is replaced by the sample standard deviation S, and the t percentage point is used instead of the standard normal percentage point.

A 100(1 − α)% Confidence Interval for a Normal Population Mean

$$\left(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \quad \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

where $t_{\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with d.f. $= n - 1$.

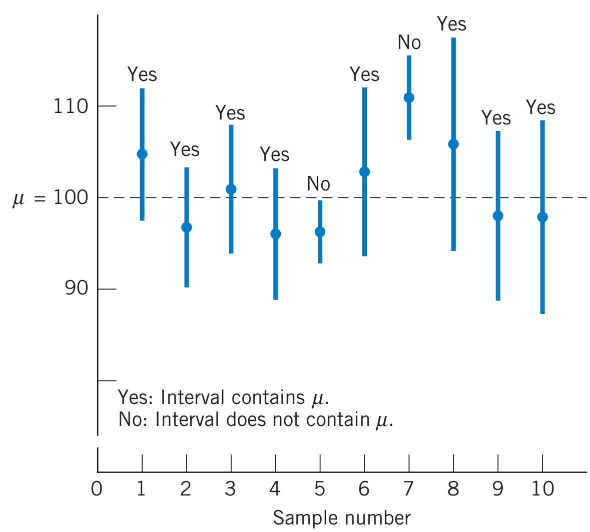


Figure 5 Behavior of confidence intervals based on the t distribution.

Let us review the meaning of a confidence interval in the present context. Imagine that random samples of size n are repeatedly drawn from a normal population and the interval $(\bar{x} - t_{\alpha/2}\, s/\sqrt{n},\ \ \bar{x} + t_{\alpha/2}\, s/\sqrt{n})$ calculated in each case. The interval is centered at $\bar{x}$ so the center varies from sample to sample. The length of an interval, $2t_{\alpha/2}\, s/\sqrt{n}$, also varies from sample to sample because it is a multiple of the sample standard deviation s. (This is unlike the fixed-length situation illustrated in Figure 4 of Chapter 8, which was concerned with a known σ.) Thus, in repeated sampling, the intervals have variable centers and variable lengths. However, our confidence statement means that if the sampling is repeated many times, about $100(1 - \alpha)\%$ of the resulting intervals would cover the true population mean μ.

Figure 5 shows the results of drawing 10 samples of size $n = 7$ from the normal population with $\mu = 100$ and $\sigma = 10$. Selecting $\alpha = .05$, we find that the value of $t_{.025}$ with 6 d.f. is 2.447, so the 95% confidence interval is $\bar{X} \pm 2.447\, S/\sqrt{7}$. In the first sample, $\bar{x} = 103.88$ and $s = 7.96$, so the interval is $(96.52, 111.24)$. The 95% confidence intervals are shown by the vertical line segments. Nine of 10 cover μ.

Example 3 Interpreting a Confidence Interval

The weights (pounds) of $n = 8$ female wolves captured in the Yukon-Charley Rivers National Reserve (see Table D.9 of the Data Bank) are

57 84 90 71 77 68 73 71

Treating these weights as a random sample from a normal distribution:

(a) Find a 90% confidence interval for the population mean weight of all female wolves living on the Yukon-Charley Rivers National Reserve.

(b) Is μ included in this interval?

SOLUTION (a) If we assume that the weights are normally distributed, a 90% confidence interval for the mean weight μ is given by

$$\left(\bar{X} - t_{.05} \frac{S}{\sqrt{n}}, \qquad \bar{X} - t_{.05} \frac{S}{\sqrt{n}} \right)$$

where $n = 8$. The t statistic is based on $n - 1 = 8 - 1 = 7$ degrees of freedom so, consulting the t table, we find $t_{.05} = 1.895$. Beforehand, this is a random interval.

To determine the confidence interval from the given sample, we first obtain the summary statistics

$$\bar{x} = \frac{591.000}{8} = 73.875 \qquad s^2 = \frac{708.875}{7} = 101.268 \qquad \text{so}$$

$$s = \sqrt{101.268} = 10.063$$

The 90% confidence interval for μ is then

$$73.875 \pm 1.895 \times \frac{10.063}{\sqrt{8}} = 73.875 \pm 6.742 \qquad \text{or } (67.13, 80.62)$$

We are 90% confident that the mean weight of all female wolves is between 67.13 and 80.62 pounds. We have this confidence because, over many occasions of sampling, approximately 90% of the intervals calculated using this procedure will contain the true mean.

(b) We will never know if a single realization of the confidence interval, such as (67.13, 80.62), covers the unknown μ. It is unknown because it is based on every female wolf in the very large reserve. Our confidence in the method is based on the high percentage of times that μ is covered by intervals in repeated samplings.

Remark: When repeated independent measurements are made on the same material and any variation in the measurements is basically due to experimental error (possibly compounded by nonhomogeneous materials), the normal model is often found to be appropriate. It is still necessary to graph the individual data points (too few here) in a dot diagram and normal-scores plot to reveal any wild observations or serious departures from normality. In all small sample situations, it is important to remember that the validity of a confidence interval rests on the reasonableness of the model assumed for the population.

Recall from the previous chapter that the length of a $100(1 - \alpha)\%$ confidence interval for a normal μ is $2\, z_{\alpha/2}\, \sigma/\sqrt{n}$ when σ is known, whereas it is $2\, t_{\alpha/2}\, S/\sqrt{n}$ when σ is unknown. Given a small sample size n and consequently a small number of degrees of freedom $(n - 1)$, the extra variability caused by estimating σ with S makes the t percentage point $t_{\alpha/2}$ much larger than the normal percentage point $z_{\alpha/2}$. For instance, with d.f. $= 4$, we have $t_{.025} = 2.776$, which is considerably larger than $z_{.025} = 1.96$. Thus, when σ is unknown, the confidence estimation of μ based on a very small sample size is expected to produce a much less precise inference (namely, a wide confidence interval) compared to the situation when σ is known. With increasing n, σ can be more closely estimated by S and the difference between $t_{\alpha/2}$ and $z_{\alpha/2}$ tends to diminish.

3.2. HYPOTHESES TESTS FOR μ

The steps for conducting a test of hypotheses concerning a population mean were presented in the previous chapter. If the sample size is small, basically the same procedure can be followed provided it is reasonable to assume that the population distribution is normal. However, in the small sample situation, our test statistic

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

has Student's t distribution with $n - 1$ degrees of freedom.

The t table (Appendix B, Table 4) is used to determine the rejection region to **test hypotheses about μ.**

Hypotheses Tests for μ—Small Samples

To test $H_0: \mu = \mu_0$ concerning the mean of a **normal population,** the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

which has Student's t distribution with $n - 1$ degrees of freedom:

$$H_1: \mu > \mu_0 \qquad R: T \geq t_\alpha$$
$$H_1: \mu < \mu_0 \qquad R: T \leq -t_\alpha$$
$$H_1: \mu \neq \mu_0 \qquad R: |T| \geq t_{\alpha/2}$$

The test is called a **Student's *t* test** or simply a ***t* test.**

Example 4 A Student's *t* Test to Confirm the Water Is Safe

A city health department wishes to determine if the mean bacteria count per unit volume of water at a lake beach is within the safety level of 200. A researcher collected 10 water samples of unit volume and found the bacteria counts to be

175	190	205	193	184
207	204	193	196	180

Do the data strongly indicate that there is no cause for concern? Test with $\alpha = .05$.

SOLUTION Let μ denote the current (population) mean bacteria count per unit volume of water. Then, the statement "no cause for concern" translates to $\mu < 200$, and the researcher is seeking strong evidence in support of this hypothesis. So the formulation of the null and alternative hypotheses should be

$$H_0: \mu = 200 \qquad \text{versus} \qquad H_1: \mu < 200$$

Since the counts are spread over a wide range, an approximation by a continuous distribution is not unrealistic for inference about the mean. Assuming further that the measurements constitute a sample from a normal population, we employ the t test with

$$T = \frac{\overline{X} - 200}{S/\sqrt{10}} \qquad \text{d.f.} = 9$$

We test at the level of significance $\alpha = .05$. Since H_1 is left-sided, we set the rejection region $T \le -t_{.05}$. From the t table we find that $t_{.05}$ with d.f. $= 9$ is 1.833, so our rejection region is $R: T \le -1.833$ as in Figure 6. Computations from the sample data yield

$$\bar{x} = 192.7$$
$$s = 10.81$$
$$t = \frac{192.7 - 200}{10.81/\sqrt{10}} = \frac{-7.3}{3.418} = -2.14$$

Because the observed value $t = -2.14$ is smaller than -1.833, the null hypothesis is rejected at $\alpha = .05$. On the basis of the data obtained from these 10 measurements, there does seem to be strong evidence that the true mean is within the safety level.

Values of T smaller than -2.14 are more extreme evidence in favor of the alternative hypothesis. Since, with 9 degrees of freedom, $t_{.05} = 1.833$ and $t_{.025} = 2.262$, the P–value is between .05 and .025. The accompanying computer output gives

$$P\text{–value} = P[\,T \le -2.14\,] = .031 \quad (\text{see Figure } 7)$$

There is strong evidence that the mean bacteria count is within the safety level.

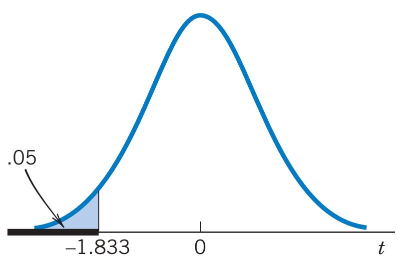


Figure 6 Left-sided rejection region $T \le -1.833$.

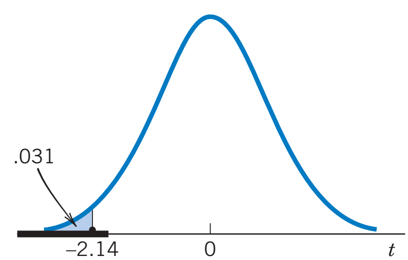


Figure 7 P–value for left-sided rejection region, T statistic.

Computer Solution to Example 4

(see the Techology section for the MINITAB COMMANDS)

T Test of the Mean

Test of mu = 200.00 vs mu < 200.00

Variable	N	Mean	StDev	SE Mean	T	P
Bac Coun	10	192.70	10.81	3.42	−2.14	0.031

Exercises

9.9 The weights from a random sample of 20 golden retriever dogs have mean 76.1 pounds and standard deviation 5.9 pounds. Assume that the weights of the dogs have a normal distribution.

(a) Construct a 98% confidence interval for the population mean.

(b) What is the length of this confidence interval? What is its center?

(c) If a 98% confidence interval were calculated from another random sample of size $n = 20$, would it have the same length as that found in part (b)? Why or why not?

9.10 Recorded here are the germination times (number of days) for seven seeds of a new strain of snap bean.

12 16 15 20 17 11 18

Stating any assumptions that you make, determine a 95% confidence interval for the true mean germination time for this strain.

9.11 A zoologist collected 20 wild lizards in the southwestern United States. The total length (mm) of each was measured.

179	157	169	146	143	131	159	142	141	130
142	116	130	140	138	137	134	114	90	114

Obtain a 95% confidence interval for the mean length.

9.12 In an investigation on toxins produced by molds that infect corn crops, a biochemist prepares extracts of the mold culture with organic solvents and then measures the amount of the toxic substance per gram of solution. From nine preparations of the mold culture, the following measurements of the toxic substance (in milligrams) are obtained:

1.2 .8 .6 1.1 1.2 .9 1.5 .9 1.0

(a) Calculate the mean $\bar{x}$ and the standard deviation s from the data.

(b) Compute a 98% confidence interval for the mean weight of toxic substance per gram of mold culture. State the assumption you make about the population.

9.13 An experimenter studying the feasibility of extracting protein from seaweed to use in animal feed makes 18 determinations of the protein extract, each based on a different 50-kilogram sample of seaweed. The sample mean and the standard deviation are found to be 3.6 and .8 kilograms, respectively. Determine a 95% confidence interval for the mean yield of protein extract per 50 kilograms of seaweed.

9.14 The monthly rent (dollars) for a two-bedroom apartment on the west side of town was recorded for a sample of ten apartments.

950 745 875 827 1030 920 840 1085 650 895

Obtain a 95% confidence interval for the mean monthly rent for two-bedroom apartments.

9.15 From a random sample of size 12, one has calculated the 95% confidence interval for μ and obtained the result (18.6, 26.2).

(a) What were the $\bar{x}$ and s for that sample?

(b) Calculate a 98% confidence interval for μ.

9.16 From a random sample of size 18, a researcher states that (12.0, 15.7) inches is a 90% confidence interval for μ, the mean length of bass caught in a small lake. A normal distribution was assumed. Using the 90% confidence interval, obtain:

(a) A point estimate of μ and its 90% margin of error.

(b) A 95% confidence interval for μ.

9.17 Henry Cavendish (1731–1810) provided direct experimental evidence of Newton's law of universal gravitation, which specifies the force of attraction between two masses. In an experiment with known masses determined by weighing, the measured force can also be used to calculate a value for the density of the earth. The values of the earth's density from Cavendish's renowned experiment in time order by column are

5.36	5.62	5.27	5.46
5.29	5.29	5.39	5.30
5.58	5.44	5.42	5.75
5.65	5.34	5.47	5.68
5.57	5.79	5.63	5.85
5.53	5.10	5.34	

[These data were published in *Philosophical Transactions*, **17** (1798), p. 469.] Find a 99% confidence interval for the density of the earth.

9.18 Refer to Exercise 9.14. Do these data support the claim that the mean monthly rent for a two-bedroom apartment differs from 775 dollars? Take $\alpha = .05$.

9.19 Refer to Exercise 9.14. Do these data provide strong evidence for the claim that the monthly rent for a two-bedroom apartment is greater than 800 dollars? Take $\alpha = .05$.

9.20 In a lake pollution study, the concentration of lead in the upper sedimentary layer of a lake bottom is measured from 25 sediment samples of 1000 cubic centimeters each. The sample mean and the standard deviation of the measurements are found to be .38 and .06, respectively. Compute a 99% confidence interval for the mean concentration of lead per 1000 cubic centimeters of sediment in the lake bottom.

9.21 The data on the lengths of anacondas on the front piece of the chapter yield a 95% confidence interval for the population mean length of all anaconda snakes in the area of the study.

```
Variable    N   Mean    StDev     95.0% CI
Length(ft) 21  12.871   2.262 ( 11.842, 13.901)
```

(a) Is the population mean length of all female anacondas living in the study area contained in this interval?

(b) Explain why you are 95% confident that it is contained in the interval.

9.22 Refer to the data on the weight (pounds) of male wolves given in Table D.9 of the Data Bank. A computer calculation gives a 95% confidence interval.

```
Variable   N    Mean    StDev    95.0% CI
Malewt    11   91.91    12.38  ( 83.59, 100.23)
```

(a) Is the population mean weight for all male wolves in the Yukon-Charley Rivers National Reserve contained in this interval?

(b) Explain why you are 95% confident that it is contained in the interval.

9.23 The following measurements of the diameters (in feet) of Indian mounds in southern Wisconsin were gathered by examining reports in *Wisconsin Archeologist* (courtesy of J. Williams).

22	24	24	30	22	20	28
30	24	34	36	15	37	

(a) Do these data substantiate the conjecture that the population mean diameter is larger than 21 feet? Test at $\alpha = .01$.

(b) Determine a 90% confidence interval for the population mean diameter of Indian mounds.

9.24 Measurements of the acidity (pH) of rain samples were recorded at 13 sites in an industrial region.

3.5	5.1	5.0	3.6	4.8	3.6	4.7
4.3	4.2	4.5	4.9	4.7	4.8	

Determine a 95% confidence interval for the mean acidity of rain in that region.

9.25 Refer to Exercise 9.11, where a zoologist collected 20 wild lizards in the southwestern United States. Do these data substantiate a claim that the mean length is greater than 128 mm? Test with $\alpha = .05$.

9.26 Geologists dating rock, using a strontium-isotope technique, provided the ages 5.2 and 4.4 million years for two specimens. Treating these as a random sample of size 2 from a normal distribution

(a) Obtain a 90% confidence interval for μ the true age of the rock formation.

(b) Geologists do get an estimate of error from the strontium-isotope analysis so they usually do not take duplicate readings. From a statistical point of view, two observations are much better than one. What happens to the length of the confidence interval if the sample size were increased from 2 to 4? Answer by considering the ratio of the lengths.

9.27 The data on the weight (lb) of female wolves, from Table D.9 of the Data Bank, are

57 84 90 71 71 77 68 73

Test the null hypothesis that the mean weight of females is 83 pounds versus a two-sided alternative. Take $\alpha = .05$.

9.28 The ability of a grocery store scanner to read accurately is measured in terms of maximum attenuation (db). In one test with 20 different products, the values of this measurement had a mean 10.7

and standard deviation 2.4. The normal assumption is reasonable.

(a) Is there strong evidence that μ, the population mean maximum attenuation for all possible products, is greater then 9.25 db?

(b) Give a 98% confidence interval for the mean maximum attenuation.

9.29 The mean drying time of a brand of spray paint is known to be 90 seconds. The research division of the company that produces this paint contemplates that adding a new chemical ingredient to the paint will accelerate the drying process. To investigate this conjecture, the paint with the chemical additive is sprayed on 15 surfaces and the drying times are recorded. The mean and the standard deviation computed from these measurements are 86 and 4.5 seconds, respectively.

(a) Do these data provide strong evidence that the mean drying time is reduced by the addition of the new chemical?

(b) Construct a 98% confidence interval for the mean drying time of the paint with the chemical additive.

(c) What did you assume about the population distribution?

9.30 A few years ago, noon bicycle traffic past a busy section of campus had a mean of $\mu = 300$. To see if any change in traffic has occurred, counts were taken for a sample of 15 weekdays. It was found that $\bar{x} = 340$ and $s = 30$.

(a) Construct an $\alpha = .05$ test of $H_0 : \mu = 300$ against the alternative that some change has occurred.

(b) Obtain a 95% confidence interval for μ.

9.31 Refer to the computer anxiety scores for female accounting students in Table D.4 of the Data Bank. A computer calculation for a test of $H_0 : \mu = 2$ versus $H_1 : \mu \neq 2$ is given below.

```
Test of mu = 2 vs mu not = 2

Variable     N    Mean    StDev
FCARS       15   2.514    0.773

Variable     95.0% CI         T      P
FCARS    ( 2.086, 2.942) 2.58 0.022
```

(a) What is the conclusion if you test with $\alpha = .05$?

(b) What mistake could you have made in part (a)?

(c) Before you collected the data, what was the probability of making the mistake in part (a)?

(d) Give a long-run relative frequency interpretation of the probability in part (c).

4. RELATIONSHIP BETWEEN TESTS AND CONFIDENCE INTERVALS

By now the careful reader should have observed a similarity between the formulas we use in testing hypotheses and in estimation by a confidence interval. To clarify the link between these two concepts, let us consider again the inferences about the mean μ of a normal population.

A $100(1 - \alpha)\%$ confidence interval for μ is

$$\left(\overline{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \qquad \overline{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right)$$

because before the sample is taken, the probability that

$$\overline{X} - t_{\alpha/2} \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}$$

is $1 - \alpha$. On the other hand, the rejection region of a level α test for $H_0 : \mu = \mu_0$ versus the two-sided alternative $H_1 : \mu \neq \mu_0$ is

$$R : \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_{\alpha/2}$$

Let us use the name "acceptance region" to mean the opposite (or complement) of the rejection region. Reversing the inequality in R, we obtain

Acceptance region

$$-t_{\alpha/2} < \frac{\overline{X} - \mu_0}{S/\sqrt{n}} < t_{\alpha/2}$$

which can also be written as

Acceptance region

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu_0 < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

The latter expression shows that any given null hypothesis μ_0 will be accepted (more precisely, will not be rejected) at level α if μ_0 lies within the $100(1 - \alpha)\%$ confidence interval. Thus, having established a $100(1 - \alpha)\%$ confidence interval for μ, we know at once that all possible null hypotheses values μ_0 lying outside this interval will be rejected at level of significance α and all those lying inside will not be rejected.

Example 5 Relation between a 95% Confidence Interval and Two-Sided $\alpha = .05$ Test

A random sample of size $n = 9$ from a normal population produced the mean $\overline{x} = 8.3$ and the standard deviation $s = 1.2$. Obtain a 95% confidence interval for μ and also test $H_0: \mu = 8.5$ versus $H_1: \mu \neq 8.5$ with $\alpha = .05$.

SOLUTION A 95% confidence interval has the form

$$\left(\overline{X} - t_{.025}\frac{S}{\sqrt{n}}, \qquad \overline{X} + t_{.025}\frac{S}{\sqrt{n}}\right)$$

where $t_{.025} = 2.306$ corresponds to $n - 1 = 8$ degrees of freedom. Here $\overline{x} = 8.3$ and $s = 1.2$, so the interval becomes

$$\left(8.3 - 2.306\frac{1.2}{\sqrt{9}}, \qquad 8.3 + 2.306\frac{1.2}{\sqrt{9}}\right) = (7.4, 9.2)$$

Turning now to the problem of testing $H_0: \mu = 8.5$, we observe that the value 8.5 lies in the 95% confidence interval we have just calculated. Using the correspondence between confidence interval and acceptance region, we can at once conclude that $H_0: \mu = 8.5$ should not be rejected at $\alpha = .05$. Alternatively, a formal step-by-step solution can be based on the test statistic

$$T = \frac{\overline{X} - 8.5}{S/\sqrt{n}}$$

The rejection region consists of both large and small values.

Rejection region

$$\left|\frac{\overline{X} - 8.5}{S/\sqrt{n}}\right| \geq t_{.025} = 2.306$$

Now the observed value $|t| = \sqrt{9}\,|8.3 - 8.5|/1.2 = .5$ does not fall in the rejection region, so the null hypothesis $H_0: \mu = 8.5$ is not rejected at $\alpha = .05$. This conclusion agrees with the one we arrived at from the confidence interval.

This relationship indicates how confidence estimation and tests of hypotheses with two-sided alternatives are really integrated in a common framework. A confidence interval statement is regarded as a more comprehensive inference procedure than testing a single null hypothesis, because a confidence interval statement in effect tests many null hypotheses at the same time.

Exercises

9.32 Based on a random sample of tail lengths for 15 male kites, an investigator calculates the 95% confidence interval (183.0, 195.0) mm based on the t distribution. The normal assumption is reasonable.

(a) What is the conclusion of the t test for $H_0 : \mu = 190.5$ versus $H_1 : \mu \neq 190.5$?

(b) What is the conclusion if $H_0 : \mu = 182.2$?

9.33 In Example 3, the 95% confidence interval for the mean weight of female wolves was found to be (67.13, 80.62) pounds.

(a) What is the conclusion of testing $H_0 : \mu = 81$ versus $H_1 : \mu \neq 81$ at level $\alpha = .05$?

(b) What is the conclusion if $H_0 : \mu = 69$?

9.34 The petal width (mm) of one kind of iris has a normal distribution. Suppose that, from a random sample of widths, the t based 90% confidence interval for the population mean width is (16.8, 19.6) mm. Answer each question "yes," "no," or "can't tell," and justify your answer. On the basis of the same sample:

(a) Would $H_0 : \mu = 20$ be rejected in favor of $H_1 : \mu \neq 20$ at $\alpha = .10$?

(b) Would $H_0 : \mu = 18$ be rejected in favor of $H_1 : \mu \neq 18$ at $\alpha = .10$?

*(c) Would $H_0 : \mu = 17$ be rejected in favor of $H_1 : \mu \neq 17$ at $\alpha = .05$?

*(d) Would $H_0 : \mu = 22$ be rejected in favor of $H_1 : \mu \neq 22$ at $\alpha = .01$?

9.35 Recorded here are the amounts of decrease in percent body fat for eight participants in an exercise program over three weeks.

1.8	10.6	−1.2	12.9	15.1	−2.0	6.2	10.8

(a) Construct a 95% confidence interval for the population mean amount μ of decrease in percent body fat over the three-week program.

(b) If you were to test $H_0 : \mu = 15$ versus $H_1 : \mu \neq 15$ at $\alpha = .05$, what would you conclude from your result in part (a)?

(c) Perform the hypothesis test indicated in part (b) and confirm your conclusion.

9.36 Refer to the data in Exercise 9.35.

(a) Construct a 90% confidence interval for μ.

(b) If you were to test $H_0 : \mu = 10$ versus $H_1 : \mu \neq 10$ at $\alpha = .10$, what would you conclude from your result in part (a)? Why?

(c) Perform the hypothesis test indicated in part (b) and confirm your conclusion.

9.37 Establish the connection between the large sample Z test, which rejects $H_0 : \mu = \mu_0$ in favor of $H_1 : \mu \neq \mu_0$, at $\alpha = .05$, if

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \geq 1.96 \qquad \text{or} \qquad \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \leq -1.96$$

and the 95% confidence interval

$$\overline{X} - 1.96\frac{S}{\sqrt{n}} \qquad \text{to} \qquad \overline{X} + 1.96\frac{S}{\sqrt{n}}$$

5. INFERENCES ABOUT THE STANDARD DEVIATION σ (THE CHI-SQUARE DISTRIBUTION)

Aside from inferences about the population mean, the population variability may also be of interest. Apart from the record of a baseball player's batting average, information on the variability of the player's performance from one game to the next may be an indicator of reliability. Uniformity is often a criterion of production quality for a manufacturing process. The quality control engineer must ensure that the variability of the measurements does not exceed a specified limit. It may also be important to ensure sufficient uniformity of the inputted raw material for trouble-free operation of the machines. In this section, we consider inferences for the standard deviation σ of a population under the assumption that the population distribution is normal. In contrast to the inference procedures concerning the population mean μ, the usefulness of the methods to be presented here is extremely limited when this assumption is violated.

To make inferences about σ^2, the natural choice of a statistic is its sample analog, which is the sample variance,

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

We take S^2 as the point estimator of σ^2 and its square root S as the **point estimator of σ.** To estimate by confidence intervals and test hypotheses, we must consider the sampling distribution of S^2. To do this, we introduce a new distribution, called the χ^2 distribution (read "chi-square distribution"), whose form depends on $n - 1$.

χ^2 Distribution

Let $X_1, \ldots, X_n$ be a random sample from a normal population $N(\mu, \sigma)$. Then the distribution of

$$\chi^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{\sigma^2} = \frac{(n - 1) S^2}{\sigma^2}$$

is called the **χ^2 distribution with $n - 1$ degrees of freedom.**

Unlike the normal or t distribution, the probability density curve of a χ^2 distribution is an asymmetric curve stretching over the positive side of the line and having a long right tail. The form of the curve depends on the value of the degrees of freedom. A typical χ^2 curve is illustrated in Figure 8.

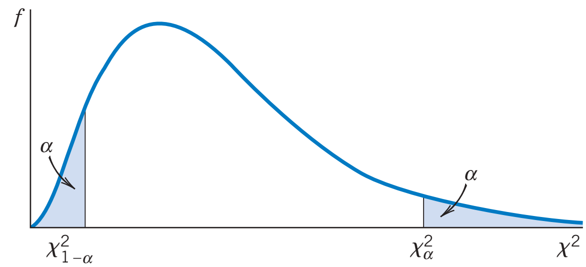


Figure 8 Probability density curve of a χ^2 distribution.

Appendix B, Table 5, provides the upper α points of χ^2 distributions for various values of α and the degrees of freedom. As in both the cases of the t and the normal distributions, the upper α point χ^2_{α} denotes the χ^2 value such that the area to the right is α. The lower α point or 100αth percentile, read from the column $\chi^2_{1-\alpha}$ in the table, has an area $1 - \alpha$ to the right. For example, the lower .05 point is obtained from the table by reading the $\chi^2_{.95}$ column, whereas the upper .05 point is obtained by reading the column $\chi^2_{.05}$.

Example 6 Finding Percentage Points of the χ^2 Distribution

Find the upper .05 point of the χ^2 distribution with 17 degrees of freedom. Also find the lower .05 point.

SOLUTION

Percentage Points of the χ^2 Distributions
(Appendix B, Table 5)

d.f. \ α	···	.95	···	.05
.		.		.
.		.		.
.		.		.
17	···	8.67	···	27.59

The upper .05 point is read from the column labeled $\alpha = .05$. We find $\chi^2_{.05} = 27.59$ for 17 d.f. The lower .05 point is read from the column $\alpha = .95$. We find $\chi^2_{.95} = 8.67$, as the lower .05 point.

The χ^2 is the basic distribution for constructing confidence intervals for σ^2 or σ. We outline the steps in terms of a 95% confidence interval for σ^2. Dividing the probability $\alpha = .05$ equally between the two tails of the χ^2 distribution and using the notation just explained, we have

$$P\left[\chi^2_{.975} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{.025}\right] = .95$$

where the percentage points are read from the χ^2 table at d.f. $= n - 1$. Because

$$\frac{(n-1)S^2}{\sigma^2} < \chi^2_{.025} \quad \text{is equivalent to} \quad \frac{(n-1)S^2}{\chi^2_{.025}} < \sigma^2$$

and

$$\chi^2_{.975} < \frac{(n-1)S^2}{\sigma^2} \quad \text{is equivalent to} \quad \sigma^2 < \frac{(n-1)S^2}{\chi^2_{.975}}$$

we have

$$P\left[\frac{(n-1)S^2}{\chi^2_{.025}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{.975}}\right] = .95$$

This last statement, concerning a random interval covering σ^2, provides a 95% confidence interval for σ^2.

A **confidence interval for σ** can be obtained by taking the square root of the endpoints of the interval. For a confidence level .95, the interval for σ becomes

$$\left(S\sqrt{\frac{n-1}{\chi^2_{.025}}}, \quad S\sqrt{\frac{n-1}{\chi^2_{.975}}}\right)$$

Example 7 A Confidence Interval for Standard Deviation of Hours of Sleep

Refer to the data on hours of sleep in Example 5, Table 4, of Chapter 2. Fifty-nine students reported the hours of sleep the previous night. Obtain 90% confidence intervals for σ^2 and σ, using the summary statistics

$$n = 59 \qquad \bar{x} = 7.181 \qquad s = 1.282$$

SOLUTION Here $n = 59$, so d.f. $= n - 1 = 58$. A computer calculation, or interpolation in the χ^2 table, gives $\chi^2_{.95} = 41.49$ and $\chi^2_{.05} = 76.78$. Using the preceding probability statements, we determine that a 90% confidence interval for the variance σ^2 is

$$\left(\frac{58 \times (1.282)^2}{76.78}, \ \frac{58 \times (1.282)^2}{41.49}\right) = (1.242, 2.298)$$

and the corresponding interval for σ is $(\sqrt{1.242}, \sqrt{2.297}) = (1.11, 1.52)$ hours. We are 90% confident that the standard deviation of number of hours slept is between 1.11 and 1.52 hours. We have this confidence because 90% of the intervals calculated by this procedure in repeated samples will cover the true σ.

You may verify that the pattern of a normal-scores plot does not contradict the assumption of normality.

It is instructive to note that the midpoint of the confidence interval for σ^2 in Example 7 is not $s^2 = 1.644 = (1.282)^2$ which is the best point estimate. This is in sharp contrast to the confidence intervals for μ, and it serves to accent the difference in logic between interval and point estimation.

For a test of the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ it is natural to employ the statistic S^2. If the alternative hypothesis is one-sided, say $H_1 : \sigma^2 > \sigma_0^2$, then the rejection region should consist of large values of S^2 or alternatively large values of the

$$\text{Test statistic} \qquad \chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \qquad \text{d.f.} = n - 1$$

To **test hypothesis about** σ, the rejection region of a level α test is,

$$R: \frac{(n-1)S^2}{\sigma_0^2} \geq \chi_\alpha^2 \qquad \text{d.f.} = n - 1$$

For a two-sided alternative $H_1 : \sigma^2 \neq \sigma_0^2$, a level α rejection region is

$$R: \frac{(n-1)S^2}{\sigma_0^2} \leq \chi_{1-\alpha/2}^2 \qquad \text{or} \qquad \frac{(n-1)S^2}{\sigma_0^2} \geq \chi_{\alpha/2}^2$$

Once again, we remind the reader that the inference procedures for σ presented in this section are extremely sensitive to departures from a normal population.

Example 8 Testing the Standard Deviation of Green Gasoline

One company, actively pursuing the making of green gasoline, starts with biomass in the form of sucrose and converts it into gasoline using catalytic reactions. At one step in a pilot plant process, the output includes carbon chains of length 3. Fifteen runs with same catalyst produced the product volumes (liter)

2.79	2.88	2.09	2.32	3.51	3.31	3.17	3.62
2.79	3.94	2.34	3.62	3.22	2.80	2.70	

While mean product volume is the prime parameter, it is also important to control variation. Conduct a test with intent of showing that the population standard deviation σ is less than .8 liter. Use $\alpha = .05$.

SOLUTION We first calculate the summary statistics $\bar{x} = 3.007$ and $s^2 = .2889$ so $s = .5374$. The claim becomes the alternative hypothesis. Since $\sigma < .8$ is equivalent to $\sigma^2 < .64 = .8^2$, we actually test

$$H_0 : \sigma^2 = .64 \qquad \text{versus} \qquad H_1 : \sigma^2 < .64$$

with $\alpha = .05$.

Because $n = 15$, the d.f. $= 14$. Also, $1 - \alpha = 1 - .05 = .95$. Consulting the χ^2 table with 14 d. f., we find $\chi_{.95}^2 = 6.57$ so the rejection region is

$$R: \frac{(n-1)S^2}{\sigma_0^2} \leq \chi_{.95}^2 = 6.57$$

The observed value of the test statistic is

$$\frac{(n-1)S^2}{\sigma_0^2} = \frac{14 \times .2889}{.64} = 6.320$$

Consequently, at significance level .05, we reject the null hypothesis and conclude that the population variance is less than .64. Equivalently, the population standard deviation is less than .80 liter.

Exercises

9.38 Using the table for the χ^2 distributions, find:

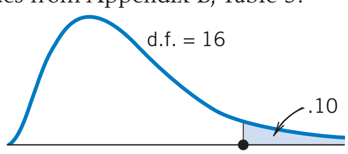
(a) The upper 5% point when d.f. = 7.

(b) The upper 1% point when d.f. = 15.

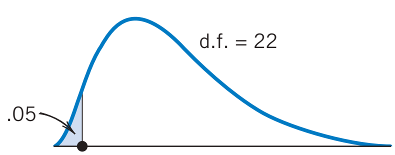
(c) The lower 2.5% point when d.f. = 9.

(d) The lower 1% point when d.f. = 24.

9.39 Name the χ^2 percentiles shown and find their values from Appendix B, Table 5.



(a)



(b)

(c) Find the percentile in part (a) if d.f. = 40.

(d) Find the percentile in part (b) if d.f. = 8.

9.40 Find the probability of

(a) $\chi^2 > 31.33$ when d.f. = 18

(b) $\chi^2 < 1.15$ when d.f. = 5

(c) $3.24 < \chi^2 < 18.31$ when d.f. = 10

(d) $3.49 < \chi^2 < 20.09$ when d.f. = 8

9.41 Beginning students in accounting took a test and a computer anxiety score (CARS) was assigned to each student on the basis of their answers to nineteen questions on the test (see Table D.4 of the Data Bank). A small population standard deviation would indicate that computer anxiety is nearly the same for all female beginning accounting students.

The scores for 15 female students are

2.90	1.00	1.90	2.37	3.32	3.79	3.26	1.90
1.84	2.58	1.58	2.90	2.42	3.42	2.53	

Assuming that the distribution of scores can be modified as a normal distribution of all college students.

(a) Obtain a point estimate of the population standard deviation σ.

(b) Construct a 95% confidence interval for σ.

(c) Examine whether or not your point estimate is located at the center of the confidence interval.

9.42 Find a 90% confidence interval for σ based on the $n = 40$ measurements of heights of red pine seedlings given in Exercise 8.4. State any assumption you make about the population. (Note: $s = .475$ for this data set.)

9.43 Refer to Exercise 9.42. A related species has population standard deviation $\sigma = .6$. Do the data provide strong evidence that the red pine population standard deviation is smaller than .6? Test with $\alpha = .05$.

9.44 Plastic sheets produced by a machine are periodically monitored for possible fluctuations in thickness. Uncontrollable heterogeneity in the viscosity of the liquid mold makes some variation in thickness measurements unavoidable. However, if the true standard deviation of thickness exceeds 1.5 millimeters, there is cause to be concerned about the product quality. Thickness measurements (in millimeters) of 10 specimens produced on a particular shift resulted in the following data.

226	228	226	225	232
228	227	229	225	230

Do the data substantiate the suspicion that the process variability exceeded the stated level on this particular shift? (Test at $\alpha = .05$.) State the assumption you make about the population distribution.

9.45 Refer to Exercise 9.44. Construct a 95% confidence interval for the true standard deviation of the thickness of sheets produced on this shift.

9.46 During manufacture, the thickness of laser printer paper is monitored. Data from several random samples each day during the year suggest that thickness follows a normal distribution. A sample of $n = 10$ thickness measurements (ten-thousandths inch) yields the 95% confidence interval (4.33, 11.50) for σ^2.

(a) What was the standard deviation s for the sample? (*Hint:* Examine how s enters the formula of a confidence interval.)

(b) Calculate a 90% confidence interval for σ.

9.47 Refer to the data of lizard lengths in Exercise 9.11.

(a) Determine a 90% confidence interval for the population standard deviation σ.

(b) Should $H_0 : \sigma = 9$ be rejected in favor of $H_1 : \sigma \neq 9$ at $\alpha = .10$? [Answer by using your result in part (a).]

9.48 Referring to the data in Exercise 9.17, determine a 99% confidence interval for the population standard deviation of the density measurements.

9.49 Referring to Exercise 9.23, construct a 95% confidence interval for the population standard deviation of the diameters of Indian mounds.

9.50 Do the data in Exercise 9.24 substantiate the conjecture that the true standard deviation of the acidity measurements is larger than 0.4? Test at $\alpha = .05$.

6. ROBUSTNESS OF INFERENCE PROCEDURES

The small sample methods for both confidence interval estimation and hypothesis testing presuppose that the sample is obtained from a normal population. Users of these methods would naturally ask:

1. What method can be used to determine if the population distribution is nearly normal?
2. What can go wrong if the population distribution is nonnormal?
3. What procedures should be used if it is nonnormal?
4. If the observations are not independent, is this serious?

1. To answer the first question, we could construct the dot diagram or normal-scores plot. These may indicate a wild observation or a striking departure from normality. If none of these is visible, the investigator would feel more secure using the preceding inference procedures. However, any plot based on a small sample cannot provide convincing justification for normality. Lacking sufficient observations to justify or refute the normal assumption, we are led to a consideration of the second question.

2. Confidence intervals and tests of hypotheses concerning μ are based on Student's t distribution. If the population is nonnormal, the actual percentage points may differ substantially from the tabulated values. When we say that $\overline{X} \pm t_{.025} S/\sqrt{n}$ is a 95% confidence interval for μ, the true probability that this random interval will contain μ may be, say, 85% or 99%. Fortunately, the effects

on inferences about μ using the t statistic are not too serious if the sample size is at least moderately large (say, 15). In larger samples, such disturbances tend to disappear due to the central limit theorem. We express this fact by saying that **inferences about μ using the t statistic are reasonably "robust."** However, this qualitative discussion should not be considered a blanket endorsement for t. When the sample size is small, a wild observation or a distribution with long tails can produce misleading results.

Unfortunately, inferences about σ using the χ^2 distribution may be seriously affected by nonnormality even with large samples. We express this by saying that **inferences about σ using the χ^2 distribution are not "robust" against departures of the population distribution from normality.**

3. We cannot give a specific answer to the third question without knowing something about the nature of nonnormality. Dot diagrams or histograms of the original data may suggest some transformations that will bring the shape of the distribution closer to normality. If it is possible to obtain a transformation that leads to reasonably normal data plots, the problem can then be recast in terms of the transformed data. Otherwise, users can benefit from consulting with a statistician.

4. A basic assumption throughout Chapters 8 and 9 is that the sample is drawn at random, so the observations are independent of one another. If the sampling is made in such a manner that the observations are dependent, however, all the inferential procedures we discussed here for small as well as large samples may be seriously in error. This applies to both the level of significance of a test and a stated confidence level. Concerned with the possible effect of a drug on the blood pressure, suppose that an investigator includes 5 patients in an experiment and makes 4 successive measurements on each. This does *not* yield a random sample of size $5 \times 4 = 20$, because the 4 measurements made on each person are likely to be dependent. This type of sampling requires a more sophisticated method of analysis. An investigator who is sampling opinions about a political issue may choose 100 families at random and record the opinions of both the husband and wife in each family. This also does *not* provide a random sample of size $100 \times 2 = 200$, although it may be a convenient sampling method. When measurements are made closely together in time or distance, there is a danger of losing independence because adjacent observations are more likely to be similar than observations that are made farther apart. Because independence is the most crucial assumption, we must be constantly alert to detect such violations. Prior to a formal analysis of the data, a close scrutiny of the sampling process is imperative.

USING STATISTICS WISELY

1. The inferences in this chapter, based on the t distribution, require that the distribution of the individual observations be normal. With very small sample sizes, say 10 or smaller, it is not possible to check this assumption. The best we can do is make dot diagrams, or normal-scores plots, to make sure there are no obvious outliers. With somewhat larger

sample sizes serious asymmetry can be recognized . Recall, however, the central limit result that the distribution of $\overline{X}$ becomes more nearly normal with increasing sample size whatever the population. This tends to make inferences based on the t statistic relatively insensitive to small or moderate departures from a normal population as long as sample size is greater than about 15.

2. Understand the interpretation of a $100(1 - \alpha)\%$ confidence interval. For normal populations, before the data are collected,

$$\left(\overline{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \qquad \overline{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right)$$

is a random interval that will cover the fixed unknown mean μ with probability $1 - \alpha$. Once a numerical value for the interval is obtained, the interval is fixed and we say we have $100(1 - \alpha)\%$ confidence that the mean is contained in the interval. After many applications of this procedure, to different samples from independent experiments, approximately proportion $1 - \alpha$ of the intervals will cover the respective population mean.

3. Formulate the assertion that the experiment seeks to confirm as the alternative hypothesis. Then, base a test of the null hypothesis $H_0 : \mu = \mu_0$ on the test statistic

$$\frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

which has a t distribution with $n - 1$ degrees of freedom. The rejection region is one-sided or two-sided corresponding to the alternative hypothesis.

KEY IDEAS AND FORMULAS

When the sample size is small, additional conditions need to be imposed on the population. In this chapter, we **assume that the population distribution is normal**.

Inferences about the mean of a normal population are based on

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

which has **Student's t distribution with $n - 1$ degrees of freedom.**

Inferences about the standard deviation of a normal population are based on $(n - 1)S^2/\sigma^2$, which has a **χ^2 distribution with $n - 1$ degrees of freedom.**

Moderate departures from a normal population distribution do not seriously affect inferences based on Student's t. These procedures are **"robust."**

Nonnormality can seriously affect inferences about σ.

Inferences about a Normal Population Mean

When n is small, we assume that the population is approximately normal. Inference procedures are derived from Student's t sampling distribution of

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

1. A $100(1 - \alpha)\%$ **confidence interval for μ** is

$$\left(\overline{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \quad \overline{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right)$$

2. **To test hypotheses about μ,** using **Student's t test,** the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

Given a level of significance α, the ***t test*** will:

Reject $H_0: \mu = \mu_0$ in favor of $H_1: \mu > \mu_0$ if $T \geq t_\alpha$
Reject $H_0: \mu = \mu_0$ in favor of $H_1: \mu < \mu_0$ if $T \leq -t_\alpha$
Reject $H_0: \mu = \mu_0$ in favor of $H_1: \mu \neq \mu_0$ if $|T| \geq t_{\alpha/2}$

Inferences about a Normal Population Standard Deviation

Inferences are derived from the χ^2 distribution for $(n - 1) S^2 / \sigma^2$.

1. A **point estimator of σ** is the sample standard deviation S.
2. A 95% **confidence interval for σ** is

$$\left(S \sqrt{\frac{n - 1}{\chi^2_{.025}}}, \quad S \sqrt{\frac{n - 1}{\chi^2_{.975}}} \right)$$

3. To **test hypotheses about** σ, the test statistic is

$$\frac{(n - 1) S^2}{\sigma_0^2}$$

Given a level of significance α,

$$\left.\begin{array}{r} \text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma < \sigma_0 \end{array}\right\} \text{ if } \quad \frac{(n - 1) S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha}$$

$$\left.\begin{array}{r} \text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma > \sigma_0 \end{array}\right\} \text{ if } \quad \frac{(n - 1) S^2}{\sigma_0^2} \geq \chi^2_{\alpha}$$

$$\left.\begin{array}{r} \text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma \neq \sigma_0 \end{array}\right\} \begin{array}{l} \text{if } \quad \dfrac{(n - 1) S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha/2} \\ \text{or } \quad \dfrac{(n - 1) S^2}{\sigma_0^2} \geq \chi^2_{\alpha/2} \end{array}$$

TECHNOLOGY

Confidence intervals and tests concerning a normal mean

MINITAB

Confidence intervals for μ

We illustrate the calculation of a 99% confidence interval for μ based on the t distribution.

Data: C1

Stat > Basic Statistics > 1-Sample t.
Type *C1* in **Samples in Columns.**
Click **Options** and type *99* in **Confidence level.**
Click **OK.** Click **OK.**

Tests of Hypotheses Concerning μ

We illustrate the calculation of an $\alpha = .01$ level test of $H_0 : \mu = 32$ versus a one-sided alternative, $H_1 : \mu > 32$.

Data: *C1*

Stat > Basic Statistics > 1-Sample t.
Type *C1* in **Samples in.**
Following **Test mean,** type *32*, the value of the mean under the null hypothesis. Click **Options** and type *99* in **Confidence level.**
In the **Alternative** cell select *greater than*, the direction of the alternative hypothesis. Click **OK.** Click **OK.**

If the sample size, mean, and standard deviation are available, instead of the second step, type these values in the corresponding cells.

EXCEL

First, obtain the summary data: sample size, $\bar{x}$, and s, as described in Chapter 2 Technology. Then substitute these values into the formulas for confidence intervals or tests.

Note that $t_{\alpha/2}$ can be obtained. For $\alpha = 0.05$ and 6 degrees of freedom:

Select **Insert** and then **Function.** Choose **Statistical** and then **TINV.**
Enter 0.025 in **Probability** and 6 in **Deg_freedom.** Click **OK.**

TI-84/-83 PLUS

Confidence intervals for μ

We illustrate the calculation of a 99% confidence interval for μ. Start with the data entered in **L1**.

Press **STAT** and select *TESTS* and then **8: Tinterval.**
Select **Data** with **List** set to **L1** and **Freq** to **1.**
Enter .99 following *C-Level*:. Select **Calculate.**
Then press **ENTER.**

If, instead, the sample size, mean, and standard deviation are available, the second step is:

Select **Stats** (instead of **Data**) and enter the sample size, mean, and standard deviation.

Tests of hypotheses concerning μ

We illustrate the calculation of an $\alpha = .01$ level test of $H_0 : \mu = 32$ versus a one-sided alternative $H_1 : \mu > 32$. Start with the data entered in **L1**.

Press **STAT** and select *TESTS* and then **2: T-Test.**
Select **Data** with **List** set to **L1** and **Freq** to **1.**
Enter 32 for μ_0. Select the direction of the alternative hypothesis.
Select **Calculate.** Press **ENTER.**

The calculator will return the *P*–value.

If, instead, the sample size, mean, and standard deviation are available, the second step is:

Select **Stats** (instead of **Data**) and enter the sample size, mean, and standard deviation.

7. REVIEW EXERCISES

9.51 Using the table of percentage points for the t distributions, find

(a) $t_{.05}$ when d.f. $= 5$

(b) $t_{.025}$ when d.f. $= 10$

(c) The lower .05 point when d.f. $= 5$

(d) The lower .05 point when d.f. $= 10$

9.52 Using the table for the t distributions, find the probability of

(a) $T > 2.720$ when d.f. $= 22$

(b) $T < 3.250$ when d.f. $= 9$

(c) $|T| < 2.567$ when d.f. $= 17$

(d) $-1.383 < T < 2.262$ when d.f. $= 9$

9.53 A t distribution assigns more probability to large values than the standard normal.

(a) Find $t_{.05}$ for d.f. $= 12$ and then evaluate $P[Z > t_{.05}]$. Verify that $P[T > t_{.05}]$ is greater than $P[Z > t_{.05}]$.

(b) Examine the relation for d.f. of 5 and 20, and comment.

9.54 Measurements of the amount of suspended solids in river water on 14 Monday mornings yield $\bar{x} = 47$ and $s = 9.4$. Obtain a 95% confidence interval for the mean amount of suspended solids. State any assumption you make about the population.

9.55 Determine a 99% confidence interval for μ using the data in Exercise 9.54.

9.56 The time to blossom of 21 plants has $\bar{x} = 38.4$ days and $s = 5.1$ days. Give a 95% confidence interval for the mean time to blossom.

9.57 Refer to Example 8 concerning the product volume for green gasoline. Obtain

(a) a point estimate of μ and it 95% error margin.

(b) a 90% confidence interval for the mean.

(c) Explain why you are 90% confident that the interval in Part (b) covers the true unknown mean.

9.58 Refer to Exercise 9.54. The water quality is acceptable if the mean amount of suspended solids is less than 49. Construct an $\alpha = .05$ test to establish that the quality is acceptable.

(a) Specify H_0 and H_1.

(b) State the test statistic.

(c) What does the test conclude?

9.59 Refer to Exercise 9.56. Do these data provide strong evidence that the mean time to blossom is less than 42 days? Test with $\alpha = .01$.

(a) Formulate the null and alternative hypotheses.

(b) Determine the test statistic.

(c) Give the form of the rejection region.

(d) What is the conclusion to your test?

(e) What error could you have made in light of your decision in Part (d)?

(f) What can you say about the P-value?

9.60 An accounting firm wishes to set a standard time μ required by employees to complete a certain audit operation. Times from 18 employees yield a sample mean of 4.1 hours and a sample standard deviation of 1.6 hours. Test $H_0 : \mu = 3.5$ versus $H_1 : \mu > 3.5$ using $\alpha = .05$.

9.61 Referring to Exercise 9.60, test $H_0 : \mu = 3.5$ versus $H_1 : \mu \neq 3.5$ using $\alpha = .02$.

9.62 Refer to Example 8 concerning the yield of green gasoline. Conduct a test of hypothesis which is intended to show that the mean product volume is greater than 2.75 liters.

(a) Formulate the null and alternative hypotheses.

(b) Determine the test statistic.

(c) Give the form of the rejection region.

(d) What is the conclusion to your test? Take $\alpha = .05$.

(e) What error could you have made in light of your decision in Part (d)?

(f) What can you say about the P–value?

9.63 The supplier of a particular brand of vitamin pills claims that the average potency of these pills after a certain exposure to heat and humidity is at least 65. Before buying these pills, a distributor wants to verify the supplier's claim is valid. To this end, the distributor will choose a random sample of 9 pills from a batch and measure their potency after the specified exposure.

(a) Formulate the hypotheses about the mean potency μ.

(b) Determine the rejection region of the test with $\alpha = .05$. State any assumption you make about the population.

(c) The data are 63, 72, 64, 69, 59, 65, 66, 64, 65. Apply the test and state your conclusion.

9.64 A weight loss program advertises "LOSE 40 POUNDS IN 4 MONTHS." A random sample of $n = 25$ customers has $\bar{x} = 32$ pounds lost and $s = 12$. To contradict this claim test $H_0 : \mu = 40$ against $H_1 : \mu < 40$ with $\alpha = .05$.

9.65 A car advertisement asserts that with the new collapsible bumper system, the average body repair cost for the damages sustained in a collision impact of 10 miles per hour does not exceed \$1500. To test the validity of this claim, 5 cars are crashed into a stone barrier at an impact force of 10 miles per hour and their subsequent body repair costs are recorded. The mean and the standard deviation are found to be \$1620 and \$90, respectively. Do these data strongly contradict the advertiser's claim?

9.66 Combustion efficiency measurements were recorded for 10 home heating furnaces of a new model. The sample mean and standard deviation were found to be 73.2 and 2.74, respectively. Do these results provide strong evidence that the average efficiency of the new model is higher than 70? (Test at $\alpha = .05$. Comment also on the P–value.)

9.67 A person with asthma took measurements by blowing into a peak-flow meter on seven consecutive days.

429	425	471	422	432	444	454

(a) Obtain a 95% confidence interval for the population mean peak-flow.

(b) Conduct an $\alpha = .10$ level test of $H_0 : \mu = 453$ versus $H_1 : \mu \neq 453$.

9.68 Times to finish a sixteen ounce bottle of mayonnaise, were recorded by a sample of 11 purchasers. It is determined that $\sum x_i = 645.7$ days and $\sum (x_i - \bar{x})^2 = 198.41$

(a) Obtain a point estimate of the population mean life μ and its 90% error margin.

(b) Obtain a 95% confidence interval for the mean.

(c) Obtain a 95% confidence interval for σ.

(d) What did you assume about the population in your answers to Part (a) and Part (b)?

9.69 Using the table of percentage points of the χ^2 distributions, find:

(a) $\chi^2_{.05}$ with d.f. $= 7$.

(b) $\chi^2_{.025}$ with d.f. $= 24$.

(c) The lower .05 point with d.f. $= 7$.

(d) The lower .025 point with d.f. $= 24$.

9.70 Using the table for the χ^2 distributions, find:

(a) The 90th percentile of χ^2 when d.f. $= 10$.

(b) The 10th percentile of χ^2 when d.f. $= 8$.

(c) The median of χ^2 when d.f. $= 20$.

(d) The 1st percentile of χ^2 when d.f. $= 50$.

9.71 Refer to Exercise 9.68. Conduct a test of hypothesis with the intent of establishing that the mean bottle life is different from 55.0 days.

(a) Formulate the null and alternative hypotheses.

(b) Determine the test statistic.

(c) Give the form of the rejection region.

(d) What is the conclusion to your test? Take $\alpha = .05$.

(e) What error could you have made in light of your decision in Part (d)?

(f) What can you say about the P–value?

9.72 Refer to the data of Exercise 9.66. Is there strong evidence that the standard deviation for the efficiency of the new model is below .30?

9.73 Refer to the water quality data in Example 4. Perform a test of hypotheses with the intent of showing that the population standard deviation is less than 18.0. Take $\alpha = .05$.

9.74 Refer to the data on the head length (cm) of male grizzly bears given in Table D.6 of the Data Bank. A computer calculation for a test of $H_0 : \mu = 21$ versus $H_1 : \mu \neq 21$ is given below.

```
Test of mu = 21 vs mu not = 21

Variable    N     Mean    StDev
Mhdln      25   18.636    3.697

Variable       95.0% CI            T      P
Mhdln      ( 17.110, 20.162)   -3.20  0.004
```

(a) What is the conclusion if you test with $\alpha = .01$?

(b) What mistake could you have made in part (a)?

(c) Before you collected the data, what was the probability of making the mistake in part (a)?

(d) Give a long-run relative frequency interpretation of the probability in part (c).

9.75 Refer to the computer output concerning the head length (cm) of male grizzly bears in Exercise 9.74.

(a) Is the population mean head length for all male bears in the study area contained in this interval?

(b) Explain why you are 95% confident that it is contained in the interval.

The Following Exercises May Require a Computer

9.76 Refer to the data on the length (cm) of male grizzly bears given in Table D.6 of the Data Bank.

(a) Find a 99% confidence interval for the population mean.

(b) Is the population mean length for all male grizzly bears in Alaska contained in this interval?

(c) Explain why you are 99% confident that it is contained in the interval.

9.77 Refer to the computer anxiety scores (CARS) for males in Table D.4 of the Data Bank. Conduct an $\alpha = .025$ level of $H_0 : \mu = 2.7$ versus $H_1 : \mu > 2.7$.

9.78 Refer to the data on the body length (cm) of male wolves given in Table D.9 of the Data Bank.

(a) Find a 98% confidence interval for the population mean.

(b) Is the population mean body length for all male wolves in the Yukon-Charley Rivers National Reserve mean contained in this interval?

(c) Explain why you are 98% confident that it is contained in the interval.

10

Comparing Two Treatments

Does Playing Action Video Games Modify Ability To See Objects?

Video game playing is now a widespread activity in our society. Two scientists[1] wondered if the game playing itself could alter players' visual abilities and improve performance on related tasks. As one part of their experiment, they briefly flashed different numbers of target squares on the screen. The count of the number of target squares flashed that could be unerringly apprehended was recorded for a sample of 13 regular game players and 13 nonplayers. (Courtesy of G. Green.)

Number of Items Apprehended

	Sample Size	Mean	Standard Deviation
Game players	$n_1 = 13$	$\bar{x}_1 = 4.89$	$s_1 = 1.58$
Non–game players	$n_2 = 13$	$\bar{x}_2 = 3.27$	$s_2 = 1.15$

The difference in sample means, $4.89 - 3.27 = 1.62$ items apprehended, is substantial. A 95% confidence interval for the difference of population means is

$$(0.50, 2.74) \qquad \text{items unerringly apprehended}$$

With 95% confidence we can assert that mean difference is greater than .5 but not greater than 2.74 items unerringly apprehended. These data provide strong evidence that regular game players can unerringly apprehend more items in this new task.

[1] *Source:* G. S. Green and D. Bavelier, "Action Video Game Modifies Visual Selective Attention," *Nature* **423** (May 29, 2003), pp. 534–537.

1. INTRODUCTION

In virtually every area of human activity, new procedures are invented and existing techniques revised. Advances occur whenever a new technique proves to be better than the old. To compare them, we conduct experiments, collect data about their performance, and then draw conclusions from statistical analyses. The manner in which sample data are collected, called an **experimental design** or **sampling design,** is crucial to an investigation. In this chapter, we introduce two experimental designs that are most fundamental to a comparative study: (1) independent samples and (2) a matched pairs sample.

As we shall see, the methods of analyzing the data are quite different for these two processes of sampling. First, we outline a few illustrative situations where a comparison of two methods requires statistical analysis of data.

Example 1 Agricultural Field Trials for Pesticide Residual in Vegetables

The amount of pesticides in the food supply is a major health problem. To ascertain whether a new pesticide will result in less residue in vegetables, field trials must be performed by applying the new and current pesticides to nearly identical farm plots. For instance, the new pesticide could be applied to about one-half of the plot and the current pesticide to the rest but leaving a sufficient buffer zone in between.

Several different fields with different climate and soil conditions and several different crops should be included in the study. A conclusion about pesticide residue in lettuce grown in sandy soil in southern Wisconsin, while of some interest, is not general enough.

Example 2 Drug Evaluation

Pharmaceutical researchers strive to synthesize chemicals to improve their efficiency in curing diseases. New chemicals may result from educated guesses concerning potential biological reactions, but evaluations must be based on their effects on diseased animals or human beings. To compare the effectiveness of two drugs in controlling tumors in mice, several mice of an identical breed may be taken as experimental subjects. After infecting them with cancer cells, some will be subsequently treated with drug 1 and others with drug 2. The data of tumor sizes for the two groups will then provide a basis for comparing the drugs. When testing the drugs on human subjects, the experiment takes a different form. Artificially infecting them with cancer cells is absurd! In fact, it would be criminal. Instead, the drugs will be administered to cancer patients who are available for the study. In contrast, with a pool of mice of an "identical breed," the available subjects may be of varying conditions of general health, prognosis of the disease, and other factors.

When discussing a comparative study, the common statistical term **treatment** is used to refer to the things that are being compared. The basic unit that is exposed to one treatment or another are called an **experimental unit** or **experimental subject,** and the characteristic that is recorded after the application of a treatment to a subject is called the **response.** For instance, the two treatments in Example 1 are the two varieties of seeds, the experimental subjects are the agricultural plots, and the response is crop yield.

The term **experimental design** refers to the manner in which subjects are chosen and assigned to treatments. For comparing two treatments, the two basic types of design are:

1. **Independent samples (complete randomization).**
2. **Matched pairs sample (randomization within each matched pair).**

The case of independent samples arises when the subjects are randomly divided into two groups, one group is assigned to treatment 1 and the other to treatment 2. The response measurements for the two treatments are then unrelated because they arise from separate and unrelated groups of subjects. Consequently, each set of response measurements can be considered a sample from a population, and we can speak in terms of a comparison between two population distributions.

With the matched pairs design, the experimental subjects are chosen in pairs so that the members in each pair are alike, whereas those in different pairs may be substantially dissimilar. One member of each pair receives treatment 1 and the other treatment 2. Example 3 illustrates these ideas.

Example 3 Independent Samples Versus Matched Pairs Design

To compare the effectiveness of two drugs in curing a disease, suppose that 8 patients are included in a clinical study. Here, the time to cure is the response of interest. Figure 1*a* portrays a design of independent samples where the 8 patients are randomly split into groups of 4, one group is treated with drug 1, and the other with drug 2. The observations for drug 1 will have no relation to those for drug 2 because the selection of patients in the two groups is left completely to chance.

To conduct a matched pairs design, one would first select the patients in pairs. The two patients in each pair should be as alike as possible in regard to their physiological conditions; for instance, they should be of the same gender and age group and have about the same severity of the disease. These preexisting conditions may be different from one pair to another. Having paired the subjects, we randomly select one member from each pair to be treated with drug 1 and the other with drug 2. Figure 1*b* shows this matched pairs design.

In contrast with the situation of Figure 1*a*, here we would expect the responses of each pair to be dependent for the reason that they are governed by the same preexisting conditions of the subjects.

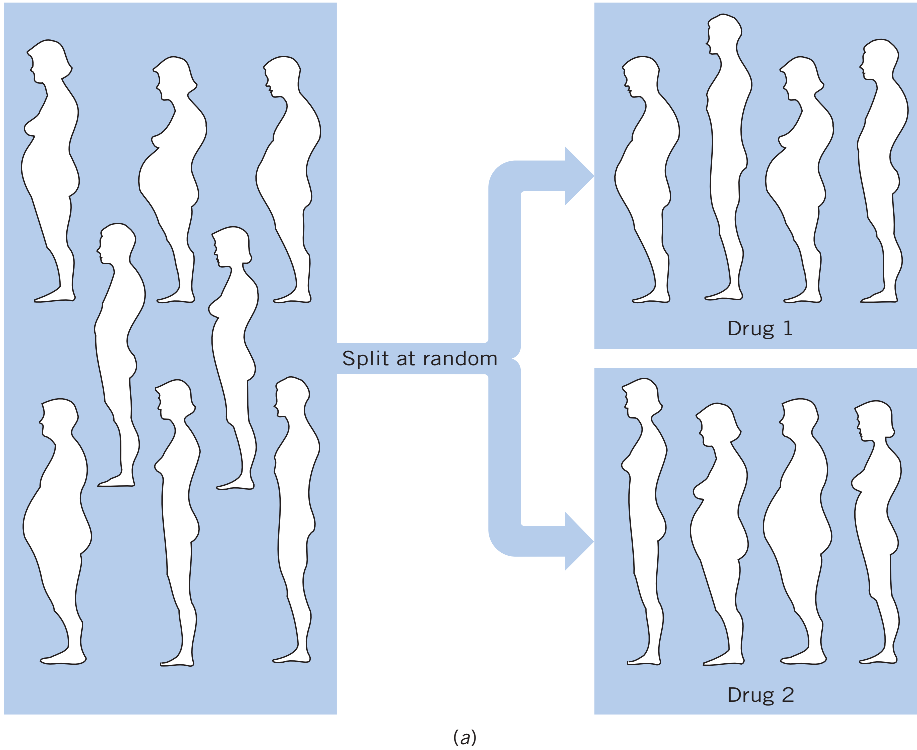


Figure 1*a* Independent samples, each of size 4.

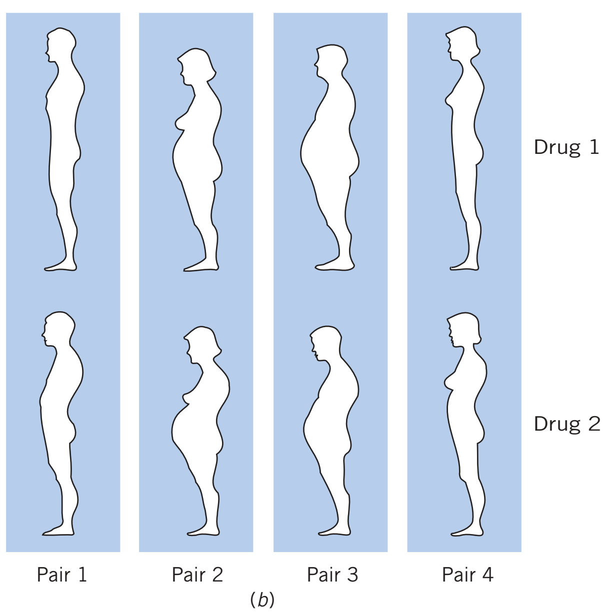


Figure 1*b* Matched pairs design with four pairs of subjects. Separate random assignment of Drug 1 each pair.

In summary, a carefully planned experimental design is crucial to a successful comparative study. The design determines the structure of the data. In turn, the design provides the key to selecting an appropriate analysis.

Exercises

10.1 Grades for first semester will be compared to those for second semester. The five one-semester courses biology, chemistry, English, history, and psychology must be taken next year. Make a list of all possible ways to split the courses into two groups where the first group has two courses to be taken the first semester and the second group has three courses to be taken the second semester.

10.2 Six mice—Alpha, Tau, Omega, Pi, Beta, and Phi—are to serve as subjects. List all possible ways to split them into two groups with the first having 4 mice and the second 2 mice.

10.3 Six students in a psychology course have volunteered to serve as subjects in a matched pairs experiment.

Name	Age	Gender
Tom	18	M
Sue	20	F
Erik	18	M
Grace	20	F
John	20	M
Roger	18	M

(a) List all possible sets of pairings if subjects are paired by age.

(b) If subjects are paired by gender, how many pairs are available for the experiment?

10.4 Identify the following as either matched pair or independent samples. Also identify the experimental units, treatments, and response in each case.

(a) Twelve persons are given a high-potency vitamin C capsule once a day. Another twelve do not take extra vitamin C. Investigators will record the number of colds in 5 winter months.

(b) One self-fertilized plant and one cross-fertilized plant are grown in each of 7 pots. Their heights will be measured after 3 months.

(c) Ten newly married couples will be interviewed. Both the husband and wife will respond to the question, "How many children would you like to have?"

(d) Learning times will be recorded for 5 dogs trained by a reward method and 3 dogs trained by a reward–punishment method.

2. INDEPENDENT RANDOM SAMPLES FROM TWO POPULATIONS

Here we discuss the methods of statistical inference for comparing two treatments or two populations on the basis of independent samples. Recall that with the **independent samples design,** a collection of $n_1 + n_2$ subjects is randomly divided into two groups and the responses are recorded. We conceptualize population 1 as the collection of responses that would result if a vast number of subjects were given treatment 1. Similarly, population 2 refers to the population of responses under treatment 2. The design of independent samples can then be viewed as one that produces unrelated random samples from two populations (see Figure 2). In other situations, the populations to be compared may be quite real entities. For instance, one may wish to compare the residential property values in the east suburb of a city to those in the west suburb. Here the issue of assigning experimental subjects to treatments does not arise. The collection of all residential properties in each suburb constitutes a population from which a sample will be drawn at random.

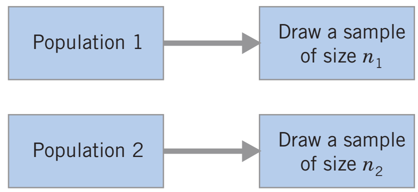


Figure 2 Independent random samples.

With the design of independent samples, we obtain

Sample	Summary Statistics	
$X_1, X_2, \ldots, X_{n_1}$ from population 1	$\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$	$S_1^2 = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2}{n_1 - 1}$
$Y_1, Y_2, \ldots, Y_{n_2}$ from population 2	$\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$	$S_2^2 = \frac{\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_2 - 1}$

To make confidence statements or to test hypotheses, we specify a statistical model for the data.

Statistical Model: Independent Random Samples

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample of size n_1 from population 1 whose mean is denoted by μ_1 and standard deviation by σ_1.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample of size n_2 from population 2 whose mean is denoted by μ_2 and standard deviation by σ_2.
3. The samples are independent. In other words, the response measurements under one treatment are unrelated to the response measurements under the other treatment.

3. LARGE SAMPLES INFERENCE ABOUT DIFFERENCE OF TWO MEANS

We now set our goal toward drawing a comparison between the mean responses of the two treatments or populations. In statistical language, we are interested in making inferences about the parameter

$$\mu_1 - \mu_2 = (\text{Mean of population 1}) - (\text{Mean of population 2})$$

When the sample sizes are large, no additional assumptions are required.

ESTIMATION

Inferences about the difference $\mu_1 - \mu_2$ are naturally based on its estimate $\overline{X} - \overline{Y}$, the difference between the sample means. When both sample sizes n_1 and n_2 are large (say, greater than 30), $\overline{X}$ and $\overline{Y}$ are each approximately normal and their difference $\overline{X} - \overline{Y}$ is approximately normal with

Mean

$$E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$$

Variance

$$\text{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Standard error

$$\text{S.E.}(\overline{X} - \overline{Y}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Note: Because the entities $\overline{X}$ and $\overline{Y}$ vary in repeated sampling and independently of each other, the distance between them becomes more variable than the individual members. This explains the mathematical fact that the variance of the difference $\overline{X} - \overline{Y}$ equals the *sum* of the variances of $\overline{X}$ and $\overline{Y}$.

When n_1 and n_2 are both large, the normal approximation remains valid if σ_1^2 and σ_2^2 are replaced by their estimators

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2}{n_1 - 1} \quad \text{and} \quad S_2^2 = \frac{\sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_2 - 1}$$

We conclude that, when the sample sizes n_1 and n_2 are large,

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \quad \text{is approximately } N(0, 1)$$

A confidence interval for $\mu_1 - \mu_2$ is constructed from this sampling distribution. As we did for the single sample problem, we obtain a confidence interval of the form

$$\text{Estimate of parameter} \quad \pm \quad (z \text{ value})\,(\text{estimated standard error})$$

Large Samples Confidence Interval for $\mu_1 - \mu_2$

When n_1 and n_2 are greater than 30, an approximate $100\,(1-\alpha)\,\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\,, \qquad \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ point of $N(0, 1)$.

Example 4 Large Samples Confidence Interval for Difference in Mean Job Satisfaction

A considerable proportion of a person's time is spent working, and satisfaction with the job and satisfaction with life tend to be related. Job satisfaction is typically measured on a four point scale

Very Dissatisfied	*A Little Dissatisfied*	*Moderately Satisfied*	*Very Satisfied*

A numerical scale is created by assigning 1 to very dissatisfied, 2 to a little dissatisfied, 3 to moderately satisfied, and 4 to very satisfied.

The responses of 226 firefighters and 247 office supervisors, presented in Exercise 10.10, yielded the summary statistics

	Fire-fighter	Office Supervisor
Mean	3.673	3.547
sd	0.7235	0.6089

Construct a 95% confidence interval for difference in mean job satisfaction scores

SOLUTION Let μ_1 be the mean job satisfaction for firefighters and μ_2 the mean for office supervisors. We have

$$\begin{array}{lll} n_1 = 226 & \overline{x} = 3.673 & s_1 = .7235 \\ n_2 = 247 & \overline{y} = 3.547 & s_2 = .6089 \end{array}$$

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(.7235)^2}{226} + \frac{(.6089)^2}{247}} = .06018$$

For a 95% confidence interval, we use $z_{.025} = 1.96$ and calculate

$$\bar{x} - \bar{y} = 3.673 - 3.547 = .126$$

$$z_{.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.96 \times .06017 = .121$$

Therefore, a 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$.126 \pm .121 \qquad \text{or} \qquad (.005,\ .247)$$

We are 95% confident that the mean score for firefighters is .005 units to .247 units higher than the mean score of office supervisors. This interval contains inconsequential positive values as well as positive values that are possibly important differences on the satisfaction scale.

HYPOTHESES TESTING

Let us turn our attention to testing hypotheses concerning $\mu_1 - \mu_2$. A test of the null hypothesis that the two population means are the same, $H_0 : \mu_1 - \mu_2 = 0$, employs the test statistic

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

which is approximately $N(0, 1)$ when $\mu_1 - \mu_2 = 0$.

Example 5 Testing Equality of Mean Job Satisfaction

Do the data in Example 4 provide strong evidence that the mean job satisfaction of firefighters is different from the mean job satisfaction of office supervisors? Test at $\alpha = .02$.

SOLUTION Because we are asked to show that the two means are different, we formulate the problem as testing

$$H_0 : \mu_1 - \mu_2 = 0 \qquad \text{versus} \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

We use the test statistic

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

and set a two-sided rejection region. Specifically, with $\alpha = .02$, we have $\alpha/2 = .01$ and $z_{\alpha/2} = 2.33$, so the rejection region is $R : |Z| \geq 2.33$.

Using the sample data given in Example 4, we calculate

$$z = \frac{3.673 - 3.547}{\sqrt{\frac{(.7325)^2}{226} + \frac{(.6089)^2}{247}}} = \frac{.126}{.06018} = 2.04$$

Because the observed value $z = 2.09$ does not lie in the rejection region, we fail to reject the null hyphothesis at level $\alpha = .02$. The evidence against equal means is only moderately strong since the $P-$value is

$$P[Z < -2.04] + P[Z > 2.04] = .0207 + .0207 = .0414 \quad \text{(see Figure 3)}$$

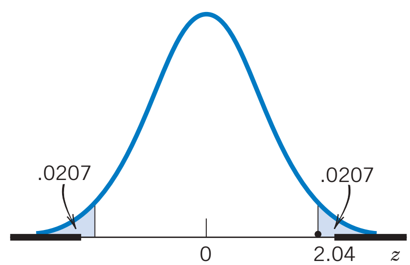


Figure 3 $P-$value with two-sided rejection region.

Example 6 Large Samples Test with a One-Sided Alternative

In June two years ago, chemical analyses were made of 85 water samples (each of unit volume) taken from various parts of a city lake, and the measurements of chlorine content were recorded. During the next two winters, the use of road salt was substantially reduced in the catchment areas of the lake. This June, 110 water samples were analyzed and their chlorine contents recorded. Calculations of the mean and the standard deviation for the two sets of data give

	Chlorine Content	
	Two Years Ago	Current Year
Mean	18.3	17.8
Standard deviation	1.2	1.8

Test the claim that lower salt usage has reduced the amount of chlorine in the lake. Base your decision on the $P-$value.

SOLUTION Let μ_1 be the population mean two years ago and μ_2 the population mean in the current year. Because the claim is that μ_2 is less than μ_1, we formulate the hypotheses

$$H_0 : \mu_1 - \mu_2 = 0 \qquad \text{versus} \qquad H_1 : \mu_1 - \mu_2 > 0$$

With the test statistic

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

the rejection region should be of the form $R: Z \geq c$ because H_1 is right-sided.

Using the data

$$\begin{array}{lll} n_1 = 85 & \bar{x} = 18.3 & s_1 = 1.2 \\ n_2 = 110 & \bar{y} = 17.8 & s_2 = 1.8 \end{array}$$

we calculate

$$z = \frac{18.3 - 17.8}{\sqrt{\dfrac{(1.2)^2}{85} + \dfrac{(1.8)^2}{110}}} = \frac{.5}{.2154} = 2.32$$

The significance probability of this observed value is (see Figure 4)

$$P\text{–value} = P[\,Z \geq 2.32\,] = .0102$$

Because the P–value is very small, we conclude that there is strong evidence to reject H_0 and support H_1.

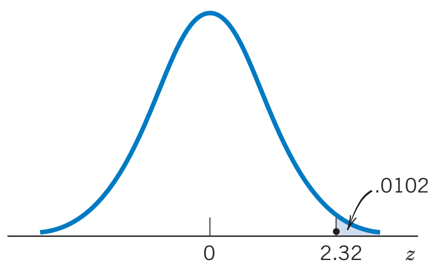


Figure 4 P–value = .0102.

We summarize the procedure for testing $\mu_1 - \mu_2 = \delta_0$ where δ_0 is specified under the null hypothesis. The case $\mu_1 = \mu_2$ corresponds to $\delta_0 = 0$.

Testing $H_0 : \mu_1 - \mu_2 = \delta_0$ with Large Samples

Test statistic:

$$Z = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

Alternative Hypothesis	Level α Rejection Region
$H_1 : \mu_1 - \mu_2 > \delta_0$	$R: Z \geq z_\alpha$
$H_1 : \mu_1 - \mu_2 < \delta_0$	$R: Z \leq -z_\alpha$
$H_1 : \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert Z \rvert \geq z_{\alpha/2}$

Finally, we would like to emphasize that with large samples we can also learn about other differences between the two populations.

Example 7 Large Samples Reveal Additional Differences between Populations

Natural resource managers have attempted to use the Satellite Landsat Multispectral Scanner data for improved landcover classification. When the satellite was flying over country known to consist of forest, the following intensities were recorded on the near-infrared band of a thermatic mapper. The sample has already been ordered.

77	77	78	78	81	81	82	82	82	82	82	83	83	84	84	84
84	85	86	86	86	86	86	87	87	87	87	87	87	87	89	89
89	89	89	89	89	90	90	90	91	91	91	91	91	91	91	91
91	91	93	93	93	93	93	93	94	94	94	94	94	94	94	94
94	94	94	94	95	95	95	95	95	96	96	96	96	96	96	97
97	97	97	97	97	97	97	97	98	99	100	100	100	100		
100	100	100	100	100	101	101	101	101	101	101	102				
102	102	102	102	102	103	103	104	104	104	105	107				

When the satellite was flying over urban areas, the intensities of reflected light on the same near-infrared band were

71	72	73	74	75	77	78	79	79	79	79	80	80	80	81	81	81
82	82	82	82	84	84	84	84	84	84	85	85	85	85	85	85	86
86	87	88	90	91	94											

If the means are different, the readings could be used to tell urban from forest area. Obtain a 95% confidence interval for the difference in mean radiance levels.

SOLUTION Computer calculations give

	Forest	Urban
Number	118	40
Mean	92.932	82.075
Standard deviation	6.9328	4.9789

and, for large sample sizes, the approximate 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\overline{X} - \overline{Y} - z_{.025} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \quad \overline{X} - \overline{Y} + z_{.025} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

Since $z_{.025} = 1.96$, the 95% confidence interval is calculated as

$$92.932 - 82.075 \pm 1.96 \sqrt{\frac{(6.9328)^2}{118} + \frac{(4.9789)^2}{40}} \quad \text{or} \quad (8.87, 12.84)$$

The mean for the forest is 8.87 to 12.84 levels of radiance higher than the mean for the urban areas.

Because the sample sizes are large, we can also learn about other differences between the two populations. The stem-and-leaf displays and boxplots in Figure 5 reveal that there is some difference in the standard deviation as well as the means. The graphs further indicate a range of high readings that are more likely to come from forests than urban areas. This feature has proven helpful in discriminating between forest and urban areas on the basis of near-infrared readings.

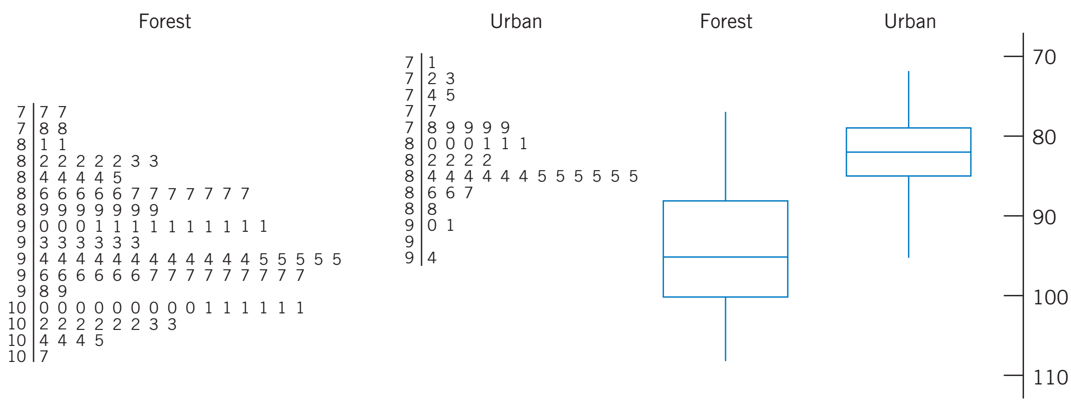


Figure 5 Stem-and-leaf displays and boxplots give additional information about population differences.

4. INFERENCES FROM SMALL SAMPLES: NORMAL POPULATIONS WITH EQUAL VARIANCES

Not surprisingly, more distributional structure is required to formulate appropriate inference procedures for small samples. Here we introduce the small samples inference procedures that are valid under the following assumptions about the population distributions. Naturally, the usefulness of such procedures depends on how closely these assumptions are realized.

Additional Assumptions When the Sample Sizes Are Small

1. Both populations are normal.
2. The population standard deviations σ_1 and σ_2 are equal.

A restriction to normal populations is not new. It was previously introduced for inferences about the mean of a single population. The second assumption, requiring equal variability of the populations, is somewhat artificial but we reserve comment until later. Letting σ denote the common standard deviation, we summarize.

Small Samples Assumptions

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample from $N(\mu_1, \sigma)$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample from $N(\mu_2, \sigma)$.
   (*Note:* σ is the same for both distributions.)
3. $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent.

ESTIMATION

Again, $\overline{X} - \overline{Y}$ is our choice for a statistic.

$$\text{Mean of } (\overline{X} - \overline{Y}) = E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$$

$$\text{Var}(\overline{X} - \overline{Y}) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

The common variance σ^2 can be estimated by combining information provided by both samples. Specifically, the sum $\sum_{i=1}^{n_1}(X_i - \overline{X})^2$ incorporates $n_1 - 1$ pieces of information about σ^2, in view of the constraint that the deviations $X_i - \overline{X}$ sum to zero. Independently of this, $\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$ contains $n_2 - 1$ pieces of information about σ^2. These two quantities can then be combined,

$$\sum(X_i - \overline{X})^2 + \sum(Y_i - \overline{Y})^2$$

to obtain a pooled estimate of the common σ^2. The proper divisor is the sum of the component degrees of freedom, or $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

Pooled Estimator of the Common σ^2

$$S^2_{\text{pooled}} = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_1 + n_2 - 2}$$

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

Example 8 Calculating the Pooled Estimate of Variance

Calculate the observed value s^2_{pooled} from these two samples.

Sample from population 1:	8	5	7	6	9	7
Sample from population 2:	2	6	4	7	6	

SOLUTION The sample means are

$$\bar{x} = \frac{\sum x_i}{6} = \frac{42}{6} = 7 \qquad \bar{y} = \frac{\sum y_i}{5} = \frac{25}{5} = 5$$

Furthermore,

$$\begin{aligned}(6-1)s_1^2 &= \sum (x_i - \bar{x})^2 \\ &= (8-7)^2 + (5-7)^2 + (7-7)^2 + (6-7)^2 \\ &\quad + (9-7)^2 + (7-7)^2 = 10 \\ (5-1)s_2^2 &= \sum (y_i - \bar{y})^2 \\ &= (2-5)^2 + (6-5)^2 + (4-5)^2 \\ &\quad + (7-5)^2 + (6-5)^2 = 16\end{aligned}$$

Thus, $s_1^2 = 2$, $s_2^2 = 4$, and the pooled variance is

$$s^2_{\text{pooled}} = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{10 + 16}{6 + 5 - 2} = 2.89$$

The pooled variance is closer to 2 than 4 because the first sample size is larger.

These arithmetic details serve to demonstrate the concept of pooling. Using a calculator with a "standard deviation" key, one can directly get the sample standard deviations $s_1 = 1.414$ and $s_2 = 2.000$. Noting that $n_1 = 6$ and $n_2 = 5$, we can then calculate the pooled variance as

$$s^2_{\text{pooled}} = \frac{5(1.414)^2 + 4(2.000)^2}{9} = 2.89$$

Employing the pooled estimator $\sqrt{S^2_{\text{pooled}}}$ for the common σ, we obtain a Student's t variable that is basic to inferences about $\mu_1 - \mu_2$.

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

has Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom.

We can now obtain confidence intervals for $\mu_1 - \mu_2$, which are of the form

$$\text{Estimate of parameter} \pm (t \text{ value}) \times (\text{Estimated standard error})$$

Confidence Interval for $\mu_1 - \mu_2$ Small Samples and $\sigma \neq \sigma_2$

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm t_{\alpha/2}\, S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$$S^2_{\text{pooled}} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

and $t_{\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with d.f. $= n_1 + n_2 - 2$.

Example 9 Calculating a Small Samples Confidence Interval

Beginning male and female accounting students were given a test and, on the basis of their answers, were assigned a computer anxiety score (CARS). Using the data given in Table D.4 of the Data Bank, obtain a 95% confidence interval for the difference in mean computer anxiety score between beginning male and female accounting students.

SOLUTION The dot diagrams of these data, plotted in Figure 6, give the appearance of approximately equal amounts of variation.

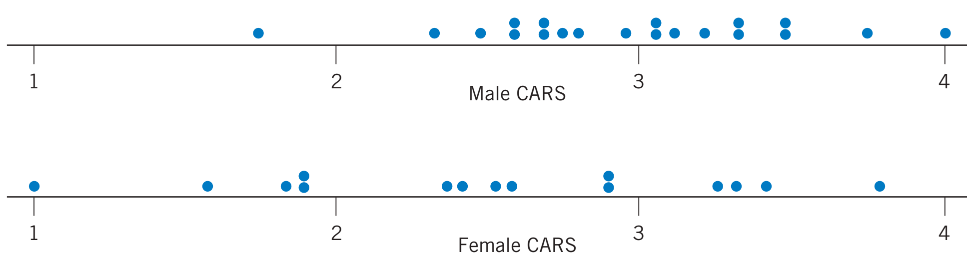


Figure 6 Dot diagrams of the computer anxiety data in Example 8.

We assume that the CARS data for both females and males are random samples from normal populations with means μ_1 and μ_2, respectively, and a common standard deviation σ. Computations from the data provide the summary statistics:

Female CARS

$$n_1 = 15 \qquad \bar{x} = 2.514 \qquad s_1 = .773$$

Male CARS

$$n_2 = 20 \qquad \bar{y} = 2.963 \qquad s_2 = .525$$

We calculate

$$s_{\text{pooled}} = \sqrt{\frac{14(.773)^2 + 19(.525)^2}{15 + 20 - 2}} = .642$$

With a 95% confidence interval $\alpha/2 = .025$ and consulting the t table, we find (interpolating) that $t_{.025} = 2.035$ for d.f. $= n_1 + n_2 - 2 = 33$. Thus a 95% confidence interval for $\mu_1 - \mu_2$ becomes

$$\bar{x} - \bar{y} \pm t_{0.25}\, s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 2.514 - 2.963 \pm 2.035 \times .642 \sqrt{\frac{1}{15} + \frac{1}{20}}$$

$$= -.449 \pm .446 \qquad \text{or} \qquad (-.895, \; -.003)$$

We can be 95% confident that the mean computer anxiety score for female beginning accounting students can be .003 to .895 units lower than the mean score for males.

This interval is quite wide. Certainly the very small values represent technically insignificant differences.

TESTS OF HYPOTHESES

Tests of hypotheses concerning the difference in means are based on a statistic having student's t distribution.

Testing $H_0: \mu_1 - \mu_2 = \delta_0$ with Small Samples and $\sigma_1 = \sigma_2$

Test statistic:

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{S_{\text{pooled}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{d.f.} = n_1 + n_2 - 2$$

Alternative Hypothesis	Level α Rejection Region
$H_1: \mu_1 - \mu_2 > \delta_0$	$R: T \geq t_\alpha$
$H_1: \mu_1 - \mu_2 < \delta_0$	$R: T \leq -t_\alpha$
$H_1: \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert T \rvert \geq t_{\alpha/2}$

Example 10 Testing the Equality of Mean Computer Anxiety Scores

Refer to the computer anxiety scores (CARS) described in Example 9 and the summary statistics

Female CARS

$$n_1 = 15 \qquad \bar{x} = 2.514 \qquad s_1 = .773$$

Male CARS

$$n_2 = 20 \qquad \bar{y} = 2.963 \qquad s_2 = .525$$

Do these data strongly indicate that the mean score for females is lower than that for males? Test at level $\alpha = .05$.

SOLUTION We are seeking strong evidence in support of the hypothesis that the mean computer anxiety score for females (μ_1) is less than the mean score for males. Therefore the alternative hypothesis should be taken as $H_1: \mu_1 < \mu_2$ or $H_1: \mu_1 - \mu_2 < 0$, and our problem can be stated as testing

$$H_0: \mu_1 - \mu_2 = 0 \qquad \text{versus} \qquad H_1: \mu_1 - \mu_2 < 0$$

We employ the test statistic

$$T = \frac{\bar{X} - \bar{Y}}{S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{d.f.} = n_1 + n_2 - 2$$

and set the left-sided rejection region $R: T \le -t_{.05}$. For d.f. $= n_1 + n_2 - 2 = 33$, we approximate the tabled value as $t_{.05} = 1.692$, so the rejection region is $R: T \le -1.692$.

With $S_{\text{pooled}} = .642$ already calculated in Example 9, the observed value of the test statistic T is

$$t = \frac{2.514 - 2.963}{.642\sqrt{\dfrac{1}{15} + \dfrac{1}{20}}} = \frac{-.449}{.2193} = -2.05$$

This value lies in the rejection region R. Consequently, at the .05 level of significance, we reject the null hypothesis in favor of the alternative hypothesis that males have a higher mean computer anxiety score.

A computer calculation gives a $P-$value of about .025 so the evidence of H_1 is moderately strong.

DECIDING WHETHER OR NOT TO POOL

Our preceding discussion of large and small sample inferences raises a few questions:

For small sample inference, why do we assume the population standard deviations to be equal when no such assumption was needed in the large samples case?

When should we be wary about this assumption, and what procedure should we use when the assumption is not reasonable?

Learning statistics would be a step simpler if the ratio

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

had a t distribution for small samples from normal populations. Unfortunately, statistical theory proves it otherwise. The distribution of this ratio is *not* a t and, worse yet, it depends on the unknown quantity σ_1 / σ_2. The assumption $\sigma_1 = \sigma_2$ and the change of the denominator to $S_{\text{pooled}} \sqrt{1/n_1 + 1/n_2}$ allow the t-based inferences to be valid. However, the $\sigma_1 = \sigma_2$ restriction and accompanying pooling are not needed in large samples where a normal approximation holds.

With regard to the second question, the relative magnitude of the two sample standard deviations s_1 and s_2 should be the prime consideration. The assumption $\sigma_1 = \sigma_2$ is reasonable if s_1 / s_2 is not very much different from 1. As a working rule, the range of values $\frac{1}{2} \leq s_1/s_2 \leq 2$ may be taken as reasonable cases for making the assumption $\sigma_1 = \sigma_2$ and hence for pooling. If s_1 / s_2 is seen to be smaller than $\frac{1}{2}$ or larger than 2, the assumption $\sigma_1 = \sigma_2$ would be suspect. In that case, some approximate methods of inference about $\mu_1 - \mu_2$ are available.

5. INFERENCES FROM SMALL SAMPLES: NORMAL POPULATIONS WITH UNEQUAL VARIANCES

We first introduce a simple conservative procedure and then a somewhat more complex approximate approach that is widely used in statistical software programs.

5.1 A CONSERVATIVE *t* TEST

We first outline a simple though conservative procedure, which treats the ratio

$$T^* = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

as a t variable with d.f. $=$ smaller of $n_1 - 1$ and $n_2 - 1$.

Small Samples Inferences for $\mu_1 - \mu_2$ When the Populations Are Normal But σ_1 and σ_2 Are Not Assumed to Be Equal

A $100(1 - \alpha)\%$ conservative confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

where $t_{\alpha/2}$ denotes the upper $\alpha/2$ point of the t distribution with d.f. = smaller of $n_1 - 1$ and $n_2 - 1$.

The null hypothesis $H_0: \mu_1 - \mu_2 = \delta_0$ is tested using the test statistic

$$T^* = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \qquad \text{d.f.} = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1$$

Here, the confidence interval is conservative in the sense that the actual level of confidence is at least $1 - \alpha$. Likewise, the level α test is conservative in the sense that the actual Type I error probability is no more than α.

Example 11 Testing Equality of Green Gas Mean Yields

One process of making green gasoline, not just a gasoline additive, takes biomass in the form of sucrose and converts it into gasoline using catalytic reactions. This research is still at the pilot plant stage. At one step in a pilot plant process, the product volume (liters) consists of carbon chains of length 3. Nine runs were made with each of two catalysts and the product volumes measured.

catalyst 1	1.86	2.05	2.06	1.88	1.75	1.64	1.86	1.75	2.13
catalyst 2	.32	1.32	.93	.84	.55	.84	.37	.52	.34

The sample sizes $n_1 = n_2 = 9$ and the summary statistics are

$$\bar{x} = 1.887, \quad s_1^2 = .0269 \quad \bar{y} = .670 \quad s_2^2 = .1133$$

Is the mean yield with catalyst 1 more than .80 liters higher than the yield with catalyst 2? Test with $\alpha = 0.05$

SOLUTION The test concerns the $\delta = \mu_1 - \mu_2$ and we wish to show that δ is greater than .80. Therefore we formulate the testing problem as

$$H_0: \mu_1 - \mu_2 = .80 \qquad \text{versus} \qquad H_1: \mu_1 - \mu_2 > .80$$

The sample sizes $n_1 = n_2 = 9$ are small and there are no outliers or obvious departures from normality. However, $s_2^2 = .1133$ is more than four times $s_1^2 = .0269$. We should not pool.

We choose the conservative test which is based on the test statistic

$$T^* = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \qquad \text{d.f.} = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1 = 8$$

For d.f. $= 8$, the tabled value is $t_{.05} = 1.860$, so we set the rejection region $R: T^* > 1.860$. The calculated value of the test statistic is

$$t^* = \frac{1.887 - .670 - .80}{\sqrt{\dfrac{.0269}{9} + \dfrac{.1133}{9}}} = 3.34$$

which falls in the rejection region. So, we reject the null hypothesis at $\alpha = .05$ and conclude that the mean product volume from catalyst 1 is more than .80 liters higher.

The associated P–value $= P[\,T \geq 3.34\,] = .0051$. Alternatively, the observed t^* is only slightly smaller than the entry $t_{.005} = 3.355$ in Table 4, Appendix B.

The conservative procedure does guarantee that α is an upper bound on the probability of falsely rejecting the null hypothesis. It is a reasonable choice when the minimum sample size, and therefore the degrees of freedom, are not too small.

5.2 AN APPROXIMATE *t* TEST—SATTERTHWAITE CORRECTION

An alternate, but more complicated approximation is available for two normal distributions whose variances seem to be unequal. This approximate t distribution is preferred over the conservative procedure which can have too few degrees of freedom. The statistic is the same T^* but, because sample sizes are small, its distribution is approximated as a t distribution. This approach is sometimes called the **Satterthwaite correction** after the person who derived the approximation.

Normal Populations with $\sigma_1 \neq \sigma_2$

Let $\delta = \mu_1 - \mu_2$. When the sample sizes n_1 and n_2 are not large and $\sigma_1 \neq \sigma_2$

$$T^* = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

is approximately distributed as a t with estimated degrees of freedom.

The estimated degrees of freedom depend on the observed values of the sample variances s_1^2 and s_2^2.

$$\text{estimated degrees of freedom} = v = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{S_1^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{S_2^2}{n_2}\right)^2}$$

The estimated degrees of freedom v are often rounded down to an integer so a t table can be consulted.

Example 12 Confidence Interval Using the Approximate t

Refer to Example 11. Use the approximate t distribution to determine a 95% confidence interval for the difference of mean product volumes.

SOLUTION From the previous example we have the summary statistics

$$\bar{x} = 1.887, \quad s_1^2 = .0269, \quad \bar{y} = .670, \quad s_2^2 = .1133$$

We first estimate the degrees of freedom

$$\frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{S_1^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{S_2^2}{n_2}\right)^2} = \frac{\left(\dfrac{0.269}{9} + \dfrac{.1133}{9}\right)^2}{\dfrac{1}{9 - 1}\left(\dfrac{.0269}{9}\right)^2 + \dfrac{1}{9 - 1}\left(\dfrac{.1133}{9}\right)^2} = 11.60$$

To use the t table, we round down to 11 and then obtain $t_{0.025} = 2.201$ for 11 degrees of freedom.

The resulting 95% confidence interval is

$$\left(1.887 - .670 - 2.201\sqrt{\frac{.02699}{9} + \frac{.1133}{9}}\,,\ 1.887 - .670 + 2.201\sqrt{\frac{.02699}{9} + \frac{.1133}{9}}\right)$$

or $(.94, 1.49)$ liters. The mean product volume for the first catalyst is greater than that of the second catalyst by .94 to 1.49 liters.

Notice that the conservative procedure would use 8 d.f. so $t_{.025} = 2.306$ and the resulting confidence interval is wider than $(.94, 1.49)$ liters.

When computer software is available to simplify the calculation, the approximate t approach is usually preferred.

Exercises

10.5 One semester, an instructor taught the same computer course at two universities located in different cities. He was able to give the same final at both locations. The student's scores provided the summary statistics.

Sample 1	Sample 2
$n_1 = 52$	$n_2 = 44$
$\bar{x} = 73$	$\bar{y} = 66$
$s_1^2 = 151$	$s_2^2 = 142$

(a) Obtain a point estimate of $\mu_1 - \mu_2$ and calculate the estimated standard error.

(b) Construct a 95% confidence interval for $\mu_1 - \mu_2$.

10.6 Rural and urban students are to be compared on the basis of their scores on a nationwide musical aptitude test. Two random samples of sizes 90 and 100 are selected from rural and urban seventh-grade students. The summary statistics from the test scores are

	Rural	Urban
Sample size	90	100
Mean	76.4	81.2
Standard deviation	8.2	7.6

Construct a 98% confidence interval for the difference in population mean scores between urban and rural students.

10.7 Perform a test to determine if there is a significant difference between the population mean scores in Exercise 10.6. Use $\alpha = .05$.

10.8 A linguist wants to compare the writing styles in two magazines and one measure is the number of words per sentence. On the basis of 50 randomly selected sentences from each source, she finds

Magazine 1 $n_1 = 50$ $\bar{x} = 12.6$ $s_1 = 4.2$

Magazine 2 $n_2 = 50$ $\bar{y} = 9.5$ $s_2 = 1.9$

Determine a 98% confidence interval for the difference in mean number of words per sentence.

10.9 Refer to Exercise 10.8. Perform a test of hypothesis that is intended to show that the mean for magazine 1 is more than 2 words larger than the mean for magazine 2.

(a) Formulate the null and alternative hypotheses. (Define any symbols you use.)

(b) State the test statistic and the rejection region with $\alpha = .05$.

(c) Perform the test at $\alpha = .05$. Also, find the P–value and comment.

10.10 Refer to Example 4. Workers in three occupations were questioned about satisfaction with their jobs. Suppose the number of responses in each of the four categories are

Occupation	*Very Dis-satisfied*	*A Little Dis-satisfied*	*Moder-ately Satisfied*	*Very Satis-fied*	Sample Size
Firefighters	6	16	24	180	226
Office Supervisors	1	12	85	149	247
Clergy	1	5	6	78	90

Assign 1 to very dissatisfied, 2 to a little dissatisfied, 3 to moderately satisfied, and 4 to very satisfied. These data have means and proportion *very satisfied* similar to those reported in a large scale survey.[2]

(a) Calculate the sample mean and variance separately for firefighters and office supervisors. The scores for clergy have mean 3.79 and standard deviation .590.

(b) Construct a 95% confidence interval for the difference in mean satisfaction score for clergy and office supervisors.

10.11 Refer to the confidence interval obtained in Exercise 10.10(b). If you were to test the null hypothesis that the mean satisfaction scores are equal versus the two-sided alternatives, what would be the conclusion of your test with $\alpha = .05$? (See Section 4 of Chapter 9.)

[2] General Social Survey 1988–2006 conducted by the National Opinion Research Center, University of Chicago.

10.12 In a study of interspousal aggression and its possible effect on child behavior, the behavior problem checklist (BPC) scores were recorded for 47 children whose parents were classified as aggressive. The sample mean and standard deviation were 7.92 and 3.45, respectively. For a sample of 38 children whose parents were classified as nonaggressive, the mean and standard deviation of the BPC scores were 5.80 and 2.87, respectively. Do these observations substantiate the conjecture that the children of aggressive families have a higher mean BPC than those of nonaggressive families? (Answer by calculating the P–value.)

10.13 Suppose that measurements of the size of butterfly wings (cm) for two related species yielded the data

Species 1	6	4	7	3
Species 2	6	9	6	

(a) Calculate s^2_{pooled}.

(b) Give an estimate of the common standard deviation for the wing size for the two species.

(c) Evaluate the t statistic for testing equality of the two population mean wing sizes.

10.14 Three male and three female students recorded the number of times they used a credit card in one week.

Male	8	4	6
Female	1	5	3

(a) Calculate s^2_{pooled}.

(b) Give an estimate of the common standard deviation for the number of uses.

(c) Evaluate that t statistic for testing equality of the two population mean number of uses.

10.15 Two gel pens, Gel-1 and Gel-2, are compared on the basis of the number of weeks before they stop writing. Out of 27 persons available, 13 are randomly chosen to receive Gel-1 and the other 14 receive Gel-2. These are the only pens they use for writing.

$$n_1 = 13, \ \bar{x} = 9, \ \ \Sigma(x_i - \bar{x})^2 = 28$$

$$n_2 = 14, \ \bar{y} = 17, \ \Sigma(y_i - \bar{y})^2 = 32$$

(a) Obtain s^2_{pooled}.

(b) Test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$ with $\alpha = .05$.

(c) Determine a 95% confidence interval for $\mu_1 - \mu_2$.

10.16 The data on the weight (lb) of male and female wolves, from Table D.9 of the Data Bank, are

Female	57	84	90	71	71	77	68	73			
Male	71	93	101	84	88	117	86	86	93	86	106

(a) Test the null hypothesis that the mean weights of males and females are equal versus a two-sided alternative. Take $\alpha = .05$.

(b) Obtain a 95% confidence interval for the difference of population mean weights.

(c) State any assumptions you make about the populations.

10.17 Psychologists have made extensive studies on the relationship between child abuse and later criminal behavior. Consider a study that consisted of the follow-ups of 52 boys who were abused in their preschool years and 67 boys who were not abused. The data of the number of criminal offenses of those boys in their teens yielded the following summary statistics.

	Abused	Nonabused
Mean	2.48	1.57
Standard deviation	1.94	1.31

Is the mean number of criminal offenses significantly higher for the abused group than that for the nonabused group? Answer by calculating the P–value.

10.18 Referring to the data in Exercise 10.17, determine a 99% confidence interval for the difference between the true means for the two groups.

10.19 To compare two programs for training industrial workers to perform a skilled job, 20 workers are included in an experiment. Of these, 10 are selected at random and trained by method 1; the remaining 10 workers are trained by method 2. After completion of training, all the workers are subjected to a time-and-motion test that records the speed of performance of a skilled job. The following data are obtained.

	Time (minutes)									
Method 1	15	20	11	23	16	21	18	16	27	24
Method 2	23	31	13	19	23	17	28	26	25	28

(a) Can you conclude from the data that the mean job time is significantly less after training with method 1 than after training with method 2? (Test with $\alpha = .05$.)

(b) State the assumptions you make for the population distributions.

(c) Construct a 95% confidence interval for the population mean difference in job times between the two methods.

10.20 Given here are the sample sizes and the sample standard deviations for independent random samples from two populations. For each case, state which of the three tests you would use in testing hypotheses about $\mu_1 - \mu_2$: (1) Z test, (2) t test with pooling, (3) conservative t test without pooling. Also, state any assumptions you would make about the population distributions.

(a) $n_1 = 35, \quad s_1 = 12.2$
$n_2 = 50, \quad s_2 = 8.6$

(b) $n_1 = 8, \quad s_1 = 0.86$
$n_2 = 7, \quad s_2 = 1.12$

(c) $n_1 = 8, \quad s_1 = 1.54$
$n_2 = 15, \quad s_2 = 5.36$

(d) $n_1 = 70, \quad s_1 = 6.2$
$n_2 = 60, \quad s_2 = 2.1$

10.21 Two different sprays for rewaterproofing parkas are compared by treating 5 parkas with spray 1 and 6 with spray 2. After one day, each parka is sprayed with water and the parka is rated on its ability to repel water. The following conclusion is reported

"Under the assumption of normal populations with equal but unknown standard deviations, the 90% confidence interval for $\mu_1 - \mu_2$ is $(3.6, 5.2)$."

From this report,

State the conclusion of testing $H_0 : \mu_1 - \mu_2 = 4.1$ versus $H_1 : \mu_1 - \mu_2 \neq 4.1$ at $\alpha = .10$. (See Section 4 of Chapter 9.)

10.22 To compare the effectiveness of isometric and isotonic exercise methods, 20 obese male students are included in an experiment: 10 are selected at random and assigned to one exercise method; the remaining 10 are assigned to the other exercise method. After five weeks, the reductions in abdomen measurements are recorded in centimeters, and the following results are obtained.

	Isometric Method A	Isotonic Method B
Mean	2.4	3.2
Standard deviation	0.8	1.0

(a) Do these data support the claim that the isotonic method is more effective?

(b) Construct a 95% confidence interval for $\mu_B - \mu_A$.

10.23 Refer to Exercise 10.22.

(a) Aside from the type of exercise method, identify a few other factors that are likely to have an important effect on the amount of reduction accomplished in a five-week period.

(b) What role does randomization play in achieving a valid comparison between the two exercise methods?

(c) If you were to design this experiment, describe how you would divide the 20 students into two groups.

10.24 Refer to Example 11.

(a) Using the conservative approach in Example 11, obtain an approximate 95% confidence interval for the difference of means.

(b) Compare the confidence interval in Part (a) with the one obtained in Example 12.

10.25 Refer to Exercise 10.24. Using the approximate t distribution as in Example 12, perform the test requested in Example 11.

10.26 The following generic computer output summarizes the data, given in Table D.5 of the Data Bank, on the pretest percent body fat in male and female students.

Gender	N	Mean	StDev
M	40	14.38	7.34
F	43	23.72	5.78

Find a 99% confidence interval for the difference of the two population means.

10.27 Refer to the data on the weight of wolves in Table D.9 of the Data Bank. A computer analysis produces the output

```
Two-sample T for Male wt vs Female wt

               N    Mean   StDev
Male wt       11    91.9    12.4
Female wt      8    73.9    10.1
T-Test of difference = 0 (vs not =):
T-Value = 3.38  P-Value = 0.004  DF = 17
```

(a) What is the conclusion to testing the equality of mean weights at level $\alpha = .05$?

(b) Test the null hypothesis that males weigh an average of 5 pounds more than females against a two-sided alternative. Take $\alpha = .05$.

6. RANDOMIZATION AND ITS ROLE IN INFERENCE

We have presented the methods of drawing inferences about the difference between two population means. Let us now turn to some important questions regarding the design of the experiment or data collection procedure. The manner in which experimental subjects are chosen for the two treatment groups can be crucial. For example, suppose that a remedial-reading instructor has developed a new teaching technique and is permitted to use the new method to instruct half the pupils in the class. The instructor might choose the most alert or the students who are more promising in some other way, leaving the weaker students to be taught in the conventional manner. Clearly, a comparison between the reading achievements of these two groups would not just be a comparison of two teaching methods. A similar fallacy can result in comparing the nutritional quality of a new lunch package if the new diet is given to a group of children suffering from malnutrition and the conventional diet is given to a group of children who are already in good health.

When the assignment of treatments to experimental units is under our control, steps can be taken to ensure a valid comparison between the two treatments. At the core lies the principle of impartial selection, or **randomization.** The choice of the experimental units for one treatment or the other must be made by a chance mechanism that does not favor one particular selection over any other. It must not be left to the discretion of the experimenters because, even unconsciously, they may be partial to one treatment.

Suppose that a comparative experiment is to be run with N experimental units, of which n_1 units are to be assigned to treatment 1 and the remaining $n_2 = N - n_1$ units are to be assigned to treatment 2. The principle of randomization tells us that the n_1 units for treatment 1 must be chosen at random from the available collection of N units—that is, in a manner such that all $\binom{N}{n_1}$ possible choices are equally likely to be selected.

Randomization Procedure for Comparing Two Treatments

From the available $N = n_1 + n_2$ experimental units, choose n_1 units at random to receive treatment 1 and assign the remaining n_2 units to treatment 2. The random choice entails that all $\binom{N}{n_1}$ possible selections are equally likely to be chosen.

As a practical method of random selection, we can label the available units from 1 to N. Then read random digits from Table 1, Appendix B, until n_1 different numbers between 1 and N are obtained. These n_1 experimental units receive treatment 1 and the remaining units receive treatment 2. For a quicker and more efficient means of random sampling, one can use the computer (see, for instance, the Technology section, page 164, Chapter 4).

Although randomization is not a difficult concept, it is one of the most fundamental principles of a good experimental design. It guarantees that uncontrolled sources of variation have the same chance of favoring the response of treatment 1 as they do of favoring the response of treatment 2. Any systematic effects of uncontrolled variables, such as age, strength, resistance, or intelligence, are chopped up or confused in their attempt to influence the treatment responses.

Randomization prevents uncontrolled sources of variation from influencing the responses in a systematic manner.

Of course, in many cases, the investigator does not have the luxury of randomization. Consider comparing crime rates of cities before and after a new law. Aside from a package of criminal laws, other factors such as poverty, inflation, and unemployment play a significant role in the prevalence of crime. As long as these contingent factors cannot be regulated during the observation period, caution should be exercised in crediting the new law if a decline in the crime rate is observed or discrediting the new law if an increase in the crime rate is observed. When randomization cannot be performed, extreme caution must be exercised in crediting an apparent difference in means to a difference in treatments. The differences may well be due to another factor.

Exercises

10.28 Randomly allocate 2 subjects from among

Al, Bob, Carol, Ellen, John

to be in the control group. The others will receive a treatment.

10.29 Randomly allocate three subjects from among 6 mice,

Alpha, Tau, Omega, Pi, Beta, Phi

to group 1.

10.30 Early studies showed a disproportionate number of heavy smokers among lung cancer patients. One scientist theorized that the presence of a particular gene could tend to make a person want to smoke and be susceptible to lung cancer.

(a) How would randomization settle this question?

(b) Would such a randomization be ethical with human subjects?

10.31 Observations on 10 mothers who nursed their babies and 8 who did not revealed that nursing mothers felt warmer toward their babies. Can we conclude that nursing affects a mother's feelings toward her child?

10.32 Suppose that you are asked to design an experiment to study the effect of a hormone injection on the weight gain of pregnant rats during gestation. You have decided to inject 6 of the 12 rats available for the experiment and retain the other 6 as controls.

(a) Briefly explain why it is important to randomly divide the rats into the two groups. What might be wrong with the experimental results if you choose to give the hormone treatment to 6 rats that are easy to grab from their cages?

(b) Suppose that the 12 rats are tagged with serial numbers from 1 through 12 and 12 marbles identical in appearance are also numbered from 1 through 12. How can you use these marbles to randomly select the rats in the treatment and control groups?

7. MATCHED PAIRS COMPARISONS

In comparing two treatments, it is desirable that the experimental units or subjects be as alike as possible, so that a difference in responses between the two groups can be attributed to differences in treatments. If some identifiable conditions vary over the units in an uncontrolled manner, they could introduce a large variability in the measurements. In turn, this could obscure a real difference in treatment effects. On the other hand, the requirement that all subjects be alike may impose a severe limitation on the number of subjects available for a comparative experiment. To compare two analgesics, for example, it would be impractical to look for a sizable number of patients who are of the same sex, age, and general health condition and who have the same severity of pain. Aside from the question of practicality, we would rarely want to confine a comparison to such a narrow group. A broader scope of inference can be attained by applying the treatments on a variety of patients of both sexes and different age groups and health conditions.

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

The concept of **matching** or **blocking** is fundamental to providing a compromise between the two conflicting requirements that the experimental units be alike and also of different kinds. The procedure consists of choosing units in pairs or blocks so that the units in each block are similar and those in different blocks are dissimilar. One of the units in each block is assigned to treatment 1, the other to treatment 2. This process preserves the effectiveness of a comparison within each block and permits a diversity of conditions to exist in different blocks. Of course, the treatments must be allotted to each pair randomly to

Matched Pairs

iStockphoto

Identical twins are the epitome of matched pair experimental subjects. They are matched with respect to not only age but also a multitude of genetic factors. Social scientists, trying to determine the influence of environment and heredity, have been especially interested in studying identical twins who were raised apart. Observed differences in IQ and behavior are then supposedly due to environmental factors.

When the subjects are animals like mice, two from the same litter can be paired. Going one step further, genetic engineers can now provide two identical plants or small animals by cloning these subjects.

avoid selection bias. This design is called a **matched pairs design** or **sampling.** For example, in studying how two different environments influence the learning capacities of preschoolers, it is desirable to remove the effect of heredity: Ideally, this is accomplished by working with twins.

In a matched pairs design, the response of an experimental unit is influenced by:

1. The conditions prevailing in the block (pair).
2. A treatment effect.

By taking the difference between the two observations in a block, we can filter out the common block effect. These differences then permit us to focus on the effects of treatments that are freed from undesirable sources of variation.

Pairing (or Blocking)

Pairing similar experimental units according to some identifiable characteristic(s) serves to remove this source of variation from the experiment.

The structure of the observations in a paired comparison is given below, where X and Y denote the responses to treatments 1 and 2, respectively. The difference between the responses in each pair is recorded in the last column, and the summary statistics are also presented.

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n - Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.
Summary statistics:

$$\overline{D} = \frac{1}{n} \sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n - 1}$$

Although the pairs (X_i, Y_i) are independent of one another, X_i and Y_i within the ith pair will usually be dependent. In fact, if the pairing of experimental units is effective, we would expect X_i and Y_i to be relatively large or small together. Expressed in another way, we would expect (X_i, Y_i) to have a high positive correlation. Because the differences $D_i = X_i - Y_i, i = 1, 2, \ldots, n$, are freed from the block effects, it is reasonable to assume that they constitute a random sample from a population with mean μ_D and variance σ_D^2, where μ_D represents the mean difference of the treatment effects. In other words,

$$\begin{aligned} E(D_i) &= \mu_D \\ \text{Var}(D_i) &= \sigma_D^2 \qquad i = 1, \ldots, n \end{aligned}$$

If the mean difference μ_D is zero, then the two treatments can be considered equivalent. A positive μ_D signifies that treatment 1 has a higher mean response than treatment 2. Considering $D_1, \ldots, D_n$ to be a single random sample from a population, we can immediately apply the techniques discussed in Chapters 8 and 9 to learn about the population mean μ_D.

7.1 INFERENCES FROM A LARGE NUMBER OF MATCHED PAIRS

As we learned in Chapter 8, the assumption of an underlying normal distribution can be relaxed when the sample size is large. The central limit theorem applied to the differences $D_1, \ldots, D_n$ suggests that when n is large, say, greater than 30,

$$\frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \qquad \text{is approximately } N(0, 1)$$

Inferences can then be based on the percentage points of the $N(0, 1)$ distribution or, equivalently, those of the t distribution, with the degrees of freedom marked "infinity."

Example 13 **Does Conditioning Reduce Percent Body Fat?**

A conditioning class is designed to introduce students to a variety of training techniques to improve fitness and flexibility. The percent body fat was measured at the start of the class and at the end of the semester. The data for 81 students are given in Table D.5 of the Data Bank.

(a) Obtain a 98% confidence interval for the mean reduction in percent body fat.

(b) Test, at $\alpha = .01$, to establish the claim that the conditioning class reduces the mean percent body fat.

SOLUTION (a) Each subject represents a block which produces one measurement of percent body fat at the start of the semester (x) and one at the end (y). The 81 paired differences $d_i = x_i - y_i$ are summarized using a computer.

	N	Mean	StDev	SE Mean
Difference	81	3.322	2.728	0.303

That is, $\bar{d} = 3.322$ and $s_D = 2.728$. The sample size 81 is large so there is no need to assume that the population is normal. Since, from the normal table, $z_{.01} = 2.33$, the 98% confidence interval becomes

$$\left(\bar{d} - 2.33 \frac{s_D}{\sqrt{81}}, \quad \bar{d} + 2.33 \frac{s_D}{\sqrt{81}} \right)$$

$$\left(3.322 - 2.33 \times \frac{2.728}{\sqrt{81}}, \quad 3.322 + 2.33 \times \frac{2.728}{\sqrt{81}} \right) = (3.322 - .706, 3.322 + .706)$$

or $(2.62, 4.03)$ percent. We are 98% confident that the mean reduction in body fat is 2.62 to 4.03 percent.

(b) Because the claim is that $\mu_D > 0$, the initial reading tends to be higher than at the end of class, we formulate:

$$H_0 : \mu_D = 0 \qquad \text{versus} \qquad H_1 : \mu_D > 0$$

The test statistic

$$Z = \frac{\bar{D}}{S_D / \sqrt{n}}$$

is approximately normally distributed so the rejection region is $R : Z \geq z_{.01} = 2.33$. The observed value of the test statistic

$$z = \frac{\bar{d}}{S_D / \sqrt{81}} = \frac{3.322}{2.728 / \sqrt{81}} = \frac{3.322}{.303} = 10.96$$

falls in the rejection region. Consequently H_0 is rejected in favor of H_1 at level $\alpha = .01$. We conclude that the conditioning class does reduce the mean percent body fat. The value of the test statistic is so far in the rejection region that the $P-$value is .0000 to at least four places. The evidence in support of H_1 is very very strong.

7.2 INFERENCES FROM A SMALL NUMBER OF MATCHED PAIRS

When the sample size is not large, we make the additional assumption that the distribution of the differences is normal.

In summary,

Small Samples Inferences about the Mean Difference μ_D

Assume that the differences $D_i = X_i - Y_i$ are a random sample from an $N(\mu_D, \sigma_D)$ distribution. Let

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} \qquad \text{and} \qquad S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n - 1}}$$

Then:

1. A $100(1 - \alpha)\%$ confidence interval for μ_D is given by

$$\left(\overline{D} - t_{\alpha/2}\frac{S_D}{\sqrt{n}}, \quad \overline{D} + t_{\alpha/2}\frac{S_D}{\sqrt{n}} \right)$$

where $t_{\alpha/2}$ is based on $n - 1$ degrees of freedom.

2. A test of $H_0 : \mu_D = \mu_{D0}$ is based on the test statistic

$$T = \frac{\overline{D} - \mu_{D0}}{S_D / \sqrt{n}} \qquad \text{d.f.} = n - 1$$

Example 14 Does a Pill Incidentally Reduce Blood Pressure?

A medical researcher wishes to determine if a pill has the undesirable side effect of reducing the blood pressure of the user. The study involves recording the initial blood pressures of 15 college-age women. After they use the pill regularly for six months, their blood pressures are again recorded. The researcher wishes to draw inferences about the effect of the pill on blood pressure from the observations given in Table 1.

(a) Calculate a 95% confidence interval for the mean reduction in blood pressure.

(b) Do the data substantiate the claim that use of the pill reduces blood pressure? Test at $\alpha = .01$.

TABLE 1 Blood-Pressure Measurements before and after Use of Pill

	Subject														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before (x)	70	80	72	76	76	76	72	78	82	64	74	92	74	68	84
After (y)	68	72	62	70	58	66	68	52	64	72	74	60	74	72	74
$d = x - y$	2	8	10	6	18	10	4	26	18	-8	0	32	0	-4	10

Courtesy of a family planning clinic.

SOLUTION (a) Here each subject represents a block generating a pair of measurements: one before using the pill and the other after using the pill. The paired differences $d_i = x_i - y_i$ are computed in the last row of Table 1, and we calculate the summary statistics

$$\bar{d} = \frac{\sum d_i}{15} = 8.80 \qquad s_D = \sqrt{\frac{\sum (d_i - \bar{d})^2}{14}} = 10.98$$

If we assume that the paired differences constitute a random sample from a normal population $N(\mu_D, \sigma_D)$, a 95% confidence interval for the mean difference μ_D is given by

$$\bar{D} \pm t_{.025} \frac{S_D}{\sqrt{15}}$$

where $t_{.025}$ is based on d.f. $= 14$. From the t table, we find $t_{.025} = 2.145$. The 95% confidence interval is then computed as

$$8.80 \pm 2.145 \times \frac{10.98}{\sqrt{15}} = 8.80 \pm 6.08 \text{ or } (2.72, 14.88)$$

This means that we are 95% confident the mean reduction of blood pressure is between 2.72 and 14.88.

(b) Because the claim is that $\mu_D > 0$, we formulate

$$H_0 : \mu_D = 0 \qquad \text{versus} \qquad H_1 : \mu_D > 0$$

We employ the test statistic $T = \dfrac{\bar{D}}{S_D / \sqrt{n}}$, d.f. $= 14$ and set a right-sided rejection region. With d.f. $= 14$, we find $t_{.01} = 2.624$, so the rejection region is $R : T \geq 2.624$.

The observed value of the test statistic

$$t = \frac{\bar{d}}{S_D / \sqrt{n}} = \frac{8.80}{10.98 / \sqrt{15}} = \frac{8.80}{2.84} = 3.10$$

falls in the rejection region. Consequently, H_0 is rejected in favor of H_1 at $\alpha = .01$. We conclude that a reduction in blood pressure following use of the pill is strongly supported by the data.

Note: To be more convinced that the pill causes the reduction in blood pressure, it is advisable to measure the blood pressures of the same subjects once again after they have stopped using the pill for a period of time. This amounts to performing the experiment in reverse order to check the findings of the first stage.

7.3 RANDOMIZATION WITH MATCHED PAIRS

Example 14 is a typical before–after situation. Data gathered to determine the effectiveness of a safety program or an exercise program would have the same structure. In such cases, there is really no way to choose how to order the experiments within a pair. The before situation must precede the after situation. If something other than the institution of the program causes performance to improve, the improvement will be incorrectly credited to the program. However, when the order of the application of treatments can be determined by the investigator, something can be done about such systematic influences. Suppose that a coin is flipped to select the treatment for the first unit in each pair. Then the other treatment is applied to the second unit. Because the coin is flipped again for each new pair, any uncontrolled variable has an equal chance of helping the performance of either treatment 1 or treatment 2. After eliminating an identified source of variation by pairing, we return to randomization in an attempt to reduce the systematic effects of any uncontrolled sources of variation.

Randomization with Pairing

After pairing, the assignment of treatments should be randomized for each pair.

Randomization within each pair chops up or diffuses any systematic influences that we are unable to control.

Exercises

10.33 Given the following matched pairs sample,

x	y
6	3
4	1
8	3
6	5
9	7
6	8

(a) Evaluate the t statistic

$$t = \frac{\bar{d}}{S_D/\sqrt{n}}.$$

(b) How many degrees of freedom does this t have?

10.34 Two sun blockers are to be compared. One blocker is rubbed on one side of a subject's back and the other blocker is rubbed on the other side. Each subject then lies in the sun for two hours. After waiting an additional hour, each side is rated according to redness.

(a) Evaluate the t statistic

$$t = \frac{\bar{d}}{S_D/\sqrt{n}}.$$

(b) How many degrees of freedom does this t have?

Subject No.	Blocker 1	Blocker 2
1	2	2
2	7	5
3	8	4
4	3	1
5	5	3

10.35 It is claimed that an industrial safety program is effective in reducing the loss of working hours due to factory accidents. The following data are collected concerning the weekly loss of working hours due to accidents in six plants both before and after the safety program is instituted.

	Plant					
	1	2	3	4	5	6
Before	12	30	15	37	29	15
After	10	29	16	35	26	16

Do the data substantiate the claim? Use $\alpha = .05$.

10.36 A manufacturer claims his boot waterproofing is better than the major brand. Five pairs of shoes are available for a test.

(a) Explain how you would conduct a paired sample test.

(b) Write down your assignment of waterproofing to each shoe. How did you randomize?

10.37 A food scientist wants to study whether quality differences exist between yogurt made from skim milk with and without the preculture of a particular type of bacteria, called Psychrotrops (PC). Samples of skim milk are procured from seven dairy farms. One-half of the milk sampled from each farm is inoculated with PC, and the other half is not. After yogurt is made with these milk samples, the firmness of the curd is measured, and those measurements are given below.

	Dairy Farm						
Curd Firmness	*A*	*B*	*C*	*D*	*E*	*F*	*G*
With PC	68	75	62	86	52	46	72
Without PC	61	69	64	76	52	38	68

(a) Do these data substantiate the conjecture that the treatment of PC results in a higher degree of curd firmness? Test at $\alpha = .05$.

(b) Determine a 90% confidence interval for the mean increase of curd firmness due to the PC treatment.

10.38 A study is to be made of the relative effectiveness of two kinds of cough medicines in increasing sleep. Six people with colds are given medicine *A* the first night and medicine *B* the second night. Their hours of sleep each night are recorded.

	Subject					
	1	2	3	4	5	6
Medicine *A*	4.8	4.1	5.8	4.9	5.1	7.4
Medicine *B*	3.9	4.2	5.0	4.9	5.2	7.1

(a) Establish a 95% confidence interval for the mean change in hours of sleep when switching from medicine *A* to medicine *B*.

(b) How and what would you randomize in this study? Briefly explain your reason for randomization.

10.39 Two methods of memorizing difficult material are being tested to determine if one produces better retention. Nine pairs of students are included in the study. The students in each pair are matched according to IQ and academic background and then assigned to the two methods at random. A memorization test is given to all the students, and the following scores are obtained:

	Pair								
	1	2	3	4	5	6	7	8	9
Method *A*	90	86	72	65	44	52	66	38	83
Method *B*	85	87	70	62	44	53	62	35	86

At $\alpha = .05$, test to determine if there is a significant difference in the effectiveness of the two methods.

10.40 In an experiment conducted to see if electrical pricing policies can affect consumer behavior, 10 homeowners in Wisconsin had to pay a premium for power use during the peak hours. They were offered lower off-peak rates. For each home, the July on-peak usage (kilowatt hours) under the pricing experiment was compared to the previous July usage.

Year	
Previous	Experimental
200	160
180	175
240	210
425	370
120	110
333	298
418	368
380	250
340	305
516	477

(a) Find a 95% confidence interval for the mean decrease.

(b) Test $H_0 : \mu_D = 0$ against $H_1 : \mu_D \neq 0$ at level $\alpha = .05$.

(c) Comment on the feasibility of randomization of treatments.

(d) Without randomization, in what way could the results in parts (a) and (b) be misleading?

(*Hint:* What if air conditioner use is a prime factor, and the July with experimental pricing was cooler than the previous July?)

10.41 To compare the crop yields from two strains of wheat, *A* and *B*, an experiment was conducted at eight farms located in different parts of a state. At each farm, strain *A* was grown on one plot and strain *B* on another; all 16 plots were of equal sizes. Given below are data of yield in pounds per plot.

	Farm							
	1	2	3	4	5	6	7	8
Strain *A*	23	39	19	43	33	29	28	42
Strain *B*	18	33	21	34	33	20	21	40

(a) Is there strong evidence that strain *A* has a higher mean yield than strain *B*? Test at $\alpha = .05$.

(b) What should be randomized in this experiment and how?

10.42 Refer to the problem stated in Exercise 10.41, but now suppose that the study was conducted at 16 farms, of which 8 were selected for planting strain *A* and the other 8 for strain *B*. Here also the plots used were all of equal sizes. Recorded below are the data of yields in pounds per plot.

Strain *A*	23	39	19	28	42	43	33	29
Strain *B*	20	21	40	34	33	18	33	21

(a) Is there strong evidence that strain *A* has a higher mean yield than strain *B*? State the assumptions you make and use $\alpha = .05$.

(b) What should be randomized in this experiment and how?

(c) Check to see that each data set here is just a scrambled form of the data set in Exercise 10.41. Briefly explain why the conclusion of your test is different in the two situations.

10.43 Laser guns for detecting speeders require an accurate measurement of distance. In one daily calibration test, police measure the known distances 164 feet and 104 feet. The difference in readings should be 60 but there is some variation. The differences for ten guns are (courtesy of Madison Police Department)

60.6 60.4 60.4 60.5 60.3 60.4 60.2 60.1 60.4 60.4

(a) Do the data provide strong evidence that the mean difference is not equal to 60 feet?

(b) Construct a 95% confidence interval for the mean difference.

(c) Check that all of the values for mean distance in the confidence interval, as well as all ten observations, are less than 1 percent of the exact 60 feet. Laser guns are much more accurate than the old radar guns.

8. CHOOSING BETWEEN INDEPENDENT SAMPLES AND A MATCHED PAIRS SAMPLE

When planning an experiment to compare two treatments, we often have the option of either designing two independent samples or designing a sample with paired observations. Therefore, some comments about the pros and cons of these two sampling methods are in order here. Because a paired sample with n pairs of observations contains $2n$ measurements, a comparable situation would be two independent samples with n observations in each. First, note that the sample mean difference is the same whether or not the samples are paired. This is because

$$\overline{D} = \frac{1}{n}\sum(X_i - Y_i) = \overline{X} - \overline{Y}$$

Therefore, using either sampling design, the confidence intervals for the difference between treatment effects have the common form

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2}\,(\text{estimated standard error})$$

However, the estimated standard error as well as the degrees of freedom for t are different between the two situations.

	Independent Samples $(n_1 = n_2 = n)$	Paired Sample (n Pairs)
Estimated standard error	$S_{\text{pooled}}\sqrt{\dfrac{1}{n} + \dfrac{1}{n}}$	$\dfrac{S_D}{\sqrt{n}}$
d.f. of t	$2n - 2$	$n - 1$

Because the length of a confidence interval is determined by these two components, we now examine their behavior under the two competing sampling schemes.

Paired sampling results in a loss of degrees of freedom and, consequently, a larger value of $t_{\alpha/2}$. For instance, with a paired sample of $n = 10$, we have $t_{.05} = 1.833$ with d.f. $= 9$. But the t value associated with independent samples, each of size 10, is $t_{.05} = 1.734$ with d.f. $= 18$. Thus, if the estimated standard errors are equal, then a loss of degrees of freedom tends to make confidence intervals larger for paired samples. Likewise, in testing hypotheses, a loss of degrees of freedom for the t test results in a loss of power to detect real differences in the population means.

The merit of paired sampling emerges when we turn our attention to the other component. If experimental units are paired so that an interfering factor is held nearly constant between members of each pair, the treatment responses X and Y within each pair will be equally affected by this factor. If the prevailing condition in a pair causes the X measurement to be large, it will also cause the corresponding Y measurement to be large and vice versa. As a result, the variance of the difference $X - Y$ will be smaller in the case of an effective pairing than it will be in the case of independent random variables. The estimated standard deviation will be typically smaller as well. With an effective pairing, the reduction in the standard deviation usually more than compensates for the loss of degrees of freedom.

In Example 14, concerning the effect of a pill in reducing blood pressure, we note that a number of important factors (age, weight, height, general health, etc.) affect a person's blood pressure. By measuring the blood pressure of the same person before and after use of the pill, these influencing factors can be held nearly constant for each pair of measurements. On the other hand, independent samples of one group of persons using the pill and a separate control group of persons not using the pill are apt to produce a greater variability in blood-pressure measurements if all the persons selected are not similar in age, weight, height, and general health.

In summary, paired sampling is preferable to independent sampling when an appreciable reduction in variability can be anticipated by means of pairing. When the experimental units are already alike or their dissimilarities cannot be linked to identifiable factors, an arbitrary pairing may fail to achieve a reduction in variance. The loss of degrees of freedom will then make a paired comparison less precise.

9. COMPARING TWO POPULATION PROPORTIONS

We are often interested in comparing two populations with regard to the rate of incidence of a particular characteristic. Comparing the jobless rates in two cities, the percentages of female employees in two categories of jobs, and infant mortality in two ethnic groups are just a few examples. Let p_1 denote the proportion of members possessing the characteristic in Population 1 and p_2 that in Population 2. Our goals in this section are to construct confidence intervals for $p_1 - p_2$ and test $H_0 : p_1 = p_2$, the null hypothesis that the rates are the same for two populations. The methods would also apply to the problems of comparison between two treatments, where the response of a subject falls into one of two possible categories that we may technically call "success" and "failure." The success rates for the two treatments can then be identified as the two population proportions p_1 and p_2.

The form of the data is displayed in Table 2, where X and Y denote the numbers of successes in independent random samples of sizes n_1 and n_2 taken from Population 1 and Population 2, respectively.

TABLE 2 Independent Samples from Two Dichotomous Populations

	No. of Successes	No. of Failures	Sample Size
Population 1	X	$n_1 - X$	n_1
Population 2	Y	$n_2 - Y$	n_2

ESTIMATION

The population proportions of successes p_1 and p_2 are estimated by the corresponding sample proportions

$$\hat{p}_1 = \frac{X}{n_1} \qquad \text{and} \qquad \hat{p}_2 = \frac{Y}{n_2}$$

Naturally, $\hat{p}_1 - \hat{p}_2$ serves to estimate the difference $p_1 - p_2$. Its standard error is given by

$$\text{S.E.}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. This formula of the standard error stems from the fact that because $\hat{p}_1$ and $\hat{p}_2$ are based on independent samples, the variance of their difference equals the sum of their individual variances.

We can calculate the estimated standard error of $\hat{p}_1 - \hat{p}_2$ by using the above expression with the population proportions replaced by the corresponding sample proportions. Moreover, when n_1 and n_2 are large, the estimator $\hat{p}_1 - \hat{p}_2$ is approximately normally distributed. Specifically,

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\text{Estimated standard error}} \quad \text{is approximately } N(0, 1)$$

and this can be the basis for constructing confidence intervals for $p_1 - p_2$.

Large Samples Confidence Interval for $p_1 - p_2$

An approximate $100(1 - \alpha)\%$ confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

provided the sample sizes n_1 and n_2 are large.

Example 15 A Confidence Interval for a Difference in Success Rates

An investigation comparing a medicated patch with the unmedicated control patch for helping smokers quit the habit was discussed on page 88. At the end of the study, the number of persons in each group who were abstinent and who were smoking are repeated in Table 3.

TABLE 3 Quitting Smoking

	Abstinent	Smoking	Total
Medicated patch	21	36	57
Unmedicated patch	11	44	55
	32	80	112

Determine a 95% confidence interval for the difference in success probabilities.

SOLUTION Let p_1 and p_2 denote the probabilities of quitting smoking with the medicated and unmedicated patches, respectively. We calculate

$$\hat{p}_1 = \frac{21}{57} = .3684 \qquad \hat{p}_2 = \frac{11}{55} = .2000$$

$$\hat{p}_1 - \hat{p}_2 = .1684$$

$$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = \sqrt{\frac{.3684 \times .6316}{57} + \frac{.2000 \times .8000}{55}} = .0836$$

A 95% confidence interval for $p_1 - p_2$ is

$$(.1684 - 1.96 \times .0836, .1684 + 1.96 \times .0836) \quad \text{or} \quad (.005, .332)$$

The confidence interval only covers positive values so we conclude that the success rate with the medicated patch is .005 to .332 higher than for the control group that received the untreated patches. The lower value is so close to 0 that it is still plausible that the medicated patch is not very effective.

Note: A confidence interval for each of the population proportions p_1 and p_2 can be determined by using the method described in Section 5 of Chapter 8. For instance, with the data of Example 15, a 90% confidence interval for p_1 is calculated as

$$\hat{p}_1 \pm 1.645\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = .3684 \pm 1.645\sqrt{\frac{.3684 \times .6316}{57}}$$
$$= .3684 \pm .1051 \quad \text{or} \quad (.263, .474)$$

TESTING STATISTICAL HYPOTHESES

In order to formulate a test of $H_0 : p_1 = p_2$ when the sample sizes n_1 and n_2 are large, we again turn to the fact that $\hat{p}_1 - \hat{p}_2$ is approximately normally distributed. But now we note that under H_0 the mean of this normal distribution is $p_1 - p_2 = 0$ and the standard deviation is

$$\sqrt{pq}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where p stands for the common probability of success $p_1 = p_2$ and $q = 1 - p$. The unknown common p is estimated by pooling information from the two samples. The proportion of successes in the combined sample provides

$$\text{Pooled estimate } \hat{p} = \frac{X + Y}{n_1 + n_2}$$

$$\text{or, alternatively,} \qquad \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Also, $$\text{Estimated S.E.}(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}\,\hat{q}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

In summary,

Testing $H_0 : p_1 = p_2$ with Large Samples

Test statistic:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{where } \hat{p} = \frac{X + Y}{n_1 + n_2}$$

The level α rejection region is $|Z| \geq z_{\alpha/2}$, $Z \leq -z_\alpha$, or $Z \geq z_\alpha$ according to whether the alternative hypothesis is $p_1 \neq p_2$, $p_1 < p_2$, or $p_1 > p_2$.

Example 16 Testing Equality of Prevalence of a Virus

A study (courtesy of R. Golubjatnikov) is undertaken to compare the rates of prevalence of CF antibody to parainfluenza I virus among boys and girls in the age group 5 to 9 years. Among 113 boys tested, 34 are found to have the antibody; among 139 girls tested, 54 have the antibody. Do the data provide strong evidence that the rate of prevalence of the antibody is significantly higher in girls than boys? Use $\alpha = .05$. Also, find the P–value.

SOLUTION Let p_1 denote the population proportion of boys who have the CF antibody and p_2 the population proportion of girls who have the CF antibody. Because we are looking for strong evidence in support of $p_1 < p_2$, we formulate the hypotheses as

$$H_0 : p_1 = p_2 \qquad \text{versus} \qquad H_1 : p_1 < p_2$$

or equivalently as

$$H_0 : p_1 - p_2 = 0 \qquad \text{versus} \qquad H_0 : p_1 - p_2 < 0$$

The sample sizes $n_1 = 113$ and $n_2 = 139$ being large, we will employ the test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\,\hat{q}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

and set a left-sided rejection region in view of the fact that H_1 is left-sided. With $\alpha = .05$, the rejection region is $R : Z \leq -1.645$. We calculate

$$\hat{p}_1 = \frac{34}{113} = .301 \qquad \hat{p}_2 = \frac{54}{139} = .388$$

$$\text{Pooled estimate } \hat{p} = \frac{34 + 54}{113 + 139} = .349$$

The observed value of the test statistic is then

$$z = \frac{.301 - .388}{\sqrt{.349 \times .651}\sqrt{\dfrac{1}{113} + \dfrac{1}{139}}} = -1.44$$

Because the value $z = -1.44$ is not in R, we do not reject H_0. Consequently, the assertion that the girls have a higher rate of prevalence of the CF antibody than boys is not substantiated at the level of significance $\alpha = .05$.

The significance probability of the observed z is

$$\begin{aligned} P\text{–value} &= P[Z \le -1.44] \\ &= .0749 \end{aligned}$$

This means that we must allow an α of at least .0749 in order to consider the result significant.

Exercises

10.44 Refer to the measurement of job satisfaction in Example 4. Rather than average score, researchers often prefer the proportion who respond "very satisfied." Using the data in Exercise 10.10, compare the proportion for firefighters, p_1, with the proportion for office supervisors, p_2.

(a) Find a 95% confidence interval for the difference of proportions $p_1 - p_2$.

(b) Perform the Z test for the null hypothesis $H_0: p_1 = p_2$ versus $H_1: p_1 < p_2$. Take $\alpha = .05$

10.45 Refer to the measurement of job satisfaction in Exercise 10.44. Using the data in Exercise 10.10, compare the proportion for clergy, p_1, with the proportion for office supervisors, p_2.

(a) Find a 95% confidence interval for the difference of proportions $p_1 - p_2$.

(b) Perform the Z test for the null hypothesis $H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$. Take $\alpha = .05$

10.46 In a comparative study of two new drugs, A and B, 120 patients were treated with drug A and 150 patients with drug B, and the following results were obtained.

	Drug A	Drug B
Cured	50	88
Not cured	70	62
Total	120	150

(a) Do these results demonstrate a significantly higher cure rate with drug B than drug A? Test at $\alpha = .05$.

(b) Construct a 95% confidence interval for the difference in the cure rates of the two drugs.

10.47 In a study of the relationship between temperament and personality, 49 female high school students who had a high level of reactivity (HRL) and 54 students who had a low level of reactivity (LRL) were classified according to their attitude to group pressure with the following results.

	Attitude		
Reactivity	Submissive	Resistant	Total
HRL	34	15	49
LRL	12	42	54

Is resistance to group pressure significantly lower in the HRL group than the LRL group? Answer by calculating the P–value.

10.48 Refer to the data in Exercise 10.47. Determine a 99% confidence interval for the difference between the proportions of resistant females in the HRL and LRL populations.

10.49 Refer to Exercise 10.17 concerning a study on the relationship between child abuse and later criminal behavior. Suppose that from follow-ups of 85 boys who were abused in their preschool years and 120 boys who were not abused, it was found that 21 boys in the abused group and 11 boys in the nonabused group were chronic offenders in their teens. Do these data substantiate the conjecture that abused boys are more prone to be chronic offenders than nonabused boys? Test at $\alpha = .01$.

10.50 Referring to the data of Exercise 10.49, determine a 95% confidence interval for the difference between the true proportions of chronic offenders in the populations of abused and nonabused boys.

10.51 The popular disinfectant Listerine is named after Joseph Lister, a British physician who pioneered the use of antiseptics. Lister conjectured that human infections might have an organic origin and thus could be prevented by using a disinfectant. Over a period of several years, he performed 75 amputations: 40 using carbolic acid as a disinfectant and 35 without any disinfectant. The following results were obtained.

	Patient Survived	Patient Died	Total
With carbolic acid	34	6	40
Without carbolic acid	19	16	35

Are the survival rates significantly different between the two groups? Test at $\alpha = .05$ and calculate the P–value.

10.52 Referring to the data of Exercise 10.51, calculate a 95% confidence interval for the difference between the survival rates for the two groups.

10.53 Random samples of 250 persons in the 30- to 40-year age group and 250 persons in the 60- to 70-year age group are asked about the average number of hours they sleep per night, and the following summary data are recorded.

Age	Hours of Sleep ≤ 8	Hours of Sleep > 8	Total
30–40	173	77	250
60–70	120	130	250
Total	293	207	500

Do these data demonstrate that the proportion of persons who have ≤ 8 hours of sleep per night is significantly higher for the age group 30 to 40 than that for the age group 60 to 70? Answer by calculating the P–value.

10.54 Referring to Exercise 10.53, denote by p_1 and p_2 the population proportions in the two groups who have ≤ 8 hours of sleep per night. Construct a 95% confidence interval for $p_1 - p_2$.

10.55 A medical researcher conjectures that smoking can result in wrinkled skin around the eyes. By observing 150 smokers and 250 nonsmokers, the researcher finds that 95 of the smokers and 103 of the nonsmokers have prominent wrinkles around their eyes.

(a) Do these data substantiate the belief that prominent wrinkles around eyes are more prevalent among smokers than nonsmokers? Answer by calculating the P–value.

(b) If the results are statistically significant, can the researcher readily conclude that smoking causes wrinkles around the eyes? Why or why not?

10.56 In the survey on which Table D.12 of the Data Bank is based, a larger number of persons was asked to respond to the statement "I would characterize my political beliefs as liberal" on a seven point Likert scale from strongly disagree (1) to strongly agree (7). A count of those who strongly disagree, for each of three age groups, yielded the summary statistics.

Age	x_i	n_i	p_i
19–24	3	47	.0638
25–34	13	64	.2031
35–44	14	93	.1505

(a) Test equality of the proportions for the second two age groups versus a two-sided alternative. Take $\alpha = .05$.

(b) Test equality of the proportions for the first two age groups versus a two-sided alternative. Take $\alpha = .05$.

(c) Explain why the large samples test in Part (b) is not valid.

10.57 A major clinical trial of a new vaccine for type-B hepatitis was conducted with a high-risk group of 1083 male volunteers. From this group, 549 men were given the vaccine and the other 534 a placebo. A follow-up of all these individuals yielded the data:

	Follow-up		
	Got Hepatitis	Did Not Get Hepatitis	Total
Vaccine	11	538	549
Placebo	70	464	534

(a) Do these observations testify that the vaccine is effective? Use $\alpha = .01$.

(b) Construct a 95% confidence interval for the difference between the incidence rates of hepatitis among the vaccinated and nonvaccinated individuals in the high-risk group.

10.58 Records of drivers with a major medical condition (diabetes, heart condition, or epilepsy) and also a group of drivers with no known health conditions were retrieved from a motor vehicle department. Drivers in each group were classified according to their driving record in the last year.

	Traffic Violations		
Medical Condition	None	One or More	Total
Diabetes	119	41	160
Heart condition	121	39	160
Epilepsy	72	78	150
None (control)	157	43	200

Let p_D, p_H, p_E, and p_C denote the population proportions of drivers having one or more traffic violations in the last year for the four groups "diabetes," "heart condition," "epilepsy," and "control," respectively.

(a) Test $H_0 : p_D = p_C$ versus $H_1 : p_D > p_C$ at $\alpha = .10$.

(b) Is there strong evidence that p_E is higher than p_C? Answer by calculating the $P-$value.

10.59 Refer to Exercise 10.58.

(a) Construct a 95% confidence interval for $p_E - p_H$.

(b) Construct a 90% confidence interval for $p_H - p_C$.

(c) Construct 95% confidence intervals for p_D, p_H, p_E, and p_C, individually.

USING STATISTICS WISELY

1. In all cases where a sample size is small and normality is assumed, the data should be graphed in a dot plot to reveal any obvious outliers which could invalidate the inferences based on the normal theory.

2. When comparing two treatments under a matched pairs design, whenever possible, assign the treatments within each pair at random. To analyze the resulting data, use the results for one sample but applied to the differences from each matched pair. For instance, if the difference of paired measurements

has a normal distribution, determine a $100(1 - \alpha)\%$ confidence interval for the mean difference μ_D as

$$\left(\bar{d} - t_{\alpha/2} \frac{S_D}{\sqrt{n}}, \ \bar{d} + t_{\alpha/2} \frac{S_D}{\sqrt{n}} \right)$$

where $t_{\alpha/2}$ is based on $n - 1$ degrees of freedom. Otherwise, with large samples use $z_{\alpha/2}$.

3. When comparing two treatments using the independent samples design, randomly assign the treatments to groups whenever possible. With the matched pairs design, randomly assign the treatments within each pair.

4. When sample sizes are large, determine the limits of a $100(1 - \alpha)\%$ confidence interval for the difference of means $\mu_1 - \mu_2$ as

$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

5. When each of the two samples are from normal populations, having the same variance, determine the limits of a $100(1 - \alpha)\%$ confidence interval for the difference of means $\mu_1 - \mu_2$ as

$$\bar{x} - \bar{y} \pm t_{\alpha/2}\, s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where the pooled estimate of variance

$$s^2_{\text{pooled}} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

and $t_{\alpha/2}$ is based on $n_1 + n_2 - 2$ degrees of freedom.

6. Do not pool the two sample variances s_1^2 and s_2^2 if they are very different. We suggest a factor of 4 as being too different. There are alternative procedures including the conservative procedure on page 400 and the approximate t on page 402.

KEY IDEAS AND FORMULAS

In any comparative study of treatments, products, methods, and so on, the term **treatment** refers to the things being compared. The basic unit or object, to which one of the treatments is applied, is called an **experimental unit** or an **experimental subject**. The **response variable** is the characteristic that is recorded on each unit.

The specification of which treatment to compare and method of assigning experimental units is called the **experimental** (or **sampling**) **design**. The choice of appropriate statistical methods for making inferences depends heavily on the experimental design chosen for data collection.

A carefully designed experiment is fundamental to the success of a comparative study.

The most basic experimental designs to compare two treatments are **independent samples** and **matched pairs sample.**

The **independent samples design** require the subjects to be randomly selected for assignment to each treatment. **Randomization** prevents uncontrolled factors from systematically favoring one treatment over the other.

With a **matched pairs design,** subjects in each pair are alike, while those in different pairs may be dissimilar. For each pair, the two treatments should be randomly allocated to the members.

The idea of **matching** or **blocking** experimental units is to remove a known source of variation from comparisons. Pairing subjects according to some feature prevents that source of variation from interfering with treatment comparisons. By contrast, random allocation of subjects according to the independent random sampling design spreads these variations between the two treatments.

Inferences with Two Independent Random Samples

1. ***Large samples.*** When n_1 and n_2 are both greater than 30, inferences about $\mu_1 - \mu_2$ are based on the fact that

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \quad \text{is approximately } N(0, 1)$$

A $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

To test $H_0 : \mu_1 - \mu_2 = \delta_0$, we use the normal test statistic

$$Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

No assumptions are needed in regard to the shape of the population distributions.

2. ***Small samples.*** When n_1 and n_2 are small, inferences using the t distribution require the assumptions:

(a) Both populations are normal.

(b) $\sigma_1 = \sigma_2$.

The common σ^2 is estimated by

$$S_{\text{pooled}}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Inferences about $\mu_1 - \mu_2$ are based on

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{d.f.} = n_1 + n_2 - 2$$

A $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2}\, S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

To test $H_0 : \mu_1 - \mu_2 = \delta_0$, the test statistic is

$$T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{d.f.} = n_1 + n_2 - 2$$

Inferences with a Matched Pair Sample

With a paired sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, the first step is to calculate the differences $D_i = X_i - Y_i$, their mean $\overline{D}$, and standard deviation S_D.

If n is small, we assume that the D_i's are normally distributed $N(\mu_D, \sigma_D)$. Inferences about μ_D are based on

$$T = \frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \qquad \text{d.f.} = n - 1$$

A $100(1 - \alpha)\%$ confidence interval for μ_D is

$$\overline{D} \pm t_{\alpha/2}\, S_D / \sqrt{n}$$

The test of $H_0 : \mu_D = \mu_D$ is performed with the test statistic:

$$\mu_D = \frac{\overline{D} - \mu_{D0}}{S_D / \sqrt{n}} \qquad \text{d.f.} = n - 1$$

If n is large, the assumption of normal distribution for the D_i's is not needed. Inferences are based on the fact that

$$Z = \frac{\overline{D} - \mu_{D0}}{S_D / \sqrt{n}} \qquad \text{is approximately } N(0, 1)$$

Summary of Inferences about Means

Table 4 summarizes all of the statistical procedures we have considered for making inferences about (1) a single mean, (2) the difference of two means, or (3) a mean difference for a pair of observations.

TABLE 4 General Formulas for Inferences about a Mean (μ), Difference of Two Means ($\mu_1 - \mu_2$)

Confidence interval = Point estimator $\pm$ (Tabled value)(Estimated or true std. dev.)

$$\textbf{Test statistic} = \frac{\text{Point estimator} - \text{Parameter value at } H_0 \text{ (null hypothesis)}}{\text{(Estimated or true) std. dev. of point estimator}}$$

	Ch. 8	Ch. 9	Ch. 10 Independent Samples			Matched Samples
Population (s)	General	Normal with unknown σ	Normal $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2)$ $\sigma_1 = \sigma_2 = \sigma$	Normal $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2)$ $\sigma_1 \neq \sigma_2$	General	Normal for the difference $D_i = X_i - Y_i$
Inference on	Mean μ	Mean μ	$\mu_1 - \mu_2$	$\mu_1 - \mu_2$	$\mu_1 - \mu_2$	$\mu_D = \mu_1 - \mu_2$
Sample(s)	$X_1, \ldots, X_n$	$X_1, \ldots, X_n$	$X_1, \ldots, X_{n_1}$ $Y_1, \ldots, Y_{n_2}$	$X_1, \ldots, X_{n_1}$ $Y_1 \ldots, Y_{n_2}$	$X_1, \ldots, X_{n_1}$ $Y_1, \ldots, Y_{n_2}$	$X_1\ Y_1 \quad D_1 = X_1 - Y_1$ $X_2\ Y_2 \quad D_2 = X_2 - Y_2$ $\vdots\ \ \vdots \qquad \vdots$ $Xn\ Yn \quad Dn = Xn - Yn$
Sample size n	Large $n > 30$	$n \geq 2$	$n_1 \geq 2$ $n_2 \geq 2$	$n_1 \geq 2$ $n_2 \geq 2$	$n_1 > 30$ $n_2 > 30$	$n \geq 2$
Point estimator	$\overline{X}$	$\overline{X}$	$\overline{X} - \overline{Y}$	$\overline{X} - \overline{Y}$	$\overline{X} - \overline{Y}$	$\overline{D} = \overline{X} - \overline{Y}$
Variance of point estimator	$\dfrac{\sigma^2}{n}$	$\dfrac{\sigma^2}{n}$	$\sigma^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$	$\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$	$\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$	$\dfrac{\sigma_D^2}{n}$
Std. dev. of point estimator	$\dfrac{\sigma}{\sqrt{n}}$	$\dfrac{\sigma}{\sqrt{n}}$	$\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$\dfrac{\sigma_D}{\sqrt{n}}$
Estimated std. dev.	$\dfrac{S}{\sqrt{n}}$	$\dfrac{S}{\sqrt{n}}$	$S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$	$\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$	$\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$	$\dfrac{S_D}{\sqrt{n}}$
Distribution	Normal	t with d.f. $= n - 1$	t with d.f. $= n_1 + n_2 - 2$	t with[3] d.f. = smaller of $n_1 - 1$ and $n_2 - 1$	Normal	t with d.f. $= n - 1$
Test statistic	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\dfrac{(\overline{X} - \overline{Y}) - \delta_0}{S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $S^2_{\text{pooled}} = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	$\dfrac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$\dfrac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$\dfrac{\overline{D} - \mu_{D0}}{S_D/\sqrt{n}}$ S_D = sample std. dev. of the D_i's

[3] or estimated by $\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2 \Bigg/ \left[\dfrac{(s_1^2\,/\,n_1)^2}{(n_1 - 1)} + \dfrac{(s_2^2\,/\,n_2)^2}{(n_2 - 1)}\right]^2$

Comparing Two Binomial Proportions—Large Samples

Data:

$$X = \text{No. of successes in } n_1 \text{ trials with success probability } P(S) = p_1$$
$$Y = \text{No. of successes in } n_2 \text{ trials with success probability } P(S) = p_2$$

To test $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$, use the Z test:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{with} \qquad R: |Z| \geq z_{\alpha/2}$$

where

$$\hat{p}_1 = \frac{X}{n_1} \qquad \hat{p}_2 = \frac{Y}{n_2} \qquad \hat{p} = \frac{X + Y}{n_1 + n_2}$$

To test $H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$, use the Z test with $R: Z \geq z_\alpha$.
A $100(1 - \alpha)\%$ confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

TECHNOLOGY

Confidence intervals and tests for comparing means

MINITAB

Matched pair samples

We illustrate with the calculation of a 98% confidence interval and .02 level test.

Enter the first sample in C*1* and second sample in C*2*.

Dialog box:

Stat > Basic Statistics > Paired t.
Type C*1* in **First Sample** and C*2* in **Second Sample.**
Click **Options.** Type *98* in **Confidence level.**
Enter the value of the mean under the null hypothesis 0 in **Test mean** and choose *not equal* under the **Alternative.** Click **OK.** Click **OK.**

Two-sample t tests and confidence intervals

We illustrate with the calculation of a 98% confidence interval for $\mu_1 - \mu_2$ and .02 level test of the null hypothesis of no difference in means.

Data:

C1 First sample
C2 Second sample

Dialog box:
Stat > **Basic Statistics** > **2-Sample t.**
Select **Samples in different columns.** Type *C1* in **First** and *C2* in **Second.**
To pool the estimates of variance, click the box **Assume equal variances.**
Click **Options.** Type *98* in **Confidence level,** the null hypothesis value of the mean 0 in **Test mean,** and select *not equal* under **Alternative**. Click **OK.** Click **OK.**

EXCEL

Matched pairs t tests

We illustrate with the calculation of a test of the null hypothesis $H_0 : \mu_D = 0$ versus a two-sided alternative. Begin with values for the first variable in column A and the second in column B.

Select **Tools** and then **Data Analysis.**
Select **t-Test: Paired Two-Sample for Means.** Click **OK.**
With the cursor in **Variable 1 Range,** highlight the data in column A.
With the cursor in **Variable 2 Range,** highlight the data in column B.
Type the hypothesized value *0* after **Hypothesized mean difference.**
Click **OK.**

The program returns a summary that includes the value of the t statistic, the $P-$value for a one-sided test (actually the smallest tail probability), and the $P-$value for a two-sided test.

Two-sample t tests

We illustrate with a test of the null hypothesis of no difference between the two means. Begin with the first sample in column A and the second in column B.

Select **Tools** and then **Data Analysis.**
Select **t-Test: Two-Sample Assuming Unequal Variances.** Click **OK** or, to pool, Select **t-Test: Two-Sample Assuming Equal Variances** and click **OK.**
With the cursor in **Variable 1 Range,** highlight the data in column A.
With the cursor in **Variable 2 Range,** highlight the data in column B.
Type the hypothesized value *0* after **Hypothesized mean difference.**
Enter *.02* for **Alpha.** Click **OK.**

The program returns the one-sided and two-sided $P-$values.

TI-84/-83 PLUS

Matched Pairs Samples

Confidence intervals

We illustrate the calculation of a 98% confidence interval. Start with the values for the first variable entered in **L1** and the second in **L2.**

Press **STAT** and select **TESTS** and then **8: Tinterval.**
Select **Data** with **List** set to **L3** and **Freq** to **1.**
Enter *.98* following C*–Level:* Select **Calculate.**
Then press **ENTER.**

Tests

We illustrate the calculation of a test of the null hypothesis of 0 mean difference. Start with the first sample entered in **L1** and the second in **L2.** Then let **L3** = **L2** − **L1** or **L3** = **L1** − **L2** depending on how the alternative is defined.

Press **STAT** and select **TESTS** and then **2: T-Test.**
Select **Data** with **List** set to **L3** and **Freq** to **1.**
Enter *0* for μ_0. Select the direction of the alternative hypothesis for the mean difference. Select **Calculate.**
Press **ENTER.**

If, instead, the sample size, mean, and standard deviation are available, the second step is

Select **Stats** and type in the sample sizes, means, and standard deviations.

Two-sample t tests

We illustrate with the calculation of a test of the null hypothesis of no difference between the two means. Start with the data entered in **L1** and **L2.**

Press **STAT** and select **TESTS** and then **4: 2-SampT-Test.**
Select **Data** with **List1** set to **L1, List2 to L2, FREQ1** to **1,** and **Freq2** to **1.**
Select the direction of the alternative hypothesis. Set **Pooled** to *NO* if you do not wish to pool.
Select **Calculate** or **Draw** and press **ENTER.**

The calculator will return the *P*–value. **Draw** will draw the *t* distribution and shade the area of the *P*–value. If, instead, the sample sizes, means, and standard deviations are available, the second step is

Select **Stats** and enter the sample sizes, means, and standard deviations.

10. REVIEW EXERCISES

10.60 Refer to Table D.12 of the Data Bank. The responses, of persons in the first and second age groups, on the frequency of credit card use question, have the summary statistics

Age	n_i	$\bar{x}_i$	s_i
19–24	18	3.000	1.609
25–34	28	4.429	1.643

(a) Obtain a 98% confidence interval for the difference in the two population mean frequency of credit card use scores.

(b) Test for equality of the two means versus a two-sided alternative. Use $\alpha = .02$.

10.61 Two versions of a new simplified tax form, A1 and A2, need to be evaluated with respect to the time, in hours, it takes to complete the form. Forty persons were selected to fill out Form A1 and 40 were selected to fill out Form A2. The summary statistics are:

Form A1: $\bar{x} = 12.2, \quad s_1 = 1.1$

Form A2: $\bar{y} = 7.2, \quad s_2 = 3.4$

(a) Find a 95% confidence interval for the difference in means.

(b) If only 80 persons are available for this study, how would you choose the 40 to use form A1?

10.62 Refer to Table D.12 of the Data Bank. The responses, of persons in the first and third age groups to the average monthly amount charged, have the summary statistics

Age	n_i	$\bar{x}_i$	s_i
19–24	18	177.2	235.0
35–44	48	563.9	831.3

Determine a 95% confidence interval for the difference of means.

10.63 Refer to the data in Exercise 10.62. Suppose you wish to establish that the mean response of 35–44 olds is larger than that of 19–24 year olds by more than 150 dollars.

(a) Formulate the null hypothesis and the alternative hypothesis.

(b) State the test statistic and the rejection region with $\alpha = .05$.

(c) Perform the test at $\alpha = .05$. Also, find the P–value and comment.

10.64 A study of postoperative pain relief is conducted to determine if drug A has a significantly longer duration of pain relief than drug B. Observations of the hours of pain relief are recorded for 55 patients given drug A and 58 patients given drug B. The summary statistics are

	A	B
Mean	4.64	4.03
Standard deviation	1.25	1.82

(a) Formulate H_0 and H_1.

(b) State the test statistic and the rejection region with $\alpha = .10$.

(c) State the conclusion of your test with $\alpha = .10$. Also, find the P–value and comment.

10.65 Consider the data of Exercise 10.64.

(a) Construct a 90% confidence interval for $\mu_A - \mu_B$.

(b) Give a 95% confidence interval for μ_A using the data of drug A alone. (*Note:* Refer to Chapter 8.)

10.66 Obtain s^2_{pooled} for the gaming data in the chapter opening.

10.67 Given the following two samples,

8 11 5 9 7 and 5 3 4 8

obtain (a) s^2_{pooled} and (b) the value of the t statistic for testing $H_0 : \mu_1 - \mu_2 = 2$. State the d.f. of the t.

10.68 A fruit grower wishes to evaluate a new spray that is claimed to reduce the loss due to damage by insects. To this end, he performs an experiment with 27 trees in his orchard by treating 12 of those trees with the new spray and the other 15 trees with the standard spray. From the data of fruit yield (in pounds) of those trees, the following summary statistics were found.

	New Spray	Standard Spray
Mean yield	249	233
Standard deviation	19	45

Do these data substantiate the claim that a higher yield should result from the use of the new spray? State the assumptions you make and test at $\alpha = .05$.

10.69 Referring to Exercise 10.68, construct a conservative 95% confidence interval for the difference in mean yields between the new spray and the standard spray.

10.70 Refer to Exercise 10.61 but suppose, instead, that the sample sizes are 8 and 7 with the same summary statistics:

Form A1: $\bar{x} = 12.2, \quad s_1 = 1.1$

Form A2: $\bar{y} = 7.2, \quad s_2 = 3.4$

(a) Test, with the conservative procedure, that the population means are different. Use $\alpha = .05$.

(b) Test, with the approximate t procedure, that the population means are different. Use $\alpha = .05$.

(c) What assumptions did you make for your answer to Parts (a) and (b)?

10.71 An investigation is conducted to determine if the mean age of welfare recipients differs between two cities A and B. Random samples of 75 and 100 welfare recipients are selected from city A and city B, respectively, and the following computations are made.

	City A	City B
Mean	37.8	43.2
Standard deviation	6.8	7.5

(a) Do the data provide strong evidence that the mean ages are different in city A and city B? (Test at $\alpha = .02$.)

(b) Construct a 98% confidence interval for the difference in mean ages between A and B.

(c) Construct a 98% confidence interval for the mean age for city A and city B individually. (*Note:* Refer to Chapter 8.)

10.72 The following generic computer output summarizes the data, given in Table D.6 of the Data Bank, on the neck size of male and female bears.

```
Sex    N    Mean   StDev
F     36   52.92    8.83
M     25   59.70    17.5
```

Test for equality of mean neck size versus a two-sided alternative. Take $\alpha = .03$.

10.73 Refer to the computer attitude scores (CAS) of students given in Table D.4 of the Data Bank. A computer analysis produces the output

```
Two-sample T for CAS
sex    N    Mean    StDev
F     15   2.643    0.554
M     20   2.945    0.390

Difference  = mu M  -  mu F
Estimate for difference: 0.302
90\% CI for the difference
(0.032, 0.571)
T-Test of difference  = 0
(vs not  =):
T-Value  = 1.89 P-Value  = 0.067
DF  = 33
Both use Pooled StDev  = 0.467
```

(a) What is the conclusion to testing the equality of mean computer attitude scores at level $\alpha = .05$?

(b) Find a 95% confidence interval for the mean attitude score for males minus the mean attitude score for females.

(c) Test the null hypothesis that, on average, males score .1 lower than females against a two-sided alternative. Take $\alpha = .05$.

10.74 In each of the following cases, how would you select the experimental units and conduct the experiment—matched pairs or independent samples?

(a) Compare the mileage obtained from two gasolines. Twelve SUVs of various sizes are available.

(b) Compare the drying times of two latex-based interior paints. Ten walls are available.

(c) Compare two methods of teaching swimming. Twenty five-year-old girls are available.

10.75 A sample of river water is divided into two specimens. One is randomly selected to be sent to Lab A and the other is sent to Lab B. This is repeated for a total of nine times. The measurement of suspended solids at Lab B is subtracted from that of Lab A to obtain the differences

12 10 15 42 11 -4 -2 10 -7

(a) Is there strong evidence that the mean difference is not zero? Test with $\alpha = .02$.

(b) Construct a 90% confidence interval for the mean difference of the suspended solids measurements.

10.76 Two scales are available at a campus athletic facility. A student wonders if, on average, they give the same reading for weight. She and four others weigh themselves on both scales. The readings are

Person	Scale 1	Scale 2
1	113.8	114.1
2	218.7	217.2
3	149.2	147.3
4	104.9	103.5
5	166.6	165.7

(a) Find a 95% confidence interval for the mean difference in scale readings.

(b) Based on your answer to Part (a), what is the conclusion to testing, with $\alpha = .05$, that the mean difference is 0 versus a two-sided alternative hypothesis? Explain your reasoning.

(c) Explain how you would randomize in this experiment.

10.77 An experiment is conducted to determine if the use of a special chemical additive with a standard fertilizer accelerates plant growth. Ten locations are included in the study. At each location, two plants growing in close proximity are treated. One is given the standard fertilizer, the other the standard fertilizer with the chemical additive. Plant growth after four weeks is measured in centimeters. Do the following data substantiate the claim that use of the chemical additive accelerates plant growth? State the assumptions that you make and devise an appropriate test of the hypothesis. Take $\alpha = .05$.

	Location									
	1	2	3	4	5	6	7	8	9	10
Without additive	20	31	17	22	19	32	25	18	21	19
With additive	23	34	16	21	22	31	29	20	25	23

10.78 Referring to Exercise 10.77, suppose that the two plants at each location are situated in the east–west direction. In designing this experiment, you must decide which of the two plants at each location—the one in the east or the one in the west—is to be given the chemical additive.

(a) Explain how, by repeatedly tossing a coin, you can randomly allocate the treatments to the plants at the 10 locations.

(b) Perform the randomization by actually tossing a coin 10 times.

10.79 Students can bike to a park on the other side of a lake by going around one side of the lake or the other. After much discussion about which was faster, they decided to perform an experiment. Among the 12 students available, 6 were randomly selected to follow Path A on one side of the lake and the rest followed Path B on the other side. They all went on different days so the conclusion would apply to a variety of conditions.

	Travel Time (minutes)					
Path A	10	12	15	11	16	11
Path B	12	15	17	13	18	16

(a) Is there a significant difference between the mean travel times between the two paths? State the assumptions you have made in performing the test.

(b) Suggest an alternative design for this study that would make the comparison more effective.

10.80 Five pairs of tests are conducted to compare two methods of making rope. Each sample batch contains enough hemp to make two ropes. The tensile strength measurements are

	Test				
	1	2	3	4	5
Method 1	14	12	18	16	15
Method 2	16	15	17	16	14

(a) Treat the data as 5 paired observations and calculate a 95% confidence interval for the mean difference in tensile strengths between ropes made by the two methods.

(b) Repeat the calculation of a 95% confidence interval treating the data as independent random samples.

(c) Briefly discuss the conditions under which each type of analysis would be appropriate.

10.81 An experiment was conducted to study whether cloud seeding reduces the occurrences of hail. At a hail-prone geographical area, seeding was done on 50 stormy days and another 165 stormy days were also observed without seeding. The following counts were obtained.

	Days	
	Seeded	Not Seeded
Hail	7	43
No hail	43	122
Total	50	165

Do these data substantiate the conjecture that seeding reduces the chance of hail? (Answer by determining the $P-$value.)

10.82 Referring to the data of Exercise 10.81, calculate a 90% confidence interval for the difference between the probabilities of hail with and without seeding.

10.83 An antibiotic for pneumonia was injected into 100 patients with kidney malfunctions (called uremic patients) and 100 patients with no kidney malfunctions (called normal patients). Some allergic reaction developed in 38 of the uremic patients and 21 of the normal patients.

(a) Do the data provide strong evidence that the rate of incidence of allergic reaction to the antibiotic is higher in uremic patients than normal patients?

(b) Construct a 95% confidence interval for the difference between the population proportions.

10.84 In the survey on which Table D.12 of the Data Bank is based, a large number of persons was each asked "Is it preferable to have 1% of your monthly charges donated to a charity of your choice rather than credited back to your account?" The full sample, larger than that in Table D.12, yielded the summary statistics for the number of yes responses.

Gender	x_i	n_i	$\hat{p}_i$
Female	48	195	0.2462
Male	52	294	0.1769

(a) Do the data provide evidence that females are more willing than males to donate to charity, in terms of having a larger proportion that would say "yes" to the question? Test with $\alpha = .05$.

(b) Obtain a 95% confidence interval for the difference in population proportions.

The Following Exercises May Require a Computer

10.85 We illustrate the MINITAB commands and output for the two-sample t test. The approximate t test is performed if you don't click **Assume equal variances.**

Data

C1: 5 2 8 3
C2: 8 9 6

Dialog box:

Stat > **Basic Statistics** > **2-Sample t**

Click **Samples in different columns.**

Type *C1* in **First.** *C2* in **Second.**

Click **OK.** Click **Assume equal variances.**
Click **OK.**

```
Two-Sample T-Test and CI: C1, C2
Two-sample T for C1 vs C2

     N    Mean    StDev    SE Mean
C1   4    4.50     2.65        1.3
C2   3    7.67     1.53       0.88

Difference  =  mu (C1)  −  mu (C2)
Estimate for difference:  −3.16667
95% CI for difference:
(−7.61492, 1.28159)
T-Test of difference  =  0 (vs
not  =): T-Value  =  −1.83
P-value  =  0.127 DF  =  5
Both use Pooled StDev  =  2.2657
```

(a) From the output, what is the conclusion to testing $H_0 : \mu_1 - \mu_2 = 0$ versus a two-sided alternative at level $\alpha = .05$?

(b) Refer to Table 12 of the Data Bank. Find a 97% confidence interval for the difference of mean monthly amount charged by the first and second age groups.

10.86 Refer to the alligator data in Table D.11 of the Data Bank. Using the data on testosterone x_4 from the Lake Apopka alligators, find a 95% confidence interval for the difference of means between males and females. There should be a large difference for healthy alligators. Comment on your conclusion.

10.87 Refer to the alligator data in Table D.11 of the Data Bank. Using the data on testosterone x_4 for male alligators, compare the means for the two lake regions.

(a) Should you pool the variances with these data?

(b) Find a 90% confidence interval for the difference of means between the two lakes. Use the conservative procedure on page 400.

(c) Which population has the highest mean and how much higher is it?

10.88 Refer to the bear data in Table D.6 of the Data Bank. Compare the head widths of males and females by obtaining a 95% confidence interval for the difference of means and also test equality versus a two-sided alternative with $\alpha = .05$.

10.89 Refer to the marine growth of salmon data in Table D.7 of the Data Bank. Compare the mean growth of males and females by obtaining a 95% confidence interval for the difference of means and also test equality versus a two-sided alternative with $\alpha = .05$.

10.90 Refer to the physical fitness data in Table D.5 of the Data Bank. Find a 95% confidence interval for the mean difference of the pretest minus posttest number of situps. Also test that the mean difference is zero versus a two-sided alternative with $\alpha = .05$.

10.91 Refer to the physical fitness data in Table D.5 of the Data Bank. Find a 95% confidence interval for the mean difference of the pretest minus posttest time to complete the rowing test. Also test that the mean difference is zero versus a two-sided alternative with $\alpha = .05$.