Chapter 12

Nonparametric Statistics

Chapter outline

12.1. Introduction **492**
12.2. Nonparametric confidence interval **493**
Exercises 12.2 **495**
12.3. Nonparametric hypothesis tests for one sample **497**
12.3.1. The sign test 497
12.3.2. Wilcoxon signed rank test 500
12.3.3. Dependent samples: paired comparison tests 504
Exercises 12.3 **505**
12.4. Nonparametric hypothesis tests for two independent samples **506**
12.4.1. Median test 507
12.4.2. The Wilcoxon rank sum test 510
Exercises 12.4 **512**
12.5. Nonparametric hypothesis tests for $k \geq 2$ samples **513**
12.5.1. The Kruskal–Wallis test 514
12.5.2. The Friedman test 516
Exercises 12.5 **519**
12.6. Chapter summary **521**
12.7. Computer examples **521**
12.7.1. Examples using R 521
12.7.2. Minitab examples 523
12.7.3. SPSS examples 526
12.7.4. SAS examples 527
Projects for Chapter 12 **527**
12A Comparison of Wilcoxon tests with normal approximation 527
12B Randomness test (Wald–Wolfowitz test) 528
Exercise **530**

Objective

In this chapter we shall introduce several classical nonparametric or distribution free tests. These tests do not require distributional assumptions about the population such as the normality.

Jacob Wolfowitz
(Source: http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Wolfowitz.html)

Jacob Wolfowitz was born on March 19, 1910, in Warsaw, Russian Empire (now Poland), and died on July 16, 1981 in Tampa, Florida, United States. Wolfowitz's earliest interest was nonparametric inference, and the first joint paper he wrote with Abraham Wald introduced methods of calculating confidence intervals that are not necessarily of fixed width. It is in this paper by Wolfowitz in 1942 that the term *nonparametric* appears for the first time. Later, he worked

Mathematical Statistics with Applications in R. https://doi.org/10.1016/B978-0-12-817815-7.00012-9

on the area of sequential analysis and published work on sequential estimators of a Bernoulli parameter and results on the efficiency of certain sequential estimators. He also studied asymptotic statistical theory and worked on many aspects of the maximum likelihood method. Information theory pioneered by Shannon was another area to which Wolfowitz made important contributions, culminating in a classic book titled *Coding Theorems of Information Theory* (third ed., 1978). After working at different places such as the Statistical Research Group at Columbia University, the University of North Carolina, and the University of Illinois at Urbana, in 1978 he joined the faculty of the University of South Florida in Tampa. Wolfowitz was elected to the National Academy of Sciences and the American Academy of Arts and Sciences. He was also elected a fellow of the Econometric Society, the International Statistics Institute, and the Institute of Mathematical Statistics. In 1979 he was Shannon Lecturer of the Institute of Electrical and Electronic Engineers.

12.1 Introduction

Most of the tests that we have learned up to this point are based on the assumption that the sample(s) came from a normal population, or at the least that the population probability distribution(s) is specified except for a set of free parameters. Such tests are called parametric tests. In general, a parametric test is known to be generally more powerful than other procedures when the underlying assumptions are met. Usually the assumption of normality or any other distributional assumption about the population is hard to verify, especially when the sample sizes are small or the data are measured on an ordinal scale such as the letter grades of a student, in which case we do not have a precise measurement. For example, incidence rates of rare diseases, data from gene-expression microarrays, and the number of car accidents in a given time interval are not normally distributed. Nonparametric tests are tests that do not make such distributional assumptions, particularly the usual assumption of normality. In situations where a distributional model for a set of data is unavailable, nonparametric tests are ideal. Even if the data are distributed normally, nonparametric methods are frequently almost as powerful as parametric methods. These tests involve only order relationships among observations and are based on ranks of the variables and analyzing the ranks instead of the original values. Nonparametric methods include tests that do not involve population parameters at all, such as testing whether the population is normal. Distribution-free tests generally do make some weak assumptions, such as equality of population variances and/or the distribution, and are of the continuous type.

Sometimes we may be required to make inferences about models that are difficult to parameterize, or we may have data in a form that makes, say, the normal theory tests unsuitable. For example, incomes of families generally follow a skewed distribution. If we do a sample survey of a large number of the families in a feeder area, the income distribution may look as in Fig. 12.1.

This distribution is clearly difficult to parameterize, that is, to identify a classical probability distribution that will characterize the data's behavior. Moreover, the mean income of this sample may be misleading. A better measure of the central tendency is the median income. At least we know that 50% of the families are below the median and 50% above. Appropriate techniques of inference in these situations are based on distribution-free methods. Most of the nonparametric methods use only the order of magnitude of observations, known as order statistics, in a random sample, rather than the observed values of the random variables.

In general, nonparametric methods are appropriate to estimation or hypothesis-testing problems when the population distributions could only be specified in general terms. The conditions may be specified as being continuous, symmetric, or identical, differing only in median or mean.

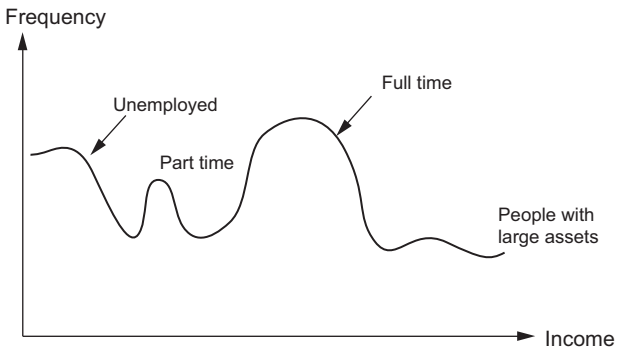


FIGURE 12.1 Income distribution of families.

The distributions need not belong to specific families such as normal or gamma, etc. Because most of the nonparametric procedures depend on a minimum number of assumptions, the chance of their being improperly used is relatively small. Most of the nonparametric procedures involve ranking data values and developing testing methods based on the ranks. Because of this, nonparametric procedures may be used when the data are measured on a weak scale such as only count data or rank data. We may ask: Why not use nonparametric methods all the time? The answer lies in the fact that when the assumptions of the parametric tests can be verified as true, parametric tests are generally more powerful than nonparametric tests. Because only ranks are used in nonparametric methods, and even though the ranks preserve information about the order of the data, because the actual values are not used some information is lost. Because of this, nonparametric procedures cannot be as powerful as their parametric counterparts when parametric tests can be used. For brevity and clarity, this chapter is presented without much theoretical explanation to focus on the methods. Theoretical developments can be found in many specialized books on the subject.

In this chapter, we study some of the commonly used classical nonparametric methods that are based on ordering, ranking, and permutations. The modern approaches are based on resampling methods such as bootstrap and will be discussed in Chapters 10 and 13.

12.2 Nonparametric confidence interval

We have seen that for a large sample, using the central limit theorem, we can obtain a confidence interval for a parameter within a well-defined probability distribution. However, for small samples, we need to make distributional assumptions that are often difficult to verify. For this reason, in practice it is often advisable to construct confidence intervals or interval estimates of population quantities that are not parameters of a particular family of distributions. In a nonparametric setting, we need procedures where the sample statistics used have distributions that do not depend on the population distribution. The median is commonly used as a parameter in nonparametric settings. We assume that the population distribution is continuous.

Let M denote the median of a distribution and X (assumed to be continuous) be any observation from that distribution. Then,

$$P(X \leq M) = P(X \geq M) = \frac{1}{2}.$$

This implies that, for a given random sample $X_1, \ldots, X_n$ from a population with median M, the distribution of the number of observations falling below M will follow a binomial distribution with parameters n and $p = 1/2$, irrespective of the population distribution. That is, let N^- be the number of observations less than M. Then the distribution of N^- is binomial with parameters n and $p = 1/2$ for a sample of size n. Hence, we can construct a confidence interval for the median using the binomial distribution.

For a given probability value α, we can determine a and b such that

$$P(N^- \leq a) = \sum_{i=0}^{a} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n-i}$$

$$= \sum_{i=0}^{a} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$$

and

$$P(N^- \geq b) = \sum_{i=b}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n-i}$$

$$= \sum_{i=b}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}.$$

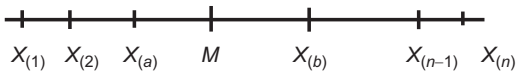


FIGURE 12.2 Ordered sample.

If exact probabilities cannot be achieved, choose a and b such that the probabilities are as close as possible to the value of $\alpha/2$. Furthermore, let $X_{(1)}, X_{(2)}, \ldots, X_{(a)}, \ldots, X_{(b)}, \ldots, X_{(n)}$ be the order statistics of $X_1, \ldots, X_n$ as in Fig. 12.2.

Then the population median will be above the order statistic, $X_{(b)}$, $\left(\frac{\alpha}{2}\right)100\%$ of the time and below the order statistic, $X_{(a)}$, $\left(\frac{\alpha}{2}\right)100\%$ of the time. Hence, a $(1-\alpha)100\%$ confidence interval for the median of a population distribution will be

$$X_{(a)} < M < X_{(b)}.$$

We can write this result as $P\left(X_{(a)} < M < X_{(b)}\right) \geq 1-\alpha$.

By dividing the upper and lower tail probabilities equally, we find that $b = n + 1 - \alpha$. Therefore, the confidence interval becomes

$$X_{(a)} < M < X_{(n+1-a)}.$$

In practice, a will be chosen so as to come as close to attaining $\frac{\alpha}{2}$ as possible.

We can summarize the nonparametric procedure for finding the confidence interval for the population median as follows.

Procedure for finding $(1-\alpha)100\%$ confidence interval for the median M

For a sample of size n:

1. Arrange the data in ascending order.
2. From the binomial table with n and $p = \frac{1}{2}$, find the value of a such that

$$p(X \leq a) = \frac{\alpha}{2} \text{ or nearest to } \frac{\alpha}{2}.$$

3. Set $b = n + 1 - a$.
4. Then the confidence interval is such that the lower limit is the ath value and the upper limit is the bth value of the observations in step 1.

Assumptions: Population distribution is continuous; the sample is a simple random sample.

We illustrate this four-step procedure with an example.

EXAMPLE 12.2.1

In a large company, the following data represent a random sample of the ages of 20 employees.

24 31 28 43 28 56 48 39 52 32
38 49 51 49 62 33 41 58 63 56.

Construct a 95% confidence interval for the population median M of the ages of the employees of this company.

Solution

For a 95% confidence interval, $\alpha = 0.05$. *Hence,* $\alpha/2 = 0.025$. *The ordered data are*

24 28 28 31 32 33 38 39 41 43
48 49 49 51 52 56 56 58 62 63.

Looking at the binomial table with $n = 20$ *and* $p = \frac{1}{2}$, *we see that* $P(X \leq 5) = 0.0207$. *Hence,* $a = 5$ *comes closest to achieving* $\alpha/2 = 0.025$. *Hence, in the ordered data, we should use the fifth observation,* 32, *for the lower confidence limit and the 16th observation* $(n + 1 - a = 21 - 5 = 16)$, 56, *for the upper confidence limit. Therefore, an approximate* 95% *confidence interval for* M *is*

$$32 < M < 56.$$

Thus, we are at least 95% certain that the true median of the employee ages of this company will be greater than 32 and less than 56, that is,

$$P(32 < M < 56) \geq 0.95.$$

The data of Example 12.2.1 passes the normality test and we can calculate the 95% parametric confidence interval as (38.40, 49.70). Comparing this to the nonparametric confidence interval, length of parametric confidence interval, in general, is smaller whenever parametric assumption can be made.

EXAMPLE 12.2.2

A drug is suspected of causing an elevated heart rate in a certain group of high-risk patients. Twenty patients from this group were given the drug. The changes in heart rates were found to be as follows,

$$\begin{array}{cccccccccc} -1 & 8 & 5 & 10 & 2 & 12 & 7 & 9 & 1 & 3 \\ 4 & 6 & 4 & 20 & 11 & 2 & -1 & 10 & 2 & 8. \end{array}$$

Construct a 98% confidence interval for the mean change in heart rate. Can we assume that the population has a normal distribution? Interpret your answer.

Solution

First testing for normality, we get the normal probability plot as shown in Fig. 12.3.

This shows that the normality assumption may not be satisfied, and thus the nonparametric method is more suitable (this conclusion is based strictly on the normal probability plot which is a visual interpretation). Using a box plot, we can also test for outliers. The ordered data are

$$\begin{array}{cccccccccc} -1 & -1 & 1 & 2 & 2 & 2 & 3 & 4 & 4 & 5 \\ 6 & 7 & 8 & 8 & 9 & 10 & 10 & 11 & 12 & 20 \end{array}$$

Looking at the binomial table with $n = 20$ *and* $p = \frac{1}{2}$, *we see that* $P(X \leq 4) = 0.006$. *Hence,* $a = 4$ *comes closest to achieving* $\alpha/2 = 0.01$. *Hence, in the ordered data, we should use the fourth observation, 2, for the lower confidence limit and the 17th observation* $(n + 1 - a = 21 - 4 = 17)$, *10, for the upper confidence limit. Therefore, an approximate* 98% *confidence interval for* M *is*

$$2 < M < 10.$$

That is, we are at least 98% certain that the true median of the mean change in heart rate will be greater than 2 and less than 10.

If we perform the usual t*-test, we will get the 98% confidence interval as (3.20, 9.0). However, such an interval is not valid, because the normality assumptions are not satisfied and will lead to misinterpretation of the facts.*

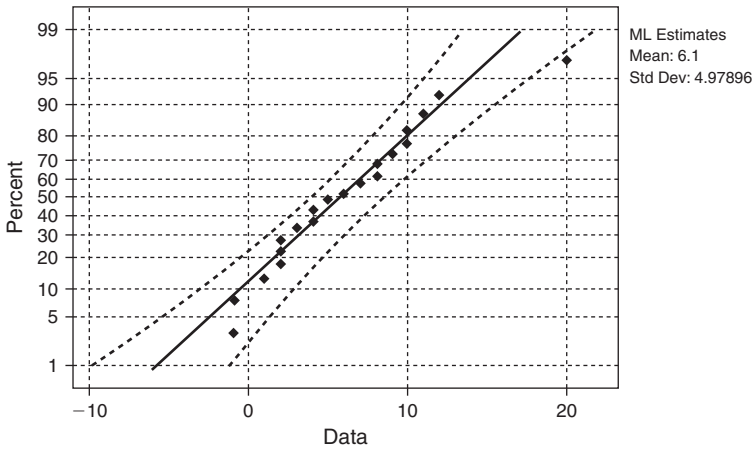


FIGURE 12.3 Normal probability plot for heart rate.

Exercises 12.2

12.2.1. For the following random sample values, construct a 95% confidence interval for the population median *M:*

$$7.2 \quad 5.7 \quad 4.9 \quad 6.2 \quad 8.5 \quad 2.7 \quad 5.9 \quad 6.0 \quad 8.2.$$

12.2.2. The following data represent a random sample of end-of-year bonuses for the lower-level managerial personnel employed by a large firm. Bonuses are expressed in percentage of yearly salary.

6.2 9.2 8.0 7.7 8.4 9.1 7.4 6.7 8.6 6.9
8.9 10.0 9.4 8.8 12.0 9.9 11.7 9.8 3.2 4.6.

Construct a 98% confidence interval for the median bonus expressed in percentage of yearly salary of this firm. Also, draw a probability plot and test for normality. Can this be considered a random sample?

12.2.3. Air pollution in large U.S. cities is monitored to see if it conforms to requirements set by the Environmental Protection Agency. The following data, expressed as an air pollution index, give the air quality of a city for 10 randomly selected days.

57.3 58.1 58.7 66.7 58.6 61.9 59.0 64.4 62.6 64.9

(a) Draw a probability plot and test for normality.
(b) Construct a 95% confidence interval for the actual median air pollution index for this city and interpret its meaning.

12.2.4. A random sample from a population yields the following 25 values:

90 87 121 96 106 107 89 107 83 92
117 93 98 120 97 109 78 87 99 79
104 85 91 107 89

Obtain a 99% confidence interval for the population median.

12.2.5. In an experiment on the uptake of solutes by liver cells, a researcher found that six determinations of the radiation, measured in counts per minute after 20 minutes of immersion, were:

2728 2585 2769 2662 2876 2777

Construct a 99% confidence interval for the population median and interpret its meaning.

12.2.6. The nominal resistance of a wire is 0.20 Ω. A testing of the wire randomly chosen from a large collection of such wires yields the following resistance data.

0.199 0.211 0.198 0.201 0.197 0.200 0.198 0.208

Obtain a 95% confidence interval for the population median.

12.2.7. In order to measure the effectiveness of a new procedure for pruning grapes, 15 workers are assigned to prune an acre of grapes. The effectiveness is measured in worker-hours per acre for each person.

5.2 5.0 4.8 4.5 3.9 6.1 4.2 4.4 5.5 5.8
4.2 5.3 4.9 4.7 4.9

Obtain a 99% confidence interval for the median time required to prune an acre of grapes for this procedure and interpret its meaning.

12.2.8. The following data give the exercise capacity (in minutes) for 10 randomly chosen patients being treated for chronic heart failure.

15 27 11 19 12 21 11 17 13 22

Obtain a 95% confidence interval for the median exercise capacity for patients being treated for chronic heart failure.

12.2.9. The data given below refer to the in-state tuition costs (in dollars) of 15 randomly selected colleges from a list of the 100 best values in public colleges (source: *Kiplinger*, October 2000).

3788 4065 2196 7360 5212 4137 4060 3956
3975 7395 4058 3683 3999 3156 4354

Obtain a 95% confidence interval for the median in-state tuition costs and interpret its meaning.

12.2.10. Sepsis is an extreme immune system response to an infection that has spread throughout the blood and tissues. Sepsis can reduce blood flow to kidneys resulting in acute renal failure (also called acute kidney injury). Relative risk of mortality associated with developing acute renal failure as of sepsis in 16 studies is given below (*Crit Care, 2002: 6(6): 509–513*).

0.75	2.03	2.29	2.11	0.80	1.50	0.79	1.01
1.23	1.48	2.45	1.02	1.03	1.30	1.54	1.27

Obtain a 95% confidence interval for the median relative risk of mortality.

12.3 Nonparametric hypothesis tests for one sample

In this section, we study two popular tests for testing hypotheses about the population location, or median using the *sign test* and the *Wilcoxon signed rank test*. The comparison of medians rather than means is a technicality that is not important unless the data are skewed substantially. In such cases, medians are somewhat more accurate than means for comparing the locations of probability distributions. Further discussions on nonparametric tests can be found in many references, such as those by W. J. Conover and by E. L. Lehmann. Before using nonparametric tests, it is desirable to test for normality of the data using normal probability plots, and for the existence of outliers using box plots, and run tests for test of randomness of the data. When we make any particular choice of method, test for the assumptions made. These assumption checks are relatively easier using statistical software packages. Many of the examples in this chapter are given more for illustration of the nonparametric methods than for assumption violations of parametric tests or for comprehensive assumption testing techniques. Also, when we use statistical software packages, generally, the p value of the test will be given in the output. In order to make a decision on a particular hypothesis, we just need to compare the p value with the chosen value of α. We are going to explain a more traditional approach instead of using the p-value approach in the discussion, however, the computer example section will illustrate the p-value approach.

12.3.1 The sign test

In this section, we describe a test that is the nonparametric alternative to the one-sample t-test and to the paired-sample t-test. Let M be the median of a certain population. Then we know that

$$P(X \leq M) = 0.5 = P(X > M).$$

We consider the problem of testing the null hypothesis

$$H_0: M = m_0 \quad \text{versus} \quad H_a: M > m_0.$$

Assume that the underlying population distribution is continuous. Let X_i be the ith observation and let N^+ be the number of observations that are greater than m_0. N^+ will be our test statistic. We will reject H_0 if, n^+ the observed value of N^+, is too large. This test is called the *sign test*. A test at a significance level α will reject H_0 if $n^+ \geq k$, where k is chosen such that

$$P(N^+ \geq k \text{ when } M = m_0) = \alpha.$$

Similarly, if the alternative is of the form H_a: $M \neq m_0$, the critical region is of the form $N^+ \leq k$ or $N^+ \geq k_1$, where $P(N^+ \leq k) + P(N^+ \geq k_1) = \alpha$.

In order to determine such a k and k_1, we need to determine the distribution of N^+. The test works on the principle that if the sample were to come from a population with a continuous distribution, then each of the observations falls above the median or below the median with probability $\frac{1}{2}$. Hence, the number of sample values falling below the median follows a binomial distribution with parameters n and $p = \frac{1}{2}$, n being the sample size. If a sample value equals the hypothesized median m_0, that observation will be discarded and the sample size will be adjusted accordingly (we remark that such values should be very few). Thus, when H_0 is true, N^+ will have a binomial distribution with parameters n and $p = \frac{1}{2}$. For this reason, some authors call this test the binomial test. The following procedure summarizes the test statistic and the corresponding critical regions.

SIGN TEST

$$H_0: M = m_0$$

Alternative hypothesis	Critical region
$H_a: M > m_0$	$N^+ \geq k$, where $\sum_{i=k}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \alpha$
$H_a: M < m_0$	$N^+ \leq k$, where $\sum_{i=0}^{k} \binom{n}{i} \left(\frac{1}{2}\right)^n = \alpha$
and	
$H_a: M \neq m_0$	$N^+ \leq k_1$, where $\sum_{i=k_1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$ or $N^+ \leq k$, where $\sum_{i=0}^{k} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$.

If α or $\alpha/2$ cannot be achieved exactly, choose k (or k and k_1) so that the probability comes as close to α (or $\alpha/2$) as possible.

We now summarize the procedure of the sign test in the case of an upper tail alternative. The other two cases are similar.

Hypothesis-testing procedure using the sign test

We test

$$H_0: M = m_0 \text{ vs. } H_1: M > m_0.$$

1. Replace each value of the observation that is greater than m_0 by a plus sign and each sample value less than m_0 by a minus sign. If the sample value is equal to m_0, discard the observation and adjust the sample size n accordingly.
2. Let n^+ be the number of +'s in the sample. For n and $p = \frac{1}{2}$, from the binomial table, find
$$\gamma = P(N^+ \geq n^+).$$
3. **Decision:** If γ is less than α, H_0 must be rejected. Based on the sample, we will conclude that the median of the population is greater than m_0 at the significance level α. Otherwise do not reject H_0.

Assumptions: The population distribution is continuous. The number of ties is small (less than 10% of the sample).

Note that the approach described in the foregoing procedure is nothing but the p-value method for hypothesis testing regarding a median using the sign test. Recall that the p value is the probability of observing a test statistic as extreme or more extreme than what was really observed, under the assumption that the null hypothesis is true. In the sign test, we had assumed that the median is $M = m_0$, so 50% of the data should be less than m_0 and 50% of the data greater than m_0. Thus, we expect half of the data to result in plus signs and half to result in minus signs. Hence, we can think of the data as following a binomial distribution with $p = 1/2$ under the null hypothesis. The p value is computed from its definition given by the formula

$$p \text{ value} = P(N^+ \geq n^+) = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \gamma.$$

The p-value method is to reject the null hypothesis if the computed p value is greater than α. These binomial probabilities can be obtained from the binomial tables, or statistical software packages. The following example illustrates how we apply the three-step procedure.

EXAMPLE 12.3.1

For the given data from an experiment

1.51 1.35 1.69 1.48 1.29 1.27 1.54 1.39 1.45

test the hypothesis that H_0: $M = 1.4$ versus H_a: $M > 1.4$ at $\alpha = 0.05$.

Solution
We test

$$H_0: M = 1.4 \text{ versus } H_a: M > 1.4.$$

Replacing each value greater than 1.4 with a plus sign and each value less than 1.4 with a minus sign, we have

$$+ - + + - - + - +.$$

Thus, $n^+ = 5$. *From the binomial table with* $n = 9$ *and* $p = \frac{1}{2}$, *we have*

$$P(N^+ \geq 5) = 0.50.$$

Hence, the p *value is 0.5. Because* $\alpha = 0.05 < 0.50$, *the null hypothesis is not rejected. We conclude that the median does not exceed* 1.4.

When the sample size n is large, we can apply the normal approximation to the binomial distribution. That is, the test statistic N^+ is approximately normally distributed. Thus, under H_0, N^+ will have approximate normal distribution with mean $np = n/2$ and variance of $np(1-p) = n/4$. By the z-transform, we have

$$Z = \frac{N^+ - n/2}{\sqrt{n/4}} = \frac{2N^+ - n}{\sqrt{n}} \sim N(0,1).$$

We could utilize this test if n is large, that is, if $np \geq 5$ and $n(1-p) \geq 5$. Hence, under H_0, because $p = 1/2$, if $n \geq 10$, we could use the large sample test. The following table summarizes the method for a large sample sign test.

A SIGN TEST FOR A LARGE RANDOM SAMPLE

When the sample size is large ($n \geq 10$), we can use the normal approximation to a binomial. This leads to the large sample sign test:

$$H_0: M = m_0$$

versus

Alternative hypothesis	Rejection region
$H_a: M > m_0$	$z \geq z_\alpha$
$H_a: M < m_0$	$z \leq -z_\alpha$
$H_a: M \neq m_0$	$\lvert z \rvert \geq z_{\alpha/2}$.

The test statistic is

$$Z = \frac{2N^+ - n}{\sqrt{n}}.$$

Decision: Reject H_0, if the test statistic falls in the rejection region, and conclude that H_a is true with at least $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0 because there is not enough evidence to conclude that H_a is true for a given α, and more data are needed.

Assumptions: (1) Population distribution is continuous. **(2)** Sample size greater than or equal to 10 (after the removal of ties). **(3)** The number of ties is small (less than 10% of the sample size).

We illustrate this procedure with the following example.

EXAMPLE 12.3.2

In order to measure the effectiveness of a new procedure for pruning grapes, 15 workers are assigned to prune an acre of grapes. The effectiveness is measured in worker-hours/acre for each person. The results are given below:

5.2 5.0 4.8 3.9 6.1 4.2 4.4 5.5 5.8 4.5
4.2 5.3 4.9 4.7 4.9

Test the null hypothesis that the median time to prune an acre of grapes with this method is 4.5 h against the alternative that it is larger. Use $\alpha = 0.05$.

Solution
We test

$$H_0: M = 4.5 \text{ versus } H_0: M > 4.5.$$

Replacing each value greater than 4.5 with a plus sign and each value less than 4.5 with a minus sign, we have

$$+ + + - + - - + + - + + + + .$$

Because there is one observation that is equal to 4.5, we must discard it and take $n = 14$.
Thus, $N^+ = 10$, *using the large sample approximation, the test statistic is*

$$Z = \frac{2N^+ - n}{\sqrt{n}} = \frac{20 - 14}{\sqrt{14}} = 1.6.$$

For $\alpha = 0.05$, *from the standard normal table, the value of* $z_{0.05} = 1.645$. *Hence, the rejection region is* $z > 1.645$. *Because the observed value of the test statistic does not fall in the rejection region, we do not reject the null hypothesis at* $\alpha = 0.05$ *and conclude that the median time to prune an acre of grapes is 4.5 hours.*

12.3.2 Wilcoxon signed rank test

In the sign test, we have considered only whether each observation is greater than m_0 or less than m_0 without giving any importance to the magnitude of the difference from m_0. Neglecting information on the magnitude of the observations is rather inefficient and may reduce the statistical power of the test. An improved version of the sign test is the Wilcoxon signed rank test, in which one replaces the observations by their ranks of the ordered magnitudes of differences, $|x_i - m_0|$. The smallest observation is ranked as 1, the next smallest will be 2, and so on. However, the Wilcoxon signed rank test requires an additional assumption that the *continuous* population distribution is *symmetric* with respect to its center. Thus, if the data are ordinal, the Wilcoxon test cannot be used.

Hypothesis testing procedure using Wilcoxon signed rank test

We wish to test

$$H_0: M = m_0 \text{ versus } H_1: M \neq m_0.$$

1. Compute the absolute differences $z_i = |x_i - m_0|$ for each observation. Replace each value of the observation that is greater than m_0 by a plus sign and each sample value that is less than m_0 by a minus sign. If the sample value is equal to m_0, discard the observation and adjust the sample size n accordingly.
2. Assign each z_i a value equal to its rank. If two values of z_i are equal, assign each z_i a rank equal to the average of the ranks each should receive if there were not a tie.
3. Let W^+ be the sum of the ranks associated with plus signs and W^- be the sums of ranks with negative signs.
4. **Decision:** If m_0 is the true median, then the observations should be evenly distributed about m_0. For a given α critical region, reject H_0 if

$$W^+ \leq c_1, \text{ where } P(W^+ \leq c_1) = \frac{\alpha}{2},$$

or

$$W^+ \geq c_2, \text{ where } P(W^+ \geq c_2) = \frac{\alpha}{2}.$$

Assumptions: The population distribution is continuous and symmetrical. The number of ties is small, less than 10% of the sample size.

The exact distribution of W^+ is considerably complicated and we will not derive it. However, for certain values of n, the distribution is given in the Wilcoxon signed rank test table.

For the Wilcoxon signed rank test, the rejections region based on the alternative hypothesis is given next.

For

$$H_a: M > m_0, \quad \text{rejection region is} \quad W^+ \geq c, \quad \text{where} \quad P(W^+ \geq c) = \alpha,$$

and for

$$H_a: M < m_0, \quad \text{rejection region is} \quad W^+ \leq c, \quad \text{where} \quad P(W^+ \leq c) = \alpha.$$

We illustrate the Wilcoxon signed rank test with the following examples.

EXAMPLE 12.3.3

For the given data that resulted from an experiment

$$1.51 \quad 1.35 \quad 1.69 \quad 1.48 \quad 1.29 \quad 1.27 \quad 1.54 \quad 1.39 \quad 1.45$$

test the hypothesis that H_0: $M = 1.4$ versus H_a: $M \neq 1.4$. Use $\alpha = 0.05$.

Solution

We wish to test

$$H_0 : M = 1.4 \ \textit{versus} \ H_a : M \neq 1.4.$$

Here, $\alpha = 0.05$, *and* $m_0 = 1.4$. *The results of steps 1 to 3 are given in* Table 12.1.

Thus, we have $W^+ = 29$ *and* $n = 9$. *From the Wilcoxon signed-rank test table in the appendix, we should reject* H_0 *if* $W^+ \leq 6$ *or* $W^+ \geq 38$ *with actual level of* $\alpha = 0.054$. *Because* $W^+ = 29$ *does not fall in the rejection region, we do not reject the null hypothesis that* $M = 1.4$.

TABLE 12.1 Data Summary for Wilcoxon Signed Rank Test.

x_i	$z_i = \lvert x_i - 1.4 \rvert$	Sign	Rank
1.51	0.11	+	5.5
1.35	0.05	−	3
1.69	0.29	+	9
1.48	0.08	+	4
1.29	0.11	−	5.5
1.27	0.13	−	7
1.54	0.14	+	8
1.39	0.01	−	1.5
1.45	0.01	+	1.5

EXAMPLE 12.3.4

Air pollution in large U.S. cities is monitored to see whether it conforms with requirements set by the Environmental Protection Agency. The following data, expressed as an air pollution index, give the air quality of a city for 10 randomly selected days.

$$57.3 \quad 58.1 \quad 58.7 \quad 66.7 \quad 58.6 \quad 61.9 \quad 59.0 \quad 64.4 \quad 62.6 \quad 64.9$$

Test the hypothesis that H_0: $M = 65$ versus H_a: $M < 65$. Use $\alpha = 0.05$.

Solution

We will test

$$H_0 : M = 65 \ \textit{versus} \ H_a : M < 65.$$

Here, $\alpha = 0.05$, *and* $m_0 = 65$.

The results of steps 1 to 3 are given in Table 12.2.

TABLE 12.2 Summary Calculations for Air Pollution Data.

x_i	$z_i = \lvert x_i - 65 \rvert$	Sign	Rank
57.3	7.7	−	10
58.1	6.9	−	9
58.7	6.3	−	8
66.7	1.7	+	3
58.8	6.2	−	7
61.9	4.1	−	5
59.0	6.0	−	6
64.4	0.6	−	2
62.6	2.4	−	4
64.9	0.1	−	1

Thus, $W^+ = 3$, *and* $n = 10$. *Using the Wilcoxon signed rank test table, we should reject* H_0 *if* $W^+ \leq 10$ *with level of significance* $\alpha = 0.042$. *Because the observed value of* W^+ *falls in the rejection region, we reject* H_0 *and conclude that the sample evidence suggests that we conclude the median air pollution index is less than 65.*

The Wilcoxon signed rank test is a nonparametric alternative to the one-sample *t*-test. The question then is, how do we decide which one to choose? Choose the one-sample *t*-test if it is reasonable to assume that the population follows a normal distribution. Otherwise, choose the Wilcoxon nonparametric test. However, the Wilcoxon test will have less power. For example, a normal probability plot of the data of Example 12.3.4 is given in Fig. 12.4. Looking at this figure, we can see that the normality assumption is suspected. It may make more sense to use the nonparametric method.

When sample size n is sufficiently large, under the assumption of H_0 being true, the distribution of W^+ is approximately normal with mean

$$E(W^+) = \frac{1}{4}n(n+1)$$

and variance

$$Var(W^+) = \frac{n(n+1)(2n+1)}{24}.$$

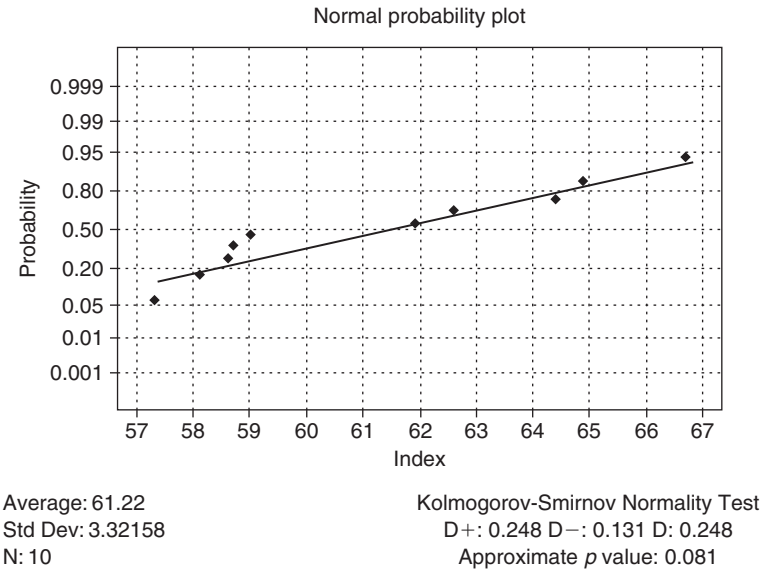


FIGURE 12.4 Normal probability for air pollution index.

Hence, the test statistic is given by

$$Z = \frac{W^+ - \frac{1}{4}n(n+1)}{\sqrt{n(n+1)(2n+1)/24}}$$

which is approximately the standard normal distribution. This approximation can be used when $n > 20$. We summarize the test procedure below.

Summary of the Wilcoxon signed rank test for large samples ($n > 20$)

We test

$$H_0: M = m_0$$

versus

$$H_a: \begin{cases} M > m_0, & \text{upper tailed test} \\ M < m_0, & \text{lower tailed test} \\ M \neq m_0, & \text{two-tailed test.} \end{cases}$$

The test statistic:

$$Z = \frac{W^+ - \frac{1}{4}n(n+1)}{\sqrt{n(n+1)(2n+1)/24}}.$$

Rejection region:

$$\begin{cases} z > z_\alpha, & \text{upper tail RR} \\ z < -z_\alpha, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR.} \end{cases}$$

Decision: Reject H_0, if the test statistic falls in the RR, and conclude that H_a is true with $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0, because there is not enough evidence to conclude that H_a is true for a given α and more data are needed.

Assumptions: (1) The population distribution is continuous and symmetric about 0. **(2)** Sample size is greater than or equal to 20. **(3)** The number of ties is small, <10% of the sample size.

We illustrate the Wilcoxon signed rank test with the following example.

EXAMPLE 12.3.5

The following data give the monthly rents (in dollars) paid by a random sample of 25 households selected from a large city.

425	960	1450	655	1025	750	670	975	660	880
1250	780	870	930	550	575	425	900	525	1800
545	840	765	950	1080					

Using the large sample Wilcoxon signed rank test, test the hypotheses that the median rent in this city is \$750 against the alternative that it is higher with $\alpha = 0.05$.

Solution

We test

$$H_0: M = 750 \text{ versus } H_a: M > 750.$$

Here, $\alpha = 0.05$, and $m_0 = 750$. The results of steps 1 to 3 are given in Table 12.3 (where the asterisk indicates $z_i = 0$).

TABLE 12.3 Summary Calculations for Monthly Rent Data.

x_i	$z_i = \lvert x_i - 750 \rvert$	Sign	Rank
425	325	−	19.5
960	210	+	15
1450	700	+	23
655	95	−	6
1025	302	+	18
750	0	*	ignore
670	80	−	3

Continued

TABLE 12.3 Summary Calculations for Monthly Rent Data.—cont'd

x_i	$z_i = \lvert x_i - 750 \rvert$	Sign	Rank
975	225	+	16.5
660	90	−	4.5
880	130	+	8
1250	500	+	22
780	30	+	2
870	120	+	7
930	180	+	11
550	200	−	12.5
575	175	−	10
425	325	−	19.5
900	150	+	9
525	225	−	16.5
1800	1050	+	24
545	205	−	14
840	90	+	4.5
765	15	+	1
950	200	+	12.5
1080	330	+	21

Here, for $n = 24$, $W^+ = 172.5$, *and the test statistic is*

$$Z = \frac{W^+ - \frac{1}{4}n(n+1)}{\sqrt{n(n+1)(2n+1)/24}}$$

$$= \frac{172.5 - \left(\frac{1}{4}\right)(24)(25)}{\sqrt{\frac{(24)(25)(49)}{24}}} = 0.64286.$$

For $\alpha = 0.05$, *the rejection region is* $z > 1.645$. *Because the observed value of the test statistic does not fall in the rejection region, we do not reject the null hypothesis. There is not enough evidence to conclude that the median rent in this city is more than* \$750.

The rank tests are useful for situations when you suspect that the data do not follow the normal population. It is important to note that ignoring the tied observations reduces the effective sample size, which in turn reduces the power of the test (see Example 7.1.4 for the effect of n on the value of β). This loss is not significant if there are only a few ties. However, if the ties are 10% or more, hypothesis testing using rank tests becomes considerably conservative. That is, they yield error probabilities that are significantly high.

12.3.3 Dependent samples: paired comparison tests

The sign test and the Wilcoxon signed rank test can also be used for paired comparisons. The experimental procedure typically consists of taking "before" and "after" types or otherwise matched as in the paired t-test case readings for each unit. Suppose there are n pairs of before and after observations and we are interested in testing the equality of the two medians. One way to test such observations is to consider the difference between the two observations for a unit to be a

single observation on that unit. Thus, we can treat the sample as being n observations on a population of differences. For this new sample of differences, the testing problem becomes

$$H_0: M = 0 \text{ versus } H_a: M > 0 (\text{or } M < 0, \text{ or } M \neq 0).$$

Hence, the basic procedure could be summarized to first find the difference between the two units for each of the observations, and then follow the testing procedures explained earlier for the sign test or the Wilcoxon signed rank test. Both small sample and large sample cases can be handled as before. In the following example, we illustrate this concept for a large sample sign test.

EXAMPLE 12.3.6

A dietary program claims that 3 months of its diet will reduce weight. In order to test this claim, a random sample of eight individuals who went through this program for 3 months is taken. The following table gives weight in pounds.

Before	180	199	175	226	189	205	169	211
After	172	191	172	230	178	199	171	201

Using a 5% significance level, is there evidence to conclude that the program really reduces the population median weight?

Solution

Let M *denote the median of the population of difference of weights. We will use the difference as "after" − "before." Then we will test*

$$H_0: M = 0 \text{ versus } H_a: M < 0.$$

We will use the large sample sign test. Replacing each value of the difference that is greater than zero by a + sign and less than zero by a − sign, we have

Difference	−8	−8	−3	4	−11	−6	2	−10
Sign	−	−	−	+	−	−	+	−

For $n = 8$ *and* $N^+ = 2$, *the test statistic is given by*

$$Z = \frac{2N^+ - n}{\sqrt{n}} = \frac{4 - 8}{\sqrt{8}} = -1.414.$$

For $\alpha = 0.05$, $z_{0.05} = 1.645$, *and the rejection region is* $z < -1.645$. *Because the observed value of the test statistic does not fall in the rejection region, we do not reject the null hypothesis. Thus, there is not enough evidence to conclude that the new program reduces the weight. Note that even though* $n = 8$ *is small, here we are using the large sample test only for demonstration purposes.*

Exercises 12.3

12.3.1. It was reported that the median interest rate on 30-year fixed mortgages in a certain large city is 7.75% on a particular day, with zero points. A random sample of nine lenders produced the following data of interest rates in percentage.

7.625 7.375 8.00 7.50 7.875 8.00 7.625 7.75 7.25

Test the hypothesis that the median interest rate in this city is different from 7.75%, using (**a**) the sign test, and (**b**) the Wilcoxon signed rank test. Use $\alpha = 0.01$. Compare the two results.

12.3.2. It is believed that a typical family spends 35% of its income on food and groceries. A sample of eight randomly selected families yielded the following data.

30 29 39 49 36 33 37 35

Test the hypothesis that the median percentage of family income spent for food and groceries is 35 against the alternative that it is less than 35. Use $\alpha = 0.05$.

12.3.3. The SAT scores (out of a maximum possible score of 1600) for a random sample of 10 students who took this test recently are:

1355 765 890 1089 986 1128 1157 1065 1224 567

Test the hypothesis that the median SAT score is 1000 against the alternative that it is greater using $\alpha = 0.05$. Use both the sign test and the Wilcoxon signed rank test. Explain if the conclusions are different.

12.3.4. The regulatory board of health in a particular state specifies that the fluoride levels in water must not exceed 1.5 parts per million (ppm). The 20 measurements given here represent the randomly selected daily early morning readings on fluoride levels in water at a certain city.

0.88	0.82	0.97	0.95	0.84	0.90	0.87	0.78	0.75	0.83
0.71	0.92	1.11	0.81	0.97	0.85	0.97	0.91	0.78	0.81

Test the hypothesis that the median fluoride level for this city is 0.90 against the alternative that the median is different from 0.9 at $\alpha = 0.01$, using **(a)** the large sample sign test, and **(b)** the Wilcoxon signed rank test. Interpret the results.

12.3.5. The following data give the weights (in pounds) for a random sample of 20 NFL players.

285	178	311	276	192	232	259	189	289	211
269	285	296	293	288	254	246	234	274	229

Test the hypothesis that the median weight of NFL players is 250 pounds against the alternative that it is greater at $\alpha = 0.05$, using **(a)** the large sample sign test and **(b)** the Wilcoxon signed rank test.

12.3.6. The following data give the amount of money (in dollars) spent on textbooks by 18 students for the last academic year at a large university.

510	425	190	298	157	260	320	615	455
490	188	115	230	610	220	155	315	110

Test the hypothesis that the median amount spent on books at this university is \$325 against the alternative that it is different using the large-sample sign test. Use $\alpha = 0.05$.

12.3.7. It is desired to study the effect of a special diet on systolic blood pressure. The following sample data are obtained for eight adults over 40 years of age before and after 6 months of this diet.

Before	185	222	235	198	224	197	228	234
After	188	217	229	190	226	185	225	231

At 95% confidence level, is there evidence to conclude that the new diet reduces the systolic blood pressure in individuals over 40 years old? Test **(a)** using the sign test, and **(b)** using the Wilcoxon signed rank test. Interpret the results.

12.3.8. In an effort to study the effect on absenteeism of having a day-care facility at the workplace for women with newborn babies (less than 1 year old), a large company compared the number of absent days for a year for seven women with newborn children before and after instituting a day-care facility.

Before	20	18	35	22	17	24	15
After	16	9	22	28	19	13	10

At 99% confidence level, is there evidence to conclude that having a day-care facility at the workplace reduces absenteeism for women with newborn children?

12.3.9. For a popular computer tablet, the user ratings (1 through 5 stars, with 5 stars being the highest rating) of 10 randomly selected are given as follows

5, 5, 1, 4, 3, 5, 4, 4, 5, 4

At the 0.05 level, is there evidence that the median rating is at least 4?

12.3.10. For the data given in Exercise 12.2.10, does the combined evidence from all 16 studies suggest that developing acute renal failure as a complication of sepsis impacts on mortality? Use $\alpha = 0.05$. Do both sign test and Wilcoxon signed rank test.

12.4 Nonparametric hypothesis tests for two independent samples

In this section we learn how to test the equality of the medians of two independent samples from two populations. This is especially useful when one studies the treatment effects, such as the effect of a certain drug to treat a given medical

condition when we have two groups—an experimental group and a control group—or the effect of a particular type of teaching method. Even though this test can be used for more than two samples, here, we will restrict it to two samples. We will describe the *median test*, which corresponds to the sign test, and the *Wilcoxon rank sum test*.

12.4.1 Median test

Let m_1 and m_2 be the medians of two populations 1 and 2, respectively, both with continuous distributions. Assume that we have a random sample of size n_1 from population 1 and a random sample of size n_2 from population 2. The median test can be summarized as follows.

HYPOTHESIS-TESTING PROCEDURE USING MEDIAN TEST

We test

$$H_0: m_1 = m_2 \text{ versus } \begin{cases} m_1 > m_2, & \text{upper tailed test} \\ H_a: m_1 < m_2, & \text{lower tailed test} \\ m_1 \neq m_2, & \text{two-tailed test.} \end{cases}$$

1. Combine the two samples into a single sample of size $n = n_1 + n_2$, keeping track of each observation's original population. Arrange the $n_1 + n_2$ observations in increasing order and find the median of this combined sample. If the median is one of the sample values, discard those observations and adjust the sample size accordingly.
2. Define N_{1b} to be the number of observations of a sample from population 1.
3. **Decision:** If H_0 is true, then we would expect N_{1b} to be equal to some number around $n_1/2$. For H_a: $m_1 > m_2$, rejection region is $N_{1b} \leq c$, where $P(N_{1b} \leq c) = \alpha$, for H_a: $m_1 < m_2$, rejection region is $N_{1b} \geq c$, where $P(N_{1b} \geq c) = \alpha$, and for H_a: $m_1 = m_2$, rejection region is $N_{1b} \geq c_1$, or $N_{1b} \leq c_2$, where

$$P(N_{1b} \geq c_1) = \frac{\alpha}{2} \text{ and } P(N_{1b} \leq c_2) = \frac{\alpha}{2}.$$

Assumptions: (**1**) Population distribution is continuous. (**2**) Samples are independent.

Note that since some observations can be equal to the overall median, and those values will be discarded, N_{1b} need not be equal to n_1. Let $n_1 + n_2 = 2k$. Under H_0, N_{1b} has a hypergeometric distribution given by

$$P(N_{1b} = n_{1b}) = \frac{\binom{n_1}{n_{1b}}\binom{n_2}{k - n_{1b}}}{\binom{n_1 + n_2}{k}}, \quad n_{1b} = 0, 1, 2, \ldots, n_1,$$

with the assumption that $\binom{i}{j} = 0$, if $j > i$. Note that the hypergeometric distribution is a discrete distribution that describes the number of "successes" in a sequence of n draws from a finite population without replacement. Thus, we can find the values of c, c_1, and c_2, required earlier. This calculation can be tedious. To overcome this, we can use the following large sample approximation valid for $n_1 > 5$ and $n_2 > 5$. First classify each observation as above or below the sample median as shown in Table 12.4.

TABLE 12.4 Data Classification With Respect to Median.

	Below	Above	Totals
Sample 1	N_{1b}	N_{1a}	n_1
Sample 2	N_{2b}	N_{2a}	n_2
Total	N_b	N_a	$n_1 + n_2 = n$

It can be verified that the expected value and variance of N_{1a} (similarly for N_{1b}) are given by

$$E(N_{1a}) = \frac{N_a n_1}{n}, \quad \text{and} \quad Var(N_{1a}) = \frac{N_a n_1 n_2 N_b}{n^2(n-1)}.$$

Thus, for a large sample we can write

$$z = \frac{N_{1a} - E(N_{1a})}{\sqrt{Var(N_{1a})}} \sim N(0,1).$$

Hence, we can follow the usual large sample rejection region procedure, which is summarized next.

Summary of large sample median sum test ($n_1 > 5$ and $n_2 > 5$)

We test

$$H_0: m_1 = m_2 \text{ versus } H_a: \begin{cases} m_1 > m_2, & \text{upper tailed test} \\ m_1 < m_2, & \text{lower tailed test} \\ m_1 \neq m_2, & \text{two-tailed test.} \end{cases}$$

The test statistic:

$$z = \frac{N_{1a} - E(N_{1a})}{\sqrt{Var(N_{1a})}},$$

where

$$E(N_{1a}) = \frac{N_a n_1}{n}$$

and

$$Var(N_{1a}) = \frac{N_a n_1 n_2 N_b}{n^2(n-1)}.$$

Rejection region:

$$\begin{cases} z > z_\alpha, & \text{upper tail RR} \\ z < -z_\alpha, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR} \end{cases}$$

Decision: Reject H_0, if the test statistic falls in the *RR*, and conclude that H_a is true with $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0, because there is not enough evidence to conclude that H_a is true for a given α and more data are needed.

Assumptions: (1) Population distributions are continuous. (2) $n_1 > 5$ and $n_2 > 5$.

We illustrate this procedure with the following example.

EXAMPLE 12.4.1

Given below are the mileages (in thousands of miles) of two samples of automobile tires of two different brands, say I and II, before they wear out.

Tire I :	34	32	37	35	42	43	47	58	59	62	69	71	78	84
Tire II :	39	48	54	65	70	76	87	90	111	118	126	127		

Use the median test to see whether the tire II gives more median mileage than tire I. Use $\alpha = 0.05$.

Solution

We will test

$$H_0: m_1 = m_2 \text{ versus } H_0: m_1 < m_2.$$

Because the sample size assumption is satisfied, we will use the large sample normal approximation. The results of steps 1 and 2, using the notation A *for above the median and* B *for below the median, are given in Table 12.5.*

The median is 63.5. Thus, we obtain Table 12.6.

Also,

$$EN_{1a} = \frac{N_a n_1}{n} = \frac{(13)(14)}{26} = 7,$$

and

$$Var(N_{1a}) = \frac{N_a n_1 n_2 N_b}{n^2(n-1)} = \frac{(13)(13)(14)(12)}{16,900} = 1.68.$$

TABLE 12.5 Mileage Data Classification.

Sample values	Population	Above/below the median
32	I	B
34	I	B
35	I	B
37	I	B
39	II	B
42	I	B
43	I	B
47	I	B
48	II	B
54	II	B
58	I	B
59	I	B
62	I	B
65	II	A
69	I	A
70	II	A
71	I	A
76	II	A
78	I	A
84	I	A
87	II	A
90	II	A
111	II	A
118	II	A
126	II	A
127	II	A

TABLE 12.6 Summary of Mileage Data for Automobile Tires.

	Below	Above	Totals
Sample 1	$N_{1b} = 10$	$N_{1a} = 4$	$n_1 = 14$
Sample 2	$N_{2b} = 3$	$N_{2a} = 9$	$n_2 = 12$
Total	$N_b = 13$	$N_a = 13$	$n_1 + n_2 = n = 26$

Hence, the test statistic is

$$z = \frac{N_{1a} - E(N_{1a})}{\sqrt{Var(N_{1a})}} = \frac{4 - 7}{\sqrt{1.68}} = -2.31.$$

For $\alpha = 0.05$, $z_{0.05} = 1.645$. *Hence, the rejection region is* $\{z < -1.645\}$. *Because the observed value of* z *does fall in the rejection region, we reject* H_0 *and conclude that there is enough evidence to conclude that there is a difference in the median mileage for the two types of tires.*

12.4.2 The Wilcoxon rank sum test

The Wilcoxon rank sum test is used for comparing the medians of two independent populations, as in the two-sample *t*-test in the parametric case. For accurate results, it is necessary to assume that the variances of the populations are equal. This test is quite similar to the Wilcoxon signed rank test. Whereas the one-sample Wilcoxon signed rank test requires an additional assumption that the population distribution is symmetric, such an assumption is not necessary for the two-sample Wilcoxon rank sum test. This test can be applied for skewed distributions. The test is almost as powerful as the parametric version when the population distributions are close to normal. Many statistical software packages do not give the Wilcoxon rank sum test; instead the Mann−Whitney test is given. It should be noted that the Wilcoxon rank sum test is equivalent to the Mann−Whitney U-test. We will not separately describe the Mann−Whitney test; however, in practice just perform the Mann−Whitney test if the software has only that test.

Assume that we have n_1 observations randomly sampled from population I and n_2 observations randomly sampled from population II with $n_1 \leq n_2$. The Wilcoxon rank sum test procedure can be summarized as follows.

Hypothesis-testing procedure using the Wilcoxon rank sum test

We test

$$H_0: m_1 = m_2 \text{ versus } H_1: m_1 \neq m_2.$$

1. Combine the two samples into a single sample of size $n_1 + n_2$, keeping track of each observation's original population. Arrange the $n_1 + n_2$ observations in ascending order and assign ranks.
2. Sum the ranks of observations from population II and call it R.
3. Let the test statistic be $W = R - \frac{1}{2}n_2(n_2 + 1)$.
4. **Decision:** If H_0 is false, one would expect that the value of W would be very small or very large. For a size α critical region reject H_0 if

$$W \leq c_1, \text{ where } P(W \leq c_1) = \frac{\alpha}{2},$$

or

$$W \geq c_2, \text{ where } P(W \geq c_2) = \frac{\alpha}{2}.$$

Note: The exact distribution of W is given in the Wilcoxon rank sum test table in the appendix for small values of n_1 and n_2.

In the Wilcoxon rank sum test, based on the alternative hypothesis, we have the following rejection regions. For

$$H_a: m_1 > m_2, \text{ rejection region is } W \geq c, \text{ where } P(W \geq c) = \alpha.$$

and for

$$H_a: m_1 < m_2, \text{ rejection region is } W \leq c, \text{ where } P(W \leq c) = \alpha.$$

We will illustrate the foregoing procedure with the following example.

EXAMPLE 12.4.2

Comparison of the prices (in dollars) of two brands of similar automobile tires resulted in the data in Table 12.7.

TABLE 12.7 Prices of Two Brands of Tires.

Tire I:	85	99	100	110	105	87		
Tire II:	67	69	70	93	105	90	110	115

Use the Wilcoxon rank sum test with $\alpha = 0.05$ to test the null hypothesis that the two population medians are the same against the alternative hypothesis that the population medians are different.

Solution

Here, we need to test

$$H_0: m_1 = m_2 \text{ versus } H_a: m_1 \neq m_2.$$

The sample sizes are $n_1 = 6$, *and* $n_2 = 8$. *Combining step 1 and step 2, we have the results shown in Table 12.8.*

TABLE 12.8 Ranking of Prices of Tires.

Value	67	69	70	85	87	90	93	99	100	105	105	110	110	115
Population	II	II	II	I	I	II	II	I	I	I	II	I	II	II
Rank	1	2	3	4	5	6	7	8	9	10.5	10.5	12.5	12.5	14

The sum of ranks of observations from population II is $R = 56$. *Hence, the test statistic is*

$$W = R - \frac{1}{2}n_2(n_2 + 1)$$

$$= 56 - \frac{1}{2}(8)(9) = 20.$$

For $\alpha = 0.05$, *the rejection region is* $W \leq 9$ *or* $W > 38$, *with the actual* α *being* 0.0592. *Because the observed value of the test statistic does not fall in the rejection region,* H_0 *is not rejected. Thus, we do not have enough evidence to conclude that the median prices are different for these two brands of automobile tires.*

When the sample sizes are large and when H_0 is true, the distribution of the Wilcoxon rank sum test can be approximated by the normal distribution. It can be shown that under H_0, when both n_1 and n_2 are greater than 10, the distribution of W is approximately normal with

$$E(W) = \frac{n_1 n_2}{2} \text{ and } Var(W) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

For a large random sample, we can summarize the test procedure as follows.

Summary of large sample median sum test ($n_1 > 10$ and $n_2 > 10$)

We test

$$H_0: m_1 = m_2 \text{ versus } H_a: \begin{cases} m_1 > m_2, & \text{upper tailed test} \\ m_1 < m_2, & \text{lower tailed test} \\ m_1 \neq m_2, & \text{two-tailed test.} \end{cases}$$

The test statistic:

$$z = \frac{W - n_1 n_2/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}.$$

Rejection region:

$$\begin{cases} z > z_\alpha, & \text{upper tail RR} \\ z < -z_\alpha, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR.} \end{cases}$$

Assumption: The samples are independent and $n_1 > 10$ and $n_2 > 10$.

Decision: Reject H_0, if the test statistic falls in the *RR*, and conclude that H_a is true with $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0, because there is not enough evidence to conclude that H_a is true for a given α and more data are needed.

We will use the foregoing procedure to solve the following problem.

EXAMPLE 12.4.3

In an effort to determine the immunoglobulin D (IgD) levels of a certain ethnic group, a large number of blood samples representing both sexes for 12-year-olds were taken. The following sample data give the IgD levels (in mg/100 mL).

Male:	9.3	0.0	12.2	8.1	5.7	6.8	3.6	9.4	8.5	7.3	9.7	
Female:	7.1	0.0	5.9	7.6	2.8	5.8	7.2	7.4	3.5	3.3	7.5	7.0

Use the large sample Wilcoxon rank sum test with the significance level $\alpha = 0.01$ to test the hypothesis that there is no difference between the sexes in the median level of IgD.

Solution
We need to test

$$H_0: m_1 = m_2 \text{ versus } H_a: m_1 \neq m_2.$$

Here, $n_1 = 11$, *and* $n_2 = 12$, *and the results of step 1 and step 2 are given in Table 12.9, where we use M or F to identify the population from which the data are coming.*

TABLE 12.9 Ranking of Immunoglobulin D (IgD) Levels.

Value	0	0	2.8	3.3	3.5	3.6	5.7	5.8	5.9	6.8	7	7.1
M or F	M	F	F	F	F	M	M	F	F	M	F	F
Rank	1.5	1.5	3	4	5	6	7	8	9	10	11	12
Value	7.2	7.3	7.4	7.5	7.6	8.1	8.5	9.3	9.4	9.7	12.2	
M or F	F	M	F	F	F	M	M	M	M	M	M	
Rank	13	14	15	16	17	18	19	20	21	22	23	

The sum of the ranks for females is $R = 114.5$, *and*

$$W = R - \frac{1}{2}n_2(n_2 + 1)$$

$$= 114.5 - \frac{1}{2}(12)(13) = 36.5.$$

Therefore, the test statistic results in

$$Z = \frac{W - n_1 n_2/2}{\sqrt{n_1 n_2(n_1 + n_2 + 1)/12}}$$

$$= \frac{36.5 - (11)(12)/2}{\sqrt{(11)(12)(24)/12}} = -1.815 \approx -1.82.$$

For $\alpha = 0.01$, *we have* $z_{\alpha/2} = z_{0.005} = 2.575$. *Hence, the rejection region is* $z < -2.575$ *or* $z > 2.575$. *Because the test statistic does not fall in the rejection region, we do not reject* H_0 *at* $\alpha = 0.01$ *and conclude that there is not enough evidence to conclude that there is any difference between the sexes in the median level of IgD.*

With a slight modification of the ranking system in the Wilcoxon rank sum test, we could test for the equality of variances when the normality assumption of the *F*-test fails.

Exercises 12.4

12.4.1. The following data give the winning proportions of the top six football teams from each of the two conferences of the NFL.

American conference	0.818	0.727	0.909	0.818	0.727	0.545
National conference	0.636	0.545	0.636	0.636	0.818	0.455

Use the Wilcoxon rank sum test at the significance level of 0.05 to test the null hypothesis that the two samples contain populations with identical medians against the alternative hypothesis that the medians are not equal. State any assumptions you have made to solve the problem.

12.4.2. Comparison of two protective methods against corrosion yielded the following maximum depths of pits (in thousandths of an inch) in pieces of similar metals subjected to the respective treatments:

Method I:	68	75	69	75	70	69	72
Method II:	61	65	57	63	58		

Use the Wilcoxon rank sum test at the significance level of 0.01 to test the null hypothesis that the two samples have identical medians against the alternative hypothesis that the medians are not equal.

12.4.3. Show that when H_0 is true, the mean and variance of the Wilcoxon rank sum test with sample sizes n_1 and n_2 are

$$E(W) = \frac{n_1 n_2}{2} \text{ and } Var(W) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

12.4.4. In order to make inferences about the temporal muscles of the cat, a certain dose of tubocurarine is injected into a random sample of nine cats. The following data give the tetanus frequency (in hertz) in the temporal (T) muscles before and after injection of tubocurarine.

T before	24	33	27	23	31	28	31	24	19
T after	27	38	34	32	37	28	35	28	41

Use the Wilcoxon rank sum test at the significance level of 0.05 to test the null hypothesis that the median tetanus frequency (in hertz) in the temporal (T) muscles is larger after injection of tubocurarine. State any assumptions you made to solve the problem.

12.4.5. In a study of the net conversion of progesterone in rat liver, the following samples were attained for the net conversion in rats 3−4 weeks old:

Male:	16.9	16.0	13.5	13.1	14.2	11.6	12.8	17.3	13.8	9.8	16.0	15.9	16.7	15.1
Female:	13.8	11.2	7.5	10.4	15.8	14.5	9.5	9.8	5.1	5.5	6.5	7.2		

Use the large sample Wilcoxon rank sum test at the significance level of 0.05 to test the hypothesis that the median net conversion of progesterone in male rats is larger than that in female rats. What would be your conclusion if you were to use the median test?

12.4.6. Two groups of randomly selected 1-acre plots were treated with two different brands of fertilizer. The following data give the yields of corn (in bushels) from each of these plots.

Fertilizer I:	89	93	105	94	92	96	93	101
Fertilizer II:	85	88	94	87	86	91		

Use the data to determine whether there is a difference in yields for two brands of fertilizers. Use $\alpha = 0.01$. State any assumptions you made to solve the problem.

12.4.7. The following information is obtained from two independent samples.

Sample 1:	15	8	12	4	10	8	13	7	12	6	14	11
Sample 2:	18	13	15	19	17	13	17	16				

Test at 1% significance level that the median for sample 1 is less than the median for sample 2 and interpret the meaning of your result.

12.4.8. In order to determine if a new hybrid seeding produces a bushier flowering plant, data are collected on shrub girth (in inches) for both current variety and hybrid plants resulted in the following values.

Current variety	27.7	25.1	35.4	36.5	22.0	30.5	
Hybrid	35.8	30.0	34.6	37.5	31.9	32.6	39.7

Test at 1% significance level that the median for sample 1 is different from the median for sample 2 and interpret the meaning of your result.

12.5 Nonparametric hypothesis tests for $k \geq 2$ samples

In this section we learn how to compare the medians of more than two independent samples and to determine whether medians of the groups differ. These tests are nonparametric alternatives to the ANOVA methods discussed in Chapter 9. We study the *Kruskal−Wallis test* and *Friedman test*. Both of these methods test the equality of the treatment medians.

12.5.1 The Kruskal–Wallis test

The Kruskal–Wallis test is a generalization of the Wilcoxon rank sum test for two independent samples to several independent samples. This test is a nonparametric alternative to one-way ANOVA. The Kruskal–Wallis test is almost as powerful as the one-way ANOVA when the data are from a normal distribution, and more powerful in the case of nonnormality or in the presence of outliers. We now describe this test.

Suppose that we have k populations, with θ_i being the median of the population i and k independent random samples from these populations. Let the samples from the ith population be n_i. We wish to test the equality of the medians of different groups—that is, to test the hypothesis

$$H_0: \theta_1 = \theta_2 = \cdots = \theta_k = 0 \quad \text{versus} \quad H_a: \text{Not all } \theta'\text{s equal } 0.$$

We shall show that the hypothesis $\theta_1 = \cdots = \theta_k$ is equivalent to the hypothesis H_0: $\theta_1 = \theta_2 = \cdots = \theta_k = 0$. Let $\theta_1 = \cdots = \theta_k = t$ (same number). Then the observations $y_{ij} - t$ ($i = 1, 2, \ldots, k$) will be from a population with median zero. Because the Kruskal–Wallis test procedure depends only on the ranks of y_{ij} values in the combined sample and the ranks of (y_{ij}–t) values are identical to those of y_{ij} values, the two hypotheses are equivalent.

We summarize the Kruskal–Wallis procedure to solve this type of problem, which is given by the following steps.

Kruskal–Wallis test procedure

1. Combine and rank all $N = \sum_{i=1}^{n} n_i$ observations y_{ij} in ascending order. Also keep track of the groups from which the observations came. Assign average ranks in case of ties. Let

$$r_{ij} = \text{rank}(y_{ij}).$$

2. Calculate the group sum,

$$r_i = \sum_{i=1}^{n_i} r_{ij}, \quad i = 1, 2, \ldots, k.$$

and the group averages

$$\overline{r}_i = \frac{r_i}{n_i}, \quad i = 1, 2, \ldots, k.$$

3. Let

$$r = \sum_{i=1}^{k} r_i = \frac{N(N+1)}{2}$$

(this can be used as a check for accuracy of your calculation of $r_i's$) and let

$$\overline{r} = \frac{r}{N} = \frac{N+1}{2}.$$

4. Calculate the Kruskal–Wallis test statistic

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i(\overline{r}_i - \overline{r})^2$$

or the convenient computational form of H,

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{r_i^2}{n_i} - 3(N+1).$$

Note that to compute the convenient form of H, there is no need to calculate $\overline{r}_i$ and $\overline{r}$.

5. Reject H_0 if

$$H \geq c,$$

where the constant c is chosen to achieve a specified value for α.

The exact distribution of H is complicated. It depends on the sample sizes, n_1, n_2, …, n_k, and so it is not practical to tabulate its values beyond a small number of cases. When k or N is large, the exact distribution of H under the null hypothesis can be approximated by the chi-square distribution with $(k - 1)$ degrees of freedom. To this effect, we state the Kruskal–Wallis theorem without proof.

Theorem 12.5.1. *When* $\mathrm{H_0}$: $\theta_1 = \theta_2 = \cdots = \theta_k$ *is true, then as* N *becomes large, the statistic*

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i(\overline{r}_i - \overline{r})^2$$

has an asymptotic distribution that is chi-square with $(k - 1)$ *degrees of freedom.*

Thus, for approximate large samples the Kruskal–Wallis test for a given α is to reject H_0 if

$$H > \chi^2_\alpha(k-1).$$

The chi-square approximation is acceptable when the group sample sizes $n_i > 5$ with $k \geq 3$. However, for convenience, we will use the chi-square approximation for all values of n_i. For this test, we follow the procedure described earlier except that for finding the rejection region, we use the chi-square table.

The following example illustrates how we use the foregoing procedure to test the appropriate hypothesis for three populations.

EXAMPLE 12.5.1

In an effort to investigate the premium charged by insurance companies for auto insurance, an agency randomly selects a few drivers who are insured from three different companies. Assume that these persons have similar autos, driving records, and level of coverage. Table 12.10 gives the premiums paid per 6 months by these drivers with these three companies. Using the 5% level of significance, test the null hypothesis that the median auto insurance premium paid per 6 months by all drivers insured with each of these companies is the same.

TABLE 12.10 Auto Insurance Premium by Company.

Company I	Company II	Company III
396	348	378
438	360	330
336	522	294
318		474
		432

Solution

Here, we need to test

$$H_0: M_1 = M_2 = M_3 = 0 \quad \textit{versus} \quad H_a: \text{Not all } M_i\text{'s equal } 0,$$

where M_i *is the true median of the auto insurance premium paid to company* i, i = 1, 2, 3.

Here $n_1 = 4$, $n_2 = 3$, *and* $n_3 = 5$. *Hence, there are* $N = \sum_{i=1}^{3} n_i = 12$ *observations. Let* Y *denote the observations in ascending order. Table 12.11 gives the combined data in ascending order while keeping track of the groups and their ranks.*

TABLE 12.11 Ranking of Auto Insurance Premiums.

Premium	294	318	330	336	348	360	378	396	432	438	474	522
Group	3	1	3	1	2	2	3	1	3	1	3	2
Rank	1	2	3	4	5	6	7	8	9	10	11	12

Thus, the group rank sums are

$$r_1 = 24, r_2 = 23, \quad \textit{and} \quad r_3 = 31.$$

As a check for accuracy of these calculations, note that

$$r_1 + r_2 + r_3 = 78 = \frac{N(N+1)}{2} = \frac{(12)(13)}{2}.$$

The test statistic is given by

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{r_i^2}{n_i} - 3(N+1)$$

$$= \frac{12}{(12)(13)}\left(\frac{(24)^2}{4} + \frac{(23)^2}{3} + \frac{(31)^2}{5}\right) - 3(13)$$

$$= 0.42564.$$

From the chi-square table, $\chi^2_{0.05}(2) = 5.991$, *and hence, the rejection region is* $H \geq 5.991$. *Because the observed value of* H *does not fall in the rejection region, we do not reject* H_0 *and conclude that there is no evidence to show that the median auto insurance premiums paid per 6 months by all drivers insured in each of these companies are different.*

12.5.2 The Friedman test

The Friedman test, named after the Nobel laureate economist Milton Friedman, tests whether several treatment effects (measured as locations) are equal for data in a two-way layout. We will assume that there are k different treatment levels and l blocks. In each block, assign one experimental unit to each treatment level. We want to test whether the true medians for different treatment levels are the same in each block—that is, to test

$$H_0: \text{True medians at different levels are all equal}$$

versus

$$H_a: \text{Not all the medians are equal.}$$

Rather than combine the entire sample as in the Kruskal–Wallis statistic, here we order the y values within each block and then assign each its rank. In order to eliminate the differences due to blocks, we take the sum of ranks for each treatment level. The following gives a summary of the procedure.

The Friedman test procedure

1. Rank observations from k treatments separately within each block. Assign average ranks in case of ties. Let $R_{ij} = \text{rank}(Y_{ij})$, the rank of the observation for treatment level i in block j.
2. Calculate the rank sums

$$R_i = \sum_{j=1}^{l} R_{ij}, \quad i = 1, 2, \ldots, k.$$

3. Calculate the Friedman statistic

$$S = \frac{12}{lk(k+1)}\sum_{i=1}^{k}\left(R_i - \frac{l(k+1)}{2}\right)^2$$

or a convenient computational form,

$$S = \frac{12}{lk(k+1)}\sum_{i=1}^{k} R_i^2 - 3l(k+1).$$

4. Reject H_0 if $S \geq c$, where the constant c is chosen to achieve a specified value for α.

The exact distribution of S is complicated. Here, for $k = 3, 4, 5$, and for various values of l, the Friedman distribution has been calculated and its values are given in the table in Appendix A7. We will illustrate this four-step procedure with an example.

EXAMPLE 12.5.2

Three classes in elementary statistics are taught by three different persons, a regular faculty member, a graduate teaching assistant, and an adjunct from outside the university. At the end of the semester, each student is given a standardized test. Five students are randomly picked from each of these classes, and their scores are given in Table 12.12. Test whether there is a difference between the scores for the three persons teaching with $\alpha = 0.05$.

TABLE 12.12 Test Grades by Instructor.

Faculty	Teaching assistant	Adjunct
93	88	86
61	90	56
87	76	73
75	82	90
92	58	47

Solution

Here, we need to test

$$H_0\text{: Median for the three persons scores are all equal}$$

$$H_a\text{: The medians are not equal}$$

We are given $\alpha = 0.05$, $k = 3$, *and* $l = 5$. *To compute the value of the statistic* S, *we first assign ranks for each student as shown in* Table 12.13. H_a: *Note that they are not all equal.*

TABLE 12.13 Ranks of Test Scores by Instructor.

Faculty	Teaching assistant	Adjunct
3	2	1
2	3	1
3	2	1
1	2	3
3	2	1

Thus, we have

$$R_1 = 12,\ \ R_2 = 11,\ \ \text{and}\ \ R_3 = 7,$$

and the test statistic is given by

$$S = \frac{12}{lk(k+1)} \sum_{i=1}^{k} R_i^2 - 3l(k+1)$$

$$= \frac{12}{(5)(3)(4)} \left((12)^2 + (11)^2 + (7)^2\right) - (3)(5)(4) = 2.8.$$

From the Friedman table, the rejection region is $S \geq 5.20$ *at an exact significance level of* 0.092. *Because the computed value of the test statistic does not fall in the rejection region, we do not reject* H_0 *and conclude that there is no difference in scores based on who teaches the course.*

When the number of blocks, l, becomes large, the Friedman test statistic has an approximate chi-square distribution under the null hypothesis. That is:

Theorem 12.5.2. *When* H_0: $\theta_1 = \theta_2 = \cdots = \theta_3$ *is true then, as* l *becomes large,*

$$S = \frac{12}{lk(k+1)} \sum_{i=1}^{k} \left(R_i - \frac{l(k+1)}{2}\right)^2$$

has an asymptotic distribution that is chi-squared with (k − 1) *degrees of freedom.*

Thus, for an approximate large random sample, the Friedman test for given α is to reject H_0 if $S > \chi^2_\alpha(k-1)$.

When the values of k and l exceed the values given in the Friedman table, we could use the chi-square approximation, which gives acceptable results. We proceed to illustrate the Friedman test with the following example.

EXAMPLE 12.5.3

In the previous example, we now randomly select 10 student grades from each class, resulting in the data shown in Table 12.14.

Test whether there is a difference between the scores for the three persons teaching. Use $\alpha = 0.05$.

TABLE 12.14 Test Grades of 10 Random Students From Each Instructor.

Faculty	Teaching assistant	Adjunct
93	88	86
61	90	56
87	76	73
75	82	90
92	58	47
45	74	88
99	23	77
86	61	18
82	60	66
74	77	55

Solution

Here we need to test

H_0: The true median scores for the three instructors are all equal

versus

H_a: They are not all equal.

We are given $\alpha = 0.05$, k = 3, *and* l = 10. *We use the chi-square approximation to solve the problem. To compute the value of the statistic* S *we first assign ranks for each student as shown in Table 12.15. The Friedman test statistic is*

TABLE 12.15 Ranks of Test Scores of 10 Random Students.

	Faculty	Teaching assistant	Adjunct
	3	2	1
	2	3	1
	3	2	1
	1	2	3
	3	2	1
	1	2	3
	3	1	2
	3	2	1
	3	1	2
	2	3	1
Total	24	20	16

$$S = \frac{12}{lk(k+1)} \sum_{i=1}^{k} R_i^2 - 3l(k+1)$$

$$= \frac{12}{(10)(3)(4)} \left((24)^2 + (20)^2 + (16)^2\right) - (3)(10)(4) = 3.2.$$

From the chi-square table, $\chi^2_{0.05}(2) = 5.992$. Hence, the rejection region is $S \geq 5.992$. The computed value of the test statistic does not fall in the rejection region, and we do not reject H_0. We conclude that there is no difference in scores based on who teaches the course.

Friedman's test is an alternative to the repeated measures ANOVA, when assumptions such as that of normality or equality of variance are not satisfied. Because this test, like many other nonparametric tests, does not make a distribution assumption, it is not as powerful as the ANOVA.

Exercises 12.5

12.5.1. Table 12.16 shows a random sample of observations on children under 10 years of age, each observation being the IgA immunoglobulin level measured in international units from a large number of blood samples, and the population is studied in blocks in terms of age groups (the upper value is not included) as I: (1−3), II: (3−6), III: (6−8), and IV: (8−10). Test for the hypothesis of equality of true medians for IgA level in each block (age level), **(a)** with the 5% level and **(b)** with the 1% level of significance. Compare the results obtained.

TABLE 12.16 IgA Immunoglobulin Level of Children.

I	6	37	19	14	51	68	27	75
II	32	65	76	42	45	41	38	63
III	73	75	59	90	37	32	63	80
IV	81	42	48	60	98	100	79	45

12.5.2. In an effort to study the effect of four different preventive maintenance programs on downtimes (in minutes) for a certain period of time in a production line, a factory runs four parallel production lines, and each line has five different types of machine. The different maintenance programs are randomly assigned to each of the four production lines so as to treat the various machines as blocks. Results are shown in Table 12.17.
Test the hypothesis at $\alpha = 0.05$, H_0: True medians of the four maintenance programs are equal versus H_a: Not all are equal. (Hint: In the Friedman test, $k = 4$, and $l = 6$.) State any assumptions you have made to solve this problem.

TABLE 12.17 Downtimes by Program.

Machine	Method 1	Method 2	Method 3	Method 4
I	181	124	126	181
II	185	122	125	160
III	67	65	68	69
IV	121	66	120	68
V	62	60	62	65

12.5.3. Show that, when $k = 2$, the Kruskal−Wallis statistics,

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{r_i^2}{n_i} - 3(N+1)$$

becomes equivalent to the Wilcoxon rank sum test.

12.5.4. A consumer testing agency is interested in determining whether there is a difference in the mileage for three brands of gasoline. To test this, four different vehicles are driven with each of these gasolines. Results are shown in Table 12.18.
Test whether there is a difference between the three gasoline medians at the 0.05 level.

TABLE 12.18 Mileage by Gasoline Type.

	Gasoline		
Vehicle	A	B	C
I	19	25	22
II	26	33	39
III	20	28	25
IV	18	30	21

12.5.5. In order to study the effect of fertilizers, five groups of 1-acre plots were randomly selected. One group was not treated with any fertilizers and the remaining four groups were treated with four different brands of fertilizers. Table 12.19 gives the yields of corn (in bushels) from each of these plots.
Use the data to determine whether there is a difference in yields for different fertilizers. Use $\alpha = 0.01$.

TABLE 12.19 Yield by Fertilizer.

None:	58	27	36	41	48	36	50	50	39
Fertilizer I:	69	67	57	63	49	65	78	69	
Fertilizer II:	95	92	92	89	100	88	79	97	75
Fertilizer III:	102	111	92	103	102	94	100	112	96
Fertilizer IV:	127	115	112	122	114	107	116	112	108

12.5.6. In order to compare grocery prices of four different grocery stores on a particular day in November 1999, 11 randomly selected items with the same brands are given in Table 12.20.
Use the data to determine whether there is a difference in prices at these four grocery store chains. Use $\alpha = 0.01$. State any assumptions you have made to solve this problem.

TABLE 12.20 Grocery Prices by Store.

Product	Store A	Store B	Store C	Store D
Bread (20 oz)	$1.39	$1.39	$1.39	$1.39
Red apples (1 lb)	1.29	1.29	0.99	0.68
Large eggs (1 dozen)	0.69	0.88	0.89	0.89
Orange juice (64 oz)	3.29	2.99	2.79	2.69
Cereal (15 oz)	3.59	3.19	3.19	3.58
Canned corn (15.25 oz)	0.50	0.53	0.50	0.49
Sugar crystals (5 lb)	1.99	2.09	1.99	1.89
2% milk (1 gal)	3.19	3.19	3.09	3.09
Frozen pizza (21.5 oz)	3.00	4.59	3.50	3.50
Puppy chow (4.4 lb)	4.59	3.69	3.69	3.99
Diapers (56-pack)	12.99	12.99	12.99	11.88

12.6 Chapter summary

In this chapter, we first learned about nonparametric approaches to interval estimation and nonparametric hypothesis tests for one sample, such as the sign test, the Wilcoxon signed rank test, and dependent sample paired comparison tests. Then nonparametric hypothesis tests for two independent samples such as the median test and Wilcoxon rank sum test were considered. Later the Kruskal–Wallis test and the Friedman test were explained for more than two samples.

It is natural to ask, "Why do we substitute a set of nonnormal numbers, such as ranks, for the original data?" Few data are truly normal. Rank tests are sometimes called "approximate" tests. They are most useful in instances when we suspect that the data are not normal, and we either cannot transform the data to make them more normal, or do not like to do so. One of the simple ways to check for appropriateness of use of nonparametric tests is to simply construct a stem-and-leaf display or a histogram for the sample data and see whether they look symmetric and approximately bell shaped. If this is not so, we may often be better off using a nonparametric approach.

Since the 1940s, many nonparametric procedures have been introduced, and the number of procedures continues to grow. The nonparametric tests presented in this chapter represent only a small portion of available nonparametric tests. There are many references available in the bibliography for further reading on the subject.

In this chapter, we have also learned the following important concepts and procedures:

- Procedure for finding $(1 - \alpha)100\%$ confidence interval for the median M
- Hypothesis-testing procedure by sign test
- A large sample sign test
- Hypothesis-testing procedure by Wilcoxon signed rank test
- Summary of large sample Wilcoxon signed rank test ($n > 20$)
- Summary of large sample median sum test ($n_1 > 5$ and $n_2 > 5$)
- Hypothesis-testing procedure by Wilcoxon rank sum test
- Summary of large sample Wilcoxon rank sum test ($n_1 > 10$ and $n_2 > 10$)
- Kruskal–Wallis test procedure
- Friedman test procedure

12.7 Computer examples

In this section, we illustrate some nonparametric procedures using statistical software packages.

12.7.1 Examples using R

EXAMPLE 12.7.1 (Sign test)

Using the following data test $H_0 : M = 1.4$ vs. $H_a : M > 1.4$, *using the sign test.*
Sample (x)*: 1.51 1.35 1.69 1.48 1.29 1.27 1.54 1.39 1.45.*

R code

```
y = length(which(x > 1.4));
n = length(x);
binom.test(y,n,alternative = "greater");
```

Output

```
        Exact binomial test.
data:   y and n
number of successes = 5, number of trials = 9, p value = 0.5
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
0.2513676 1.0000000
sample estimates:
probability of success
        0.5555556
```

Our p *value suggests that we fail to reject the null hypothesis for any reasonable level of significance and that the medians are equal.*

EXAMPLE 12.7.2 (Wilcoxon test)

Using the data from the previous example test $H_0 : M = 1.4$ vs. $H_a : M \neq 1.4$, *using one-sample Wilcoxon test.*

R code

wilcox.test(x,mu = 1.4);

Output

Wilcoxon signed rank test

We fail to reject the null hypothesis suggesting that the true mean is equal to 1.4 for any reasonable level of significance.

data: x

V = 30, *p* value = 0.4258

alternative hypothesis: true location is not equal to 1.4

EXAMPLE 12.7.3 (Two-sample sign test)

Using the following data, test $H_0 : M = 0$ vs. $H_a : M < 0$, *using the two-sample sign test, where* M *is the median difference. Use* $\alpha = 0.05$.

Sample (x)	180	199	175	226	189	205	169	211
Sample (y)	172	191	172	230	178	199	171	201

R code

z = x-y;

y = length(which(z < 0));

n = length(z);

binom.test(y,n,alternative = "less");

Fail to reject the null hypothesis since the p *value is larger than our alpha. This suggests the median difference is zero.*

Output

Exact binomial test.

data: y and n

number of successes = 2, number of trials = 8, *p* value = 0.1445

alternative hypothesis: true probability of success is less than 0.5

95% confidence interval:

0.0000000 0.5996894

sample estimates:

probability of success

0.25

EXAMPLE 12.7.4 (Wilcoxon two-sample test)

Use the Wilcoxon rank sum test with $\alpha = 0.05$ *to test the null hypothesis that the two population medians are the same against the alternative hypothesis that the population medians are different.*

Sample (x): 85 99 100 110 105 87

Sample (y): 67 69 70 93 105 90 110 115

R code

wilcox.test(x,y);

Output

Wilcoxon rank sum test with continuity correction

Fail to reject the null hypothesis

data: x and y

W = 28, *p* value = 0.6507

alternative hypothesis: true location shift is not equal to 0

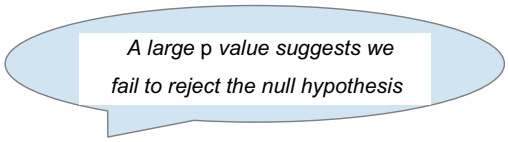
EXAMPLE 12.7.5 (Kruskal–Wallis test)

In an effort to investigate the premium charged by insurance companies for auto insurance, an agency randomly selects a few drivers who are insured by three different companies. Assume that these persons have similar cars, driving records, and levels of coverage. The following data are the premiums paid per 6 months by these drivers with these three companies. Using $\alpha = 0.05$*, test the null hypothesis that the median auto insurance premium paid per 6 months by all drivers insured in each of these companies is the same.*

Company	C1	C1	C1	C1	C2	C2	C2	C3	C3	C3	C3	C3
Value	396	438	336	318	348	360	522	378	330	294	474	432

R code

kruskal.test(value, company);



Output

Kruskal–Wallis rank sum test

data: value and company

Kruskal–Wallis chi-squared = 0.4256, df = 2, p value = 0.8083

EXAMPLE 12.7.6 (Friedman test)

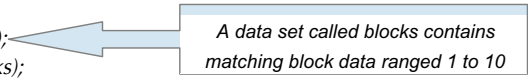
Using the following data conduct a Friedman test.

C1	93	61	87	75	92	45	99	86	82	74
C2	88	90	76	82	58	74	23	61	60	77
C3	86	56	73	90	47	88	77	18	66	55

R code

blocks = c(c(1:10),c(1:10),c(1:10));

friedman.test(values, groups,blocks);



Output

Friedman rank sum test

data: values, groups and blocks

Friedman chi-squared = 3.2, df = 2, p value = 0.2019

12.7.2 Minitab examples

EXAMPLE 12.7.7

(One-sample sign): For the data

1.51 1.35 1.69 1.48 1.29 1.27 1.54 1.39 1.45

test H_0: $M = 1.4$ versus H_a: $M > 1.4$, using sign test.

Solution

Enter data in ***C1.*** *Then*

Stat > ***Nonparametric*** > ***1-Sample Sign*** … > in ***Variables:*** type ***C1*** > click ***Test median:*** type ***1.4*** > in ***Alternative:*** click ***greater than*** > click ***OK***

We can obtain the nonparametric confidence interval using the following procedure. Enter in variable, **C1,** and then ***Stat*** > ***Nonparametric*** > ***1-Sample Sign*** … > in ***Variables:*** type ***C1*** > click ***Confidence interval*** > in ***Level:*** enter appropriate, say, ***95.0*** > click ***OK***

EXAMPLE 12.7.8

(One-sample Wilcoxon): For the data

1.51 1.35 1.69 1.48 1.29 1.27 1.54 1.39 1.45

test H_0: $M = 1.4$ versus H_a: $M \neq 1.4$, using one-sample Wilcoxon test.

Solution

We will give only Sessions commands; the Windows procedure is similar to the previous example.

Stat > ***Nonparametric*** > ***1-Sample Wilcoxon*** … > in ***Variables:*** type ***C1*** > click ***Test median:*** type ***1.4*** > in ***Alternative:*** click ***Not equal*** > click ***OK***

EXAMPLE 12.7.9

(Two-sample sign test): For the data

Sample 1	180	199	175	226	189	205	169	211
Sample 2	172	191	172	230	178	199	171	201

test H_0: $M = 0$ versus H_a: $M < 0$, using the two-sample sign test, where M is the median of the difference. Use $\alpha = 0.05$.

Solution

After entering sample 1 data in ***C1*** *and sample 2 data in* ***C2,*** *we can use the following sequence:*

Calc > ***Calculator*** … > in ***Store result in variable:*** type ***C3*** > in ***Expression:*** type ***C2−C3*** > click ***OK***

We will get the pairwise difference of the two samples. For these values, we will apply the one-sample sign test.

Stat > ***Nonparametric*** > ***1-sample sign*** … > in ***Variables:*** type ***C3*** > click ***Test median:*** and in ***Alternative:*** choose ***Less than*** > click ***OK***

EXAMPLE 12.7.10

(Kruskal−Wallis test): In an effort to investigate the premium charged by insurance companies for auto insurance, an agency randomly selects a few drivers who are insured by three different companies. Assume that these persons have similar cars, driving records, and levels of coverage. Table 12.21 gives the premiums paid per 6 months by these drivers with these three companies.

Using the 5% significance level, test the null hypothesis that the median auto insurance premium paid per 6 months by all drivers insured in each of these companies is the same. Use Minitab.

TABLE 12.21 Auto Insurance Premium by Company.

Company I	Company II	Company III
396	348	378
438	360	330
336	522	294
318		474
		432

Solution

Enter data for company I in ***C1,*** *for company II in* ***C2,*** *and for company III in* ***C3****. First stack the data while keeping track of the companies in the following way.*

Manip > ***Stack/Unstack*** > ***Stack Columns*** … > in ***Stack the following columns:*** type ***C1 C2 C3*** > in ***Stored data in:*** type ***C4*** > in ***Store subscripts in:*** type ***C5*** > click ***OK***

Now we can use Kruskal−Wallis as follows.

Stat > ***Nonparametric*** > ***Kruskal−Wallis*** … > in ***Response:*** type ***C4*** > in ***Factor:*** type ***C5*** > click ***OK***

We will get the output shown in Table 12.22.
Because the p *value of* 0.808 *is larger than* $\alpha = 0.05$, *we cannot reject the null hypothesis.*

TABLE 12.22 Kruskal–Wallis Test.

Kruskal–Wallis test on C4				
C5	*N*	Median	Ave rank	*Z*
1	4	366.0	6.0	−0.34
2	3	360.0	7.7	0.65
3	5	378.0	6.2	−0.24
Overall	12		6.5	
H = 0.43; DF = 2; $p = 0.808$				
* NOTE * one or more small samples				

EXAMPLE 12.7.11

(Friedman test): For the following data, conduct a Friedman test.

93	61	87	75	92	45	99	86	82	74
88	90	76	82	58	74	23	61	60	77
86	56	73	90	47	88	77	18	66	55

Solution
Enter each row of data in ***C1, C2,*** *and* ***C3,*** *respectively. Then stack the data in* ***C1, C2, C3*** *in the following way.*
Manip > ***Stack/Unstack*** > ***Stack Columns*** … > in ***Stack the following columns:*** type ***C1 C2 C3*** > in ***Stored data in:*** type ***C4*** > in ***Store subscripts in:*** type ***C5*** > click ***OK***
In ***C6,*** *enter numbers 1 through 10 in the first 10 rows, enter numbers 1 through 10 in the next 10 rows, and enter numbers 1 through 10 in the following 10 rows. Now we can use the Friedman test as follows.*
Stat > ***Nonparametric*** > ***Friedman*** … > in ***Response:*** type ***C4*** > in ***Treatment: C5*** > in ***Blocks:*** type ***C6*** > click **OK**
We will get the output shown in Table 12.23.

TABLE 12.23 Friedman Test for C4 by C5 Blocked by C6.

C5	*N*	Est median	Sum of ranks
1	10	81.500	24.0
2	10	72.000	20.0
3	10	68.000	16.0
Grand median = 73.833			
S = 3.20; DF = 2; $p = 0.202$.			

Because the p *value is* 0.202, *for any value of* $\alpha < 0.202$, *we cannot reject the null hypothesis.*

12.7.3 SPSS examples

EXAMPLE 12.7.12

(Wilcoxon rank sum test): For the data of Example 12.4.2, use the Wilcoxon rank sum test at the significance level of 0.05 to test the null hypothesis that the two population medians are the same against the alternative hypothesis that the population medians are different. Use an SPSS procedure.

Solution

Because the SPSS pull-down menu does not have the Wilcoxon rank sum test, we will use the Mann–Whitney U-test. The Mann–Whitney U-test is equivalent to the Wilcoxon rank sum test, although we calculate it in a slightly different way. For the same data set, any p *values generated from one test will be identical to those generated from the other. The following gives the steps to follow. Enter tire brands as* ***1*** *to identify brand* ***1*** *and* ***2*** *to identify brand* ***2****, in* ***C1****. Enter the corresponding prices in* ***C2****. Name* ***C1*** *as* ***Brand*** *and* ***C2*** *as* ***Price****. Then click*

Analyze > ***Nonparametric tests*** > ***2 Independent samples*** … > move ***Brand*** to ***Grouping Variable:*** and ***Price*** to ***Test Variable list:*** > click ***Define Groups***… > enter ***1*** in ***Group 1:***, and ***2*** in ***Group 2:*** > click ***continue*** > choose ***Mann–Whitney U*** > ***OK***

We obtained the following output:

Mann–Whitney Test

Ranks

	BRAND	N	Mean rank	Sum of ranks
Price	1.00	6	8.17	49.00
	2.00	8	7.00	56.00
	Total	14		

Test Statistics

	Price
Mann–Whitney U	20.000
Wilcoxon W	56.000
Z	−0.518
Asymp. Sig. (2-tailed)	0.605
Exact Sig. [2*(1-tailed Sig.)]	0.662

(a) *Not corrected for ties.*

(b) *Grouping Variable: BRAND*

In the first table just shown, ranks show the mean ranking of tire brand I and tire brand II. The Mann–Whitney test is used to assess whether the distribution of ranks is statistically significant. Under the null hypothesis, the distribution of ranks *should be the same for both groups. Looking at the second table, the calculated value of the Mann–Whitney U is 20. The value U represents the amount by which the ranks for tire brand I and tire brand II deviate from what we would expect under the null hypothesis. For a 0.05 significance level, we can reject the null hypothesis if the 2-tailed significance (see Asymp. sig in the second table) is less than 0.05. In this case, because Asymp. Sig. (2-tailed) = 0.605, we do reject the null hypothesis.*

EXAMPLE 12.7.13

(Kruskal–Wallis test): For the data of Example 12.5.1, conduct the Kruskal–Wallis test using SPSS.

Solution

Enter insurance companies as ***1*** *to identify company I,* ***2*** *to identify company II, and* ***3*** *to identify company III, in* ***C1****. Enter the corresponding premiums in* ***C2****. Name* ***C1*** *as* ***Company****, and* ***C2*** *as* ***Premium****. Then:*

Analyze > ***Nonparametric Tests*** > ***K Independent samples*** … > move ***Premium*** to ***Test Variable List:*** and ***Company*** to ***Grouping variable:*** > click ***Define Rage*** … > enter ***1*** in ***Minimum***, and ***3*** in ***Maximum*** > click ***Continue*** > click ***Kruskal–Wallis H*** > ***OK***

If we need to do a Friedman test, say for the data of Example 12.7.5, enter each row of data in **C1, C2,** and **C3,** respectively. Then use the following sequence to obtain the appropriate output.

Analyze > ***Nonparametric Tests*** > ***K Related Samples*** … > move each of the three columns to ***Test Variables:*** > check in ***Test Type Friedman*** > ***OK***

12.7.4 SAS examples

To perform the nonparametric tests, use the SAS statement PROC NPAR1WAY. In the procedure, if we include the EXACT statement, the program will compute the exact p value computations for the Wilcoxon rank sum test.

EXAMPLE 12.7.14

(Wilcoxon rank sum test): Comparison of the prices (in dollars) of two brands of similar tires gave the following data.

Tire I:	85	99	100	110	105	87		
Tire II:	67	69	70	93	105	90	110	115

Use the Wilcoxon rank sum test at the significance level of 0.05 to test the null hypothesis that the two population medians are the same against the alternative hypothesis that the population medians are different. Use the SAS procedure.

Solution

We can use the following procedure:

```
options nodate nonumber;
DATA tprice;
INPUT Brand Price @@;
CARDS;
1 85 1 99 1 100 1 110 1 105 1 87
2 67 2 69 2 70 2 93 2 105 2 90 2 110 2 115
;
/* Nonparametric statistics/Wilcoxon Rank-
Sum */
PROC NPAR1WAY DATA = tprice WILCOXON;
CLASS Brand;
VAR Price;
EXACT WILCOXON;
run;
```

EXAMPLE 12.7.15

(Kruskal–Wallis test): For the data of Example 12.7.4, perform the Kruskal–Wallis test using SAS.

Solution

We can use the following code;

```
options nodate nonumber;
DATA insprice;
INPUT Company Price @@;
CARDS;
1 396 1 438 1 336 1 318
2 348 2 360 2 522
3 378 3 330 3 294 3 474 3 432
;
proc npar1way data = insprice;
     class company;
     var Price;
run;
```

Projects for Chapter 12

12A Comparison of Wilcoxon tests with normal approximation

(i) For the Wilcoxon signed rank test, compare the results from the Wilcoxon signed rank test table with the normal approximation using several sets of data of various sample sizes. Also, if the sample size is very small, compare the results from the Wilcoxon signed rank test with a small sample t-test.

(ii) For the Wilcoxon rank sum test, compare the results from the Wilcoxon rank sum test table with the normal approximation using several sets of data (from pairs of samples) of various sample sizes. Also, if the sample sizes are very small, compare the results from the Wilcoxon rank sum test with small sample *t*-test for two samples.

12B Randomness test (Wald–Wolfowitz test)

When we have no control over the way in which the data are selected, it is useful to have a technique for testing whether the sample may be looked on as random. The condition of randomness is essential for all of the analyses explained in this book: that is, whether a sequence of random variables $X_1, \ldots, X_n$ are independent based on a set of observations $x_1, \ldots, x_n$ of these random variables. Here we will give a method based on the number of runs displayed in the sample events. This is a nonparametric procedure. The run test is used to test the randomness of a sample at $100(1 - \alpha)\%$ confidence level.

Given a sequence of two symbols, say H and T, a run is defined as a succession of identical symbols contained between different symbols or none at all. The total number of runs in a sequence of n trials serves as an indication whether the arrangement is random or not. If a sequence contains n_1 symbols of one kind and n_2 symbols of another kind and both n_1 and n_2 are greater than 10 (this is a rule of thumb; for more accuracy we can also take both n_1 and n_2 as greater than 20), then the sampling distribution of the total number of runs, R, has an asymptotic normal distribution with mean

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1$$

and variance

$$\sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

For example, if we have the following symbols

$$\underline{HHH}\ \ \underline{T}\ \ \underline{HH}\ \ \underline{TTT}\ \ \underline{HH}\ \ \underline{TTT}$$

there are six runs indicated by the underlines and $n_1 = 7$ and $n_2 = 8$. If the sample contains numerical data, the run test is used by counting runs above and below the median. Denoting the observations above the median by the letter A and observations below the median by the letter B, we can determine the run as before. For example, if we have data values

$$2 \quad 5 \quad 11 \quad 13 \quad 7 \quad 22 \quad 6 \quad 8 \quad 15 \quad 9$$

then the median is 8.5. Hence, we get the following arrangement of values above and below the median:

$$\underline{BB}\ \ \underline{AA}\ \ \underline{B}\ \ \underline{A}\ \ \underline{BB}\ \ \underline{AA}.$$

Hence, there are six runs with $n_1 = 5$ and $n_2 = 5$.

Now we can formulate the test of randomness as a hypothesis-testing problem as described in the following procedure.

Procedure for test of randomness using the run test

To test

$$H_0 : \text{Arrangement of sample values is random}$$

versus

$$H_a : \text{Data is not random.}$$

1. Compute the median of the sample.
2. Going through the sample values, replace any observation with A if the value is above the median, or B if the value is below the median. Discard any ties.
3. Compute n_1, n_2, and R. Also, compute the mean and variance of R.

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1,$$

and

$$\sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

4. Compute the test statistic:

$$Z = \frac{R - \mu_R}{\sigma_R}.$$

5. Rejection region:

$$|Z| > Z_{\alpha/2}.$$

6. Decision: If the test statistic falls in the rejection region, reject H_0 and conclude that the sample is not random with $(1 - \alpha)$ 100% confidence.
Assumption: $n_1 \geq 10$ and $n_2 \geq 10$.

Note 1: Sometimes the same procedure is used with the median replaced by the mean of the sample. That is, if the observation is above the sample, use A, and if it is below the sample, use B. We use this procedure for large samples. For small sample sizes, to determine the upper and lower critical values, a special table is needed. Some statistical software packages have the ability to use the run test for randomness. For example, in Minitab we can use the following procedure.

Enter the data that we want to test for randomness in C1. Then:

Stat > ***Nonparametric*** > ***Runs Test*** … > In ***variables:*** enter ***C1*** > ***OK***

Default in Minitab is a run test with the mean. If we prefer median, type the value of the median by first clicking ***Above and below:*.**

EXAMPLE 12.B.1

The following table gives the radon concentration in pCi/L obtained from 40 houses in a certain area.

2.9	0.6	13.5	17.1	2.8	3.8	16.0	2.1	6.4	17.2
7.9	0.5	13.7	11.5	2.9	3.6	6.1	8.8	2.2	9.4
15.9	8.8	9.8	11.5	12.3	3.7	8.9	13.0	7.9	11.7
6.2	6.9	12.8	13.7	2.7	3.5	8.3	15.9	5.1	6.0

Test using Minitab (or some other software) whether the data are random at 95% confidence level.

Solution

Running the data with Minitab, we get the following output.

```
radon
                    K = 8.3400
        The observed number of runs = 17
The expected number of runs = 20.9500
        19 Observations above K 21 below
       The test is significant at 0.2046
         Cannot reject at alpha = 0.05
```

Thus the data set is a random sample at 95% confidence level.

Note 2: If the large samples assumption is not satisfied (that is, $n_1 < 10$ and $n_2 < 10$, for more accuracy use 20 instead of 10), then use the total number of runs, R, itself as the test statistic and we can find lower and upper critical values for a given α (from Frieda S. Swed and C. Eisenhart. Tables for testing randomness of grouping in a sequence of alternatives, *Annals of Mathematical Statistics*, 14, 83−86, 1943). We will not be giving this table in this book.

Exercise

Pick a couple of data sets from this book or your own and test for randomness using **(1)** hand calculations, and **(2)** a statistical software package.