Treatment Group	Placebo Group
$n_1 = 22$	$n_2 = 22$
$\bar{x}_1 = 0.049$	$\bar{x}_2 = 0.000$
$s_1 = 0.015$	$s_2 = 0.000$

***30.** Effect of No Variation in Sample An experiment was conducted to test the effects of alcohol. The breath alcohol levels were measured for a treatment group of people who drank ethanol and another group given a placebo. The results are given in the accompanying table. Use a 0.05 significance level to test the claim that the two sample groups come from populations with the same mean. The given results are based on data from "Effects of Alcohol Intoxication on Risk Taking, Strategy, and Error Rate in Visuomotor Performance," by Streufert et al., *Journal of Applied Psychology*, Vol. 77, No. 4.

8-4 Inferences from Matched Pairs

In Section 8-3 we defined two samples to be *independent* if the sample values selected from one population are not related to or somehow paired or matched with the sample values selected from the other population. Section 8-3 dealt with inferences about the means of two independent populations, and this section focuses on dependent samples, which we refer to as *matched pairs*. With matched pairs, there is some relationship so that each value in one sample is paired with a corresponding value in the other sample. Here are some typical examples of matched pairs:

- When conducting an experiment to test the effectiveness of a low-fat diet, the weight of each subject is measured once before the diet and once after the diet.
- In a test of the effects of a fertilizer on heights of trees, sample trees are planted in pairs, with one tree given the fertilizer treatment while the other tree is not given the treatment.

When dealing with inferences about the means of matched pairs, summaries of the relevant requirements, notation, hypothesis test statistic, and confidence interval are given below. Because the hypothesis test and confidence interval use the same distribution and standard error, they are equivalent in the sense that they result in the same conclusions. Consequently, we can test the null hypothesis that the mean difference equals zero by determining whether the confidence interval includes zero. [For two-tailed hypothesis tests construct a confidence interval with a confidence level of $1 - \alpha$; but for a one-tailed hypothesis test with significance level α, construct a confidence interval with a confidence level of $1 - 2\alpha$. (See Table 7-2 for common cases.) For example, the claim that the mean difference is greater than 0 can be tested with a 0.05 significance level by constructing a 90% confidence interval.]

Requirements

1. The sample data consist of matched pairs.
2. The samples are simple random samples.
3. Either or both of these conditions is satisfied: The number of matched pairs of sample data is large ($n > 30$) or the pairs of values have differences that

are from a population having a distribution that is approximately normal. (If there is a radical departure from a normal distribution, we should not use the methods given in this section, but we may be able to use nonparametric methods discussed in Chapter 12.)

Notation for Matched Pairs

d = individual difference between the two values in a single matched pair

μ_d = mean value of the differences d for the *population* of all matched pairs

$\bar{d}$ = mean value of the differences d for the paired *sample* data (equal to the mean of the $x - y$ values)

s_d = standard deviation of the differences d for the paired *sample* data

n = number of *pairs* of data

Hypothesis Test Statistic for Matched Pairs

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

where degrees of freedom = $n - 1$.

***P*-values and Critical values:** Table A-3 (t distribution) with $n - 1$ degrees of freedom.

Confidence Intervals for Matched Pairs

$$\bar{d} - E < \mu_d < \bar{d} + E$$

where $E = t_{\alpha/2} \frac{s_d}{\sqrt{n}}$

Critical values of $t_{\alpha/2}$: Use Table A-3 with $n - 1$ degrees of freedom.

Exploring the Data Sets

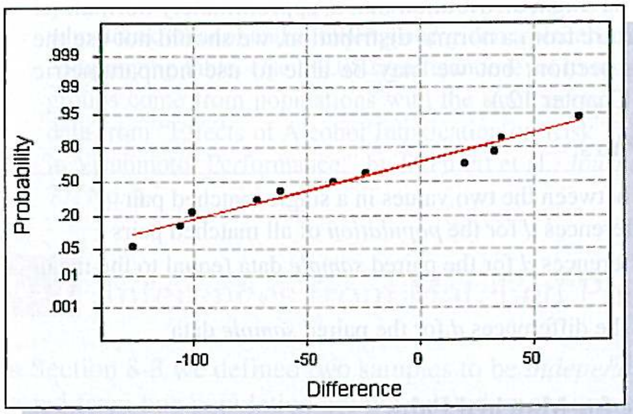
As always, we should avoid blind and thoughtless rote application of any statistical procedure. We should begin by exploring the data to see what might be learned. We should consider the center, variation, distribution, outliers, and any changes that take place over time. The requirements listed above require that "the number of matched pairs of sample data is large ($n > 30$) or the pairs of values have differences that are from a population having a distribution that is approximately normal." In the following example, we have 11 matched pairs of data, so we should check that the differences have a distribution that is approximately normal. Checking for normality with a histogram might be difficult with only 11 values, but shown below is a Minitab-generated normal probability plot of the 11 differences. This graph serves the same purpose as a normal quantile plot introduced in Section 5-7. Because the pattern of plotted points is reasonably close to a straight line and there is no systematic pattern that is not a straight-line pattern, we conclude that the differences appear to come from a population that has a normal distribution.

Research in Twins

Identical twins occur when a single fertilized egg splits in two, so that both twins share the same genetic makeup. There is now an explosion in research focused on those twins. Speaking for the Center for Study of Multiple Birth, Louis Keith noted that now "we have far more ability to analyze the data on twins using computers with new, built-in statistical packages." A common goal of such studies is to explore the classic issue of "nature versus nurture." For example, Thomas Bouchard, who runs the Minnesota Study of Twins Reared Apart, has found that IQ is 50%–60% inherited, while the remainder is the result of external forces.

Identical twins are matched pairs that provide better results by allowing us to reduce the genetic variation that is inevitable with unrelated pairs of people.

Minitab **Normal Probability Plot**



EXAMPLE Does the Type of Seed Affect Corn Growth? In 1908, William Gosset published the article "The Probable Error of a Mean" under the pseudonym of "Student" (*Biometrika*, Vol. 6, No. 1). He included the data listed below for two different types of corn seed (regular and kiln dried) that were used on adjacent plots of land. The listed values in Table 8-2 are the yields of head corn in pounds per acre. Use a 0.05 significance level to test the claim that the type of seed affects the yield.

SOLUTION

REQUIREMENT The data consist of matched pairs because they are corn yields from *adjacent* plots of land. Based on the design of the original experiment, it is reasonable to assume that the matched pairs constitute a simple random sample. The preceding Minitab display of a normal quantile plot suggests that the differences are from a population with a normal distribution. Note also that the list of differences does not appear to contain an outlier. The requirements are therefore satisfied.

We will follow the same basic method of hypothesis testing that was introduced in Chapter 7, but we will use the above test statistic for matched pairs.

Step 1: The claim that the type of seed affects the yield is a claim that there is a difference between the yields from regular seed and kiln-dried seed. This difference can be expressed as $\mu_d \neq 0$.

Step 2: If the original claim is not true, we have $\mu_d = 0$.

Table 8-2 Yields of Corn from Different Seeds

Regular	1903	1935	1910	2496	2108	1961	2060	1444	1612	1316	1511
Kiln Dried	2009	1915	2011	2463	2180	1925	2122	1482	1542	1443	1535
Difference d	−106	20	−101	33	−72	36	−62	−38	70	−127	−24

Step 3: The null hypothesis must be a statement of equality and the alternative hypothesis cannot include equality, so we have

$$H_0: \mu_d = 0 \qquad H_1: \mu_d \neq 0 \qquad \text{(original claim)}$$

Step 4: The significance level is $\alpha = 0.05$.

Step 5: We use the Student t distribution because the requirements are satisfied. (We are testing a claim about matched pairs of data, we have two simple random samples, and a normal quantile plot of the sample differences shows that they have a distribution that is approximately normal.)

Step 6: Before finding the value of the test statistic, we must first find the values of $\bar{d}$, and s_d. Refer to Table 8-2 and use the 11 sample differences to find these sample statistics: $\bar{d} = -33.727$ and $s_d = 66.171$. Using these sample statistics and the assumption of the hypothesis test that $\mu_d = 0$, we can now find the value of the test statistic:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{-33.727 - 0}{\frac{66.171}{\sqrt{11}}} = -1.690$$

The critical values of $t = \pm 2.228$ are found from Table A-3 as follows: Use the column for 0.05 (Area in Two Tails), and use the row with degrees of freedom of $n - 1 = 10$. Figure 8-4 shows the test statistic, critical values, and critical region.

Step 7: Because the test statistic does not fall in the critical region, we fail to reject the null hypothesis.

INTERPRETATION The sample data in Table 8-2 do not provide sufficient evidence to support the claim that yields from regular seeds and kiln-dried seeds are different. This does *not* prove that the actual yields are equal. Perhaps

continued

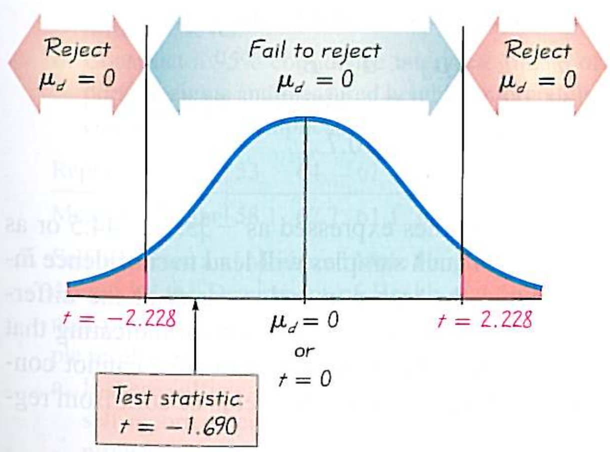


FIGURE 8-4
Distribution of Differences d between Values in Matched Pairs

additional sample data might provide the necessary evidence to conclude that the yields from the two types of seed are different. However, based on the available sample data, there does not appear to be a significant difference, so farmers should not treat the seeds by kiln drying them. Kiln drying appears to have no significant effect.

P-Value Method The preceding example used the traditional method, but the P-value approach could be used by modifying Steps 6 and 7. In Step 6, use the test statistic of $t = -1.690$ and refer to the 10th row of Table A-3 to find that the test statistic (without the negative sign) is between 1.372 and 1.812, indicating that the P-value is between 0.10 and 0.20. Using STATDISK, Excel, Minitab, and a TI-83/84 Plus calculator, the P-value is found to be 0.1218. We again fail to reject the null hypothesis, because the P-value is greater than the significance level of $\alpha = 0.05$.

Twins in Twinsburg

During the first weekend in August of each year, Twinsburg, Ohio celebrates its annual "Twins Days in Twinsburg" festival. Thousands of twins from around the world have attended this festival in the past. Scientists saw the festival as an opportunity to study identical twins. Because they have the same basic genetic structure, identical twins are ideal for studying the different effects of heredity and environment on a variety of traits, such as male baldness, heart disease, and deafness—traits that were recently studied at one Twinsburg festival. A study of twins showed that myopia (nearsightedness) is strongly affected by hereditary factors, not by environmental factors such as watching television, surfing the Internet, or playing computer or video games.

EXAMPLE Confidence Interval for Differences Using the same sample matched pairs in Table 8-2, construct a 95% confidence interval estimate of μ_d, which is the mean of the differences between the yields from regular seeds and kiln-dried seeds. Interpret the result.

SOLUTION

REQUIREMENT ✓ See the preceding example which includes the same requirement check that applies here. We can now proceed with the construction of the confidence interval. ✓

We use the values of $\bar{d} = -33.727$, $s_d = 66.171$, $n = 11$, and $t_{\alpha/2} = 2.228$ (found from Table A-3 with $n - 1 = 10$ degrees of freedom and an area of 0.05 in two tails). We first find the value of the margin of error E.

$$E = t_{\alpha/2}\frac{s_d}{\sqrt{n}} = 2.228 \cdot \frac{66.171}{\sqrt{11}} = 44.452$$

The confidence interval can now be found. Note that the final result is rounded using one more decimal place than the original sample differences.

$$\bar{d} - E < \mu_d < \bar{d} + E$$
$$-33.727 - 44.452 < \mu_d < -33.727 + 44.452$$
$$-78.2 < \mu_d < 10.7$$

INTERPRETATION The result is sometimes expressed as -33.7 ± 44.5 or as $(-78.2, 10.7)$. In the long run, 95% of such samples will lead to confidence interval limits that actually do contain the true population mean of the differences. Note that the confidence interval limits do contain zero, indicating that the true value of μ_d is not significantly different from zero. We cannot conclude that there is a significant difference between the yields of corn from regular seed and kiln-dried seed.

8-4 Exercises

Calculations for Matched Pairs. *In Exercises 1 and 2, assume that you want to use a 0.05 significance level to test the claim that the paired sample data come from a population for which the mean difference is $\mu_d = 0$. Find (a) $\overline{d}$, (b) s_d, (c) the t test statistic, and (d) the critical values.*

1.

x	1	1	3	5	4
y	0	2	5	8	0

2.

x	5	3	7	9	2	5
y	5	1	2	6	6	4

3. Using the sample paired data in Exercise 1, construct a 95% confidence interval for the population mean of all differences $x - y$.

4. Using the sample paired data in Exercise 2, construct a 99% confidence interval for the population mean of all differences $x - y$.

5. Testing Corn Seeds In 1908, William Gosset published the article "The Probable Error of a Mean" under the pseudonym of "Student" (*Biometrika*, Vol. 6, No. 1). He included the data listed below for yields from two different types of seed (regular and kiln dried) that were used on adjacent plots of land. The listed values are the yields of straw in cwts per acre, where cwt represents 100 pounds.

a. Using a 0.05 significance level, test the claim that there is no difference between the yields from the two types of seed.

b. Construct a 95% confidence interval estimate of the mean difference between the yields from the two types of seed.

c. Does it appear that either type of seed is better?

Regular	19.25	22.75	23	23	22.5	19.75	24.5	15.5	18	14.25	17
Kiln dried	25	24	24	28	22.5	19.5	22.25	16	17.25	15.75	17.25

6. Self-Reported and Measured Female Heights As part of the National Health Survey conducted by the Department of Health and Human Services, self-reported heights and measured heights were obtained for females aged 12–16. Listed below are sample results.

a. Is there sufficient evidence to support the claim that there is a difference between self-reported heights and measured heights of females aged 12–16? Use a 0.05 significance level.

b. Construct a 95% confidence interval estimate of the mean difference between reported heights and measured heights. Interpret the resulting confidence interval, and comment on the implications of whether the confidence interval limits contain 0.

Reported height	53	64	61	66	64	65	68	63	64	64	64	67
Measured height	58.1	62.7	61.1	64.8	63.2	66.4	67.6	63.5	66.8	63. 9	62.1	68.5

7. Self-Reported and Measured Male Heights As part of the National Health Survey conducted by the Department of Health and Human Services, self-reported heights and measured heights were obtained for males aged 12–16. Listed on the next page are sample results.

a. Is there sufficient evidence to support the claim that there is a difference between self-reported heights and measured heights of males aged 12–16? Use a 0.05 significance level.