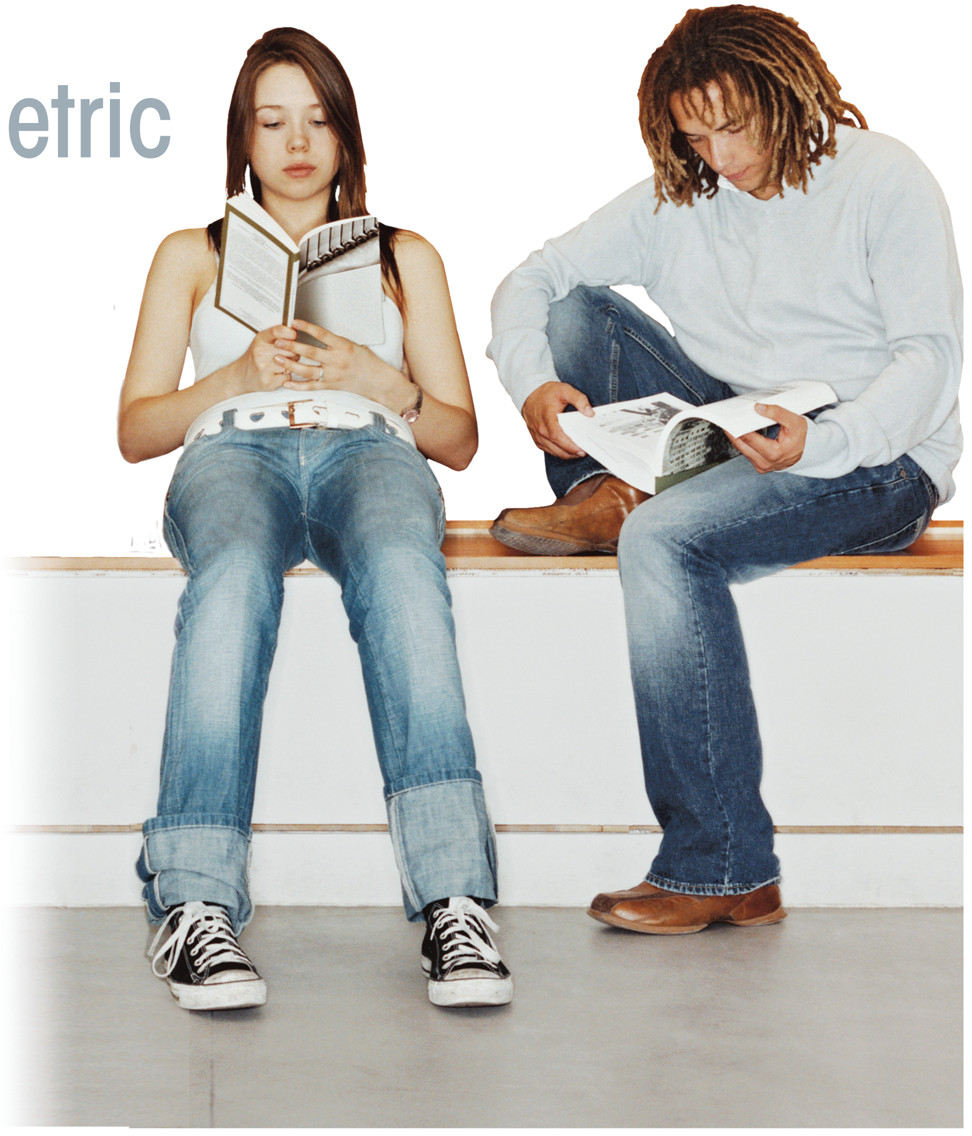
CHAPTER 14

Nonparametric Statistics



CONTENTS

STATISTICS IN ACTION

USING TECHNOLOGY

WHERE WE'VE BEEN

- Presented methods for making inferences about means (Chapters 5–8) and for making inferences about the correlation between two quantitative variables (Chapter 10)
- These methods required that the data be normally distributed or that the sampling distributions of the relevant statistics be normally distributed

WHERE WE'RE GOING

- Develop the need for inferential techniques that require fewer, or less stringent, assumptions than the methods of Chapters 5–8 and 10
- Introduce *nonparametric* tests that are based on ranks (i.e., on an ordering of the sample measurements according to their relative magnitudes).

STATISTICS IN ACTION

Deadly Exposure: Agent Orange and Vietnam Vets

Agent Orange was the code name for a herbicide developed for the U.S. Armed Forces, primarily for use in tropical climates. The purpose of the product was to deny an enemy cover and concealment in dense terrain by defoliating trees and shrubbery where the enemy could hide. (The code name comes from the orange band that was used to mark the drums the herbicide was stored in.) Agent Orange was tested in Southeast Asia in the early 1960s and brought into ever-widening use during the height of the Vietnam War (1967–1968); its use was diminished and eventually discontinued in 1971.

Agent Orange was a 50-50 mix of two chemicals, known conventionally as 2,4,D and 2,4,5,T. The combined product was mixed with kerosene or diesel fuel and dispersed by aircraft, vehicle, and hand spraying. As an unwanted byproduct of the chemical manufacturing process, Agent Orange was found to be extremely contaminated with TCDD, or dioxin. In laboratory tests on animals, TCDD has caused a wide variety of diseases (including cancer), many of them fatal.

During the Vietnam War, an estimated 19 million gallons of Agent Orange were used to destroy the dense plant and tree cover of the Asian jungle. As a result of this exposure, many Vietnam veterans have dangerously high levels of TCDD in their blood and adipose (fatty) tissue. A study published in *Chemosphere* (Vol. 20, 1990) reported on the TCDD levels of 20 Massachusetts Vietnam vets who were possibly exposed to Agent Orange. The TCDD amounts (measured in parts per trillion) in both plasma and fat tissue of the 20 vets are listed in Table SIA14.1. These data are saved in the **TCDD** file.

TCDD

TABLE SIA14.1 TCDD Measurements for 20 Vietnam Vets

Vet	Fat	Plasma
1	4.9	2.5
2	6.9	3.5
3	10.0	6.8
4	4.4	4.7
5	4.6	4.6
6	1.1	1.8
7	2.3	2.5
8	5.9	3.1
9	7.0	3.1
10	5.5	3.0
11	7.0	6.9
12	1.4	1.6
13	11.0	20.0
14	2.5	4.1
15	4.4	2.1
16	4.2	1.8
17	41.0	36.0
18	2.9	3.3
19	7.7	7.2
20	2.5	2.0

Source: Schecter, A. et al. "Partitioning of 2,3,7, 8-Chlorinated Dibenzo-*p*-Dioxins and Dibenzofurans between Adipose Tissue and Plasma Lipid of 20 Massachusetts Vietnam Veterans." *Chemosphere*, Vol. 20, Nos. 7–9, 1990, pp. 954–955 (Tables I and II).

What do the data tell us about the levels of TCDD in fat and plasma of Vietnam veterans? Is there a relationship between the fat and plasma TCDD levels? In the Statistics in Action Revisited sections listed below, we apply the nonparametric tests of this chapter to answer these questions.

Statistics in Action Revisited

- Testing the Median TCDD Level of Vietnam Vets (p. 14-8)
- Comparing the TCDD Levels in Fat and Plasma of Vietnam Vets (p. 14-25)
- Testing whether the TCDD Levels in Fat and Plasma of Vietnam Vets Are Correlated (p. 14-43)

14.1 Introduction: Distribution-Free Tests

The confidence interval and testing procedures developed in Chapters 5–8 all involve making inferences about population parameters. Consequently, they are often referred to as **parametric statistical tests.** Many of these parametric methods (e.g., the small sample t-test of Chapter 6 or the ANOVA F-test of Chapter 8) rely on the assumption that the data are sampled from a normally distributed population. When the data are normal, these tests are *most powerful*—that is, the use of these parametric tests maximizes power—the probability of the researcher correctly rejecting the null hypothesis.

Consider a population of data that is decidedly nonnormal. For example, the distribution might be very flat, peaked, or strongly skewed to the right or left (see Figure 14.1). Applying the small sample t-test to such a data set may result in serious consequences. Since the normality assumption is clearly violated, the results of the t-test are unreliable: (1) The probability of a Type I error (i.e., rejecting H_0 when it is true) may be larger than the value of α selected; and (2) the power of the test, $1 - \beta$, is not maximized.

A host of *nonparametric* techniques are available for analyzing data that do not follow a normal distribution. Nonparametric tests do not depend on the distribution of the sampled population; thus, they are called *distribution-free tests*. Also, nonparametric methods focus on the location of the probability distribution of the population, rather than on specific parameters of the population, such as the mean (hence, the name *nonparametrics*).

Definition 14.1

Distribution-free tests are statistical tests that do not rely on any underlying assumptions about the probability distribution of the sampled population.

Definition 14.2

The branch of inferential statistics devoted to distribution-free tests is called **nonparametrics.**

Nonparametric tests are also appropriate when the data are nonnumerical in nature but can be ranked.* For example, when taste-testing foods or in other types of consumer product evaluations, we can say we like product A better than product B, and B better than C, but we cannot obtain exact quantitative values for the respective measurements. Nonparametric tests based on the ranks of measurements are called *rank tests*.

Definition 14.3

Nonparametric statistics (or tests) based on the ranks of measurements are called **rank statistics** (or **rank tests**).

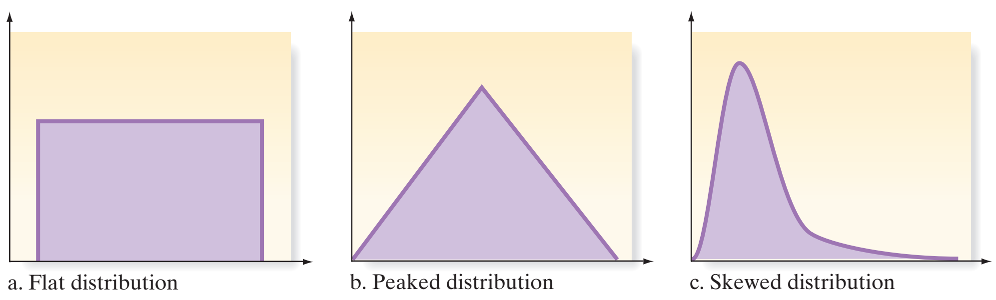


FIGURE 14.1
Some nonnormal distributions for which the t statistic is invalid

*Qualitative data that can be ranked in order of magnitude are called *ordinal data*.

In this chapter, we present several useful nonparametric methods. Keep in mind that these nonparametric tests are more powerful than their corresponding parametric counterparts in those situations where either the data are nonnormal or the data are ranked.

In Section 14.2, we develop a test to make inferences about the central tendency of a single population. In Sections 14.3 and 14.5, we present rank statistics for comparing two or more probability distributions using independent samples. In Sections 14.4 and 14.6, the matched-pairs and randomized block designs are used to make nonparametric comparisons of populations. Finally, in Section 14.7, we present a nonparametric measure of correlation between two variables.

14.2 Single Population Inferences

In Chapter 6 we utilized the z- and t-statistics for testing hypotheses about a population mean. The z-statistic is appropriate for large random samples selected from "general" populations—that is, with few limitations on the probability distribution of the underlying population. The t-statistic was developed for small-sample tests in which the sample is selected at random from a *normal* distribution. The question is, How can we conduct a test of hypothesis when we have a small sample from a *nonnormal* distribution?

The **sign test** is a relatively simple nonparametric procedure for testing hypotheses about the central tendency of a nonnormal probability distribution. Note that we used the phrase *central tendency* rather than *population mean.* This is because the sign test, like many nonparametric procedures, provides inferences about the population *median* rather than the population mean μ. Denoting the population median by the Greek letter, η, we know (Chapter 2) that η is the 50th percentile of the distribution (Figure 14.2) and as such is less affected by the skewness of the distribution and the presence of outliers (extreme observations). Since the nonparametric test must be suitable for all distributions, not just the normal, it is reasonable for nonparametric tests to focus on the more robust (less sensitive to extreme values) measure of central tendency, the median.

For example, increasing numbers of both private and public agencies are requiring their employees to submit to tests for substance abuse. One laboratory that conducts such testing has developed a system with a normalized measurement scale, in which values less than 1.00 indicate "normal" ranges and values equal to or greater than 1.00 are indicative of potential substance abuse. The lab reports a normal result as long as the median level for an individual is less than 1.00. Eight independent measurements of each individual's sample are made. One individual's results are shown in Table 14.1.

If the objective is to determine whether the *population* median (that is, the true median level if an indefinitely large number of measurements were made on the same individual sample) is less than 1.00, we establish that as our alternative hypothesis and test

$$H_0\colon \eta = 1.00$$
$$H_a\colon \eta < 1.00$$

The one-tailed sign test is conducted by counting the number of sample measurements that "favor" the alternative hypothesis—in this case, the numbers that are less than 1.00. If the null hypothesis is true, we expect approximately half of the measurements to fall on each side of the hypothesized median, and if the alternative is true, we

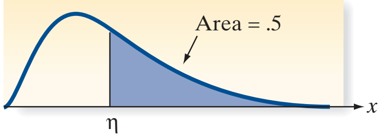


FIGURE 14.2
Location of the population median, η

SUBABUSE

TABLE 14.1 Substance Abuse Test Results

.78	.51	3.79	.23	.77	.98	.96	.89

expect significantly more than half to favor the alternative—that is, to be less than 1.00. Thus,

$$\textit{Test statistics: } S = \text{Number of measurements less than 1.00, the null hypothesized median}$$

If we wish to conduct the test at the $\alpha = .05$ level of significance, the rejection region can be expressed in terms of the observed significance level, or p-value of the test:

$$\textit{Rejection region: } p\text{-value} \leq .05$$

In this example, $S = 7$ of the 8 measurements are less than 1.00. To determine the observed significance level associated with this outcome, we note that the number of measurements less than 1.00 is a binomial random variable (check the binomial characteristics presented in Chapter 4), and *if* H_0 *is true*, the binomial probability p that a measurement lies below (or above) the median 1.00 is equal to .5 (Figure 14.2). What is the probability that a result is *as contrary to or more contrary to* H_0 than the one observed if H_0 is true? That is, what is the probability that 7 *or more* of 8 binomial measurements will result in Success (be less than 1.00) if the probability of Success is .5? Binomial Table II in Appendix B (using $n = 8$ and $p = .5$) indicates that

$$P(x \geq 7) = 1 - P(x \leq 6) = 1 - .965 = .035$$

Thus, the probability that at least 7 of 8 measurements would be less than 1.00 *if the true median were* 1.00 is only .035. The p-value of the test is therefore .035.

This p-value can also be obtained using a statistical software package. The MINITAB printout of the analysis is shown in Figure 14.3, with the p-value highlighted on the printout. Since $p = .035$ is less than $\alpha = .05$, we conclude that this sample provides sufficient evidence to reject the null hypothesis. The implication of this rejection is that the laboratory can conclude at the $\alpha = .05$ level of significance that the true median level for the tested individual is less than 1.00. However, we note that one of the measurements greatly exceeds the others, with a value of 3.79, and deserves special attention. Note that this large measurement is an outlier that would make the use of a t-test and its concomitant assumption of normality dubious. The only assumption necessary to ensure the validity of the sign test is that the probability distribution of measurements is continuous.

Sign Test for Median: READING

```
Sign test of median =  1.000 versus < 1.000

         N  Below  Equal  Above       P  Median
READING  8      7      0      1  0.0352  0.8350
```

FIGURE 14.3
MINITAB printout of sign test

The use of the sign test for testing hypotheses about population medians is summarized in the box.

Sign Test for a Population Median η

One-Tailed Test	Two-Tailed Test
$H_0: \eta = \eta_0$ $H_a: \eta > \eta_0$ [or $H_a: \eta < \eta_0$]	$H_0: \eta = \eta_0$ $H_a: \eta \neq \eta_0$
Test statistic:	
S = Number of sample measurements greater than η_0 [or S = number of measurements less then η_0]	S = Larger of S_1 and S_2, where S_1 is the number of measurements less than η_0 and S_2 is the number of measurements greater than η_0
Observed significance level:	
p-value $= P(x \geq S)$	p-value $= 2P(x \geq S)$

where x has a binomial distribution with parameters n and $p = .5$. (Use Table II, Appendix B.)

Rejection region: Reject H_0 if p-value $\leq .05$.

Conditions Required for Valid Application of the Sign Test

The sample is selected randomly from a continuous probability distribution. [*Note:* No assumptions need to be made about the shape of the probability distribution.]

Recall that the normal probability distribution provides a good approximation for the binomial distribution when the sample size is large. For tests about the median of a distribution, the null hypothesis implies that $p = .5$, and the normal distribution provides a good approximation if $n \geq 10$. Thus, we can use the standard normal z-distribution to conduct the sign test for large samples. The large-sample sign test is summarized in the next box.

Large-Sample Sign Test for a Population Median η

One-Tailed Test	Two-Tailed Test
H_0: $\eta = \eta_0$	H_0: $\eta = \eta_0$
H_a: $\eta > \eta_0$ [or H_a: $\eta < \eta_0$]	H_a: $\eta \neq \eta_0$

$$\textit{Test statistic}: z = \frac{(S - .5) - .5n}{.5\sqrt{n}}$$

[*Note: S* is calculated as shown in the previous box. We subtract .5 from S as the "correction for continuity." The null hypothesized mean value is $np = .5n$, and the standard deviation is

$$\sqrt{npq} = \sqrt{n(.5)(.5)} = .5\sqrt{n}$$

See Chapter 4 for details on the normal approximation to the binomial distribution.]

Rejection region: $z > z_\alpha$ | *Rejection region:* $z > z_{\alpha/2}$

where tabulated z values can be found in Table IV of Appendix B.

EXAMPLE 14.1

Testing Failure Time of CDs Using the Sign Test

Problem A manufacturer of compact disk (CD) players has established that the median time to failure for its players is 5,250 hours of utilization. A sample of 20 CDs from a competitor is obtained, and they are continuously tested until each fails. The 20 failure times range from 5 hours (a "defective" player) to 6,575 hours, and 14 of the 20 exceed 5,250 hours. Is there evidence that the median failure time of the competitor differs from 5,250 hours? Use $\alpha = .10$.

Solution The null and alternative hypotheses of interest are

$$H_0: \eta = 5{,}250 \text{ hours}$$
$$H_a: \eta \neq 5{,}250 \text{ hours}$$

Test statistic: Since $n \geq 10$, we use the standard normal z statistic:

$$z = \frac{(S - .5) - .5n}{.5\sqrt{n}}$$

where S is the maximum of S_1, the number of measurements greater than 5,250, and S_2, the number of measurements less than 5,250.

Rejection region: $z > 1.645$ where $z_{\alpha/2} = z_{.05} = 1.645$

Assumptions: The distribution of the failure times is continuous (time is a continuous variable), but nothing is assumed about the shape of its probability distribution.

Since the number of measurements exceeding 5,250 is $S_2 = 14$ and thus the number of measurements less than 5,250 is $S_1 = 6$, then $S = 14$, the greater of S_1 and S_2. The calculated z statistic is therefore

$$z = \frac{(S - .5) - .5n}{.5\sqrt{n}} = \frac{13.5 - 10}{.5\sqrt{20}} = \frac{3.5}{2.236} = 1.565$$

The value of z is not in the rejection region, so we cannot reject the null hypothesis at the $\alpha = .10$ level of significance.

Look Back The CD manufacturer should not conclude, on the basis of this sample, that its competitor's CDs have a median failure time that differs from 5,250 hours. The manufacturer will not "Accept H_0", however, since the probability of a Type II error is unknown.

Now Work Exercise 14.5

The one-sample nonparametric sign test for a median provides an alternative to the t-test for small samples from nonnormal distributions. However, if the distribution is approximately normal, the t-test provides a more powerful test about the central tendency of the distribution.

STATISTICS IN ACTION REVISITED

Testing the Median TCDD Level of Vietnam Vets

Recall that during the Vietnam War, U.S. soldiers were exposed to the herbicide Agent Orange (p. 14-3). As a result of this exposure, many Vietnam veterans have dangerously high levels of the dioxin TCDD in their plasma and fat tissue. Some medical researchers consider a TCDD level of 3 parts per trillion (ppt) to be dangerously high. Do the data in Table SIA14.1 provide evidence to indicate that the median level of TCDD in both plasma and fat tissue of Vietnam vets exceeds 3 ppt? To answer this question, we applied the sign test to the data saved in the TCDD file. The MINITAB printout is shown in Figure SIA14.1.

We want to test H_0: $\eta = 3$ versus H_a: $\eta > 3$ for both variables, TCDD in fat and TCDD in plasma. According to the printout, 14 of the 20 Vietnam vets had TCDD levels in fat above 3 ppt, and 12 of the 20 Vietnam vets had TCDD levels in plasma above 3 ppt. Consequently, the two test statistic values are $S = 14$ and $S = 12$, respectively. The one-tailed p-values for the tests (highlighted on the printout) are .0577 and .1796, respectively.

At $\alpha = .10$, the test for TCDD in fat is significant, but the test for TCDD in plasma is not significant. Therefore, the data provide sufficient evidence to say that the median TCDD in fat exceeds 3 ppt; however, the evidence is not as strong for TCDD in plasma.

Why apply the nonparametric sign test to the data rather than the more familiar Student's t-test? The MINITAB histograms in Figure SIA14.2 illustrate the problem. Clearly the sample TCDD levels are not normally distributed. Consequently, the assumption required for the t-test to yield valid inferences is violated.

Figure SIA14.1
MINITAB sign tests for TCDD data

Sign Test for Median: FAT, PLASMA

Sign test of median = 3.000 versus > 3.000

	N	Below	Equal	Above	P	Median
FAT	20	6	0	14	0.0577	4.750
PLASMA	20	7	1	12	0.1796	3.200

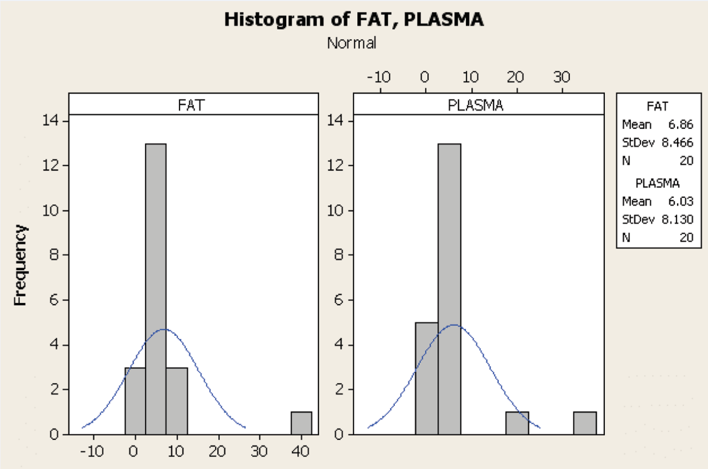


Figure SIA14.2
MINITAB histograms for TCDD data

Exercises 14.1–14.12

Learning the Mechanics

14.1 Under what circumstances is the sign test preferred to the t-test for making inferences about the central tendency of a population?

14.2 What is the probability that a randomly selected observation exceeds the
- **a.** Mean of a normal distribution?
- **b.** Median of a normal distribution?
- **c.** Mean of a nonnormal distribution?
- **d.** Median of a nonnormal distribution?

14.3 Use Table II of Appendix B to calculate the following binomial probabilities:
- **a.** $P(x \geq 7)$ when $n = 8$ and $p = .5$
- **b.** $P(x \geq 5)$ when $n = 8$ and $p = .5$
- **c.** $P(x \geq 8)$ when $n = 8$ and $p = .5$
- **d.** $P(x \geq 10)$ when $n = 15$ and $p = .5$. Also use the normal approximation to calculate this probability, then compare the approximation with the exact value.
- **e.** $P(x \geq 15)$ when $n = 25$ and $p = .5$. Also use the normal approximation to calculate this probability, then compare the approximation with the exact value.

14.4 Consider the following sample of 10 measurements:

LM14_4

8.4	16.9	15.8	12.5	10.3	4.9	12.9	9.8	23.7	7.3

Use these data to conduct each of the following sign tests using the binomial tables (Table II, Appendix B) and $\alpha = .05$:
- **a.** H_0: $\eta = 9$ versus H_a: $\eta > 9$
- **b.** H_0: $\eta = 9$ versus H_a: $\eta \neq 9$
- **c.** H_0: $\eta = 20$ versus H_a: $\eta < 20$
- **d.** H_0: $\eta = 20$ versus H_a: $\eta \neq 20$
- **e.** Repeat each of the preceding tests using the normal approximation to the binomial probabilities. Compare the results.
- **f.** What assumptions are necessary to ensure the validity of each of the preceding tests?

14.5 NW Suppose you wish to conduct a test of the research hypothesis that the median of a population is greater than 75. You randomly sample 25 measurements from the population and determine that 17 of them exceed 75. Set up and conduct the appropriate test of hypothesis at the .10 level of significance. Be sure to specify all necessary assumptions.

Applying the Concepts—Basic

14.6 **Caffeine in Starbucks' coffee.** Researchers at the University of Florida College of Medicine investigated the level of caffeine in 16-ounce cups of Starbucks' coffee (*Journal of Analytical Toxicology*, Oct. 2003). In one phase of the experiment, cups of Starbucks Breakfast Blend (a mix of Latin American coffees) were purchased on 6 consecutive days from a single specialty coffee shop. The amount of caffeine in each of the six cups (measured in milligrams) is provided in the table.

STARBUCKS

564	498	259	303	300	307

- **a.** Suppose the scientists are interested in determining whether the median amount of caffeine in Breakfast Blend coffee exceeds 300 milligrams. Set up the null and alternative hypotheses of interest.

b. How many of the cups in the sample have a caffeine content that exceeds 300 milligrams?
c. Assuming $p = .5$, use the binomial table in Appendix B to find the probability that at least 4 of the 6 cups have caffeine amounts that exceed 300 milligrams.
d. Based on the probability, part **c,** what do you conclude about H_0 and H_a? (Use $\alpha = .05$.)

14.7 Salaries of experienced MBA graduates. One way to assess the benefits of an MBA degree is to investigate the salaries received by MBA students several years after graduation. The Graduate Management Admission Council estimates that the median earnings for graduates of full-time, highly ranked MBA programs 4 years after graduating is $96,000 (*Selections,* Winter 1999). A random sample of 50 recent graduates from a particular highly ranked MBA program were mailed a questionnaire and asked to report their annual earnings. Fifteen useable responses were received; 9 indicated earnings greater than $96,000 and 6 indicated earnings below $96,000.
a. Specify the null and alternative hypotheses that should be used in testing whether the median income of graduates of the MBA program is more than $96,000 in 2006.
b. Conduct the test of part **a** using $\alpha = .05$ and draw your conclusion in the context of the problem.
c. What assumptions must hold to ensure the validity of your hypothesis test?

14.8 Quality of white shrimp. In *The American Statistician* (May 2001), the nonparametric sign test was used to analyze data on the quality of white shrimp. One measure of shrimp quality is cohesiveness. Since freshly caught shrimp are usually stored on ice, there is concern that cohesiveness will deteriorate after storage. For a sample of 20 newly caught white shrimp, cohesiveness was measured both before storage and after storage on ice for 2 weeks. The difference in the cohesiveness measurements (before minus after) was obtained for each shrimp. If storage has no effect on cohesiveness, the population median of the differences will be 0. If cohesiveness deteriorates after storage, the population median of the differences will be positive.
a. Set up the null and alternative hypotheses to test whether cohesiveness will deteriorate after storage.
b. In the sample of 20 shrimp, there were 13 positive differences. Use this value to find the p-value of the test.
c. Make the appropriate conclusion (in the words of the problem) if $\alpha = .05$.

Applying the Concepts—Intermediate

14.9 RIF plan to fire older employees. Reducing the size of a company's workforce in order to reduce costs is referred to as *corporate downsizing* or *reductions in force* (RIF) by the business community and media (*Business Week,* Feb. 24, 1997). Following RIFs, companies are often sued by former employees who allege that the RIFs were discriminatory with regard to age. Federal law protects employees over 40 years of age against such discrimination. Suppose one large company's employees have a median age of 37. Its RIF plan is to fire 15 employees with ages listed in the next table.
a. Calculate the median age of the employees who are being terminated.

FIRE15

43	32	39	28	54	41	50	62
22	45	47	54	43	33	59	

b. What are the appropriate null and alternative hypotheses to test whether the population from which the terminated employees were selected has a median age that exceeds the entire company's median age?
c. Conduct the test of part **b.** Find the significance level of the test and interpret its value.
d. Assuming that courts generally require statistical evidence at the .10 level of significance before ruling that age discrimination laws were violated, what do you advise the company about its planned RIF? Explain.

14.10 Reviewing aircraft maintenance procedures. The Federal Aviation Administration (FAA) increased the frequency and thoroughness of its review of aircraft maintenance procedures in response to the admission by ValuJet Airlines that it had not met some maintenance requirements. Suppose that the FAA samples the records of six aircraft currently utilized by one airline and determines the number of flights between the last two complete engine maintenances for each, with the results shown in the table. The FAA requires that this maintenance be performed at least every 30 flights. Although it is obvious that not all aircraft are meeting the requirement, the FAA wishes to test whether the airline is meeting this particular maintenance requirement "on average."

FAA6

24	27	25
94	29	28

a. Would you suggest the t-test or sign test to conduct the test? Why?
b. Set up the null and alternative hypotheses such that the burden of proof is on the airline to show it is meeting the "on-average" requirement.
c. What are the test statistic and rejection region for this test if the level of significance is $\alpha = .01$? Why would the level of significance be set at such a low value?
d. Conduct the test, and state the conclusion in terms of this application.

14.11 Technology Fast 500. In Exercise 5.31 (p. 310), the average 5-year revenue growth for the 500 fastest-growing technology companies in 2005 (i.e., *Forbes' Technology Fast 500*) was investigated. The data are reproduced in the table.

FAST500

Rank	Company	5-Year Revenue Growth Rate (%)
4	CaseStack	39,071
22	Go2call.com	12,514
88	Active Motif	2,789
160	Netspoke	1,185
193	Conduant Corp.	860
268	Argon St.	576
274	Immtech Int'l	555
323	Open Solutions	430
359	eCopy	384
397	iBasis	331
444	Espial Group	281
485	Pacific Biometrics	249

Source: Technology Fast 500, Deloitte & Touche, 2005.

a. Recall that in Exercise 5.31a, the t-distribution was employed to make an inference about the true mean 5-year revenue growth rate for the *Technology Fast 500*. Explain why the resulting inference may be invalid.

b. Give the null and alternative hypotheses for a nonparametric test designed to determine if the "average" 5-year revenue growth rate is less than 5,000 percent.

c. Conduct the test of part **b** using $\alpha = .05$. Interpret your result in the context of the problem.

14.12 Surface roughness of pipe. Refer to the *Anti-Corrosion Methods and Materials* (Vol. 50, 2003) study of the surface roughness of coated interior pipe used in oil fields, Exercise 5.32 (p. 310). The data (in micrometers) for 20 sampled pipe sections are reproduced in the table. Conduct a nonparametric test to determine whether the median surface roughness of coated interior pipe, η, differs from 2 micrometers. Test using $\alpha = .05$.

ROUGHPIPE

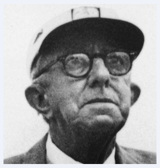
1.72	2.50	2.16	2.13	1.06	2.24	2.31	2.03	1.09	1.40
2.57	2.64	1.26	2.05	1.19	2.13	1.27	1.51	2.41	1.95

Source: Farshad, F., and Pesacreta, T. "Coated Pipe Interior Surface Roughness as Measured by Three Scanning Probe Instruments." *Anti-Corrosion Methods and Materials*, Vol. 50, No. 1, 2003 (Table III).

14.3 Comparing Two Populations: Independent Samples

Suppose two independent random samples are to be used to compare two populations and the t-test of Chapter 7 is inappropriate for making the comparison. We may be unwilling to make assumptions about the form of the underlying population probability distributions or we may be unable to obtain exact values of the sample measurements. If the data can be ranked in order of magnitude for either of these situations, the **Wilcoxon rank sum test** (developed by Frank Wilcoxon) can be used to test the hypothesis that the probability distributions associated with the two populations are equivalent.

Biography



FRANK WILCOXON (1892–1965) Wilcoxon Rank Tests

Frank Wilcoxon was born in Ireland, where his wealthy American parents were vacationing. He grew up in the family home in Catskill, New York, then spent time working as an oil worker and tree surgeon in the back country of West Virginia. At age 25, Wilcoxon's parents sent him to Pennsylvania Military College, but he dropped out due to the death of his twin sister. Later, Wilcoxon earned degrees in chemistry from Rutgers (master's) and Cornell University (PhD). After receiving his doctorate, Wilcoxon began work as a chemical researcher at the Boyce Thompson Institute for Plant Research. There, he began studying R. A. Fisher's newly issued *Statistical Methods for Research Workers*. In a now-famous 1945 paper, Wilcoxon presented the idea of replacing the actual sample data in Fisher's tests by their ranks and called the tests the *rank-sum test* and *signed-rank test*. These tests proved to be inspirational to the further development of nonparametrics. After retiring from industry, Wilcoxon accepted a Distinguished Lectureship position at the newly created Department of Statistics at Florida State University.

For example, suppose six economists who work for the federal government and seven university economists are randomly selected, and each is asked to predict next year's percentage change in cost of living as compared with this year's figure. The objective of the study is to compare the government economists' predictions to those of the university economists. The data are shown in Table 14.2.

Experience has shown that the populations of predicted percentage changes often possess probability distributions that are skewed, as shown in Figure 14.4. Consequently, a t-test should not be used to compare the mean predictions of the two groups of economists because the normality assumption that is required for the t-test may not be valid.

The two populations of predictions are those that would be obtained from *all* government and *all* university economists if they could all be questioned. To compare their probability distributions using a nonparametric test, we first *rank the sample observations as though they were all drawn from the same population*—that is, we pool the measurements from both samples and then rank the measurements from the smallest

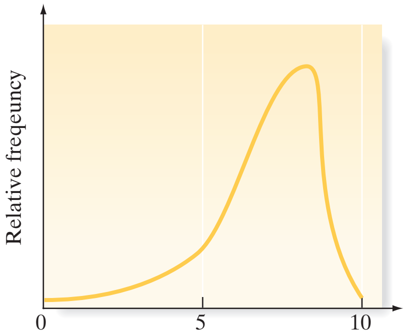


FIGURE 14.4
Typical probability distribution of predicted cost-of-living changes

COSTLIVING

TABLE 14.2 Percentage Cost of Living Change, as Predicted by Government and University Economists

GOVERNMENT ECONOMIST (1)		UNIVERSITY ECONOMIST (2)	
Prediction	Rank	Prediction	Rank
3.1	4	4.4	6
4.8	7	5.8	9
2.3	2	3.9	5
5.6	8	8.7	11
0.0	1	6.3	10
2.9	3	10.5	12
		10.8	13

(a rank of 1) to the largest (a rank of 13). The ranks of the 13 economists' predictions are indicated in Table 14.2.

If the two populations were identical, we would expect the ranks to be *randomly mixed* between the two samples. If, on the other hand, one population tends to have larger percentage changes than the other, we would expect the larger ranks to be mostly in one sample and the smaller ranks mostly in the other. Thus, the test statistic for the Wilcoxon test is based on the totals of the ranks for each of the two samples—that is, on the **rank sums.** The greater the difference in rank sums, the greater the evidence to indicate a difference between the populations.

For the economists predictions, we arbitrarily denote the rank sum for government economists by T_1 and that for university economists by T_2. Then

$$T_1 = 4 + 7 + 2 + 8 + 1 + 3 = 25$$
$$T_2 = 6 + 9 + 5 + 11 + 10 + 12 + 13 = 66$$

The sum of T_1 and T_2 will always equal $n(n + 1)/2$, where $n = n_1 + n_2$. So, for this example, $n_1 = 6$, $n_2 = 7$, and

$$T_1 + T_2 = \frac{13(13 + 1)}{2} = 91$$

Since $T_1 + T_2$ is fixed, a small value for T_1 implies a large value for T_2 (and vice versa) and a large difference between T_1 and T_2. Therefore, the smaller the value of one of the rank sums, the greater the evidence to indicate that the samples were selected from different populations.

The test statistic for this test is the rank sum for the smaller sample; or, in the case where $n_1 = n_2$, either rank sum can be used. Values that locate the rejection region for this rank sum are given in Table XIV of Appendix B. A partial reproduction of this table is shown in Table 14.3. The columns of the table represent n_1, the first sample size, and the rows represent n_2, the second sample size. *The* $\mathrm{T_L}$ *and* $\mathrm{T_U}$ *entries in the table are the boundaries of the lower and upper regions, respectively, for the rank sum associated with the sample that has fewer measurements*. If the sample sizes n_1 and n_2 are the same, either rank sum may be used as the test statistic. To illustrate, suppose $n_1 = 6$ and $n_2 = 7$. For a two-tailed test with $\alpha = .05$, we consult part **a** of the table and find that the null hypothesis will be rejected if the rank sum of sample 1 (the sample with fewer measurements), T, is less than or equal to $T_L = 28$ or greater than or equal to $T_U = 56$. (These values are highlighted in Table 14.3.) The Wilcoxon rank sum test is summarized in the next box.

TABLE 14.3 Reproduction of Part of Table XIV in Appendix B: Critical Values for the Wilcoxon Rank Sum Test

$\alpha = .025$ one-tailed; $\alpha = .05$ two-tailed

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

Wilcoxon Rank Sum Test: Independent Samples*

Let D_1 and D_2 represent the probability distributions for populations 1 and 2, respectively.

One-Tailed Test	Two-Tailed Test
H_0: D_1 and D_2 are identical	H_0: D_1 and D_2 are identical
H_a: D_1 is shifted to the right of D_2 [or H_a: D_1 is shifted to the left of D_2]	H_a: D_1 is shifted either to the left or to the right of D_2
Test statistic:	*Test statistic:*
T_1, if $n_1 < n_2$; T_2, if $n_2 < n_1$	T_1, if $n_1 < n_2$; T_2, if $n_2 < n_1$
(Either rank sum can be used if $n_1 = n_2$.)	(Either rank sum can be used if $n_1 = n_2$.) We will denote this rank sum as T.
Rejection region:	*Rejection region:*
T_1: $T_1 \geq T_U$ [or $T_1 \leq T_L$] T_2: $T_2 \leq T_L$ [or $T_2 \geq T_U$]	$T \leq T_L$ or $T \geq T_U$

where T_L and T_U are obtained from Table XIV of Appendix B.

Ties: Assign tied measurements the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign each a rank of $(3 + 4)/2 = 3.5$.

Conditions Required for a Valid Wilcoxon Rank Sum Test

1. The two samples are random and independent.
2. The two probability distributions from which the samples are drawn are continuous.

Note that the assumptions necessary for the validity of the Wilcoxon rank sum test do not specify the shape or type of probability distribution. However, the distributions are assumed to be continuous so that the probability of tied measurements is 0 (see Chapter 4), and each measurement can be assigned a unique rank. In practice, however, rounding of continuous measurements will sometimes produce ties. As long

*Another statistic used for comparing two populations based on independent random samples is the *Mann-Whitney U-statistic*. The U-statistic is a simple function of the rank sums. It can be shown that the Wilcoxon rank sum test and the Mann-Whitney U-test are equivalent.

as the number of ties is small relative to the sample sizes, the Wilcoxon test procedure will still have an approximate significance level of α. The test is not recommended to compare discrete distributions for which many ties are expected.

EXAMPLE 14.2 Comparing Economists' Predictions with the Rank Sum Test

Problem Test the hypothesis that the government economists' predictions of next year's percentage change in cost of living tend to be lower than the university economists'—that is, test to determine if the probability distribution of government economists' predictions is *shifted to the left* of the probability distribution of university economists' predictions. Conduct the test using the data in Table 14.2 and $\alpha = .05$.

Solution

H_0: The probability distributions corresponding to the government and university economists' predictions of inflation rate are identical

H_a: The probability distribution for the government economists' predictions lies below (to the left of) the probability distribution for the university economists' predictions*

Test statistic: Since fewer government economists ($n_1 = 6$) than university economists ($n_2 = 7$) were sampled, the test statistic is T_1, the rank sum of the government economists' predictions.

Rejection region: Since the test is one-sided, we consult part **b** of Table XIV for the rejection region corresponding to $\alpha = .05$. We reject H_0 only for $T_1 \leq T_L$, the lower value from Table XIV, since we are specifically testing that the distribution of government economists' predictions lies *below* the distribution of university economists' predictions, as shown in Figure 14.5. Thus, we reject H_0 if $T_1 \leq 30$.

Since T_1, the rank sum of the government economists' predictions in Table 14.2, is 25, it is in the rejection region (see Figure 14.5). Therefore, we can conclude that the university economists' predictions tend, in general, to exceed the government economists' predictions. This same conclusion can be reached using a statistical software package. The SPSS printout of the analysis is shown in Figure 14.6. Both the test statistic ($T_1 = 25$) and two-tailed p-value ($p = .014$) are highlighted on the printout. The one-tailed p-value, $p = .014/2 = .007$, is less than $\alpha = .05$, leading us to reject H_0.

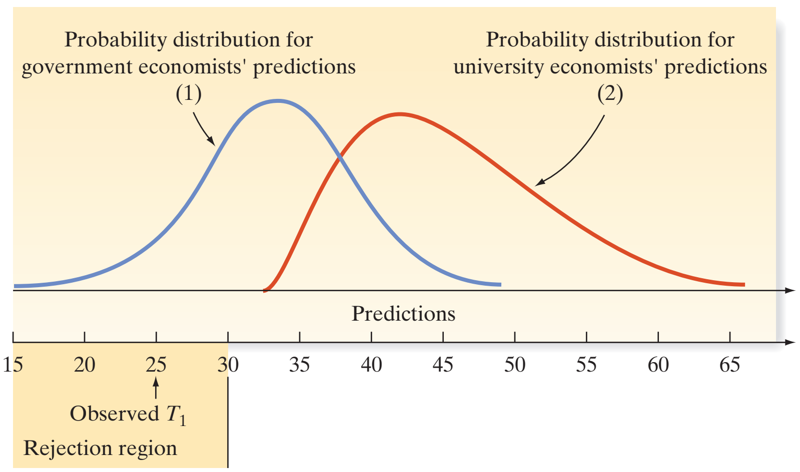


FIGURE 14.5
Alternative hypothesis and rejection region for Example 14.2.

*The alternative hypotheses in this chapter will be stated in terms of a difference in the *location* of the distributions. However, since the shapes of the distributions may also differ under H_a, some of the figures (e.g., Figure 14.5) depicting the alternative hypothesis will show probability distributions with different shapes.

Mann-Whitney Test

Ranks

	ECONOMST	N	Mean Rank	Sum of Ranks
PCTCHNG	1	6	4.17	25.00
	2	7	9.43	66.00
	Total	13		

Test Statistics[b]

	PCTCHNG
Mann-Whitney U	4.000
Wilcoxon W	25.000
Z	-2.429
Asymp. Sig. (2-tailed)	.015
Exact Sig. [2*(1-tailed Sig.)]	.014[a]

a. Not corrected for ties.

b. Grouping Variable: ECONOMST

FIGURE 14.6
SPSS printout of rank sum test

Now Work Exercise 14.14

Table XIV in Appendix B gives values of T_L and T_U for values of n_1 and n_2 less than or equal to 10. When both sample sizes n_1 and n_2 are 10 or larger, the sampling distribution of T_1 can be approximated by a normal distribution with mean and variance

$$E(T_1) = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \text{and} \quad \sigma^2_{T_1} = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Therefore, for $n_1 \geq 10$ and $n_2 \geq 10$ we can conduct the Wilcoxon rank sum test using the familiar z-test of Chapters 6 and 7. The test is summarized in the next box.

The Wilcoxon Rank Sum Test for Large Samples ($n_1 \geq 10$ and $n_2 \geq 10$)

Let D_1 and D_2 represent the probability distributions for populations 1 and 2, respectively.

One-Tailed Test	Two-Tailed Test
H_0: D_1 and D_2 are identical	H_0: D_1 and D_2 are identical
H_a: D_1 is shifted to the right of D_2 (or H_a: D_1 is shifted to the left of D_2)	H_a: D_1 is shifted either to the right or to the left of D_2

Continued

$$\textit{Test statistic: } z = \frac{T_1 - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Rejection region:	*Rejection region:*
$z > z_\alpha$ (or $z < -z_\alpha$)	$\lvert z \rvert > z_{\alpha/2}$

ACTIVITY 14.1: *Keep the Change:* Nonparametric Statistics

In this activity, you will refer to your results from Activity 6.2: *Keep the Change: Tests of Hypotheses* and Activity 7.2: *Keep the Change: Inferences Based on Two Samples*.

1. Referring to Exercises 1 and 2 of Activity 6.2, explain why a sign test of the population median might be a better fit than the hypothesis test of the population mean, especially if the sample is small. Perform the corresponding sign test assuming once again that your data set *Amounts Transferred* represents a random sample from all Bank of America customers' transfer amounts. Are your conclusions similar? Explain.

2. Referring to Exercises 4 and 5 of Activity 6.2, explain why a sign test of the population median might be a better fit than the hypothesis test of the population mean, especially if the sample is small. Perform the corresponding sign test assuming once again that your data set *Bank Matching* represents a random sample from all Bank of America customers' bank matching. Are your conclusions similar? Explain.

3. Refer to Exercise 2 of Activity 7.2 where you designed a study to compare the mean amounts for bank matching in California and bank matching in Florida. Design a Wilcoxon rank sum test to compare the corresponding probability distributions. Be specific about sample sizes, hypotheses, and how a conclusion will be reached. Under what conditions might the rank sum test provide more useful information than the mean comparison test?

Exercises 14.13–14.25

Learning the Mechanics

14.13 Specify the test statistic and the rejection region for the Wilcoxon rank sum test for independent samples in each of the following situations:

a. $n_1 = 10, n_2 = 6, \alpha = .10$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the right or left of the probability distribution for population 2

b. $n_1 = 5, n_2 = 7, \alpha = .05$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the right of the probability distribution for population 2

c. $n_1 = 9, n_2 = 8, \alpha = .025$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the left of the probability distribution for population 2

d. $n_1 = 15, n_2 = 15, \alpha = .05$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the right or left of the probability distribution for population 2

14.14 NW Suppose you wish to compare two treatments, A and B, based on independent random samples of 15 observations selected from each of the two populations. If $T_1 = 173$, do the data indicate that distribution A is shifted to the left of distribution B? Test using $\alpha = .05$

14.15 Suppose you want to compare two treatments, A and B. In particular, you wish to determine whether the distribution for population B is shifted to the right of the distribution for population A. You plan to use the Wilcoxon rank sum test.

a. Specify the null and alternative hypotheses you would test.

b. Suppose you obtained the following independent random samples of observations on experimental units subjected to the two treatments. Conduct a test of the hypotheses described in part **a.** Test using $\alpha = .05$.

LM14_15

Sample A:	37,	40,	33,	29,	42,	33,	35,	28,	34
Sample B:	65,	35,	47,	52					

14.16 Independent random samples are selected from two populations. The data are shown in the table.

LM14_16

SAMPLE 1		SAMPLE 2		
15	16	5	9	5
10	13	12	8	10
12	8	9	4	

a. Use the Wilcoxon rank sum test to determine whether the data provide sufficient evidence to indicate a shift in the locations of the probability distributions of the sampled populations. Test using $\alpha = .05$.

b. Do the data provide sufficient evidence to indicate that the probability distribution for population 1 is shifted to the right of the probability distribution for population 2? Use the Wilcoxon rank sum test with $\alpha = .05$.

Applying the Concepts—Basic

14.17 Bursting strength of bottles. Polyethylene terephthalate (PET) bottles are used for carbonated beverages. A critical property of PET bottles is their bursting strength (i.e., the pressure at which bottles filled with water burst when pressurized). In the *Journal of Data Science* (May 2003), researchers measured the bursting strength of PET bottles made from two different designs—an old design and a new design. The data (pounds per square inch) for 10 bottles of each design are shown in the table. Suppose you want to compare the distributions of bursting strengths for the two designs.

PET

Old Design	210	212	211	211	190	213	212	211	164	209
New Design	216	217	162	137	219	216	179	153	152	217

a. Rank all 20 observed pressures from smallest to largest, and assign ranks from 1 to 20.

b. Sum the ranks of the observations from the old design.

c. Sum the ranks of the observations from the new design.

d. Compute the Wilcoxon rank sum statistic.

e. Carry out a nonparametric test (at $\alpha = .05$) to compare the distribution of bursting strengths for the two designs.

14.18 Information systems and technology expenditures. University of Queensland researchers sampled private sector and public sector organizations in Australia to study the planning undertaken by their information systems departments (*Management Science,* July 1996). They asked each sample organization how much it had spent on information systems and technology in the previous fiscal year as a percentage of the organization's total revenues. The results are reported in the table.

INFOSYS

Private Sector	Public Sector
2.58%	5.40%
5.05	2.55
.05	9.00
2.10	10.55
4.30	1.02
2.25	5.11
2.50	12.42
1.94	1.67
2.33	3.33

Source: Adapted from Hann, J., and Weber, R. "Information Systems Planning: A Model and Empirical Tests." *Management Science,* Vol. 42, No. 2. July 1996, pp. 1043–1064.

a. Find the rank sums for the two sectors using the Wilcoxon method.

b. Do the two sampled populations have identical probability distributions or is the distribution for public sector organizations in Australia located to the right of Australia's private sector firms? Test using $\alpha = .05$.

c. Is the p-value for the test less than or greater than .05? Justify your answer.

d. What assumptions must be met to ensure the validity of the test you conducted in part **a?**

14.19 Computer-mediated communication study. Refer to the *Journal of Computer-Mediated Communication* (Apr. 2004) study to compare those who interact via computer-mediated communication (CMC) to those who meet face-to-face (FTF), Exercise 7.18. Recall that 48 undergraduate students were randomly divided into the two groups. Those in the CMC group communicated using the "chat" mode of instant-messaging software; those in the FTF group met in a conference room. The relational intimacy scores (measured on a 7-point scale) of the participants are reproduced in the table. The researchers hypothesized that the relational intimacy scores for participants in the CMC group will tend to be lower than the relational intimacy scores for participants in the FTF group.

a. Which nonparametric procedure should be used to analyze the data?

b. Specify the null and alternative hypotheses of the test.

c. Give the rejection region for the test using $\alpha = .10$.

d. Conduct the test and give the appropriate conclusion in the context of the problem.

INTIMACY

CMC:	4	3	3	4	3	3	3	3	4	4	3	4
	3	3	2	4	2	4	5	4	4	4	5	3
FTF:	5	4	4	4	3	3	3	4	3	3	3	3
	4	4	4	4	4	3	3	3	4	4	2	4

Applying the Concepts—Intermediate

14.20 Comparing ethics of American and Mexican purchasing managers. In Mexico, the United State's third largest trading partner, purchasing has not fully evolved into a profession with its own standards of ethical behavior. Researchers at Xavier University investigated the question, Do American and Mexican purchasing managers perceive ethical situations differently (*Industrial Marketing Management,* July 1999)? As part of their study, 15 Mexican purchasing managers and 15 American purchasing managers were asked to consider different ethical situations and respond on a 100-point scale with end points "strongly disagree" (1) and "strongly agree" (100). For the situation "accepting free trips from salespeople is okay," the responses shown in the table on the next page were obtained.

a. Consider a Wilcoxon rank sum test to determine whether American and Mexican purchasing managers perceive the given ethical situation differently. Find the rank sums for the test.

b. Conduct the test at $\alpha = .05$.

c. Under what circumstances could the two-sample t-test of Chapter 7 be used to analyze the data? Check to see whether the t-test is appropriate in this situation.

ETHICS

American Purchasing Managers			Mexican Purchasing Managers		
50	15	19	10	15	5
10	8	11	90	60	55
35	40	5	65	80	40
30	80	25	50	85	45
20	75	30	20	35	95

Source: Adapted from Tadepalli, R., Moreno, A., and Trevino, S., "Do American and Mexican Purchasing Managers Perceive Ethical Situations Differently? An Empirical Investigation." *Industrial Marketing Management,* Vol. 28, No. 4, July 1999, pp. 369–380.

14.21 Patent infringement case. Refer to the *Chance* (Fall 2002) study of a patent infringement case brought against Intel Corp., Exercise 7.17. Recall that the case rested on whether a patent witness' signature was written on top of key text in a patent notebook or under the key text. Using an X-ray beam, zinc measurements were taken at several spots on the notebook page. The zinc measurements for three notebook locations—on a text line, on a witness line, and on the intersection of the witness and text line—are reproduced in the table.

PATENT

Text Line	.335	.374	.440			
Witness Line	.210	.262	.188	.329	.439	.397
Intersection	.393	.353	.285	.295	.319	

a. Why might the Student's t-procedure you applied in Exercise 7.17 be inappropriate for analyzing this data?

b. Use a nonparametric test (at $\alpha = .05$) to compare the distribution of zinc measurements for the text line with the distribution for the intersection.

c. Use a nonparametric test (at $\alpha = .05$) to compare the distribution of zinc measurements for the witness line with the distribution for the intersection.

d. From the results, parts **b** and **c,** what can you infer about the median zinc measurements at the three notebook locations?

GASTURBINE

14.22 Cooling method for gas turbines. Refer to the *Journal of Engineering for Gas Turbines and Power* (Jan. 2005) study of gas turbines augmented with high-pressure inlet fogging, Exercise 7.23. The data on engine heat rate (kilojoules per kilowatt per hour) are saved in the **GASTURBINE** file. Recall that the researchers classified gas turbines into three categories: traditional, advanced, and aeroderivative. Suppose you want to compare the heat-rate distributions for traditional and aeroderivative turbine engines.

a. Demonstrate that the assumptions required to compare the mean heat rates using a t-test are likely to be violated.

b. A MINITAB printout of the nonparametric test to compare the two heat-rate distributions is shown at the bottom of the page. Interpret the p-value of the test shown at the bottom of the printout.

14.23 Preference for type of disposable razor. A major razor blade manufacturer advertises that its twin-blade disposable razor "gets you lots more shaves" than any single-blade disposable razor on the market. A rival company that has been very successful in selling single-blade razors plans to test this claim. Independent random samples of eight single-blade users and eight twin-blade users are taken, and the number of shaves that each gets before indicating a preference to change blades is recorded. The results are shown in the table.

RAZOR

Twin Blades		Single Blades	
8	15	10	13
17	10	6	14
9	6	3	5
11	12	7	7

a. Do the data support the twin-blade manufacturer's claim? Use $\alpha = .05$.

b. Do you think this experiment was designed in the best possible way? If not, what design might have been better?

c. What assumptions are necessary for the validity of the test you performed in part **a?** Do the assumptions seem reasonable for this application?

14.24 Satisfaction with MIS implementation. A *management information system* (MIS) is a computer-based information-processing system designed to support the operations, management, and decision functions of an organization. The development of an MIS involves three stages: definition, physical design, and implementation of the system (*Managing Information,* 1993). Thirty firms that recently implemented an MIS were surveyed: 16 were satisfied with the implementation results, 14 were not. Each firm was asked to rate the quality of the planning and negotia-

MINITAB Output for Exercise 14.22

Mann-Whitney Test and CI: TRAD-HR, AERO-HR

```
          N  Median
TRAD-HR  39   11183
AERO-HR   7   12414


Point estimate for ETA1-ETA2 is -1125
95.3 Percent CI for ETA1-ETA2 is (-2358,1448)
W = 885.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.3431
The test is significant at 0.3431 (adjusted for ties)
```

tion stages of the development process, using a scale of 0 to 100, with higher numbers indicating better quality. (A score of 100 indicates that all the problems that occurred in the planning and negotiation stages were successfully resolved, while 0 indicates that none were resolved.) The results are shown in the table below.

MIS

Firms with a Good MIS			Firms with a Poor MIS		
52	59	95	60	40	90
70	60	90	50	55	85
40	90	86	55	65	80
80	75	95	70	55	90
82	80	93	41	70	
65					

a. Use the Wilcoxon rank sum test to compare the quality of the development processes of successfully and unsuccessfully implemented MISs. Test using $\alpha = .05$.

b. Under what circumstances could you use the two-sample t-test of Chapter 7 to conduct the same test?

Applying the Concepts—Advanced

14.25 Turnover rates in the United States and Japan. In Exercise 7.16, you used a Student's t-test to compare the mean annual percentages of labor turnover between U.S. and Japanese manufacturers of air conditioners. The annual percentage turnover rates for five U.S. and five Japanese plants are shown in the table.

TURNOVER

U.S. Plants	Japanese Plants
7.11%	3.52%
6.06	2.02
8.00	4.91
6.87	3.22
4.77	1.92

a. Recall that the variance of a binomial sample proportion, $\hat{p}$, depends on the value of the population parameter, p. As a consequence, the variance of a sample percentage, $(100\hat{p})\%$, also depends on p. If you conduct an unpaired t-test (Section 7.2) to compare the means of two populations of percentages, which assumption may be violated?

b. Do the data provide sufficient evidence to indicate that the mean annual percentage turnover for American plants exceeds the corresponding mean for Japanese plants? Test using the Wilcoxon rank sum test with $\alpha = .05$. Do your test conclusions agree with those of the t-test in Exercise 7.16?

14.4 Comparing Two Populations: Paired Difference Experiment

Nonparametric techniques can also be employed to compare two probability distributions when a paired difference design is used. For example, consumer preferences for two competing products are often compared by having each of a sample of consumers rate both products. Thus, the ratings have been paired on each consumer. Here is an example of this type of experiment.

For some paper products, softness is an important consideration in determining consumer acceptance. One method of determining softness is to have judges give a sample of the products a softness rating. Suppose each of 10 judges is given a sample of two products that a company wants to compare. Each judge rates the softness of each product on a scale from 1 to 10, with higher ratings implying a softer product. The results of the experiment are shown in Table 14.4.

TABLE 14.4 Softness Ratings of Paper

	PRODUCT		DIFFERENCE		
Judge	A	B	(A − B)	Absolute Value of Difference	Rank of Absolute Value
1	6	4	2	2	5
2	8	5	3	3	7.5
3	4	5	−1	1	2
4	9	8	1	1	2
5	4	1	3	3	7.5
6	7	9	−2	2	5
7	6	2	4	4	9
8	5	3	2	2	5
9	6	7	−1	1	2
10	8	2	6	6	10

T_+ = Sum of positive ranks = 46
T_- = Sum of negative ranks = 9

Since this is a paired difference experiment, we analyze the differences between the measurements (see Section 7.3). However, a nonparametric approach developed by Wilcoxon requires that we calculate the ranks of the absolute values of the differences between the measurements—that is, the ranks of the differences after removing any minus signs. *Note that tied absolute differences are assigned the average of the ranks they would receive if they were unequal but successive measurements.* After the absolute differences are ranked, the sum of the ranks of the positive differences of the original measurements, T_+, and the sum of the ranks of the negative differences of the original measurements, T_-, are computed.

We are now prepared to test the nonparametric hypotheses:

H_0: The probability distributions of the ratings for products A and B are identical.

H_a: The probability distributions of the ratings differ (in location) for the two products. (Note that this is a two-sided alternative and that it implies a two-tailed test.)

Test statistic: T = Smaller of the positive and negative rank sums T_+ and T_-.

The smaller the value of T, the greater the evidence to indicate that the two probability distributions differ in location. The rejection region for T can be determined by consulting Table XV in Appendix B (part of the table is shown in Table 14.5). This table gives a value T_0 for both one-tailed and two-tailed tests for each value of n, the number of matched pairs. For a two-tailed test with $\alpha = .05$, we will reject H_0 if $T \leq T_0$. You can see in Table 14.5 that the value of T_0 that locates the boundary of the rejection region for $\alpha = .05$ and $n = 10$ pairs of observations is 8. Thus, the rejection region for the test (see Figure 14.7) is

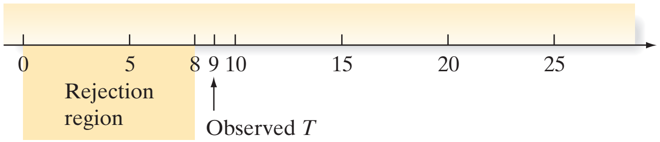
$$\textit{Rejection region}: T \leq 8 \text{ for } \alpha = .05$$

Since the smaller rank sum for the paper data, $T_- = 9$, does not fall within the rejection region, the experiment has not provided sufficient evidence to indicate that the two paper products differ with respect to their softness ratings at the $\alpha = .05$ level.

TABLE 14.5 Reproduction of Part of Table XV of Appendix B: Critical Values for the Wilcoxon Paired Difference Signed Rank Test

One-Tailed	Two-Tailed	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$\alpha = .05$	$\alpha = .10$	1	2	4	6	8	11
$\alpha = .025$	$\alpha = .05$		1	2	4	6	8
$\alpha = .01$	$\alpha = .02$			0	2	3	5
$\alpha = .005$	$\alpha = .01$				0	2	3
		$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$\alpha = .05$	$\alpha = .10$	14	17	21	26	30	36
$\alpha = .025$	$\alpha = .05$	11	14	17	21	25	30
$\alpha = .01$	$\alpha = .02$	7	10	13	16	20	24
$\alpha = .005$	$\alpha = .01$	5	7	10	13	16	19
		$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
$\alpha = .05$	$\alpha = .10$	41	47	54	60	68	75
$\alpha = .025$	$\alpha = .05$	35	40	46	52	59	66
$\alpha = .01$	$\alpha = .02$	28	33	38	43	49	56
$\alpha = .005$	$\alpha = .01$	23	28	32	37	43	49
		$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
$\alpha = .05$	$\alpha = .10$	83	92	101	110	120	130
$\alpha = .025$	$\alpha = .05$	73	81	90	98	107	117
$\alpha = .01$	$\alpha = .02$	62	69	77	85	93	102
$\alpha = .005$	$\alpha = .01$	55	61	68	76	84	92

FIGURE 14.7
Rejection region for paired difference experiment



Note that if a significance level of $\alpha = .10$ had been used, the rejection region would have been $T \leq 11$ and we would have rejected H_0. In other words, the samples do provide evidence that the probability distributions of the softness ratings differ at the $\alpha = .10$ significance level.

The **Wilcoxon signed rank test** is summarized in the box. Note that the difference measurements are assumed to have a continuous probability distribution so that the absolute differences will have unique ranks. Although tied (absolute) differences can be assigned ranks by averaging, the number of ties should be small relative to the number of observations to ensure the validity of the test.

Wilcoxon Signed Rank Test for a Paired Difference Experiment

Let D_1 and D_2 represent the probability distributions for populations 1 and 2, respectively.

One-Tailed Test	Two-Tailed Test
H_0: D_1 and D_2 are identical	H_0: D_1 and D_2 are identical
H_a: D_1 is shifted to the right of D_2 [or H_a: D_1 is shifted to the left of D_2]	H_a: D_1 is shifted either to the left or to the right of D_2

Calculate the difference within each of the n matched pairs of observations. Then rank the absolute value of the n differences from the smallest (rank 1) to the highest (rank n) and calculate the rank sum T_- of the negative differences and the rank sum T_+ of the positive differences. [*Note:* Differences equal to 0 are eliminated, and the number n of differences is reduced accordingly.]

Test statistic:	*Test statistic:*
T_-, the rank sum of the negative differences [or T_+, the rank sum of the positive differences]	T, the smaller of T_+ or T_-
Rejection region:	*Rejection region:*
$T_- \leq T_0$ [or $T_+ \leq T_0$]	$T \leq T_0$

where T_0 is given in Table XV in Appendix B.

Ties: assign tied absolute differences the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked differences are tied, assign both a rank of $(3 + 4)/2 = 3.5$.

Conditions Required for a Valid Signed Rank Test

1. The sample of differences is randomly selected from the population of differences.
2. The probability distribution from which the sample of paired differences is drawn is continuous.

EXAMPLE 14.3
Comparing Electrical Safety Ratings Using the Signed Rank Test

Problem Suppose the U.S. Consumer Product Safety Commission (CPSC) wants to test the hypothesis that New York City electrical contractors are more likely to install unsafe electrical outlets in urban homes than in suburban homes. A pair of homes, one urban and one suburban and both serviced by the same electrical contractor, is chosen for each of 10 randomly selected electrical contractors. A CPSC inspector assigns each of the 20 homes a safety

SAFETY

TABLE 14.6 Electrical Safety Ratings for 10 Pairs of New York City Homes

	LOCATION		DIFFERENCE	
Contractor	Urban A	Suburban B	(A − B)	Rank of Absolute Difference
1	7	9	−2	4.5
2	4	5	−1	2
3	8	8	0	(Eliminated)
4	9	8	1	2
5	3	6	−3	6
6	6	10	−4	7.5
7	8	9	−1	2
8	10	8	2	4.5
9	9	4	5	9
10	5	9	−4	7.5
				Positive rank sum $= T_+ = 15.5$

rating between 1 and 10, with higher numbers implying safer electrical conditions. The results are shown in Table 14.6. Use the Wilcoxon signed rank test to determine whether the CPSC hypothesis is supported at the $\alpha = .05$ level.

Solution The null and alternative hypotheses are

H_0: The probability distributions of home electrical ratings are identical for urban and suburban homes

H_a: The electrical ratings for suburban homes tend to exceed the electrical ratings for urban homes

Since a paired difference design was used (the homes were selected in urban-suburban pairs so that the electrical contractor was the same for both), we first calculate the difference between the ratings for each pair of homes, and then rank the absolute values of the differences (see Table 14.6). Note that one pair of ratings was the same (both 8), and the resulting 0 difference contributes to neither the positive nor the negative rank sum. Thus, we eliminate this pair from the calculation of the test statistic.

Test statistic: T_+, the positive rank sum

In Table 14.6, we compute the urban minus suburban rating differences, and if the alternative hypothesis is true, we would expect most of these differences to be negative. Or, in other words, we would expect the *positive* rank sum T_+ to be small if the alternative hypothesis is true (see Figure 14.8).

Rejection region: For $\alpha = .05$, from Table XV of Appendix B, we use $n = 9$ (remember, one pair of observations was eliminated) to find the rejection region for this one-tailed test: $T_+ \leq 8$

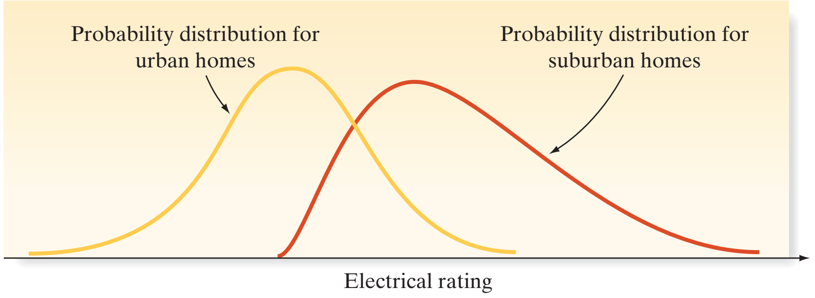


FIGURE 14.8
The alternative hypothesis for Example 14.3: We expect T_+ to be small

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
SUBURB - URBAN	Negative Ranks	3[a]	5.17	15.50
	Positive Ranks	6[b]	4.92	29.50
	Ties	1[c]		
	Total	10		

a. SUBURB < URBAN

b. SUBURB > URBAN

c. SUBURB = URBAN

Test Statistics[b]

	SUBURB - URBAN
Z	-.834[a]
Asymp. Sig. (2-tailed)	.404

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

FIGURE 14.9
SPSS printout of signed rank test

Since the computed value $T_+ = 15.5$ exceeds the critical value of 8, we conclude that this sample provides insufficient evidence at $\alpha = .05$ to support the alternative hypothesis. We *cannot* conclude on the basis of this sample information that suburban homes have safer electrical outlets than urban homes.

An SPSS printout of the analysis, shown in Figure 14.9, confirms this conclusion. The 2-tailed p-value of the test (highlighted) is .404. Since the 1-tailed p-value, $.404/2 = .202$, exceeds $\alpha = .05$, we fail to reject H_0.

Now Work Exercise 14.28

As is the case for the rank sum test for independent samples, the sampling distribution of the signed rank statistic can be approximated by a normal distribution when the number n of paired observations is large (say, $n \geq 25$). The large-sample z-test is summarized in the next box.

Wilcoxon Signed Rank Test for Large Samples ($n \geq 25$)

Let D_1 and D_2 represent the probability distributions for populations 1 and 2, respectively.

One-Tailed Test	Two-Tailed Test
H_0: D_1 and D_2 are identical	H_0: D_1 and D_2 are identical
H_a: D_1 is shifted to the right of D_2 [or H_a: D_1 is shifted to the left of D_2]	H_a: D_1 is shifted either to the left or to the right of D_2

$$\textit{Test statistic: } z = \frac{T_+ - [n(n+1)/4]}{\sqrt{[n(n+1)(2n+1)]/24}}$$

Continued

Rejection region:	*Rejection region:*
$z > z_\alpha$ [or $z < -z_\alpha$]	$\lvert z \rvert > z_{\alpha/2}$

Assumptions: The sample size n is greater than or equal to 25. Differences equal to 0 are eliminated and the number n of differences is reduced accordingly. Tied absolute differences receive ranks equal to the average of the ranks they would have received had they not been tied.

Exercises 14.26–14.36

Learning the Mechanics

14.26 Specify the test statistic and the rejection region for the Wilcoxon signed rank test for the paired difference design in each of the following situations:

a. $n = 30, \alpha = .10$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the right or left of probability distribution for population 2

b. $n = 20, \alpha = .05$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the right of the probability distribution for population 2

c. $n = 8, \alpha = .005$
H_0: Two probability distributions, 1 and 2, are identical
H_a: Probability distribution for population 1 is shifted to the left of the probability distribution for population 2

14.27 Suppose you want to test a hypothesis that two treatments, A and B, are equivalent against the alternative hypothesis that the responses for A tend to be larger than those for B. You plan to use a paired difference experiment and to analyze the resulting data (shown below) using the Wilcoxon signed rank test.

a. Specify the null and alternative hypotheses you would test.

b. Suppose the paired difference experiment yielded the data in the table. Conduct the test of part **a.** Test using $\alpha = .025$.

LM14_27

	TREATMENT			TREATMENT	
Pair	A	B	Pair	A	B
1	54	45	6	77	75
2	60	45	7	74	63
3	98	87	8	29	30
4	43	31	9	63	59
5	82	71	10	80	82

14.28 NW Suppose you wish to test a hypothesis that two treatments, A and B, are equivalent against the alternative that the responses for A tend to be larger than those for B.

a. If the number of pairs equals 25, give the rejection region for the large-sample Wilcoxon signed rank test for $\alpha = .05$.

b. Suppose that $T_+ = 273$. State your test conclusions.

c. Find the p-value for the test and interpret it.

14.29 A paired difference experiment with $n = 30$ pairs yielded $T_+ = 354$.

a. Specify the null and alternative hypotheses that should be used in conducting a hypothesis test to determine whether the probability distribution for population 1 is located to the right of that for population 2.

b. Conduct the test of part **a** using $\alpha = .05$.

c. What is the approximate p-value of the test of part **b?**

d. What assumptions are necessary to ensure the validity of the test you performed in part **b?**

Applying the Concepts—Basic

14.30 NHTSA new car crash tests. Refer to the National Highway Traffic Safety Administration (NHTSA) crash test data for new cars saved in the **CRASH** file. In Exercise 7.34 you compared the chest injury ratings of drivers and front-seat passengers using the Student's t-procedure for matched pairs. Suppose you want to make the comparison for only those cars that have a driver's star rating of 5 stars (the highest rating). The data for these 18 cars are listed in the table. Now consider analyzing these data using the Wilcoxon signed rank test.

a. State the null and alternative hypothesis.

b. Use a statistical software package to find the signed rank test statistic.

c. Give the rejection region for the test using $\alpha = .01$.

d. State the conclusion in practical terms. Report the p-value of the test.

CRASH5

	CHEST INJURY RATING	
Car	Driver	Passenger
1	42	35
2	42	35
3	34	45
4	34	45
5	45	45
6	40	42
7	42	46
8	43	58
9	45	43
10	36	37
11	36	37
12	43	58
13	40	42
14	43	58
15	37	41
16	37	41
17	44	57
18	42	42

14.31 Computer-mediated communication study. Refer to the *Journal of Computer-Mediated Communication* (Apr. 2004) study to compare people who interact via computer-mediated communication (CMC) with those who meet face-to-face (FTF), Exercise 14.19 (p. 14-17). Relational intimacy scores (measured on a 7-point scale) were obtained

STATISTICS IN ACTION REVISITED

Comparing the TCDD Levels in Fat and Plasma of Vietnam Vets

Medical researchers used the sample data in Table SIA14.1 to compare the TCDD levels in fat tissue and plasma for Vietnam veterans. Specifically, they wanted to determine if the distribution of TCDD levels in fat is shifted above or below the distribution of TCDD levels is plasma. To answer this question, first note that the data are collected as matched pairs—for each Vietnam vet in the sample the researchers recorded both the TCDD level in fat and in plasma. Therefore, the correct test to apply is the Wilcoxon signed rank test. The MINITAB printout for this analysis is shown in Figure SIA14.3.

Both the test statistic and p-value are highlighted on the printout. Since p-value $= .073$, at $\alpha = .05$ there is insufficient evidence to conclude that the distribution of TCDD levels in fat differs from the distribution of TCDD levels in plasma.

[*Note:* A histogram of the differences between the TCDD levels in fat and plasma, shown in Figure SIA14.4, illustrates that the sample differences are unlikely to have come from a normal population.]

Wilcoxon Signed Rank Test: DIFF

```
Test of median = 0.000000 versus median not = 0.000000

               N
              for   Wilcoxon              Estimated
        N    Test  Statistic        P        Median
DIFF   20      19      140.0    0.073         1.175
```

Figure SIA14.3
MINITAB signed rank test for TCDD data

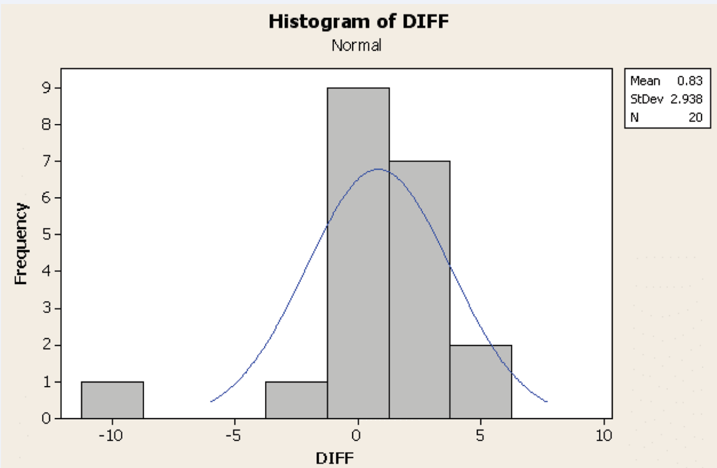


Figure SIA14.4
MINITAB histogram for differences in TCDD levels

for each participant after each of three different meeting sessions. The researchers hypothesized that relational intimacy scores for participants in the CMC group will tend to be higher at the third meeting than at the first meeting; however, they hypothesize that there are no differences in scores between the first and third meetings for the FTF group.

a. Explain why a nonparametric Wilcoxon signed ranks test is appropriate for analyzing the data.

b. For the CMC group comparison, give the null and alternative hypotheses of interest.

c. Give the rejection region for conducting the test (at $\alpha = .05$), part **b.** Recall that there were 24 participants assigned to the CMC group.

d. For the FTF group comparison, give the null and alternative hypotheses of interest.

e. Give the rejection region for conducting the test (at $\alpha = .05$), part **d.** Recall that there were 24 participants assigned to the FTF group.

Applying the Concepts—Intermediate

14.32 Designing an atlas. An atlas is a compendium of geographic, economic, and social information that describes one or more geographic regions. Atlases are used by the sales and marketing functions of businesses, local chambers of commerce, and educators. One of the most critical aspects of a new atlas design is its thematic content. In a survey of atlas users (*Journal of Geography,* May/June 1995), a large sample of high school teachers in British Columbia ranked 12 thematic atlas topics for usefulness. The consensus rankings of the teachers (based on the percentage of teachers who responded they "would definitely use" the topic) are given in the table. These teacher rankings were compared to the rankings of a group of university geography alumni made 3 years earlier. Compare the distributions of theme rankings for the two groups with an appropriate nonparametric test. Use $\alpha = .05$. Interpret the results practically.

ATLAS

	RANKINGS	
Theme	High School Teachers	Geography Alumni
Tourism	10	2
Physical	2	1
Transportation	7	3
People	1	6
History	2	5
Climate	6	4
Forestry	5	8
Agriculture	7	10
Fishing	9	7
Energy	2	8
Mining	10	11
Manufacturing	12	12

Source: Keller, C. P., et al. "Planning the Next Generation of Regional Atlases: Input from Educators." *Journal of Geography,* Vol. 94, No. 3, May/June 1995, p. 413 (Table 1).

14.33 Taking "power naps" during work breaks. According to the National Sleep Foundation, companies are encouraging their workers to take "power naps" (*Athens Daily News,* Jan. 9, 2000). In Exercise 7.35, you analyzed data collected by a major airline that recently began encouraging reservation agents to nap during their breaks. The number of complaints received about each of a sample of 10 reservation agents during the 6 months before naps were encouraged and during the 6 months after the policy change are reproduced in the accompanying table. Compare the distributions of number of complaints for the two time periods using the Wilcoxon signed rank test. Use $\alpha = .05$ to make the appropriate inference.

POWERNAP

Agent	Before Policy	After Policy
1	10	5
2	3	0
3	16	7
4	11	4
5	8	6
6	2	4
7	1	2
8	14	3
9	5	5
10	6	1

14.34 Flexible working hours program. A job-scheduling innovation that has helped managers overcome motivation and absenteeism problems associated with a fixed 8-hour workday is a concept called *flextime.* This flexible working hours program permits employees to design their own 40-hour workweek to meet their personal needs (*New York Times,* Mar. 31, 1996). The management of a large manufacturing firm may adopt a flextime program depending on the success or failure of a pilot program. Ten employees were randomly selected and given a questionnaire designed to measure their attitude toward their job. Each was then permitted to design and follow a flextime workday. After 6 months, attitudes toward their jobs were again measured. The resulting attitude scores are displayed in the table. The higher the score, the more favorable the employee's attitude toward his or her work. Use a nonparametric test procedure to evaluate the success of the pilot flextime program. Test using $\alpha = .05$.

FLEXTIME

Employee	Before	After	Employee	Before	After
1	54	68	6	82	88
2	25	42	7	94	90
3	80	80	8	72	81
4	76	91	9	33	39
5	63	70	10	90	93

14.35 Testing electronic circuits. Refer to the *IEICE Transactions on Information & Systems* (Jan. 2005) comparison of two methods of testing electronic circuits, Exercise 7.33. Recall that each of 11 circuits was tested using the standard compression/depression method and the new Huffman-based coding method and the compression ratio recorded. The data are reproduced in the next table.

a. In Exercise 7.33, you used a parametric procedure to determine whether the Huffman-coding method will yield a smaller mean compression ratio than the standard method. Perform the alternative nonparametric test, using $\alpha = .05$.

b. Do the conclusions of the two procedures agree?

CIRCUITS

Circuit	Standard Method	Huffman-Coding Method
1	.80	.78
2	.80	.80
3	.83	.86
4	.53	.53
5	.50	.51
6	.96	.68
7	.99	.82
8	.98	.72
9	.81	.45
10	.95	.79
11	.99	.77

Source: Ichihara, H., Shintani, M., and Inoue, T. "Huffman-Based Test Response Coding." *IEICE Transactions on Information & Systems*, Vol. E88-D, No. 1, Jan. 2005 (Table 3).

14.36 Teachers who involve parents. Teachers Involve Parents in Schoolwork (TIPS) is an interactive homework process designed to improve the quality of homework assignments for elementary, middle, and high school students. TIPS homework assignments require students to conduct interactions with family partners (parents, guardians, etc.) while completing the homework. Frances Van Voorhis (Johns Hopkins University) conducted a study to investigate the effects of TIPS in science and mathematics homework assignments (April 2001). Each large group of middle school students was assigned to complete TIPS homework assignments. At the end of the study, all students reported on the level of family involvement in their homework on a 4-point scale (0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Frequently, 4 = Always). The data for science and math for a random sample of 10 students selected from the large group is shown in the table. Conduct a nonparametric analysis to compare the level of family involvement in science and math homework assignments of TIPS students. Use $\alpha = 05$.

HWSTUDY10

Student	Science	Math	Student	Science	Math
1	0	2	6	2	3
2	4	3	7	4	0
3	3	0	8	2	1
4	1	1	9	3	1
5	3	1	10	4	1

Source: Van Voorhis, F. L. "Teachers' Use of Interactive Homework and Its Effects on Family Involvement and Science Achievement of Middle Grade Students." Paper presented at the annual meeting of the American Educational Research Association, Seattle, April 2001.

14.5 Comparing Three or More Populations: Completely Randomized Design

In Chapter 8 we used an analysis of variance and the F-test to compare the means of p populations (treatments) based on random sampling from populations that were normally distributed with a common variance σ^2. We now present a nonparametric technique—**Kruskal-Wallis *H*-test**—for comparing the populations that requires no assumptions concerning the population probability distributions.

Suppose a health administrator wants to compare the unoccupied bed space for three hospitals in the same city. She randomly selects 10 different days from the records of each hospital and lists the number of unoccupied beds for each day (see Table 14.7). Because the number of unoccupied beds per day may occasionally be quite large, it is conceivable that the population distributions of data may be skewed to the right and that this type of data may not satisfy the assumptions necessary for a parametric comparison of the population means. We therefore use a nonparametric analysis and base our comparison on the rank sums for the three sets of sample data. Just as with two independent samples (Section 14.3), the ranks are computed for each observation according to the relative magnitude of the measurements *when the data for all the samples are combined* (see Table 14.7). Ties are treated as they were for the Wilcoxon rank sum and signed rank tests by assigning the average value of the ranks to each of the tied observations.

We test

H_0: The probability distributions of the number of unoccupied beds are the same for all three hospitals

H_a: At least two of the three hospitals have probability distributions of the number of unoccupied beds that differ in location

If we denote the three sample rank sums by R_1, R_2, and R_3, the test statistic is given by

$$H = \frac{12}{n(n+1)}\sum\frac{R_j^2}{n_j} - 3(n+1)$$

HOSPBEDS

TABLE 14.7 Number of Available Beds

HOSPITAL 1		HOSPITAL 2		HOSPITAL 3	
Beds	Rank	Beds	Rank	Beds	Rank
6	5	34	25	13	9.5
38	27	28	19	35	26
3	2	42	30	19	15
17	13	13	9.5	4	3
11	8	40	29	29	20
30	21	31	22	0	1
15	11	9	7	7	6
16	12	32	23	33	24
25	17	39	28	18	14
5	4	27	18	24	16
$R_1 = 120$		$R_2 = 210.5$		$R_3 = 134.5$	

where n_j is the number of measurements in the jth sample and n is the total sample size ($n = n_1 + n_2 + \cdots + n_k$). For the data in Table 14.7, we have $n_1 = n_2 = n_3 = 10$ and $n = 30$. The rank sums are $R_1 = 120$, $R_2 = 210.5$, and $R_3 = 134.5$. Thus,

$$H = \frac{12}{30(31)}\left[\frac{(120)^2}{10} + \frac{(210.5)^2}{10} + \frac{(134.5)^2}{10}\right] - 3(31)$$
$$= 99.097 - 93 = 6.097$$

The H-statistic measures the extent to which the k samples differ with respect to their relative ranks. This is more easily seen by writing H in an alternative but equivalent form:

$$H = \frac{12}{n(n+1)}\sum n_j(\overline{R}_j - \overline{R})^2$$

where $\overline{R}_j$ is the mean rank corresponding to sample j, and $\overline{R}$ is the mean of all the ranks [that is, $\overline{R} = \frac{1}{2}(n + 1)$]. Thus, the H-statistic is 0 if all samples have the same mean rank and becomes increasingly large as the distance between the sample mean ranks grows.

If the null hypothesis is true, the distribution of H in repeated sampling is approximately a χ^2 (chi-square) distribution. This approximation for the sampling distribution of H is adequate as long as each of the k sample sizes exceeds 5. (See the references for more detail.) The degrees of freedom corresponding to the approximate sampling distribution of H will always be $(k - 1)$—one less than the number of probability distributions being compared. Because large values of H support the alternative hypothesis that the populations have different probability distributions, the rejection region for the test is located in the upper tail of the χ^2 distribution.

For the data of Table 14.7, the test statistic H has a χ^2 distribution with $(k - 1) = 2$ df. To determine how large H must be before we will reject the null hypothesis, we consult Table VI in Appendix B. For $\alpha = .05$ and df $= 2$, $\chi^2_{.05} = 5.99147$. Therefore, we can reject the null hypothesis that the three probability distributions are the same if $H > 5.99147$.

The rejection region is pictured in Figure 14.10. Since the calculated $H = 6.097$ exceeds the critical value of 5.99147, we conclude that at least one of the three hospitals tends to have a larger number of unoccupied beds than the others.

Now Work Exercise 14.39

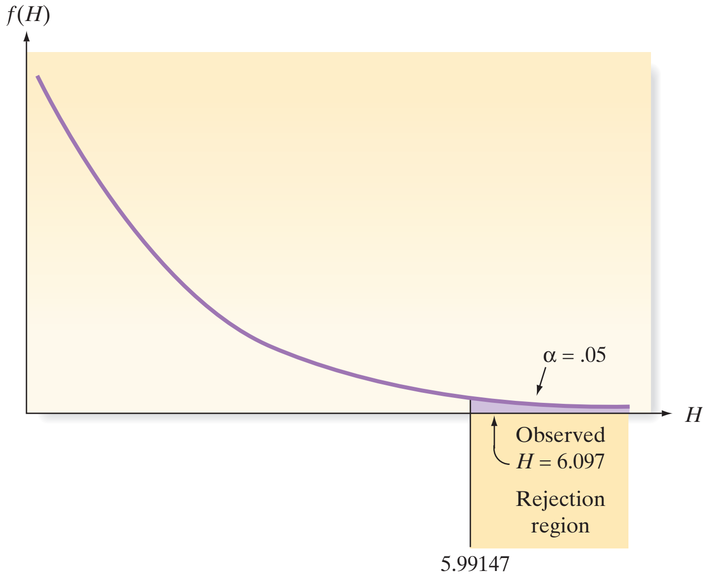


FIGURE 14.10
Rejection region for the comparison of three probability distributions

	A	B
1	**Kruskal-Wallis Rank Test**	
2		
3	**Data**	
4	**Level of Significance**	**0.05**
5		
6	Group 1	
7	Sum of Ranks	120
8	Sample Size	10
9	Group 2	
10	Sum of Ranks	210.5
11	Sample Size	10
12	Group 3	
13	Sum of Ranks	134.5
14	Sample Size	10
15		
16	Intermediate Calculations	
17	Sum of Squared Ranks/Sample Size	7680.05
18	Sum of Sample Sizes	30
19	Number of groups	3
20	*H* Test Statistic	6.097419
21		
22	**Test Result**	
23	**Critical Value**	**5.991476**
24	***p*-Value**	**0.04742**
25	**Reject the null hypothesis**	

FIGURE 14.11
Excel/PHStat2 printout of Kruskal-Wallis test

The same conclusion can be reached from a computer printout of the analysis. The rank sums, test statistic, and p-value of the nonparametric test are highlighted on the Excel printout in Figure 14.11. Since $\alpha = .05$ exceeds p-value $= .04742$, there is sufficient evidence to reject H_0.

The Kruskal-Wallis H-test for comparing more than two probability distributions is summarized in the box. Note that we can use the Wilcoxon rank sum test of Section 14.3 to compare the separate pairs of populations if the Kruskal-Wallis H-test supports the alternative hypothesis that at least two of the probability distributions differ.*

Kruskal-Wallis *H*-Test for Comparing *k* Probability Distributions

H_0: The k probability distributions are identical

H_a: At least two of the k probability distributions differ in location

$$\textit{Test statistic: } H = \frac{12}{n(n+1)}\sum\frac{R_j^2}{n_j} - 3(n+1)$$

where

n_j = Number of measurements in sample j

R_j = Rank sum for sample j, where the rank of each measurement is computed according to its relative magnitude in the totality of data for the k samples

n = Total sample size $= n_1 + n_2 + \cdots + n_k$

Rejection region: $H > \chi^2_\alpha$ with $(k - 1)$ degrees of freedom

Ties: Assign tied measurements the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign both a rank of $(3 + 4)/2 = 3.5$. The number of ties should be small relative to the total number of observations.

Conditions Required for the Validity of the Kruskal-Wallis *H*-Test

1. The k samples are random and independent.
2. There are five or more measurements in each sample.
3. The k probability distributions from which the samples are drawn are continuous.

Exercises 14.37–14.46

Learning the Mechanics

14.37 Under what circumstances does the χ^2 distribution provide an appropriate characterization of the sampling distribution of the Kruskal-Wallis H-statistic?

14.38 Data were collected from three populations, A, B, and C, using a completely randomized design. The following describes the sample data:

$$n_A = n_B = n_C = 15$$
$$R_A = 230 \quad R_B = 440 \quad R_C = 365$$

a. Specify the null and alternative hypotheses that should be used in conducting a test of hypothesis to determine whether the probability distributions of populations A, B, and C differ in location.

b. Conduct the test of part **a.** Use $\alpha = .05$.

c. What is the approximate p-value of the test of part **b?**

d. Calculate the mean rank for each sample, and compute H according to the formula that utilizes these means. Verify that this formula yields the same value of H that you obtained in part **b.**

14.39 NW Suppose you want to use the Kruskal-Wallis H-test to compare the probability distributions of three populations. The following are independent random samples selected from the three populations:

LM14_39

I:	66, 23, 55, 88, 58, 62, 79, 49
II:	19, 31, 16, 29, 30, 33, 40
III:	75, 96, 102, 75, 98, 78

a. What type of experimental design was used?

b. Specify the null and alternative hypotheses you would test.

c. Specify the rejection region you would use for your hypothesis test, at $\alpha = .01$.

*The multiple comparisons procedure of Chapter 8 can be used to rank the treatment medians. This nonparametric multiple comparisons of medians will control the experimentwise error rate selected by the analyst. Consult the references [Daniel (1990) and Dunn (1964)] for details.

d. Conduct the test at $\alpha = .01$.

Applying the Concepts—Basic

14.40 Forums for tax litigation. In litigating tax disputes with the IRS, taxpayers are permitted to choose the court forum. Three trial courts are available: (1) U.S. Tax Court, (2) Federal District Court, and (3) U.S. Claims Court. Accounting professors B. A. Billings (Wayne State University) and B. P. Green (University of Michigan-Dearborn) and business law professor W. H. Volz (Wayne State University) conducted a study of taxpayers' choice of forum in litigating tax issues (*Journal of Applied Business Research,* Fall 1996). A random sample of 161 court decisions were obtained for analysis. Two of the many variables measured for each case were taxpayer's choice of forum (Tax, District, or Claims Court) and tax deficiency DEF (i.e., the disputed amount, in dollars).

a. The researchers applied a nonparametric test rather than a parametric test to compare the DEF distributions of the three tax litigation forums. Give a plausible reason for their choice.

b. What nonparametric test is appropriate for this analysis? Explain.

c. The table below summarizes the data analyzed by the researchers. Use the information in the table to compute the appropriate test statistic.

d. The observed significance level (p-value) of the test was reported as $p = .0037$. Fully interpret this result.

Court Selected by Taxpayer	Sample Size	Sample Mean DEF	Rank Sum of DEF Values
Tax	67	$ 80,357	5,335
District	57	74,213	3,937
Claims	37	185,648	3,769

Source: Billings, B. A., Green, B. P., and Volz, W. H. "Selection of Forum for Litigated Tax Issues." *Journal of Applied Business Research,* Vol. 12, Fall 1996, p. 38 (Table 2).

TVADRECALL

14.41 Study of recall of TV commercials. Refer to the *Journal of Applied Psychology* (June 2002) study of recall of television commercials, Exercise 8.23. In a designed experiment, 324 adults were randomly assigned to one of three viewer groups: (1) watch a TV program with a violent content code (V) rating, (2) watch a show with a sex content code (S) rating, and (3) watch a neutral TV program. The number of brand names recalled in the commercial messages was recorded for each participant and the data are saved in the **TVADRECALL** file.

a. Give the null and alternative hypotheses for a Kruskal-Wallis test applied to the data.

b. The results of the nonparametric test are shown in the accompanying MINITAB printout. Locate the test statistic and p-value on the printout.

c. Interpret the results, part **b**, using $\alpha = .01$. What can the researchers conclude about the three groups of TV ad viewers?

MINITAB Output for Exercise 14.41

Kruskal-Wallis Test: RECALL versus GROUP

```
Kruskal-Wallis Test on RECALL

GROUP       N  Median  Ave Rank      Z
N         108   3.000     205.1   5.79
S         108   1.000     131.2  -4.26
V         108   2.000     151.2  -1.53
Overall   324             162.5

H = 36.04  DF = 2  P = 0.000
H = 37.15  DF = 2  P = 0.000  (adjusted for ties)
```

14.42 Office rental growth rates. Real estate market cycles are commonly divided into four phases that are based on the rate of change of the demand for and supply of properties: I—Recovery, II—Expansion, III—Hypersupply, and IV—Recession. Glenn Mueller of Johns Hopkins University studied the office market cycles of U.S. real estate markets (*Journal of Real Estate Research*, July/Aug. 1999). For each of the four market cycles, office rental growth rates (i.e., growth rates for asking rents) were measured for a sample of six different real estate markets. These data (in percentages) are presented in the table below.

MKTCYCLE

Phase I	Phase II	Phase III	Phase IV
2.7	10.5	6.1	−1.0
−1.0	11.5	1.2	6.2
1.1	9.4	11.4	−10.8
3.4	12.2	4.4	2.0
4.2	8.6	6.2	−1.1
3.5	10.9	7.6	−2.3

Source: Adapted from Mueller, G. R. "Real Estate Rental Growth Rates at Different Points in the Physical Market Cycle." *Journal of Real Estate Research,* Vol. 18, No. 1, July/Aug. 1999, pp. 131–150.

a. Specify the null hypothesis for a Kruskal-Wallis test.

b. Rank the 24 measurements in the data set.

c. Find the rank sums and calculate the test statistic.

d. Give the rejection region for the test at $\alpha = .05$.

e. Is there sufficient evidence to conclude that the distributions of office rental growth rates differ among the four market cycle phases?

f. What are the advantages and disadvantages of applying the Kruskal-Wallis H-test in part **a** rather than the parametric F-test of Chapter 10?

Applying the Concepts—Intermediate

14.43 Rates of return for mutual funds. Three of the categories *Business Week* uses to classify mutual funds are growth, blend, and value. Those with significantly lower-than-average price-earnings ratios (p-e) and price-to-book ratios (p-b) are *value funds;* those with higher-than-average p-e's and p-b's are called *growth funds;* those in the middle are called *blend funds.* The next table lists the pre-tax rate of return to investors for samples of five mutual funds in each of these three categories.

a. Do the data provide sufficient evidence to conclude that the rate-of-return distributions differ among the three types of mutual funds? Test using $\alpha = .05$.

b. What assumptions must hold for your test of part **a** to be valid?

c. Describe the Type I and II errors associated with the test of part **a** in the context of the problem.

d. Under what circumstances could the ANOVA *F*-test of Chapter 8 help answer part **a?**

MFUNDS

Category	Fund	12-Month Return (%)
Growth	Baron Asset	11.5
	North Track Geneva	8.6
	T Rowe Price	5.5
	Janus Enterprise	9.2
	Seligman Capital	2.2
Blend	Legg Mason	2.5
	Columbia	6.2
	Wells Fargo	12.9
	Saratoga	1.0
	J.P. Morgan	7.9
Value	First American	10.7
	Profunds	6.0
	Volumetric	3.2
	Riversource	13.2
	Rydex	6.6

Source: Business Week, Mutual Fund Scoreboard, Sep. 30, 2006.

14.44 Hazardous organic solvents. The *Journal of Hazardous Materials* (July 1995) published a study of the chemical properties of three different types of hazardous organic solvents used to clean metal parts. Independent samples of solvents from each of the three types—aromatics, chloralkanes, and esters—were collected. The data on sorption rates for the three solvents are listed in the table. Use a nonparametric test to compare the sorption rate distributions at $\alpha = .01$..

HAZARDS

Aromatics		Chloralkanes		Esters		
1.06	.95	1.58	1.12	.29	.43	.06
.79	.65	1.45	.91	.06	.51	.09
.82	1.15	.57	.83	.44	.10	.17
.89	1.12	1.16	.43	.61	.34	.60
1.05				.55	.53	.17

Source: Ortego, J. D., et al. "A Review of Polymeric Geosynthetics Used in Hazardous Waste Facilities." *Journal of Hazardous Materials,* Vol. 42, No. 2, July 1995, p. 142 (Table 9).

14.45 Debt/capital ratios of firms. A firm's debt/capital ratio is a measure of its long-term debt divided by total invested capital. To potential lenders to the firm, a high debt/capital ratio signals that in case of default, the lender is unlikely to recover outstanding loans to the firm. *Forbes* (March 2000) reported the debt/capital ratios (expressed as a percentage) for over 400 firms. The debt/capital ratios of selected firms in four industries are listed in the next table.

a. Use the Kruskal-Wallis *H*-test to investigate whether debt/capital ratios differ among the four industries. Be sure to specify your null and alternative hypotheses and to state your conclusion in the context of the problem. Use $\alpha = .05$.

b. Assuming the Kruskal-Wallis *H*-test indicates that differences exist among the four industries, which nonparametric procedure could be employed to compare the distribution of debt/capital ratios for the retailing and chemical industries? If appropriate, conduct the analysis.

DEBTCAP

Field	Firm	Debt/Capital %
Aerospace & Defense	Raytheon	45.3
	Boeing	37.0
	Textron	64.6
	Northrop Grumman Corp	40.6
	Lockheed Martin	63.9
	Rohr	63.3
Electric Utilities	PacificCorp	56.4
	Avista Corp.	59.9
	Florida Progress	58.6
	Duke Energy Corp.	46.9
	Northern States Power	49.8
	Consolidated Edison, Inc.	41.7
	Hawaiian Electric	36.5
Retailing	Rite Aid	62.2
	Kmart	31.2
	Bradlees	75.6
	Spiegel	48.8
	Fay's Incorporated	42.1
Chemical	Olin	22.6
	Valspar	47.2
	Dow Chemical	44.2
	FMC	67.0
	Union Carbide	47.6

Source: Forbes Current Statistics, March 2000 *(http://www.forbes.com/tool/toolbox/mktguide).*

14.46 Public defenders' salaries. Random samples of seven lawyers employed as public defenders were selected from each of three major cities. Their salaries were determined and are recorded in the table. You have been hired to determine whether differences exist among the salary distributions for public defenders in the three cities.

PUBDEF

Atlanta	Los Angeles	Washington, D.C.
$34,600	$ 42,400	$38,000
84,900	135,000	76,900
61,700	63,000	48,000
38,900	43,700	72,600
77,200	69,400	73,200
83,600	97,000	51,800
59,800	49,500	55,000

Source: Adapted from *American Almanac of Jobs and Salaries, 1997–1998 Edition,* New York: Avon Books, 1996, pp. 246–260.

a. Under what circumstances would it be appropriate to use the *F*-test for a completely randomized design to perform the required analysis?

b. Which assumptions required by the *F*-test are likely to be violated in this problem? Explain.

c. Use the Kruskal-Wallis *H*-test to determine whether the salary distributions differ among the three cities. Specify your null and alternative hypotheses, and state your conclusions in the context of the problem. Use $\alpha = .05$.

d. What assumptions are necessary to ensure the validity of the nonparametric test in part **c?**

14.6 Comparing Three or More Populations: Randomized Block Design (Optional)

In Section 8.4 we employed an analysis of variance to compare p population (treatment) means when the data were collected using a randomized block design. The *Friedman F_r-test* provides another method for testing to detect a shift in location of a set of p populations.* Like other nonparametric tests, it requires no assumptions concerning the nature of the populations other than the capacity of individual observations to be ranked.

Consider the problem of comparing the reaction times of subjects under the influence of different drugs. When the effect of a drug is short-lived (there is no carry over effect) and when the drug effect varies greatly from person to person, it may be useful to employ a *randomized block design*. Using the subjects as blocks, we would hope to eliminate the variability among subjects and thereby increase the amount of information in the experiment. Suppose that three drugs, A, B, and C, are to be compared using a randomized block design. Each of the three drugs is administered to the *same subject* with suitable time lags between the three doses. The order in which the drugs are administered is randomly determined for each subject. Thus, one drug would be administered to a subject and its reaction time would be noted; then after a sufficient length of time, the second drug administered; etc.

Suppose six subjects are chosen and that the reaction times for each drug are as shown in Table 14.8. To compare the three drugs, we rank the observations within each subject (block) and then compute the rank sums for each of the drugs (treatments). Tied observations within blocks are handled in the usual manner by assigning the average value of the ranks to each of the tied observations.

The null and alternative hypotheses are

H_0: The populations of reaction times are identically distributed for all three drugs

H_a: At least two of the drugs have probability distributions of reaction times that differ in location

The **Friedman *F_r*-statistic,** which is based on the rank sums for each treatment, is

$$F_r = \frac{12}{bk(k+1)}\sum R_j^2 - 3b(k+1)$$

where b is the number of blocks, k is the number of treatments, and R_j is the jth rank sum. For the data in Table 14.8,

$$F_r = \frac{12}{(6)(3)(4)}[(7)^2 + (12)^2 + (17)^2] - 3(6)(4) = 80.33 - 72 = 8.33$$

REACTION

TABLE 14.8 Reaction Time for Three Drugs

Subject	Drug A	Rank	Drug B	Rank	Drug C	Rank
1	1.21	1	1.48	2	1.56	3
2	1.63	1	1.85	2	2.01	3
3	1.42	1	2.06	3	1.70	2
4	2.43	2	1.98	1	2.64	3
5	1.16	1	1.27	2	1.48	3
6	1.94	1	2.44	2	2.81	3
		$R_1 = 7$		$R_2 = 12$		$R_3 = 17$

*The Friedman F_r-test was developed by the Nobel prize–winning economist Milton Friedman.

The Friedman F_r-statistic measures the extent to which the k samples differ with respect to their relative ranks within the blocks. This is more easily seen by writing F_r in an alternative, but equivalent, form:

$$F_r = \frac{12}{bk(k+1)}\sum b(\overline{R}_j - \overline{R})^2$$

where $\overline{R}_j$ is the mean rank corresponding to treatment j and $\overline{R}$ is the mean of all the ranks [(i.e., $\overline{R} = \frac{1}{2}(k+1)$)]. Thus, the F_r-statistic is 0 if all treatments have the same mean rank and becomes increasingly large as the distance between the sample mean ranks grows.

As for the Kruskal-Wallis H statistic, the Friedman F_r-statistic has approximately a χ^2 sampling distribution with $(k-1)$ degrees of freedom. Empirical results show the approximation to be adequate if either b (the number of blocks) or k (the number of treatments) exceeds 5. The Friedman F_r-test for a randomized block design is summarized in the next box.

Friedman F_r-Test for a Randomized Block Design

H_0: The probability distributions for the k treatments are identical

H_a: At least two of the probability distributions differ in location

$$\textit{Test statistic:}\quad F_r = \frac{12}{bk(k+1)}\sum R_j^2 - 3b(k+1)$$

where

b = Number of blocks

k = Number of treatments

R_j = Rank sum of the jth treatment; where the rank of each measurement is computed relative to its position *within its own block*

Rejection region: $F_r > \chi^2_\alpha$ with $(k-1)$ degrees of freedom

Ties: Assign tied measurements within a block the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign each a rank of $(3+4)/2 = 3.5$. The number of ties should be small relative to the total number of observations.

Conditions Required for a Valid Friedman F_r-Test

1. The treatments are randomly assigned to experimental units within the blocks.
2. The measurements can be ranked within blocks.
3. The k probability distributions from which the samples within each block are drawn are continuous.

For the drug example, we will use $\alpha = .05$ to form the rejection region:

$$F_r > \chi^2_{.05} = 5.99147 \text{ (see Figure 14.12)}$$

where $\chi^2_{.05}$ is based on $(k-1) = 2$ degrees of freedom. Consequently, because the observed value, $F_r = 8.33$, exceeds 5.99147, we conclude that at least two of the three drugs have probability distributions of reaction times that differ in location.

A MINITAB printout of the nonparametric analysis, shown in Figure 14.13, confirms our inference. Both the test statistic and p-value are highlighted on the printout. Since p-value = .016 is less than our selected $\alpha = .05$, there is evidence to reject H_0.

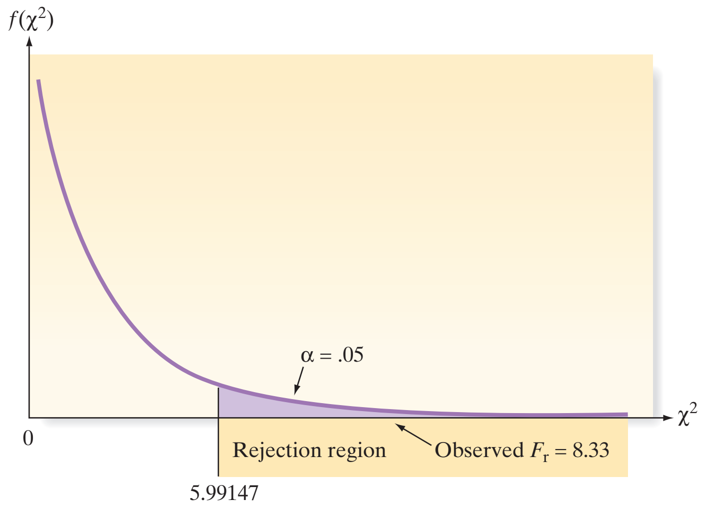


FIGURE 14.12
Rejection region for reaction time example

Friedman Test: REACTIME versus DRUG blocked by SUBJECT

```
S = 8.33    DF = 2    P = 0.016

                              Sum
                               of
DRUG   N   Est Median      Ranks
A      6       1.5283        7.0
B      6       1.7417       12.0
C      6       1.8950       17.0

Grand median = 1.7217
```

FIGURE 14.13
MINITAB printout of Friedman test

Now Work Exercise 14.48

Clearly, the assumptions for this test—that the measurements are ranked within blocks and that the number of blocks (subjects) is greater than 5—are satisfied. However, we must be sure that the treatments are randomly assigned to blocks. For the procedure to be valid, we assume that the three drugs are administered in a random order to each subject. If this were not true, the difference in the reaction times for the three drugs might be due to the order in which the drugs are given.

Exercises 14.47–14.56

Learning the Mechanics

14.47 Data were collected using a randomized block design with four treatments (A, B, C, and D) and $b = 6$. The following rank sums were obtained:

$$R_A = 11 \quad R_B = 21 \quad R_C = 21 \quad R_D = 7$$

a. How many blocks were used in the experimental design?

b. Specify the null and alternative hypotheses that should be used in conducting a hypothesis test to determine whether the probability distributions for at least two of the treatments differ in location.

c. Conduct the test of part **b.** Use $\alpha = .10$.

d. What is the approximate p-value of the test of part **c**?

e. Calculate the mean rank for each of the four treatments and compute the value of the F_r-test statistic according to the formula that utilizes those means. Verify that the test statistic is the same as that you obtained in part **c.**

14.48 NW Suppose you have used a randomized block design to help you compare the effectiveness of three different treatments, A, B, and C. You obtained the data given in the next table and plan to conduct a Friedman F_r-test.

a. Specify the null and alternative hypotheses you will test.

b. Specify the rejection region for the test. Use $\alpha = .10$.

c. Conduct the test and interpret the results.

LM14_48

Block	TREATMENT A	B	C
1	9	11	18
2	13	13	13
3	11	12	12
4	10	15	16
5	9	8	10
6	14	12	16
7	10	12	15

14.49 An experiment was conducted using a randomized block design with four treatments and six blocks. The ranks of the measurements within each block are shown in the table. Use the Friedman F_r-test for a randomized block design to determine whether the data provide sufficient evidence to indicate that at least two of the treatment probability distributions differ in location. Test using $\alpha = .05$.

LM14_49

Treatment	BLOCK 1	2	3	4	5	6
1	3	3	2	3	2	3
2	1	1	1	2	1	1
3	4	4	3	4	4	4
4	2	2	4	1	3	2

Applying the Concepts—Basic

14.50 Conditions impeding farm production A review of farmer involvement in agricultural research was presented in the *Journal of Agricultural, Biological, and Environmental Statistics* (Mar. 2001). In one study, each of six farmers ranked the level of farm production constraint imposed by five conditions: drought, pest damage, weed interference, farming costs, and labor shortage. The rankings, ranging from 1 (least severe) to 5 (most severe), and rank sums for the five conditions are listed in the table at the bottom of the page.

a. Use the rank sums shown in the table to compute the Friedman F_r statistic.

b. At $\alpha = .05$, find the rejection region for a test to compare the farmer opinion distributions for the five conditions.

c. Make the proper conclusion, in the words of the problem.

14.51 Designing an atlas. Refer to the *Journal of Geography*'s published rankings of regional atlas theme topics, Exercise 14.32 (p. 14-26). In addition to high school teachers and university geography alumni, university geography students and representatives of the general public also ranked the 12 thematic topics. The rankings of all four groups are shown in the table at the top of the next page. A MINITAB analysis is also provided below to compare the atlas theme ranking distributions of the four groups.

a. Locate the rank sums on the printout.

b. Use the rank sums to find the Friedman F_r statistic.

c. Locate the test statistic and associated p-value on the printout.

d. Conduct the test and give the conclusion in the words of the problem.

Friedman Test: RANK versus GROUP blocked by THEME

```
S = 0.93  DF = 3  P = 0.819
S = 1.08  DF = 3  P = 0.782 (adjusted for ties)

                              Sum
                               of
GROUP    N  Est Median      Ranks
1       12      5.1250       27.0
2       12      6.1250       32.5
3       12      5.6250       29.0
4       12      6.1250       31.5

Grand median = 5.7500
```

14.52 Rotary oil rigs. Refer to Exercise 8.51 and the *World Oil* (Jan. 2002) study of rotary oil rigs. Three months were randomly selected and the number of oil rigs running in each of three states—California, Utah, and Alaska—was recorded. The data for the randomized block design are reproduced on the next page. Consider a nonparametric test to compare the distributions of rotary oil rigs running in the three states.

a. State the null and alternative hypothesis for the test.

FARM6

		CONDITION Drought	Pest Damage	Weed Interference	Farming Costs	Labor Shortage
Farmer	**1**	5	4	3	2	1
	2	5	3	4	1	2
	3	3	5	4	2	1
	4	5	4	1	2	3
	5	4	5	3	2	1
	6	5	4	3	2	1
	Rank sum	27	25	18	11	9

Source: Riley, J., and Fielding, W. J. "An Illustrated Review of Some Farmer Participatory Research Techniques." *Journal of Agricultural, Biological, and Environmental Statistics,* Vol. 6, No. 1, Mar. 2001 (Table 1).

ATLAS2

Theme	RANKINGS			
	High School Teachers	Geography Alumni	Geography Students	General Public
Tourism	10	2	5	1
Physical	2	1	1	5
Transportation	7	3	7	2
People	1	6	2	3
History	2	5	9	4
Climate	6	4	4	8
Forestry	5	8	2	7
Agriculture	7	10	6	9
Fishing	9	7	10	6
Energy	2	8	7	10
Mining	10	11	11	11
Manufacturing	12	12	12	12

Source: Keller, C. P., et al. "Planning the Next Generation of Regional Atlases: Input from Educators." *Journal of Geography,* Vol. 94, No. 3, May/June 1995, p. 413 (Table 1).

b. Rank the data within each month, then sum the ranks for each state.

c. Use the rank sums from part **b** to find the value of the test statistic.

d. Give the rejection region of the test at $\alpha = .05$.

e. State the conclusion in the words of the problem. Does the conclusion agree with that for the ANOVA F-test conducted in Exercise 8.51?

OILRIGS

Month/Year	California	Utah	Alaska
Nov. 2000	27	17	11
Oct. 2001	34	20	14
Nov. 2001	36	15	14

Applying the Concepts—Intermediate

14.53 Reducing on-the-job stress. Refer to the Kansas State study designed to investigate the effects of plants on human stress levels, Exercise 8.54. The data (next column) are given as finger temperatures for each of 10 students in a dimly lit room under three experimental conditions: presence of a live plant, presence of a plant photo, and absence of a plant (either live or photo). Analyze the data using a nonparametric procedure. Do the students' finger temperatures depend on the experimental condition?

PLANTS

Student	Live Plant	Plant Photo	No Plant (control)
1	91.4	93.5	96.6
2	94.9	96.6	90.5
3	97.0	95.8	95.4
4	93.7	96.2	96.7
5	96.0	96.6	93.5
6	96.7	95.5	94.8
7	95.2	94.6	95.7
8	96.0	97.2	96.2
9	95.6	94.8	96.0
10	95.6	92.6	96.6

Source: Elizabeth Schreiber, Department of Statistics, Kansas State University, Manhattan, Kansas.

14.54 Effectiveness of a mosquito insecticide. The *Journal of the American Mosquito Control Association* (Mar. 1995) reported on a study of the effectiveness of five different types of insecticides in controlling a species of Caribbean mosquito. The resistance ratios (i.e., the dosage of insecticide required to kill 50% of the larvae divided by the known dosage for a susceptible mosquito strain) of the insecticides at each of seven Caribbean locations are reproduced in the table at the bottom of the page. Compare the resistance ratio distributions of the five insecticides using a nonparametric procedure. Are any of the insecticides more effective than any of the others?

MOSQUITO

		INSECTICIDE				
		Temephos	Malathion	Fenitrothion	Fenthion	Chlorpyrifos
Location	**Anguilla**	4.6	1.2	1.5	1.8	1.5
	Antigua	9.2	2.9	2.0	7.0	2.0
	Dominica	7.8	1.4	2.4	4.2	4.1
	Guyana	1.7	1.9	2.2	1.5	1.8
	Jamaica	3.4	3.7	2.0	1.5	7.1
	St. Lucia	6.7	2.7	2.7	4.8	8.7
	Suriname	1.4	1.9	2.0	2.1	1.7

Source: Rawlins, S. C., and Oh Hing Wan, J. "Resistance in Some Caribbean Populations of Aedes aegypti to Several Insecticides." *Journal of the American Mosquito Control Association,* Vol. 11, No. 1, Mar. 1995 (Table 1).

14.55 Sealers used to retard metal corrosion. Corrosion of different metals is a problem in many mechanical devices. Three sealers used to help retard the corrosion of metals were tested to see whether there were any differences among them. Samples of 10 different metal compositions were treated with each of the three sealers, and the amount of corrosion was measured after exposure to the same environmental conditions for 1 month. The data are given in the table below. Is there any evidence of a difference in the probability distributions of the amounts of corrosion among the three types of sealer? Use $\alpha = .05$.

CORRODE

	SEALER		
Metal	1	2	3
1	4.6	4.2	4.9
2	7.2	6.4	7.0
3	3.4	3.5	3.4
4	6.2	5.3	5.9
5	8.4	6.8	7.8
6	5.6	4.8	5.7
7	3.7	3.7	4.1
8	6.1	6.2	6.4
9	4.9	4.1	4.2
10	5.2	5.0	5.1

14.56 Absentee rates at a jeans plant. Refer to Exercise 8.55 and the *New Technology, Work, and Employment* (July 2001) study of daily worker absentee rates at a jeans plant. Nine weeks were randomly selected and the absentee rate (percentage of workers absent) determined for each day (Monday through Friday) of the workweek. The data are reproduced in the table. Conduct a nonparametric analysis of the data to compare the distributions of absentee rates for the 5 days of the workweek.

JEANS

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	5.3	0.6	1.9	1.3	1.6
2	12.9	9.4	2.6	0.4	0.5
3	0.8	0.8	5.7	0.4	1.4
4	2.6	0.0	4.5	10.2	4.5
5	23.5	9.6	11.3	13.6	14.1
6	9.1	4.5	7.5	2.1	9.3
7	11.1	4.2	4.1	4.2	4.1
8	9.5	7.1	4.5	9.1	12.9
9	4.8	5.2	10.0	6.9	9.0

Source: Boggis, J. J. "The Eradication of Leisure." *New Technology, Work, and Employment,* Volume 16, Number 2, July 2001 (Table 3).

14.7 Rank Correlation

Suppose 10 new car models are evaluated by two consumer magazines and each magazine ranks the braking systems of the cars from 1 (best) to 10 (worst). We want to determine whether the magazines' ranks are related. Does a correspondence exist between their ratings? If a car is ranked high by magazine 1, is it likely to be ranked high by magazine 2? Or do high rankings by one magazine correspond to low rankings by the other? That is, are the rankings of the magazines *correlated?*

If the rankings are as shown in the "Perfect Agreement" columns of Table 14.9, we immediately notice that the magazines agree on the rank of every car. High ranks correspond to high ranks and low ranks to low ranks. This is an example of *perfect positive correlation* between the ranks. In contrast, if the rankings appear as shown in the "Perfect Disagreement" columns of Table 14.9, high ranks for one magazine correspond to low ranks for the other. This is an example of *perfect negative correlation*.

TABLE 14.9 Brake Rankings of 10 New Car Models by Two Consumer Magazines

	PERFECT AGREEMENT		PERFECT DISAGREEMENT	
Car Model	Magazine 1	Magazine 2	Magazine 1	Magazine 2
1	4	4	9	2
2	1	1	3	8
3	7	7	5	6
4	5	5	1	10
5	2	2	2	9
6	6	6	10	1
7	8	8	6	5
8	3	3	4	7
9	10	10	8	3
10	9	9	7	4

TABLE 14.10 Brake Rankings of New Car Models: Less Than Perfect Agreement

	MAGAZINE		DIFFERENCE BETWEEN RANK 1 AND RANK 2	
Car Model	1	2	d	d^2
1	4	5	−1	1
2	1	2	−1	1
3	9	10	−1	1
4	5	6	−1	1
5	2	1	1	1
6	10	9	1	1
7	7	7	0	0
8	3	3	0	0
9	6	4	2	4
10	8	8	0	0
				$\Sigma d^2 = 10$

In practice, you will rarely see perfect positive or negative correlation between the ranks. In fact, it is quite possible for the magazines' ranks to appear as shown in Table 14.10. You will note that these rankings indicate some agreement between the consumer magazines, but not perfect agreement, thus indicating a need for a measure of rank correlation.

Spearman's rank correlation coefficient, r_s, provides a measure of correlation between ranks. The formula for this measure of correlation is given in the next box. We also give a formula that is identical to r_s when there are no ties in rankings; this provides a good approximation to r_s when the number of ties is small relative to the number of pairs.

Note that if the ranks for the two magazines are identical, as in the second and third columns of Table 14.9, the differences between the ranks, *d,* will all be 0. Thus,

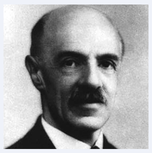
$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(0)}{10(99)} = 1$$

That is, *perfect positive correlation* between the pairs of ranks is characterized by a Spearman correlation coefficient of $r_s = 1$. When the ranks indicate perfect disagreement, as in the fourth and fifth columns of Table 14.9, $\Sigma d_i^2 = 330$ and

$$r_s = 1 - \frac{6(330)}{10(99)} = -1$$

Thus, *perfect negative correlation* is indicated by $r_s = -1$.

Biography



CHARLES E. SPEARMAN (1863–1945)
Spearman's Correlation

London-born Charles Spearman was educated at Leamington College before joining the British Army. After 20 years as a highly decorated officer, Spearman retired from the army and moved to Germany to begin his study of experimental psychology at the University of Liepzig. At the age of 41 he earned his PhD and ultimately become one of the most influential figures in the field of psychology. Spearman was the originator of the classical theory of mental tests and developed the "two-factor" theory of intelligence. These theories were used to develop and support the "Plus-Elevens" tests in England—exams administered to British 11-year-olds that predict whether they should attend a university or technical school. Spearman was greatly influenced by the works of Francis Galton; consequently, he developed a strong statistical background. While conducting his research on intelligence, he proposed the rank order correlation coefficient—now called *Spearman's correlation coefficient.* During his career, Spearman spent time at various universities, including University College (London), Colombia University, Catholic University, and the University of Cairo (Egypt).

Spearman's Rank Correlation Coefficient

$$r_s = \frac{SS_{uv}}{\sqrt{SS_{uu}SS_{vv}}}$$

where

$$SS_{uv} = \sum(u_i - \bar{u})(v_i - \bar{v}) = \sum u_i v_i - \frac{\left(\sum u_i\right)\left(\sum v_i\right)}{n}$$

$$SS_{uu} = \sum(u_i - \bar{u})^2 = \sum u_i^2 - \frac{\left(\sum u_i\right)^2}{n}$$

$$SS_{vv} = \sum(v_i - \bar{v})^2 = \sum v_i^2 - \frac{\left(\sum v_i\right)^2}{n}$$

u_i = Rank of the ith observation in sample 1
v_i = Rank of the ith observation in sample 2
n = Number of pairs of observations (number of observations in each sample)

Shortcut Formula for r_s*

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

where

$d_i = u_i - v_i$ (difference in ranks of ith observations for samples 1 and 2)

For the data of Table 14.10,

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(10)}{10(99)} = 1 - \frac{6}{99} = .94$$

The fact that r_s is close to 1 indicates that the magazines tend to agree, but the agreement is not perfect.

The value of r_s always falls between -1 and $+1$, with $+1$ indicating perfect positive correlation and -1 indicating perfect negative correlation. The closer r_s falls to $+1$ or -1, the greater the correlation between the ranks. Conversely, the nearer r_s is to 0, the less the correlation.

Note that the concept of correlation implies that two responses are obtained for each experimental unit. In the consumer magazine example, each new car model received two ranks (one for each magazine) and the objective of the study was to determine the degree of positive correlation between the two rankings. Rank correlation methods can be used to measure the correlation between any pair of variables. If two variables are measured on each of n experimental units, we rank the measurements associated with each variable separately. Ties receive the average of the ranks of the tied observations. Then we calculate the value of r_s for the two rankings. This value measures the rank correlation between the two variables. We illustrate the procedure in Example 14.4.

*The shortcut formula is not exact when there are tied measurements, but it is a good approximation when the total number of ties is not large relative to n.

EXAMPLE 14.4 Spearman's Rank Correlation Applied to a Food Preservation Study

Problem Manufacturers of perishable foods often use preservatives to retard spoilage. One concern is that too much preservative will change the flavor of the food. Suppose an experiment is conducted using samples of a food product with varying amounts of preservative added. Both length of time until the food shows signs of spoiling and a taste rating are recorded for each sample. The taste rating is the average rating for three tasters, each of whom rates each sample on a scale from 1 (good) to 5 (bad). Twelve sample measurements are shown in Table 14.11.

a. Calculate Spearman's rank correlation coefficient between spoiling time and taste rating.

b. Use a nonparametric test to find out whether the spoilage times and taste ratings are negatively correlated. Use $\alpha = .05$.

Solution

a. We first rank the days until spoilage, assigning a 1 to the smallest number (26) and a 12 to the largest (109). Similarly, we assign ranks to the 12 taste ratings. [*Note:* The tied taste ratings receive the average of their respective ranks.] Since the number of ties is relatively small, we will use the shortcut formula to calculate r_s. The differences d between the ranks of days until spoilage and the ranks of taste rating are shown in Table 14.11. The squares of the differences, d^2, are also given. Thus,

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6(536.5)}{12(12^2 - 1)} = 1 - 1.88 = -.88$$

The value of r_s can also be obtained using a computer. An SPSS printout of the analysis is shown in Figure 14.14. The value of r_s, highlighted on the printout, is $-.879$ and agrees (except for rounding) with our hand-calculated value. This negative correlation coefficient indicates that in this sample an increase in the number of days until spoilage is *associated with* (but is not necessarily the *cause of*) a decrease in the taste rating.

b. If we define ρ as the **population rank correlation coefficient** [i.e., the rank correlation coefficient that could be calculated from all (x, y) values in the population], this question can be answered by conducting the test

SPOILAGE

TABLE 14.11 Data and Correlations for Example 14.4

Sample	Days Until Spoilage	Rank	Taste Rating	Rank	d	d^2
1	30	2	4.3	11	−9	81
2	47	5	3.6	7.5	−2.5	6.25
3	26	1	4.5	12	−11	121
4	94	11	2.8	3	8	64
5	67	7	3.3	6	1	1
6	83	10	2.7	2	8	64
7	36	3	4.2	10	−7	49
8	77	9	3.9	9	0	0
9	43	4	3.6	7.5	−3.5	12.25
10	109	12	2.2	1	11	121
11	56	6	3.1	5	1	1
12	70	8	2.9	4	4	16
					Total = 536.5	

Note: *Tied measurements are assigned the average of the ranks they would be given if they were different but consecutive.*

Correlations

			DAYS	TASTE
Spearman's rho	DAYS	Correlation Coefficient	1.000	-.879**
		Sig. (2-tailed)	.	.000
		N	12	12
	TASTE	Correlation Coefficient	-.879**	1.000
		Sig. (2-tailed)	.000	.
		N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

FIGURE 14.14
SPSS printout of Spearman's rank correlation test

H_0: $\rho = 0$ (no population correlation between ranks)
H_a: $\rho < 0$ (negative population correlation between ranks)
Test statistic: r_s (the *sample* Spearman rank correlation coefficient)

To determine a rejection region, we consult Table XVI in Appendix B, which is partially reproduced in Table 14.12. Note that the left-hand column gives values of *n*, the number of pairs of observations. The entries in the table are values for an upper-tail rejection region, since only positive values are given. Thus, for $n = 12$ and $\alpha = .05$, the value .497 is the boundary of the upper-tailed rejection region, so that $P(r_s > .497) = .05$ if H_0: $\rho = 0$ is true. Similarly, for negative values of r_s, we have $P(r_s < -.497) = .05$ if $\rho = 0$—that is, we expect to see $r_s < -.497$ only 5% of the time if there is really no relationship between the ranks of the variables. The lower-tailed rejection region is therefore

$$\textit{Rejection region } (\alpha = .05)\text{: } r_s < -.497$$

Since the calculated $r_s = -.876$ is less than $-.497$, we reject H_0 at the $\alpha = .05$ level of significance—that is, this sample provides sufficient evidence to conclude that a negative correlation exists between number of days until spoilage and taste rating of the food product. It appears that the preservative does affect the taste of this food adversely.

TABLE 14.12 Reproduction of Part of Table XVI in Appendix B: Critical Values of Spearman's Rank Correlation Coefficient

n	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
5	.900	—	—	—
6	.829	.886	.943	—
7	.714	.786	.893	—
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.745	.794
11	.523	.623	.736	.818
12	.497	.591	.703	.780
13	.475	.566	.673	.745
14	.457	.545	.646	.716
15	.441	.525	.623	.689
16	.425	.507	.601	.666
17	.412	.490	.582	.645
18	.399	.476	.564	.625
19	.388	.462	.549	.608
20	.377	.450	.534	.591

Look Back The two-tailed *p*-value of the test is given on the SPSS printout, Figure 14.14, below the value of r_s. Since the lower-tailed *p*-value, $p = .000/2 = .000$, is less than $\alpha = .05$, our conclusion is the same: reject H_0.

Now Work Exercise 14.60

A summary of Spearman's nonparametric test for correlation is given in the box.

Spearman's Nonparametric Test for Rank Correlation

One-Tailed Test	Two-Tailed Test
H_0: $\rho = 0$	H_0: $\rho = 0$
H_a: $\rho > 0$ (or H_a: $\rho < 0$)	H_a: $\rho \neq 0$

Test statistic: r_s, the sample rank correlation (see the formulas for calculating r_s)

One-Tailed Test	Two-Tailed Test
Rejection region: $r_s > r_{s,\alpha}$ (or $r_s < -r_{s,\alpha}$ when H_a: $\rho_s < 0$)	*Rejection region:* $\lvert r_s \rvert > r_{s,\alpha/2}$
where $r_{s,\alpha}$ is the value from Table XVI corresponding to the upper-tail area α and n pairs of observations	where $r_{s,\alpha/2}$ is the value from Table XVI corresponding to the upper-tail area $\alpha/2$ and n pairs of observations

Ties: Assign tied measurements the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign each a rank of $(3 + 4)/2 = 3.5$. The number of ties should be small relative to the total number of observations.

Conditions Required for a Valid Spearman's Test

1. The sample of experimental units on which the two variables are measured is randomly selected.
2. The probability distributions of the two variables are continuous.

STATISTICS IN ACTION REVISITED

Testing whether the TCDD Levels in Fat and Plasma of Vietnam Vets are Correlated

The medical researchers who examined the Vietnam War veterans exposed to Agent Orange also wanted an estimate of the correlation between the TCDD level in fat tissue and the TCDD level in plasma. Since the two variables were not normally distributed, they employed Spearman's rank correlation method. The SPSS printout for this analysis is shown in Figure SIA14.5.

The value of r_s (highlighted on the printout) is .774; thus, there appears to be a fairly strong positive correlation between fat and plasma TCDD levels in the sample of Vietnam vets. The one-tailed p-value of the test (also highlighted) is .000, indicating that there is sufficient evidence (at $\alpha = .01$) of a positive association between the two TCDD measures.

Correlations

			FAT	PLASMA
Spearman's rho	FAT	Correlation Coefficient	1.000	.774**
		Sig. (1-tailed)	.	.000
		N	20	20
	PLASMA	Correlation Coefficient	.774**	1.000
		Sig. (1-tailed)	.000	.
		N	20	20

**. Correlation is significant at the 0.01 level (1-tailed).

Figure SIA14.5
SPSS Spearman rank correlation test for TCDD data

Exercises 14.57–14.68

Learning the Mechanics

14.57 Use Table XVI of Appendix B to find each of the following probabilities:

a. $P(r_s > .508)$ when $n = 22$

b. $P(r_s > .448)$ when $n = 28$

c. $P(r_s \leq .648)$ when $n = 10$

d. $P(r_s < -.738$ or $r_s > .738)$ when $n = 8$

14.58 Specify the rejection region for Spearman's nonparametric test for rank correlation in each of the following situations:

a. H_0: $\rho = 0$; H_a: $\rho \neq 0, n = 10, \alpha = .05$

b. H_0: $\rho = 0$; H_a: $\rho > 0, n = 20, \alpha = .025$

c. H_0: $\rho = 0$; H_a: $\rho < 0, n = 30, \alpha = .01$

14.59 Compute Spearman's rank correlation coefficient for each of the following pairs of sample observations:

a.

x	33	61	20	19	40
y	26	36	65	25	35

b.

x	89	102	120	137	41
y	81	94	75	52	136

c.

x	2	15	4	10
y	11	2	15	21

d.

x	5	20	15	10	3
y	80	83	91	82	87

14.60 The following sample data were collected on variables x and y:
NW

x	0	3	0	−4	3	0	4
y	0	2	2	0	3	1	2

a. Specify the null and alternative hypotheses that should be used in conducting a hypothesis test to determine whether the variables x and y are correlated.

b. Conduct the test of part **a** using $\alpha = .05$.

c. What is the approximate p-value of the test of part **b?**

d. What assumptions are necessary to ensure the validity of the test of part **b?**

Applying the Concepts—Basic

14.61 Extending the life of an aluminum smelter pot. Refer to the *American Ceramic Society Bulletin* (Feb. 2005) study of the life length of an aluminum smelter pot, Exercise 8.16. Since the life of a smelter pot depends on the porosity of the brick lining, the researchers measured the apparent porosity and the mean pore diameter of each of six bricks. The data are reproduced in the next table.

a. Rank the apparent porosity values for the six bricks. Then rank the six pore diameter values.

b. Use the ranks, part **a**, to find the rank correlation between apparent porosity (y) and mean pore diameter (x). Interpret the result.

c. Conduct a test for positive rank correlation. Use $\alpha = .01$.

SMELTPOT

Brick	Apparent Porosity (%)	Mean Pore Diameter (micro-meters)
A	18.8	12.0
B	18.3	9.7
C	16.3	7.3
D	6.9	5.3
E	17.1	10.9
F	20.4	16.8

Source: Bonadia, P., et al. "Aluminosilicate Refractories for Aluminum Cell Linings." *The American Ceramic Society Bulletin*, Vol. 84, No. 2, Feb. 2005 (Table II).

14.62 Wine-tasting experiment. Two expert wine tasters were asked to rank six brands of wine. Their rankings are shown in the table.

a. Use the rankings to compute r_s.

b. Do the data present sufficient evidence to indicate a positive correlation in the rankings of the two experts? Test using $\alpha = .05$.

WINETASTE

Brand	Expert 1	Expert 2
A	6	5
B	5	6
C	1	2
D	3	1
E	2	4
F	4	3

14.63 Organizational use of the Internet. Researchers from the United Kingdom and Germany attempted to develop a theoretically grounded measure of organizational Internet use (OIU) and published their results in *Internet Research* (Vol. 15, 2005). Using data collected from a sample of 77 Web sites, they investigated the link between OIU level (measured on a 7-point scale) and several observation-based indicators. Spearman's rank correlation coefficient (and associated p-values) for several indicators are shown in the table.

	CORRELATION WITH OIU LEVEL	
Indicator	r_s	p-value
Navigability	.179	.148
Transactions	.334	.023
Locatability	.590	.000
Information richness	−.115	.252
Number of files	.114	.255

Source: Brock, J. K., and Zhou, Y. "Organizational Use of the Internet." *Internet Research*, Vol. 15, No. 1, 2005 (Table IV).

a. Interpret each of the values of r_s given in the table.

b. Interpret each of the p-values given in the table. (Use $\alpha = .10$ to conduct each test.)

Applying the Concepts—Intermediate

14.64 Sweetness of orange juice. Refer to the orange juice quality study, Exercise 10.18. Recall that a manufacturer that has developed a quantitative index of the "sweetness" of orange juice is investigating the relationship between the sweetness index and the amount of water soluble pectin in the orange juice it produces. The data for 24 production runs at a juice manufacturing plant are reproduced in the table below.

a. Calculate Spearman's rank correlation coefficient between the sweetness index and the amount of pectin. Interpret the result.

b. Conduct a nonparametric test to determine whether there is a negative association between the sweetness index and the amount of pectin. Use $\alpha = .01$.

OJUICE

Run	Sweetness Index	Pectin (ppm)	Run	Sweetness Index	Pectin (ppm)
1	5.2	220	13	5.8	306
2	5.5	227	14	5.5	259
3	6.0	259	15	5.3	284
4	5.9	210	16	5.3	383
5	5.8	224	17	5.7	271
6	6.0	215	18	5.5	264
7	5.8	231	19	5.7	227
8	5.6	268	20	5.3	263
9	5.6	239	21	5.9	232
10	5.9	212	22	5.8	220
11	5.4	410	23	5.8	246
12	5.6	256	24	5.9	241

Note: The data in the table are authentic. For confidentiality reasons, the manufacturer cannot be disclosed.

14.65 Assessmemt of biometric recognition methods. Biometric technologies have been developed to detect or verify an individual's identity. These methods are based on physiological characteristics (called *biometric signatures*) such as facila features, eye irises, fingerprints, voice, hand shape, and gait. In *Chance* (Winter 2004), four biometric recognition algorithms were compared. All four methods were applied to 1,196 biometric signatures and "match" scores were obtained. The Spearman correlation between match scores for each possible algorithm pair was determined. The rank correlation matrix is shown below. Interpret the results.

Method	I	II	III	IV
I	1	.189	.592	.340
II		1	.205	.324
III			1	.314
IV				1

14.66 Parent companies and subsidiaries. Metropolitan areas with many corporate headquarters are finding it easier to transition from a manufacturing economy to a service economy through job growth in small companies and subsidiaries that service the corporate parent. James O. Wheeler of the University of Georgia studied the relationship between the number of corporate headquarters in 11 metropolitan areas and the number of subsidiaries located there (*Growth and Change*, Spring 1988). He hypothesized that there would be a positive relationship between the variables.

METRO

Metropolitan Area	No. of Parent Companies	No. of Subsidiaries
New York	643	2,617
Chicago	381	1,724
Los Angeles	342	1,867
Dallas–Ft. Worth	251	1,238
Detroit	216	890
Boston	208	681
Houston	192	1,534
San Francisco	141	899
Minneapolis	131	492
Cleveland	128	579
Denver	124	672

Source: Wheeler, J. O. "The Corporate Role of Large Metropolitan Areas in the United States." *Growth and Change,* Spring 1988, pp. 75–88.

a. Calculate Spearman's rank correlation coefficient for the data in the table above. What does it indicate about Wheeler's hypothesis?

b. To conduct a formal test of Wheeler's hypothesis using Spearman's rank correlation coefficient, certain assumptions must hold. What are they? Do they appear to hold? Explain.

14.67 Company reputations. *The Wall Street Journal* (Feb. 7, 2001) reported on a Harris Interactive, Inc., survey of consumers to rate the reputations of America's most visible companies. The 1999 and 2000 ranks of 15 randomly selected companies are listed in the accompanying table. Conduct a test to determine if the 1999 and 2000 reputation ranks are positively correlated. Use $\alpha = .05$.

COREP

Company	1999 Rank	2000 Rank
Johnson & Johnson	1	1
Anheuser-Busch	14	6
Disney	10	8
Microsoft	15	9
FedEx	21	13
Wal-Mart	6	14
Coca-Cola	2	16
McDonalds	24	24
Yahoo!	19	27
Sears	29	35
America Online	26	39
Kmart	37	41
Toyota	28	19
Home Depot	8	4
IBM	17	7

Source: Harris Interactive, the Reputation Institute.

14.68 Employee suggestion system. An *employee suggestion system* is a formal process for capturing, analyzing, implementing, and recognizing employee-proposed organizational improvements. (The first known system was implemented by the Yale and Towne Manufacturing Company of

Stamford, Connecticut, in 1880.) Using data from the National Association of Suggestion Systems, D. Carnevale and B. Sharp examined the strengths of the relationships between the extent of employee participation in suggestion plans and cost savings realized by employers (*Review of Public Personnel Administration,* Spring 1993). The data in the table at right are representative of the data they analyzed for a sample of federal, state, and local government agencies. Savings are calculated from the first year measurable benefits were observed.

a. Explain why the savings data used in this study may understate the total benefits derived from the implemented suggestions.

b. Carnevale and Sharp concluded that a significant moderate positive relationship exists between participation rates and cost savings rates in public sector suggestion systems. Do you agree? Test using $\alpha = .01$.

c. Justify the statistical methodology you used in part **b.**

SUGGEST

Employee Involvement (% of all employees submitting suggestions)	Savings Rate (% of total budget)
10.1%	8.5%
6.2	6.0
16.3	9.0
1.2	0.0
4.8	5.1
11.5	6.1
.6	1.2
2.8	4.5
8.9	5.4
20.2	15.3
2.7	3.8

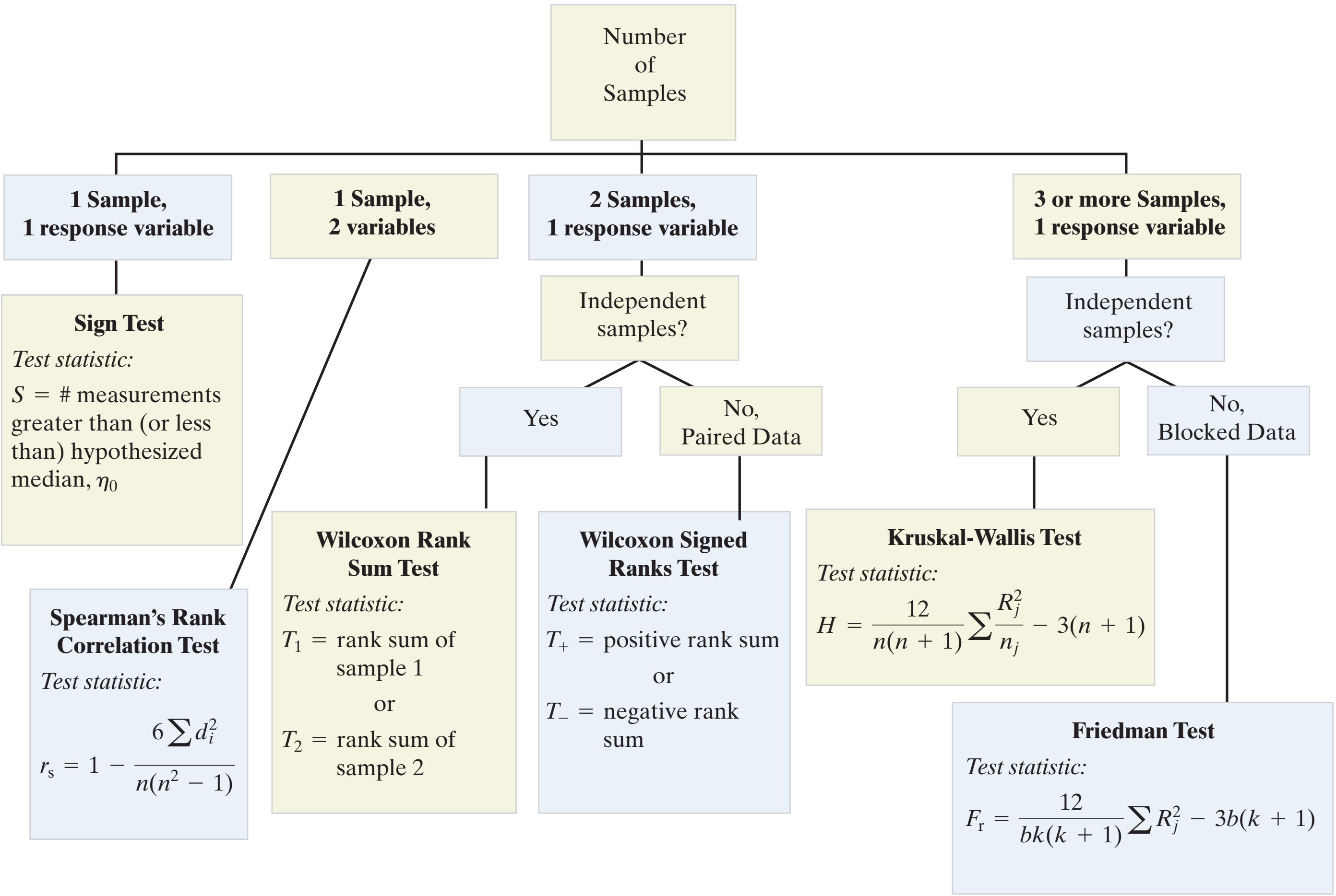
Source: Data adapted from Carnevale, D. G., and Sharp, B. S. "The Old Employee Suggestion Box." *Review of Public Personnel Administration,* Spring 1993, pp. 82–92.

KEY TERMS

Note: Items marked with an asterisk (*) are from the optional section in this Chapter.

Distribution-free tests 14-4
*Friedman F_r-statistic 14-33
Kruskal-Wallis *H*-test 14-30
Nonparametrics 14-4
Parametric tests 14-4
Population rank correlation coefficient 14-41
Rank statistics 14-4
Rank sum 14-12
Sign test 14-6
Spearman's rank correlation coefficient 14-39
Wilcoxon rank sum test 14-13
Wilcoxon signed rank test 14-21

GUIDE TO SELECTING A NONPARAMETRIC METHOD



CHAPTER NOTES

Distribution-free Tests

Do not rely on assumptions about the probability distribution of the sampled population

Nonparametrics

Distribution-free tests that are based on **rank statistics**

One-sample nonparametric test for the population median—**sign test**

Nonparametric test for *two independent samples*—**Wilcoxon rank sum test**

Nonparametric test for *matched pairs*—**Wilcoxon signed rank test**

Nonparametric test for a *completely randomized design*—**Kruskal-Wallis test**

Nonparametric test for a *randomized block design*—**Friedman test**

Nonparametric test for *rank correlation*—**Spearman's test**

Key Formulas

Sign Test, large sample test statistic:

$$z = \frac{(S - .5) - .5n}{.5\sqrt{n}}$$

Wilcoxon Rank Sum Test, large-sample test statistic:

$$z = \frac{T_1 - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}}$$

Wilcoxon Signed Rank Test, large-sample test statistic:

$$z = \frac{T_+ - \dfrac{n(n + 1)}{4}}{\sqrt{\dfrac{n(n + 1)(2n + 1)}{24}}}$$

Key Symbols

η	. . . Population median
S	. . . Test statistic for sign test
T_1	. . . Sum of ranks of observations in sample 1
T_2	. . . Sum of ranks of observations in sample 2
T_L	. . . Critical lower Wilcoxon rank sum value
T_U	. . . Critical upper Wilcoxon rank sum value
T_+	. . . Sum of ranks of positive differences of paired observations
T_-	. . . Sum of ranks of negative differences of paired observations
T_0	. . . Critical value of Wilcoxon signed ranks test
R_j	. . . Rank sum of observations in sample j
H	. . . Test statistic for Kruskal-Wallis test
F_r	. . . *Test statistic for Friedman test
r_s	. . . Spearman's rank correlation coefficient
ρ	. . . Population correlation coefficient

SUPPLEMENTARY EXERCISES 14.69–14.90

Note: Exercises marked with an asterisk (*) are from the optional section in this chapter.

Learning the Mechanics

14.69 A random sample of nine pairs of observations are recorded on two variables, x and y. The data are shown in the table below.

a. Do the data provide sufficient evidence to indicate that ρ, the rank correlation between x and y, differs from 0? Test using $\alpha = .05$.

b. Do the data provide sufficient evidence to indicate that the probability distribution for x is shifted to the right of that for y? Test using $\alpha = .05$.

LM14_69

Pair	x	y	Pair	x	y
1	19	12	6	29	10
2	27	19	7	16	16
3	15	7	8	22	10
4	35	25	9	16	18
5	13	11			

14.70 The data for three independent random samples are shown in the table. It is known that the sampled populations are not normally distributed. Use an appropriate test to determine whether the data provide sufficient evidence to indicate that at least two of the populations differ in location. Test using $\alpha = .05$.

LM14_70

Sample from Population 1		Sample from Population 2		Sample from Population 3	
18	15	12	34	87	50
32	63	33	18	53	64
43		10		65	77

*__14.71__ An experiment was conducted using a randomized block design with five treatments and four blocks. The data are shown in the table below. Do the data provide sufficient evidence to conclude that at least two of the treatment probability distributions differ in location? Test using $\alpha = .05$.

LM14_71

	BLOCK			
Treatment	1	2	3	4
1	75	77	70	80
2	65	69	63	69
3	74	78	69	80
4	80	80	75	86
5	69	72	63	77

14.72 Two independent random samples produced the measurements listed in the next table. Do the data provide sufficient evidence to conclude that there is a difference between the locations of the probability distributions for the sampled populations? Test using $\alpha = .05$.

LM14_72

Sample from Population 1		Sample from Population 2	
1.2	1.0	1.5	1.9
1.9	1.8	1.3	2.7
.7	1.1	2.9	3.5
2.5			

Applying the Concepts—Basic

14.73 Comparing print types for ease of reading. An experiment was conducted to compare two print types, A and B, to determine whether type A is easier to read. Ten subjects were randomly divided into two groups of five. Each subject was given the same material to read, one group receiving the material printed with type A, the other group receiving print type B. The times necessary for each subject to read the material (in seconds) are shown in the table. Do the data provide sufficient evidence to indicate that print type A is easier to read? Test using $\alpha = .05$.

PRINTAB

Type A:	95	122	101	99	108
Type B:	110	102	115	112	120

14.74 Insecticide for marketing. The biting rate of a particular species of fly was investigated in a study reported in the *Journal of the American Mosquito Control Association* (Mar. 1995). Biting rate was defined as the number of flies biting a volunteer during 15 minutes of exposure. This species of fly is known to have a median biting rate of 5 bites per 15 minutes on Stanbury Island, Utah. However, it is theorized that the median biting rate is higher in bright, sunny weather. (This information is of interest to marketers of pesticides.) To test this theory, 122 volunteers were exposed to the flies during a sunny day on Stanbury Island. Of these volunteers, 95 experienced biting rates greater than 5.

a. Set up the null and alternative hypotheses for the test.

b. Calculate the approximate p-value of the test. [*Hint:* Use the normal approximation for a binomial probability.]

c. Make the appropriate conclusion at $\alpha = .01$.

14.75 Insecticide for marketing (cont'd). Refer to the *Journal of the American Mosquito Control Association* study of biting flies, Exercise 14.74. The effect of wind speeds in kilometers per hour (kph) on the biting rate of the fly on Stanbury Island, Utah, was investigated by exposing samples of volunteers to one of six wind speed conditions. The distributions of the biting rates for the six wind speeds were compared using the Kruskal-Wallis test. The rank sums of the biting rates for the six conditions are shown in the next table.

a. The researchers reported the test statistic as $H = 35.2$. Verify this value.

b. Find the rejection region for the test using $\alpha = .01$.

c. Make the proper conclusions.

d. The researchers reported the p-value of the test as $p < .01$. Does this value support your inference in part **c?** Explain.

Data for Exercise 14.75

Wind Speed (kph)	Number of Volunteers (n_j)	Rank Sum of Biting Rates (R_j)
<1	11	1,804
1–2.9	49	6,398
3–4.9	62	7,328
5–6.9	39	4,075
7–8.9	35	2,660
9–20	21	1,388
Totals	217	23,653

Source: Strickman, D., et al. "Meteorological Effects on the Biting Activity of Leptoconops americanus (Diptera: Ceratopogonidae)." *Journal of the American Mosquito Control Association,* Vol. II, No. 1, Mar. 1995, p. 17 (Table 1).

14.76 A state highway patrol was interested in knowing whether frequent patrolling of highways substantially reduces the number of speeders. Two similar interstate highways were selected for the study—one heavily patrolled and the other only occasionally patrolled. After 1 month, random samples of 100 cars were chosen on each highway, and the number of cars exceeding the speed limit was recorded. This process was repeated on five randomly selected days. The data are shown in the following table.

HWPATROL

Day	Highway 1 (heavily patrolled)	Highway 2 (occasionally patrolled)
1	35	60
2	40	36
3	25	48
4	38	54
5	47	63

a. Do the data provide evidence to indicate that the heavily patrolled highway tends to have fewer speeders per 100 cars than the occasionally patrolled highway? Test using $\alpha = .05$.

b. Use the paired t-test with $\alpha = .05$ to compare the population mean number of speeders per 100 cars for the two highways. What assumptions are necessary for this procedure to be valid?

14.77 Length of TV commercials. Refer to the *Nutrition & Food Science* (Vol. 30, 2000) study of the trend in prime-time television advertising, Exercise 10.81. The rate per hour of prime-time TV commercials for food products during 1971, 1977, 1988, 1992, and 1998 are listed in the table. Consider the correlation between the rate of food ads per hour (y) and the number of years since 1970 (x).

FOODADS

Year	Number of Years since 1970, x	Food Ads, y (rate per hour)
1971	1	5.4
1977	7	3.0
1988	18	6.5
1992	22	6.0
1998	28	6.0

Source: Byrd-Bredbenner, C., and Grasso, D. "Trends in US Prime-Time Television Food Advertising across Three Decades." *Nutrition & Food Science,* Vol. 30, No. 2, 2000, p. 61 (Table 1).

a. Rank the five x-values.

b. Rank the five y-values.

c. Use the ranks, parts **a** and **b,** to find the value of Spearman's rank correlation coefficient. Interpret the result.

d. Give the rejection region for testing whether rate of food ads and number of years since 1970 are rank correlated. Use $\alpha = .10$.

e. State the conclusion of the test in the words of the problem.

Applying the Concepts—Intermediate

14.78 Number-one selling fast-food item. According to the National Restaurant Association, hamburgers are the number-one-selling fast-food item in the United States. An economist studying the fast-food buying habits of Americans paid graduate students to stand outside two suburban McDonald's restaurants near Boston and ask departing customers whether they spent more or less than \$2.25 on hamburger products for their lunch. Twenty answered "less than"; 50 said "more than"; and 10 refused to answer the question.

a. Is there sufficient evidence to conclude that the median amount spent for hamburgers at lunch at McDonald's is less than \$2.25?

b. Does your conclusion apply to all Americans who eat lunch at McDonald's? Justify your answer.

c. What assumptions must hold to ensure the validity of your test in part **a?**

14.79 Effectiveness of methods of solving complex group problems. A study published in the *Journal of Business Communications* (Fall 1985) found that video teleconferencing may be a more effective method of dealing with complex group problem-solving tasks than face-to-face meetings. Ten groups of four people each were randomly assigned both to a specific communication setting (face-to-face or video teleconferencing) and to one of two specific complex problems. Upon completion of the problem-solving task, the same groups were placed in the alternative communication setting and asked to complete the second problem-solving task. The percentage of each problem task correctly completed was recorded for each group, with the results given in the table.

FACEVT

Group	Face-to-Face	Video Teleconferencing
1	65%	75%
2	82	80
3	54	60
4	69	65
5	40	55
6	85	90
7	98	98
8	35	40
9	85	89
10	70	80

a. What type of experimental design was used in this study?

b. Specify the null and alternative hypotheses that should be used in determining whether the data provide sufficient evidence to conclude that the problem-solving performance of video teleconferencing groups is superior to that of groups that interact face-to-face.

c. Conduct the hypothesis test of part **b.** Use $\alpha = .05$. Interpret the results of your test in the context of the problem.

d. What is the p-value of the test in part **c?**

14.80 School property tax rates. An economist is interested in knowing whether property tax rates differ among three types of school districts—urban, suburban, and rural. A random sample of several districts of each type produced the data in the table at the top of the next column (rate is in mills, where 1 mill = $1/1,000). Do the data indicate a difference in the level of property taxes among the three types of school districts? Use $\alpha = .05$.

PROPTAX

Urban	Suburban	Rural
4.3	5.9	5.1
5.2	6.7	4.8
6.2	7.6	3.9
5.6	4.9	6.2
3.8	5.2	4.2
5.8	6.8	4.3
4.7		

14.81 Patients' responses to a new painkiller. The length of time required for a human subject to respond to a new painkiller was tested in the following manner. Seven randomly selected subjects were assigned to receive both aspirin and the new drug. The two treatments were spaced in time and assigned in random order. The length of time (in minutes) required for a subject to indicate that he or she could feel pain relief was recorded for both the aspirin and the drug. The data are shown in the next table. Do the data provide sufficient evidence to indicate that the probability distribution of the times required to obtain relief with aspirin is shifted to the right of the probability distribution of the times required to obtain relief with the new drug?

PAINKILL

Subject	Aspirin	New Drug
1	15	7
2	20	14
3	12	13
4	20	11
5	17	10
6	14	16
7	17	11

14.82 Evaluating a truck weigh-in-motion scale. Exercise 10.89 described a calibration study undertaken by the Minnesota Department of Transportation to evaluate a new weigh-in-motion scale. Pearson's product moment correlation coefficient was used to measure the strength of the relationship between the static weight of a truck and the truck's weight as measured by the weigh-in-motion equipment. The data (in thousands of pounds) are repeated in the table at the bottom of the page.

a. Calculate Spearman's rank correlation coefficient for x and y_1 and for x and y_2. Interpret, in the context of the problem, the values you obtain. Compare your results with those of Exercise 10.89, part **c.**

b. In the context of this problem, what circumstances would result in Spearman's rank correlation coefficient being exactly 1? Being exactly 0?

14.83 Study of solid waste generation. In Exercise 7.103, the solid waste generation rates for cities in industrialized countries and cities in middle-income countries were investigated. In this exercise, the focus is on middle-income countries versus low-income countries. The table, extracted from the *International Journal of Environmental Health Research* (1994) reports waste generation values (kg per capita per day) for two independent samples. Do the rates differ for the two categories of countries?

a. Which nonparametric hypothesis-testing procedures could be used to answer the question?

b. Specify the null and alternative hypotheses of the test.

c. Give the rejection region for the test using $\alpha = .05$.

d. Conduct the test and give the appropriate conclusion in the context of the problem.

SOLWAST2

Cities of Low-Income Countries		Cities of Middle-Income Countries	
Jakarta	.60	Singapore	.87
Surabaya	.52	Hong Kong	.85
Bandung	.55	Medellin	.54
Lahore	.60	Kano	.46
Karachi	.50	Manila	.50
Calcutta	.51	Cairo	.50
Kanpur	.50	Tunis	.56

Source: Al-Momani, A. H. "Solid-Waste Management: Sampling, Analysis and Assessment of Household Waste in the City of Amman." *International Journal of Environmental Health Research,* Vol. 4, 1994, pp. 208–222.

14.84 Fluoride in drinking water. Many water treatment facilities supplement the natural fluoride concentration with

TRUCKWTS

Truck	Static Weight, x	Weigh-in-Motion Reading Prior to Calibration Adjustment, y_1	Weigh-in-Motion Reading After Calibration Adjustment, y_2
1	27.9	26.0	27.8
2	29.1	29.9	29.1
3	38.0	39.5	37.8
4	27.0	25.1	27.1
5	30.3	31.6	30.6
6	34.5	36.2	34.3
7	27.8	25.1	26.9
8	29.6	31.0	29.6
9	33.1	35.6	33.0
10	35.5	40.2	35.0

hydrofluosilicic acid in order to reach a target concentration of fluoride in drinking water. Certain levels are thought to enhance dental health, but very high concentrations can be dangerous. Suppose that one such treatment plant targets .75 milligrams per liter (mg/L) for their water. The plant tests 25 samples each day to determine whether the median level differs from the target.

a. Set up the null and alternative hypotheses.
b. Set up the test statistic and rejection region using $\alpha = .10$.
c. Explain the implication of a Type I error in the context of this application. A Type II error.
d. Suppose that one day's samples result in 18 values that exceed .75 mg/L. Conduct the test and state the appropriate conclusion in the context of this application.
e. When it was suggested to the plant's supervisor that a *t*-test should be used to conduct the daily test, she replied that the probability distribution of the fluoride concentrations was "heavily skewed to the right." Show graphically what she meant by this, and explain why this is a reason to prefer the sign test to the *t*-test.

14.85 Hotel no-show study. A hotel had a problem with people reserving rooms for a weekend and then not honoring their reservations (no-shows). As a result, the hotel developed a new reservation and deposit plan that it hoped would reduce the number of no-shows. One year after the policy was initiated, management evaluated its effect in comparison with the old policy. Compare the records given in the table on the number of no-shows for the 10 nonholiday weekends preceding the institution of the new policy and the 10 nonholiday weekends preceding the evaluation time. Has the situation improved under the new policy? Test at $\alpha = .05$.

NOSHOWS

Before		After	
10	11	4	4
5	8	3	2
3	9	8	5
6	6	5	7
7	5	6	1

14.86 Does fatigue lead to more defectives? A manufacturer wants to determine whether the number of defectives produced by its employees tends to increase as the day progresses. Unknown to the employees, a complete inspection is made of every item that was produced on one day, and the hourly fraction defective is recorded. The resulting data are given in the accompanying table. Do they provide evidence that the fraction defective increases as the day progresses? Test at the $\alpha = .05$ level.

DEFECTS

Hour	Fraction Defective
1	.02
2	.05
3	.03
4	.08
5	.06
6	.09
7	.11
8	.10

*** 14.87 Union negotiation preferences.** A union wants to determine the preferences of its members before negotiating with management. Ten union members are randomly selected, and an extensive questionnaire is completed by each member. The responses to the various aspects of the questionnaire will enable the union to rank in order of importance the items to be negotiated. The rankings are shown in the next table. Conduct a nonparametric test to determine whether evidence exists that the probability distributions of ratings differ for at least two of the four negotiable items. Use $\alpha = .05$.

UNION

Person	More Pay	Job Stability	Fringe Benefits	Shorter Hours
1	2	1	3	4
2	1	2	3	4
3	4	3	2	1
4	1	4	2	3
5	1	2	3	4
6	1	3	4	2
7	2.5	1	2.5	4
8	3	1	4	2
9	1.5	1.5	3	4
10	2	3	1	4

14.88 Selecting a savings and loan office site. A savings and loan association is considering three locations in a large city as potential office sites. The company has hired a marketing firm to compare the incomes of people living in the area surrounding each site. The market researchers interview 10 households chosen at random in each area to determine the type of job, length of employment, etc., of those in the households who work. This information enables them to estimate the annual income of each household, as shown in the table (in thousands of dollars).

OFFICES

Site 1		Site 2		Site 3	
34.3	36.2	39.3	42.2	34.5	38.3
35.5	43.5	45.5	103.5	29.3	43.3
32.1	34.7	50.2	47.9	37.2	36.7
28.3	38.0	72.1	41.2	33.2	40.0
40.5	35.1	48.6	44.0	32.6	35.2

a. What type of design was utilized for this experiment?
b. Use the appropriate nonparametric test to compare the treatments. Specify the hypotheses and interpret the results in terms of this experiment. Use $\alpha = .05$.
c. Does the result of your test warrant further comparison of the pairs of treatments? If so, compare the pairs using the appropriate nonparametric technique and $\alpha = .05$ for each comparison. Interpret the results in terms of this experiment.
d. What assumptions are necessary to ensure the validity of the nonparametric procedures you employed? How do they compare with the assumptions that would have to be satisfied in order to use the appropriate parametric technique to analyze this experiment?

14.89 HMO crisis intervention. To investigate the standards used by health maintenance organizations (HMOs) in interpreting crisis intervention, a comprehensive questionnaire survey of 145 national HMOs was conducted

(*Medical Care,* Jan. 1985). Each HMO was asked to "write a brief, descriptive definition of situations or states you would include as qualifying for 'crisis intervention.'" The researchers sorted these situations into multiple categories and then asked three experienced clinicians to rate each category for two criteria: validity of crisis intervention (i.e., is the situation defined really a "crisis") and clarity of guidelines for offering service. A four-point rating scale was provided for both criteria. The mean ratings for the 10 randomly selected categories on both the crisis intervention and clarity scales are given in the table below. Is there evidence of a positive relationship between the mean crisis intervention and mean clarity ratings? Test using $\alpha = .05$.

Critical Thinking Challenge

14.90 Self-managed work teams and family life. Refer to the *Quality Management Journal* (Summer 1995) study of self-managed work teams (SMWTs), discussed in the Statistics in Action, Chapter 7. Recall that the researchers investigated the connection between SMWT work characteristics and workers' perceptions of positive spillover into family life (one group of workers reported positive spillover of work skills to family life while another group did not report positive work spillover). The data collected on 114 AT&T employees, saved in the **SPILLOVER** file, are described below. In the Statistics in Action for Chapter 7, we compared the two groups of workers on each characteristic using parametric methods. Reanalyze the data using nonparametrics. Are the job-related characteristics most highly associated with positive work spillover the same as those identified in the Statistics in Action, Chapter 7? Comment on the validity of the parametric and nonparametric results.

SPILLOVER

Variables Measured in the SMWT Survey

Characteristic	Variable
Information flow	Use of creative ideas (7-point scale)
Information flow	Utilization of information (7-point scale)
Decision-making	Participation in decisions regarding personnel matters (7-point scale)
Job	Good use of skills (7-point scale)
Job	Task identity (7-point scale)
Demographic	Age (years)
Demographic	Education (years)
Demographic	Gender (male or female)
Comparison	Group (positive spillover or no spillover)

CRISIS

Category (Situation)	Crisis Intervention Rating (1 = definitely a crisis, 4 = definitely not a crisis)	Clarity Rating (1 = very clear guideline, 4 = very unclear guideline)
Psychosis	1.31	1.33
Drug/alcohol abuse	1.33	1.29
Depression/anxiety	1.48	1.59
Emphasis on acuteness	1.76	2.50
Insistence on "short-term" response	2.48	3.22
Suicide	1.13	1.32
Family problems	2.59	2.30
Violence/harm	1.06	1.86
Miscellaneous	2.60	2.33
Nondefinition	3.57	3.57

Source: Cheifetz, D. I., and Salloway, J. C. "Crisis Intervention: Interpretation and Practice by HMO." *Medical Care,* Vol. 23, No. 1, Jan. 1985, pp. 89–93.

REFERENCES

Agresti, A., and Agresti, B. F. *Statistical Methods for the Social Sciences*, 2nd ed. San Francisco: Dellen, 1986.

Conover, W. J. *Practical Nonparametric Statistics*, 2nd ed. New York: Wiley, 1980.

Daniel, W. W. *Applied Nonparametric Statistics*, 2nd ed. Boston: PWS-Kent, 1990.

Dunn, O. J. "Multiple Comparisons Using Rank Sums." *Technometrics*, Vol. 6, 1964.

Friedman, M. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Journal of the American Statistical Association*, Vol. 32, 1937.

Gibbons, J. D. *Nonparametric Statistical Inference*, 2nd ed. New York: McGraw-Hill, 1985.

Hollander, M., and Wolfe, D. A. *Nonparametric Statistical Methods*. New York: Wiley, 1973.

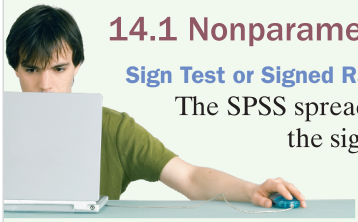
Kruskal, W. H., and Wallis, W. A. "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association*, Vol. 47, 1952.

Lehmann, E. L. *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day, 1975.

Marascuilo, L. A., and McSweeney, M. *Nonparametric and Distribution-Free Methods for the Social Sciences*. Monterey, CA: Brooks/Cole, 1977.

Wilcoxon, F., and Wilcox, R. A. "Some Rapid Approximate Statistical Procedures." The American Cyanamid Co., 1964.

Using Technology



14.1 Nonparametric Tests Using SPSS

Sign Test or Signed Ranks Test:

The SPSS spreadsheet file that contains the sample data should contain two quantitative variables. (*Note*: For the sign test, one variable is the variable to be analyzed and the other variable will have the value of the hypothesized median for all cases. For the signed ranks test, the two variables represent the two variables in the paired difference.) Click on the "Analyze" button on the SPSS menu bar, then click on "Nonparametric Tests" and "2 Related Samples", as shown in Figure 14.S.1.

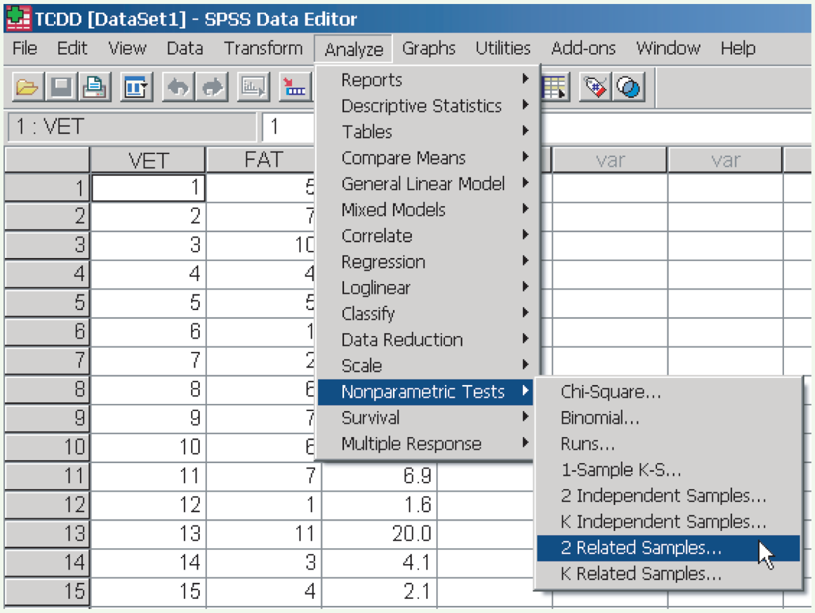


Figure 14.S.1
SPSS menu options for a sign or signed ranks test

The resulting dialog box appears as shown in Figure 14.S.2. Select the two quantitative variables of interest so that the difference appears in the "Test Pair(s) List" box. Under "Test Type", select the "Sign" option for a sign test or the "Wilcoxon" option for a signed ranks test. Click "OK" to generate the SPSS printout.

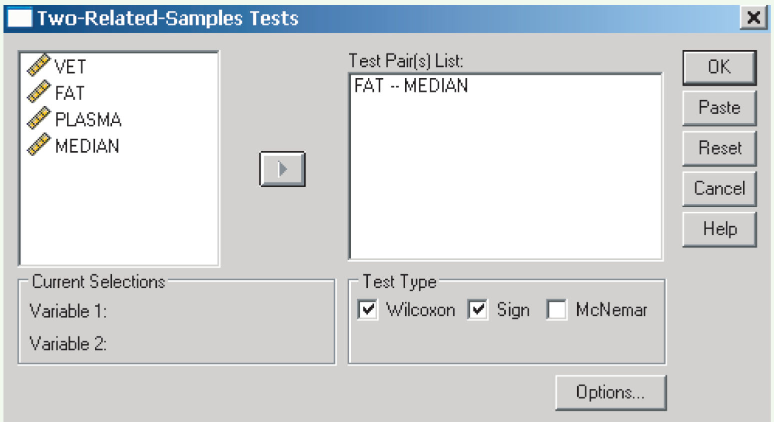


Figure 14.S.2
SPSS Two-related-samples dialog box

Rank Sum Test:

The SPSS spreadsheet file that contains the sample data should contain one quantitative variable and one qualitative variable with either two numerical coded values (e.g., 1 and 2) or two short categorical levels (e.g., "yes" and "no"). These two values represent the two groups or populations to be compared. Click on the "Analyze" button on the SPSS menu bar, then click on "Nonparametric Tests" and "2 Independent Samples" (see Figure 14.S.1). The resulting dialog box appears as shown in Figure 14.S.3.

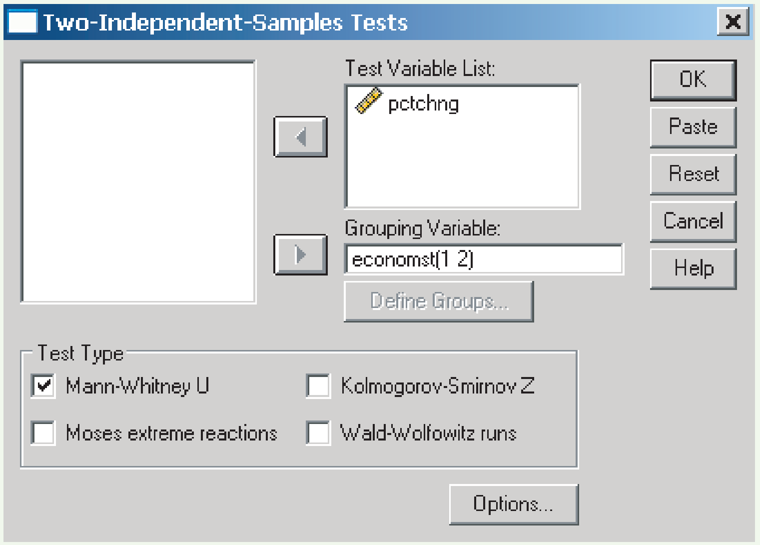


Figure 14.S.3
SPSS two-independent-samples dialog box

Specify the quantitative variable of interest in the "Test Variable List" box and the qualitative variable in the "Grouping Variable" box. Click the "Define Groups" button, and specify the values of the two groups in the resulting dialog box. Then click "Continue" to return to the "Two-Independent-Samples" dialog screen. Select the "Mann-Whitney U" option under "Test Type", then click "OK" to generate the SPSS printout.

Kruskal-Wallis Test:

The SPSS spreadsheet file that contains the completely randomized design data should contain one quantitative variable (the response, or dependent, variable) and one factor variable with at least two levels. (These values must be numbers, e.g., 1, 2, 3, etc.) Click on the "Analyze" button on the SPSS menu bar, then click on "Nonparametric Tests" and "K Related Samples" (see Figure 14.S.1). The resulting dialog box appears as shown in Figure 14.S.4.

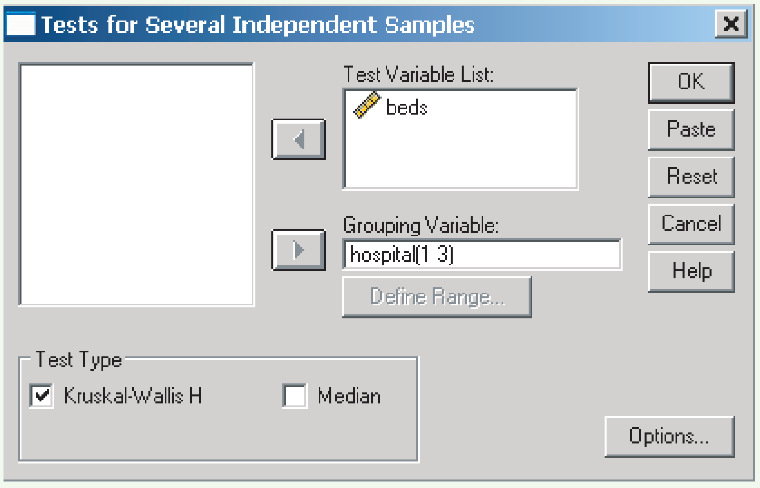


Figure 14.S.4
SPSS K-independent-samples box

Specify the response variable in the "Test Variable List" box and the factor variable in the "Grouping Variable" box. Click the "Define Range" button, and specify the values of the grouping factor in the resulting dialog box. Then click "Continue" to return to the "K-Independent-Samples" dialog screen. Select the "Kruskal-Wallis" option under "Test Type", then click "OK" to generate the SPSS printout.

Friedman Test:

The SPSS spreadsheet file that contains the randomized block design data should contain k quantitative variables, representing the k treatments to be compared. (*Note*: The cases in the rows represent the blocks.) Click on the "Analyze" button on the SPSS menu bar, then click on “Nonparametric Tests” and “K Related Samples” (see Figure 14.S.1). The resulting dialog box appears as shown in Figure 14.S.5. Specify the treatment variables in the “Test Variables” box and select the “Friedman” option under “Test Type”. Click “OK” to generate the SPSS printout.

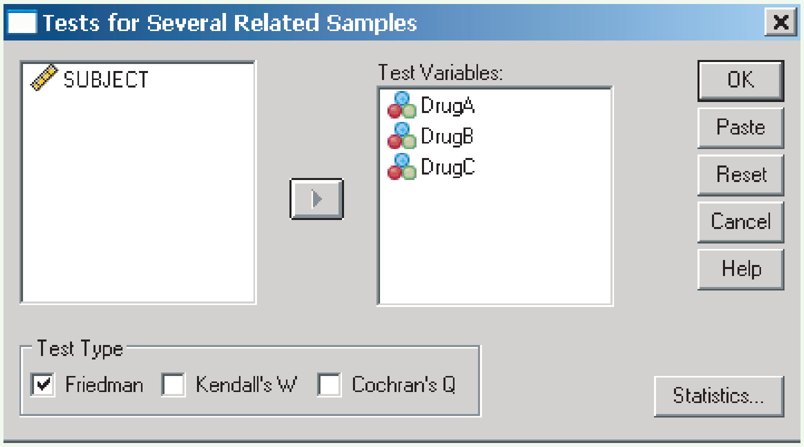


Figure 14.S.5
SPSS K-related-samples dialog box

Rank Correlation:
To obtain Spearman’s rank correlation coefficient for the two quantitative variables of interest, click on the “Analyze” button on the main menu bar, then click on “Correlate” and “Bivariate”. The resulting dialog box appears in Figure 14.S.6. Enter the variables of interest in the “Variables” box, check the “Spearman” option under “Correlation Coefficients”, then click “OK” to obtain the SPSS printout.

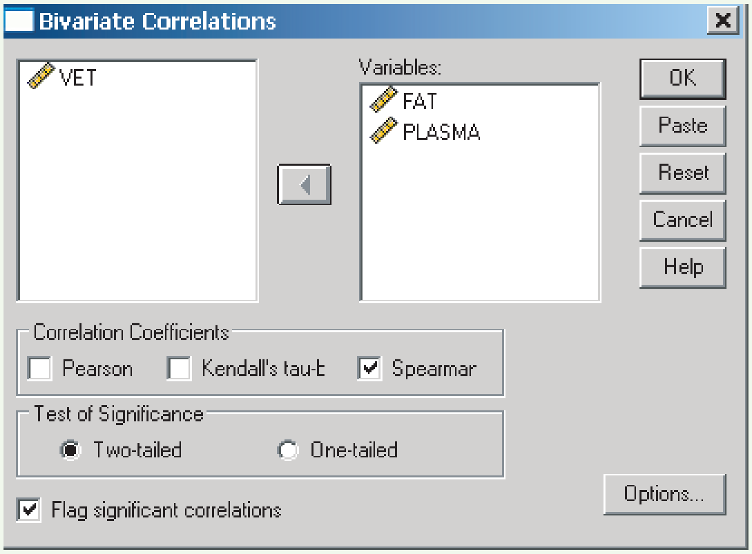


Figure 14.S.6
SPSS correlation dialog box

14.2 Nonparametric Tests Using MINITAB

Sign Test:
The MINITAB worksheet file with the sample data should contain a single quantitative variable. Click on the “Stat” button on the MINITAB menu bar, then click on “Nonparametrics” and “1-Sample Sign”, as shown in Figure 14.M.1.

The resulting dialog box appears as shown in Figure 14.M.2. Enter the quantitative variable to be analyzed in the “Variables” box. Select the “Test median” option and specify the hypothesized value of the median and the form of the alternative hypothesis (“not equal”, “less than”, or “greater than”). Click “OK” to generate the MINITAB printout.

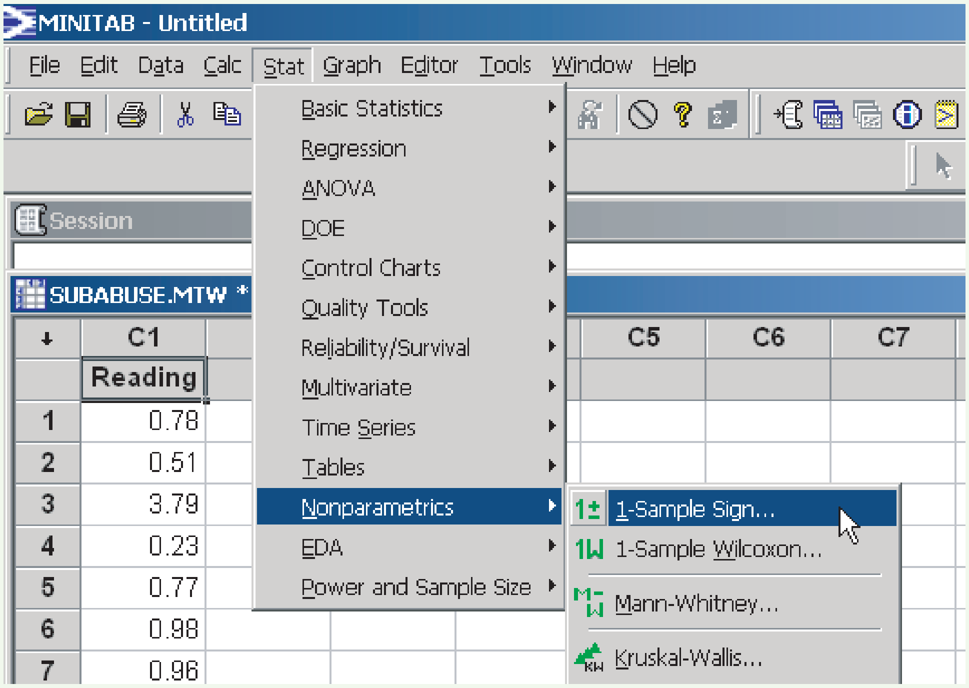


Figure 14.M.1
MINITAB nonparametric menu options

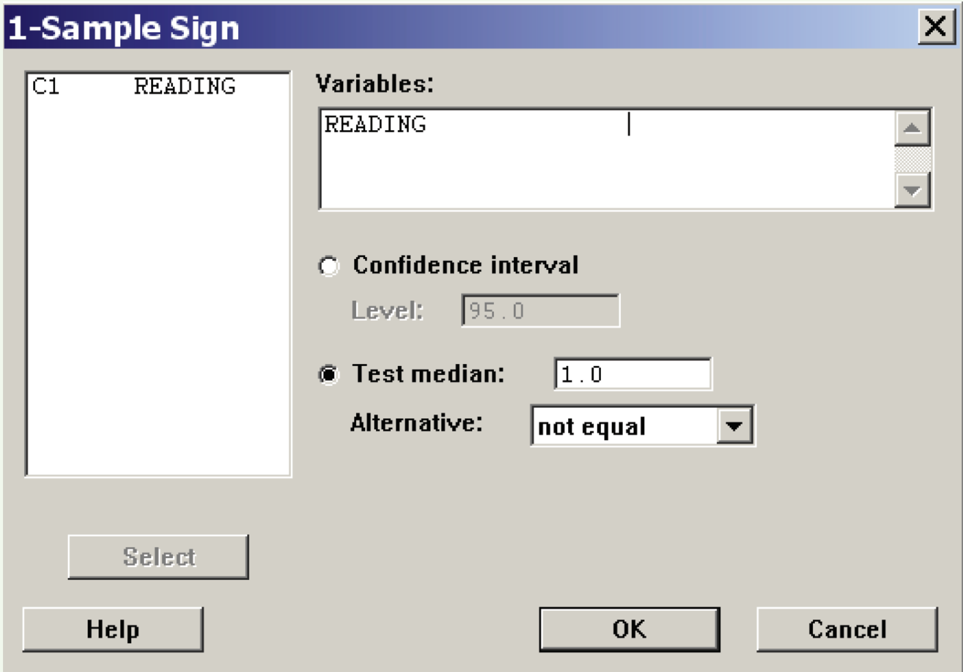


Figure 14.M.2
MINITAB 1-sample sign dialog box

Rank Sum Test:

The MINITAB worksheet file with the sample data should contain two quantitative variables, one for each of the two samples being compared. Click on the "Stat" button on the MINITAB menu bar, then click on "Nonparametrics" and "Mann-Whitney" (see Figure 14.M.1). The resulting dialog box appears as shown in Figure 14.M.3.

Specify the variable for the first sample in the "First Sample" box and the variable for the second sample in the "Second Sample" box. Specify the form of the alternative hypothesis ("not equal", "less than", or "greater than"), then click "OK" to generate the MINITAB printout.

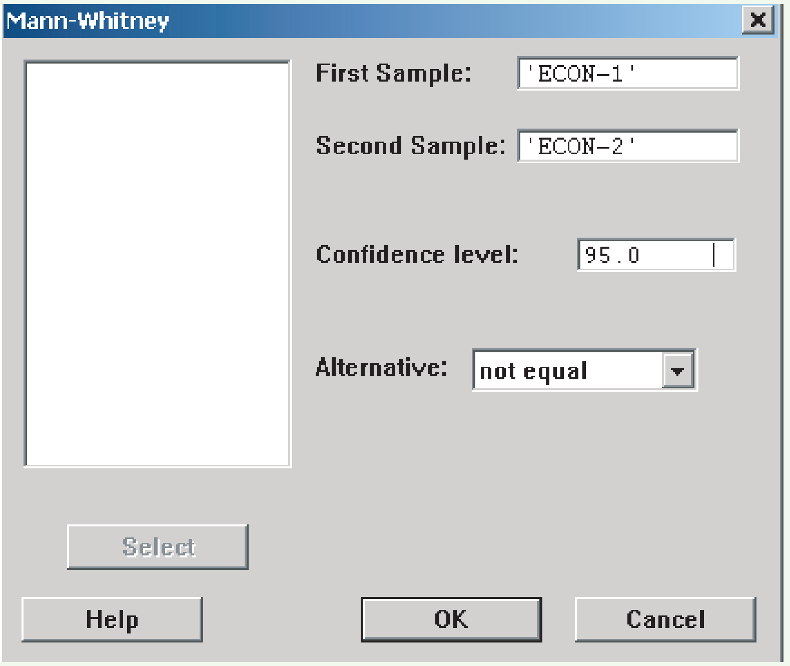


Figure 14.M.3
MINITAB Mann-Whitney dialog box

Signed Ranks Test:

The MINITAB worksheet file with the matched-pairs data should contain two quantitative variables, one for each of the two groups being compared. You will need to compute the difference between these two variables and save it in a column on the worksheet. (Use the "Calc" button on the MINITAB menu bar.) Next, click on the "Stat" button on the MINITAB menu bar, then click on "Nonparametrics" and "1-Sample Wilcoxon" (see Figure 14.M.1). The resulting dialog box appears as shown in Figure 14.M.4.

Enter the variable representing the paired differences in the "Variables" box. Select the "Test median" option and specify the hypothesized value of the median as "0". Select the form of the alternative hypothesis ("not equal", "less than", or "greater than"), then click "OK" to generate the MINITAB printout.

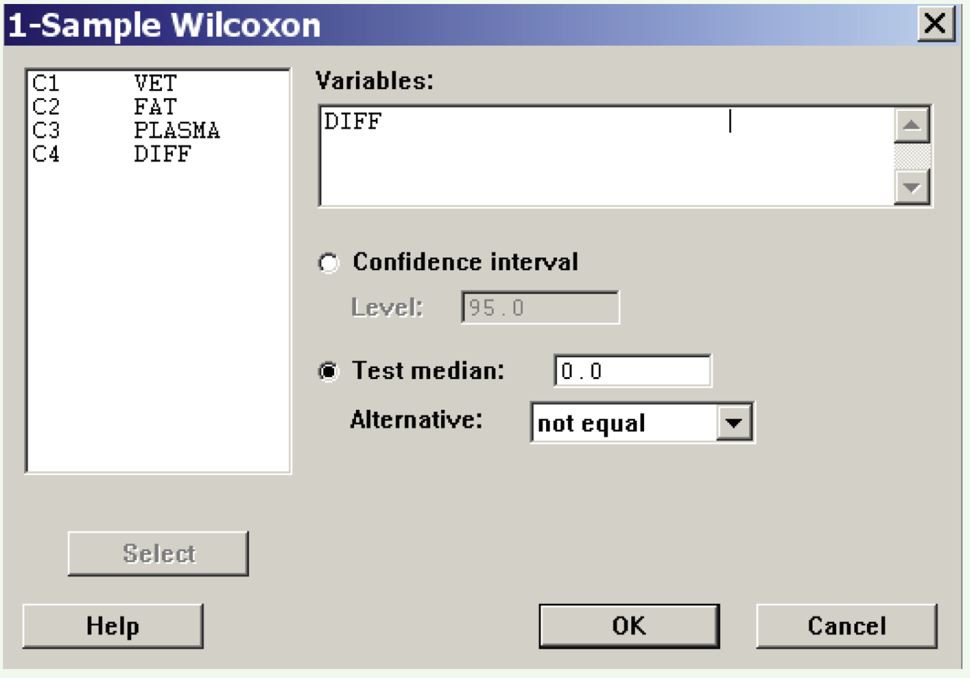


Figure 14.M.4
MINITAB 1-sample Wilcoxon dialog box

Kruskal-Wallis Test:

The MINITAB worksheet file that contains the completely randomized design data should contain one quantitative variable (the response, or dependent, variable) and one factor variable with at least two levels. Click on the "Stat" button on the MINITAB menu bar, then click on "Nonparametrics" and "Kruskal-Wallis" (see Figure 14.M.1). The resulting dialog box appears as shown in Figure 14.M.5. Specify the response variable in the "Response" box and the factor variable in the "Factor" box. Click "OK" to generate the MINITAB printout.

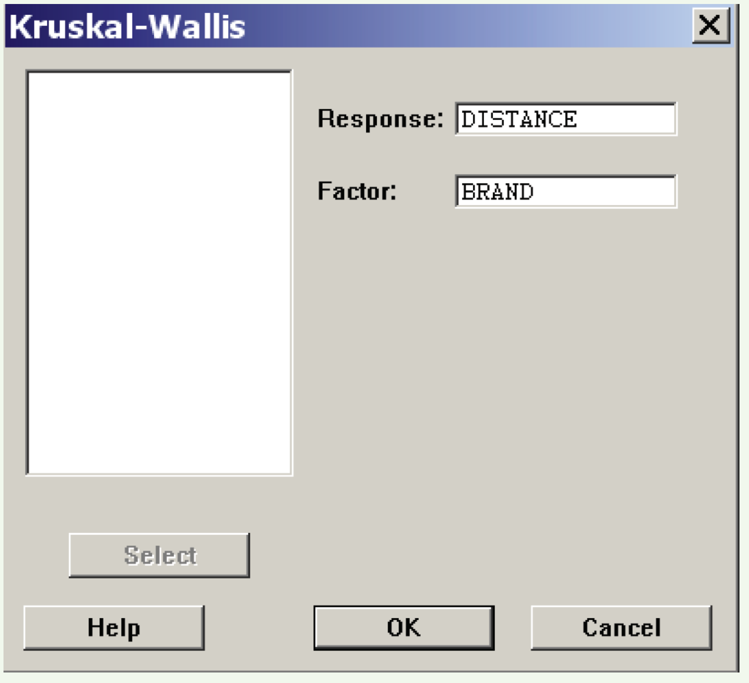


Figure 14.M.5
MINITAB Kruskal-Wallis dialog box

Friedman Test:

The MINITAB spreadsheet file that contains the randomized block design data should contain one quantitative variable (the response, or dependent, variable) and one factor variable and one blocking variable. Click on the "Stat" button on the MINITAB menu bar, then click on "Nonparametrics" and "Friedman" (see Figure 14.M.1). The resulting dialog box appears as shown in Figure 14.M.6. Specify the response, treatment, and blocking variables in the appropriate boxes, then click "OK" to generate the MINITAB printout.

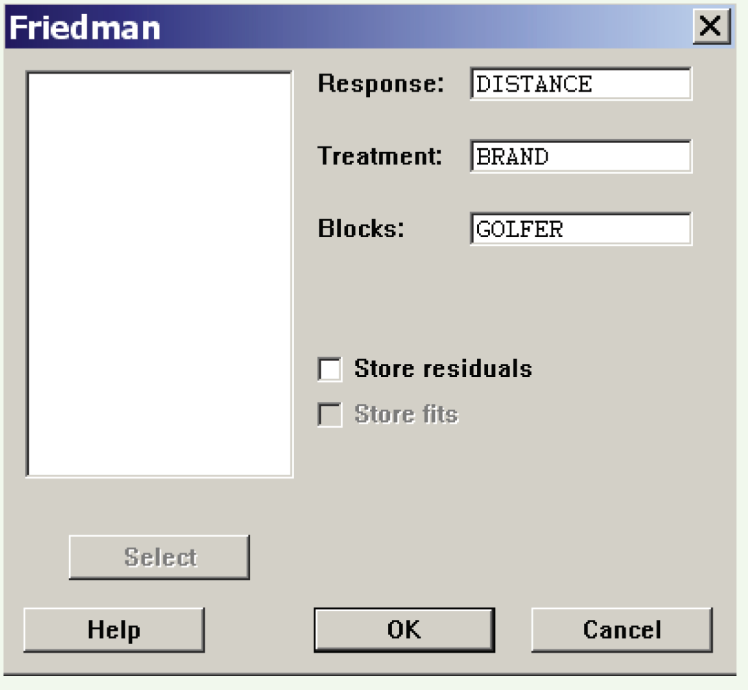


Figure 14.M.6
MINITAB Friedman dialog box

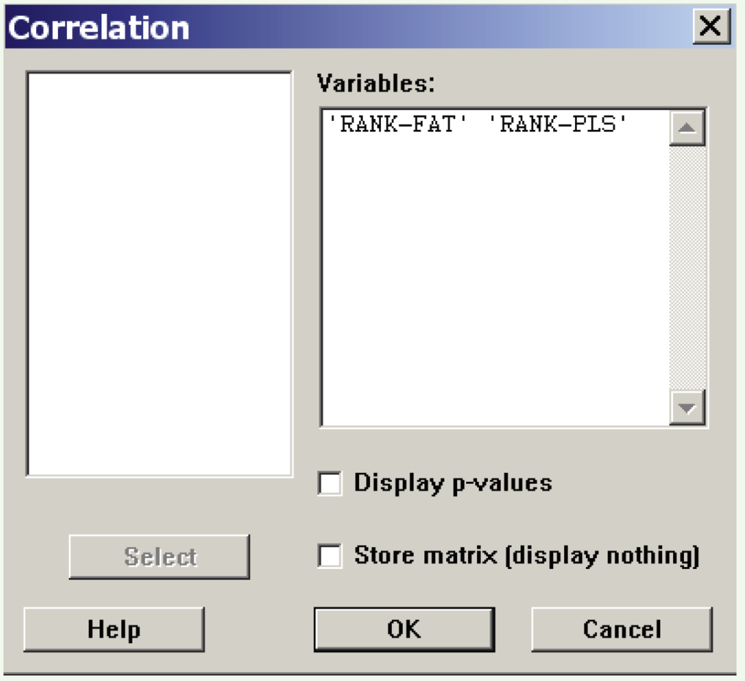


Figure 14.M.7
MINITAB correlation dialog box

Rank Correlation:
To obtain Spearman's rank correlation coefficient in MINITAB, you must first rank the values of the two quantitative variables of interest. Click the "Calc" button on the MINITAB menu bar, and create two additional columns, one for the ranks of the x-variable and one for the ranks of the y-variable. (Use the "Rank" function on the MINITAB calculator.) Next, click on the "Stat" button on the main menu bar, then click on "Basic Statistics" and "Correlation". The resulting dialog box appears in Figure 14.M.7. Enter the ranked variables in the "Variables" box and unselect the "Display p-values" option. Click "OK" to obtain the MINITAB printout. (You will need to look up the critical value of Spearman's rank correlation to conduct the test.)

14.3 Nonparametric Tests Using Excel/PHStat2

Note: Only the Wilcoxon rank sum test and Kruskal-Wallis H-test are available in Excel with the PHStat add-in.

Rank Sum Test:
The Excel workbook with the sample data should contain two quantitative variables (columns), one for each of the two samples being compared. Click on the "PHStat" button on the Excel menu bar, then click on "Two-Sample Tests" and "Wilcoxon Rank Sum Test", as shown in Figure 14.E.1. The resulting dialog box appears as shown in Figure 14.E.2.

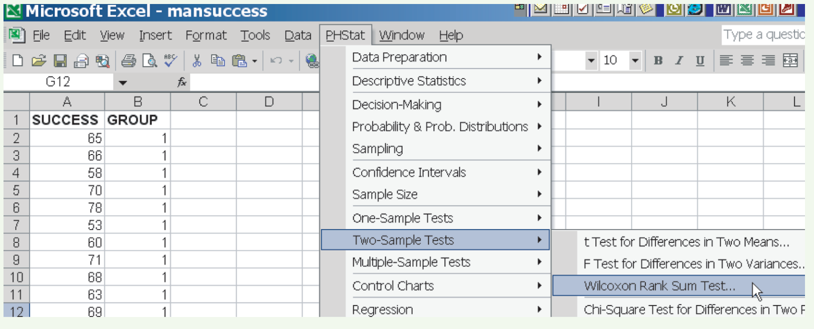


Figure 14.E.1
Excel menu options for the rank sum test

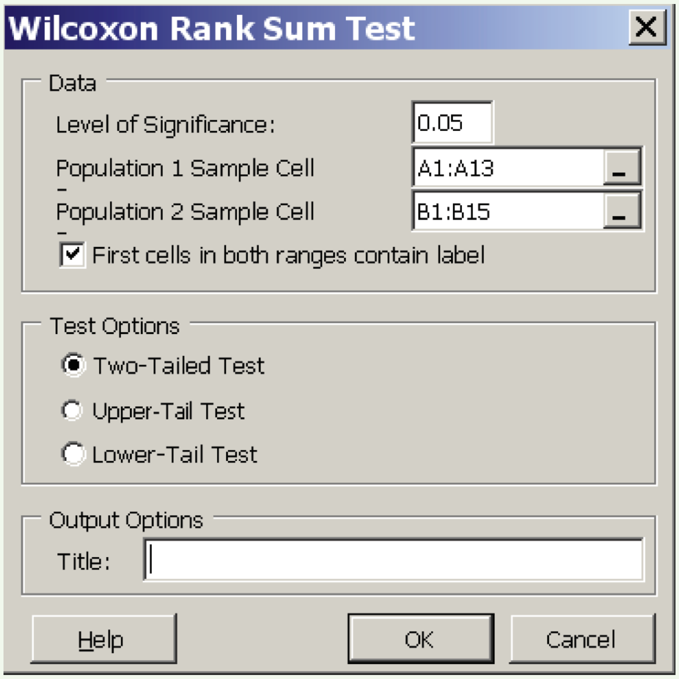


Figure 14.E.2
Excel Wilcoxon rank sum test dialog box

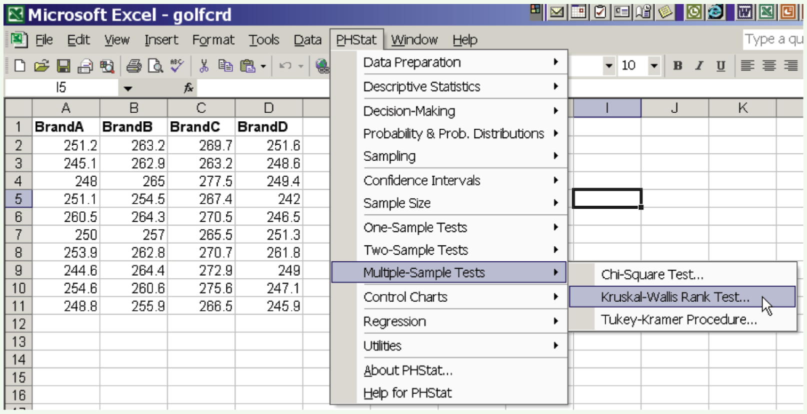


Figure 14.E.3
Excel menu options for the Kruskal-Wallis test

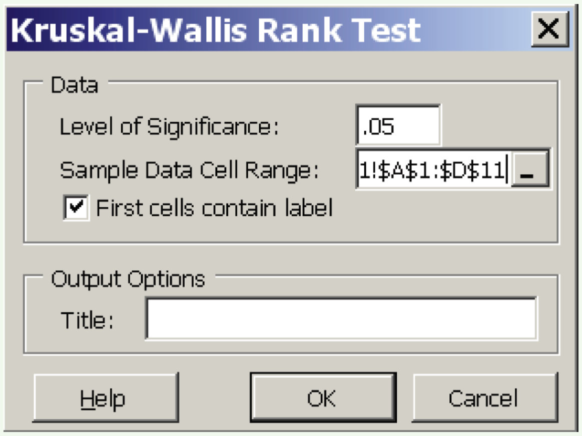


Figure 14.E.4
Excel Kruskal-Wallis rank test dialog box

On the dialog box, specify the significance level of the test, enter the cell ranges for the two samples, and select the type of test ("Two-tailed", "Upper-tailed", or "Lower-tailed"). Click "OK" to generate the Excel printout.

Kruskal-Wallis Test:
The Excel workbook that contains the completely randomized design data should contain k quantitative variables (columns), one for each of the k samples being compared. Click on the "PHStat" button on the Excel menu bar, then click on "Multi-Sample Tests" and "Kruskal-Wallis Rank Test", as shown in Figure 14.E.3. The resulting dialog box appears as shown in Figure 14.E.4.

On the dialog box, specify the significance level of the test and enter the cell range for the k samples. Click "OK" to generate the Excel printout.

ANSWERS TO SELECTED EXERCISES

14.3 a. .035 **b.** .363 **c.** .004 **d.** .151; .151 **e.** .2122; .2119 **14.5** p-value = .054; reject H_0 **14.7 a.** H_0: $\eta = 96{,}000$, H_a: $\eta > 96{,}000$ **b.** $S = 9$, p-value = .304, do not reject H_0 **c.** random sample **14.9 a.** 43 **b.** H_0: $\eta = 37$, H_a: $\eta > 37$ **c.** $S = 11$, p-value = .059, reject H_0 at $\alpha = .10$ **14.11 a.** data not normally distributed **b.** H_0: $\eta = 5{,}000$, H_a: $\eta < 5{,}000$ **c.** $S = 10$, p-value = .019, reject H_0 **14.13 a.** T_2; $T_2 < 35$ or $T_2 > 67$ **b.** T_1; $T_1 > 43$ **c.** T_2; $T_2 > 93$ **d.** z; $|z| > 1.96$ **14.15 b.** $T_2 = 42.5$, reject H_0 **14.17 b.** $T_1 = 104$ **c.** $T_2 = 106$ **d.** either T_1 or T_2 **e.** do not reject H_0 **14.19 a.** Wilcoxon rank sum test **b.** H_0: CMC and FTF groups have identical probability distributions **c.** $z < -1.28$ **d.** $z = -.21$, do not reject H_0 **14.21 a.** data not normal or variances not equal **b.** $T_1 = 18$, do not reject H_0 **c.** $T_2 = 32$, do not reject H_0 **d.** no differences **14.23 a.** No, $T_1 = 82$ **b.** No, used paired difference design **14.25 a.** equal population variances **b.** yes, $T_1 = 39$; yes **14.27 a.** H_0: Two sampled populations have identical probability distributions **b.** $T_- = 3.5$, reject H_0 **14.29 a.** H_0: Two sampled populations have identical probability distributions **b.** $z = 2.50$, reject H_0 **c.** .0062 **14.31 a.** data likely not normal **b.** H_a: Probability distribution of scores at 3rd meeting shifted to the right of probability distribution of scores at 1st meeting **c.** $T_- \le 92$ **d.** H_a: Probability distribution of scores at 3rd meeting shifted to the right or left of probability distribution of scores at 1st meeting **e.** $T_- \le 81$ **14.33** $T_- = 3.5$, reject H_0 **14.35 a.** $T_- = 4$, reject H_0 **b.** yes **14.39 a.** completely randomized **b.** H_0: Three probability distributions are identical **c.** $H > 9.21034$ **d.** $H = 13.85$, reject H_0 **14.41 a.** H_0: The three probability distributions are identical **b.** H = 36.04, p-value = .000 **c.** reject H_0 **14.43 a.** no, $H = .62$ **c.** Type I error: conclude at least two of the rate-of-return distributions differ when they do not; Type II error: conclude the three rate-of-return distributions are identical when they differ **d.** Normal distributions with equal variances **14.45 a.** $H = 0.354$, do not reject H_0 **b.** Wilcoxon rank sum test **14.47 a.** 6 **b.** H_0: Four treatment probability distributions are identical **c.** $F_r = 15.2$, reject H_0 **d.** p-value < .005 **14.49** $F_r = 13$, reject H_0 **14.51 a.** 27.0, 32.5, 29.0, 31.5 **b.** $F_r = .93$ **c.** p-value = .819 **d.** do not reject H_0 **14.53** No, $F_r = .20$ **14.55** Yes, $F_r = 6.35$ **14.57 a.** .01 **b.** .01 **c.** .975 **d.** .05 **14.59 a.** .4 **b.** $-.9$ **c.** $-.2$ **d.** .2 **14.61 b.** $r_s = .943$ **c.** do not reject H_0 **14.63 b.** Navigability: do not reject H_0; transactions: reject H_0; locatability : reject H_0; information: do not reject H_0; files: do not reject H_0 **14.65** moderate positive rank correlation between Methods I and III; all other pairs have weak positive rank correlation **14.67** $r_s = .757$, reject H_0 **14.69 a.** no, $r_s = .40$ **b.** yes, $T_- = 1.5$ **14.71** Yes, $F_r = 14.9$ **14.73** no, $T_A = 21$ **14.75 b.** $H > 15.0863$ **c.** reject H_0 **d.** yes **14.77 c.** $r_s = .564$ **d.** $|r_s| > .9$ **e.** do not reject H_0 **14.79 a.** paired difference **b.** H_0: Two sampled populations have identical probability distributions, H_a: Distribution for A (face-to-face) is shifted to the left of distribution for B (video telecon.) **c.** $T_+ = 3.5$, reject H_0 **d.** $.01 < p$-value $< .025$ **14.81** yes, $T_- = 3$ **14.83 a.** rank sum test **b.** H_0: Two populations have identical probability distributions **c.** $T_2 \le 37$ or $T_2 \ge 68$ **d.** $T_2 = 53$, do not reject H_0 **14.85** yes, $T_{\text{before}} = 132.5$ **14.87** $F_r = 6.21$, do not reject H_0 **14.89** yes, $r_s = .745$