CHAPTER\_3\_NONPARAMETRIC\_DISTRIBUTION-FREE\_STATISTICS

CHAPTER OVERVIEW

This chapter explores a wide variety of techniques that are useful when the

underlying assumptions of traditional hypothesis tests are violated or one

wishes to perform a test without making assumptions about the sampled

population.

TOPICS

13.1 INTRODUCTION

13.2 MEASUREMENT SCALES

13.3 THE SIGN TEST

13.4 THE WILCOXON SIGNED-RANK TEST FOR LOCATION

13.5 THE MEDIAN TEST

13.6 THE MANN-WHITNEY TEST

13.7 THE KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST

13.8 THE KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE BY RANKS

13.9 THE FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE BY RANKS

13.10 THE SPEARMAN RANK CORRELATION COEFFICIENT

13.11 NONPARAMETRIC REGRESSION ANALYSIS

13.12 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand the rank transformation and how nonparametric procedures can be used for weak measurement scales.
2. be able to calculate and interpret a wide variety of nonparametric tests commonly used in practice.
3. understand which nonparametric tests may be used in place of traditional parametric statistical tests when various test assumptions are violated.

13.1 INTRODUCTION

Most of the statistical inference procedures we have discussed up to this point are classified

as parametric statistics. One exception is our use of chi-square—as a test of

goodness-of-fit and as a test of independence. These uses of chi-square come under

the heading of nonparametric statistics.

The obvious question now is, “What is the difference?” In answer, let us recall the

nature of the inferential procedures that we have categorized as parametric. In each case,

our interest was focused on estimating or testing a hypothesis about one or more population

parameters. Furthermore, central to these procedures was a knowledge of the functional form

of the population from which were drawn the samples providing the basis for the inference.

An example of a parametric statistical test is the widely used t test. The most common

uses of this test are for testing a hypothesis about a single population mean or the

difference between two population means. One of the assumptions underlying the valid

use of this test is that the sampled population or populations are at least approximately

normally distributed.

As we will learn, the procedures that we discuss in this chapter either are not concerned

with population parameters or do not depend on knowledge of the sampled population.

Strictly speaking, only those procedures that test hypotheses that are not statements

about population parameters are classified as nonparametric, while those that make no

assumption about the sampled population are called distribution-free procedures. Despite

this distinction, it is customary to use the terms nonparametric and distribution-free interchangeably

and to discuss the various procedures of both types under the heading nonparametric

statistics. We will follow this convention.

The above discussion implies the following four advantages of nonparametric

statistics.

1. They allow for the testing of hypotheses that are not statements about population parameter values. Some of the chi-square tests of goodness-of-fit and the tests of independence are examples of tests possessing this advantage.
2. Nonparametric tests may be used when the form of the sampled population is unknown.
3. Nonparametric procedures tend to be computationally easier and consequently more quickly applied than parametric procedures. This can be a desirable feature in certain cases, but when time is not at a premium, it merits a low priority as a criterion for choosing a nonparametric test. Indeed, most statistical software packages now include a wide variety of nonparametric analysis options, making considerations about computation speed unnecessary.
4. Nonparametric procedures may be applied when the data being analyzed consist merely of rankings or classifications. That is, the data may not be based on a measurement scale strong enough to allow the arithmetic operations necessary for carrying out parametric procedures. The subject of measurement scales is discussed in more detail in the next section.

Although nonparametric statistics enjoy a number of advantages, their disadvantages must also be recognized.

1. The use of nonparametric procedures with data that can be handled with a parametric procedure results in a waste of data.
2. The application of some of the nonparametric tests may be laborious for large samples.

13.2 MEASUREMENT SCALES

As was pointed out in the previous section, one of the advantages of nonparametric statistical

procedures is that they can be used with data that are based on a weak measurement

scale. To understand fully the meaning of this statement, it is necessary to know

and understand the meaning of measurement and the various measurement scales most

frequently used. At this point the reader may wish to refer to the discussion of measurement

scales in Chapter 1.

Many authorities are of the opinion that different statistical tests require different

measurement scales. Although this idea appears to be followed in practice, there are alternative

points of view.

Data based on ranks, as will be discussed in this chapter, are commonly encountered

in statistics. We may, for example, simply note the order in which a sample of

subjects complete an event instead of the actual time taken to complete it. More often,

however, we use a rank transformation on the data by replacing, prior to analysis, the

original data by their ranks. Although we usually lose some information by employing

this procedure (for example, the ability to calculate the mean and variance), the transformed

measurement scale allows the computation of most nonparametric statistical procedures.

In fact, most of the commonly used nonparametric procedures, including most

of those presented in this chapter, can be obtained by first applying the rank transformation

and then using the standard parametric procedure on the transformed data instead

of on the original data. For example, if we wish to determine whether two independent

samples differ, we may employ the independent samples t test if the data are approximately

normally distributed. If we cannot make the assumption of normal distributions,

we may, as we shall see in the sections that follow, employ an appropriate nonparametric

test. In lieu of these procedures, we could first apply the rank transformation on the

data and then use the independent samples t test on the ranks. This will provide an

equivalent test to the nonparametric test, and is a useful tool to employ if a desired nonparametric

test is not available in your available statistical software package.

Readers should also keep in mind that other transformations (e.g., taking the logarithm

of the original data) may sufficiently normalize the data such that standard parametric

procedures can be used on the transformed data in lieu of using nonparametric

methods.

13.3 THE SIGN TEST

The familiar t test is not strictly valid for testing (1) the null hypothesis that a population

mean is equal to some particular value, or (2) the null hypothesis that the mean of

a population of differences between pairs of measurements is equal to zero unless the

relevant populations are at least approximately normally distributed. Case 2 will be recognized as a situation that was analyzed by the paired comparisons test in Chapter 7.

When the normality assumptions cannot be made or when the data at hand are ranks

rather than measurements on an interval or ratio scale, the investigator may wish for an

optional procedure. Although the t test is known to be rather insensitive to violations of

the normality assumption, there are times when an alternative test is desirable.

A frequently used nonparametric test that does not depend on the assumptions of

the t test is the sign test. This test focuses on the median rather than the mean as a measure

of central tendency or location. The median and mean will be equal in symmetric

distributions. The only assumption underlying the test is that the distribution of the variable

of interest is continuous. This assumption rules out the use of nominal data.

The sign test gets its name from the fact that pluses and minuses, rather than numerical

values, provide the raw data used in the calculations. We illustrate the use of the sign

test, first in the case of a single sample, and then by an example involving paired samples.

EXAMPLE 13.3.1

Researchers wished to know if instruction in personal care and grooming would improve

the appearance of mentally retarded girls. In a school for the mentally retarded, 10 girls

selected at random received special instruction in personal care and grooming. Two

weeks after completion of the course of instruction the girls were interviewed by a nurse

and a social worker who assigned each girl a score based on her general appearance.

The investigators believed that the scores achieved the level of an ordinal scale. They

felt that although a score of, say, 8 represented a better appearance than a score of 6,

they were unwilling to say that the difference between scores of 6 and 8 was equal to

the difference between, say, scores of 8 and 10; or that the difference between scores of

6 and 8 represented twice as much improvement as the difference between scores of 5

and 6. The scores are shown in Table 13.3.1. We wish to know if we can conclude that

the median score of the population from which we assume this sample to have been

drawn is different from 5.

Sign Test: Paired Data

When the data to be analyzed consist of observations

in matched pairs and the assumptions underlying the t test are not met, or the measurement

scale is weak, the sign test may be employed to test the null hypothesis that the

median difference is 0. An alternative way of stating the null hypothesis is

P\*Xi > Yi) = P(Xi < Yi) = .5

One of the matched scores, say Yi, is subtracted from the other score,Xi. If Yi is

less than Xi the sign of the difference is +, and if Yi is greater than Xi, the sign of the

difference is -. If the median difference is 0, we would expect a pair picked at random

to be just as likely to yield a + as a - when the subtraction is performed. We may state

the null hypothesis, then, as

Ho:P(+) = P(-) = .5

In a random sample of matched pairs, we would expect the number of +’s and -’s to

be about equal. If there are more +’s or more -’s than can be accounted for by chance

alone when the null hypothesis is true, we will entertain some doubt about the truth of

our null hypothesis. By means of the sign test, we can decide how many of one sign

constitutes more than can be accounted for by chance alone.

EXAMPLE 13.3.2

A dental research team wished to know if teaching people how to brush their teeth would

be beneficial. Twelve pairs of patients seen in a dental clinic were obtained by carefully

matching on such factors as age, sex, intelligence, and initial oral hygiene scores. One

member of each pair received instruction on how to brush his or her teeth and on other

oral hygiene matters. Six months later all 24 subjects were examined and assigned an

oral hygiene score by a dental hygienist unaware of which subjects had received the

instruction. A low score indicates a high level of oral hygiene. The results are shown

in Table 13.3.3.

TABLE 13.3.3 Oral Hygiene Scores of 12 Subjects

Receiving Oral Hygiene Instruction and

12 Subjects Not Receiving Instruction

Score

Pair Instructed Not Instructed

Number

1 1.5 2.0

2 2.0 2.0

3 3.5 4.0

4 3.0 2.5

5 3.5 4.0

6 2.5 3.0

7 2.0 3.5

8 1.5 3.0

9 1.5 2.5

10 2.0 2.5

11 3.0 2.5

12 2.0 2.5

Sign Test with “Greater Than” Tables

As has been demonstrated, the

sign test may be used with a single sample or with two samples in which each member

of one sample is matched with a member of the other sample to form a sample of

matched pairs. We have also seen that the alternative hypothesis may lead to either a

one-sided or a two-sided test. In either case we concentrate on the less frequently occurring

sign and calculate the probability of obtaining that few or fewer of that sign.

We use the least frequently occurring sign as our test statistic because the binomial

probabilities in Appendix Table B are “less than or equal to” probabilities. By using

the least frequently occurring sign, we can obtain the probability we need directly from

Table B without having to do any subtracting. If the probabilities in Table B were

“greater than or equal to” probabilities, which are often found in tables of the binomial

distribution, we would use the more frequently occurring sign as our test statistic in

order to take advantage of the convenience of obtaining the desired probability directly

from the table without having to do any subtracting. In fact, we could, in our present

examples, use the more frequently occurring sign as our test statistic, but because Table B

contains “less than or equal to” probabilities we would have to perform a subtraction

operation to obtain the desired probability. As an illustration, consider the last example.

If we use as our test statistic the most frequently occurring sign, it is 9, the number

of minuses. The desired probability, then, is the probability of nine or more minuses,

when and That is, we want

P(k=9 | 11,.5)

However, since Table B contains “less than or equal to” probabilities, we must obtain

this probability by subtraction. That is,

which is the result obtained previously.

Sample Size

We saw in Chapter 5 that when the sample size is large and when

p is close to .5, the binomial distribution may be approximated by the normal distribution.

The rule of thumb used was that the normal approximation is appropriate when

both np and nq are greater than 5. When as was hypothesized in our two examples,

a sample of size 12 would satisfy the rule of thumb. Following this guideline, one

could use the normal approximation when the sign test is used to test the null hypothesis

that the median or median difference is 0 and n is equal to or greater than 12. Since

the procedure involves approximating a continuous distribution by a discrete distribution,

the continuity correction of .5 is generally used. The test statistic then is

(13.3.2)

which is compared with the value of z from the standard normal distribution corresponding

to the chosen level of significance. In Equation 13.3.2, is used when

and k - .5 is used when k Ú n>2.

13.4 THE WILCOXON SIGNED-RANK

TEST FOR LOCATION

Sometimes we wish to test a null hypothesis about a population mean, but for some reason

neither z nor t is an appropriate test statistic. If we have a small sample from

a population that is known to be grossly nonnormally distributed, and the central limit theorem

is not applicable, the z statistic is ruled out. The t statistic is not appropriate because the sampled population does not sufficiently approximate a normal distribution. When confronted

with such a situation we usually look for an appropriate nonparametric statistical

procedure. As we have seen, the sign test may be used when our data consist of a single

sample or when we have paired data. If, however, the data for analysis are measured on at

least an interval scale, the sign test may be undesirable because it would not make full use

of the information contained in the data. A more appropriate procedure might be the

Wilcoxon (1) signed-rank test, which makes use of the magnitudes of the differences

between measurements and a hypothesized location parameter rather than just the signs of

the differences.

Assumptions

The Wilcoxon test for location is based on the following assumptions

about the data.

1. The sample is random.

2. The variable is continuous.

3. The population is symmetrically distributed about its mean

4. The measurement scale is at least interval.

Hypotheses The following are the null hypotheses (along with their alternatives)

that may be tested about some unknown population mean muo.

(a) (b) (c)

When we use the Wilcoxon procedure, we perform the following calculations.

1. Subtract the hypothesized mean muo from each observation xi, to obtain

di = xi / muo

If any xi is equal to the mean, so that di=0 eliminate that di from the calculations and reduce n accordingly.

1. Rank the usable di from the smallest to the largest without regard to the sign of

. That is, consider only the absolute value of the designated when ranking

them. If two or more of the are equal, assign each tied value the mean of

the rank positions the tied values occupy. If, for example, the three smallest

are all equal, place them in rank positions 1, 2, and 3, but assign each a rank of

3. Assign each rank the sign of the that yields that rank.

4. Find the sum of the ranks with positive signs, and the sum of the ranks

with negative signs.

The Test Statistic

The Wilcoxon test statistic is either T+ or T- depending on

the nature of the alternative hypothesis. If the null hypothesis is true, that is, if the true

population mean is equal to the hypothesized mean, and if the assumptions are met, the

probability of observing a positive difference di = x i - m0 of a given magnitude is equal

to the probability of observing a negative difference of the same magnitude. Then, in

repeated sampling, when the null hypothesis is true and the assumptions are met, the

expected value of T+ is equal to the expected value of T- . We do not expect T+ and T-

computed from a given sample to be equal. However, when H0 is true, we do not expect

a large difference in their values. Consequently, a sufficiently small value of T+ or a suf-

ficiently small value of T- will cause rejection of H0.

When the alternative hypothesis is two-sided 1m Z m02, either a sufficiently small

value of T+ or a sufficiently small value of T- will cause us to reject H0: m = m0. The

test statistic, then, is T+ or T- , whichever is smaller. To simplify notation, we call the

smaller of the two T.

When H0: m Ú m0 is true, we expect our sample to yield a large value of T+ . There-

fore, when the one-sided alternative hypothesis states that the true population mean is

less than the hypothesized mean 1m 6 m02, a sufficiently small value of T+ will cause

rejection of H0, and T+ is the test statistic.

When H0: m … m0 is true, we expect our sample to yield a large value of T- . There-

fore, for the one-sided alternative HA: m 7 m0, a sufficiently small value of T- will cause

rejection of H0 and T- is the test statistic.

Critical Values Critical values of the Wilcoxon test statistic are given in Appendix

Table K. Exact probability levels (P) are given to four decimal places for all possible rank

totals (T ) that yield a different probability level at the fourth decimal place from .0001 up

through .5000. The rank totals (T ) are tabulated for all sample sizes from n = 5 through

n = 30. The following are the decision rules for the three possible alternative hypotheses:

(a) HA: m Z m0. Reject H0 at the a level of significance if the calculated T is smaller

than or equal to the tabulated T for n and preselected a>2. Alternatively, we may

enter Table K with n and our calculated value of T to see whether the tabulated P

associated with the calculated T is less than or equal to our stated level of signif-

icance. If so, we may reject H0.

(b) HA: m 6 m0. Reject H0 at the a level of significance if T+ is less than or equal to

the tabulated T for n and preselected a.

(c) HA: m 7 m0. Reject H0 at the a level of significance if T- is less than or equal to

the tabulated T for n and preselected a.

EXAMPLE 13.4.1

Cardiac output (liters/minute) was measured by thermodilution in a simple random sam-

ple of 15 postcardiac surgical patients in the left lateral position. The results were as

follows:

4.91

5.98

4.10

3.14

6.74

3.23

7.27

5.80

7.42

6.17

7.50

5.39

6.56

5.77

4.64

We wish to know if we can conclude on the basis of these data that the population mean

is different from 5.05.

Wilcoxon Matched-Pairs Signed-Ranks Test The Wilcoxon test

may be used with paired data under circumstances in which it is not appropriate to

use the paired-comparisons t test described in Chapter 7. In such cases obtain each of

the n di values, the difference between each of the n pairs of measurements. If we

let mD = the mean of a population of such differences, we may follow the procedure

described above to test any one of the following null hypotheses: H0: mD = 0, H0: mD Ú 0,

and H0: mD Ú 0.

13.5

THE MEDIAN TEST

A nonparametric procedure that may be used to test the null hypothesis that two inde-

pendent samples have been drawn from populations with equal medians is the median

test. The test, attributed mainly to Mood (2) and Westenberg (3), is also discussed by

Brown and Mood (4).

We illustrate the procedure by means of an example.

EXAMPLE 13.5.1

Do urban and rural male junior high school students differ with respect to their level of

mental health?

Handling Values Equal to the Median Sometimes one or more observed

values will be exactly equal to the common median and, hence, will fall neither above

nor below it. We note that if n1 + n 2 is odd, at least one value will always be exactly

equal to the median. This raises the question of what to do with observations of this

kind. One solution is to drop them from the analysis if n1 + n 2 is large and there are

only a few values that fall at the combined median. Or we may dichotomize the scores

into those that exceed the median and those that do not, in which case the observations

that equal the median will be counted in the second category.

Median Test Extension The median test extends logically to the case where

it is desired to test the null hypothesis that k Ú 3 samples are from populations with

equal medians. For this test a 2 \* k contingency table may be constructed by using the

frequencies that fall above and below the median computed from combined samples. If

conditions as to sample size and expected frequencies are met, X 2 may be computed and

compared with the critical x2 with k - 1 degrees of freedom.

13.6

THE MANN–WHITNEY TEST

The median test discussed in the preceding section does not make full use of all the

information present in the two samples when the variable of interest is measured on at

least an ordinal scale. Reducing an observation’s information content to merely that of

whether or not it falls above or below the common median is a waste of information. If,

for testing the desired hypothesis, there is available a procedure that makes use of more

of the information inherent in the data, that procedure should be used if possible. Such

a nonparametric procedure that can often be used instead of the median test is the

Mann–Whitney test (5), sometimes called the Mann–Whitney–Wilcoxon test. Since this

test is based on the ranks of the observations, it utilizes more information than does the

median test.

Assumptions The assumptions underlying the Mann–Whitney test are as follows:

1. The two samples, of size n and m, respectively, available for analysis have been

independently and randomly drawn from their respective populations.

2. The measurement scale is at least ordinal.

3. The variable of interest is continuous.

4. If the populations differ at all, they differ only with respect to their medians.

Hypotheses When these assumptions are met we may test the null hypothesis that

the two populations have equal medians against either of the three possible alternatives:

(1) the populations do not have equal medians (two-sided test), (2) the median of popu-

lation 1 is larger than the median of population 2 (one-sided test), or (3) the median of

population 1 is smaller than the median of population 2 (one-sided test). If the two pop-

ulations are symmetric, so that within each population the mean and median are the same,

the conclusions we reach regarding the two population medians will also apply to the two

population means. The following example illustrates the use of the Mann–Whitney test.

EXAMPLE 13.6.1

A researcher designed an experiment to assess the effects of prolonged inhalation of cad-

mium oxide. Fifteen laboratory animals served as experimental subjects, while 10 similar

animals served as controls. The variable of interest was hemoglobin level following the

experiment. The results are shown in Table 13.6.1. We wish to know if we can conclude

that prolonged inhalation of cadmium oxide reduces hemoglobin level.

Large-Sample Approximation When either n or m is greater than 20 we

cannot use Appendix Table L to obtain critical values for the Mann–Whitney test. When

this is the case we may compute

z =

T - mn>2

1nm 1n + m + 12>12

(13.6.2)

and compare the result, for significance, with critical values of the standard normal

distribution.

Mann–Whitney Statistic and the Wilcoxon Statistic As was noted

at the beginning of this section, the Mann–Whitney test is sometimes referred to as the

Mann–Whitney-Wilcoxon test. Indeed, many computer packages give the test value of

both the Mann–Whitney test (U) and the Wilcoxon test (W). These two tests are alge-

braically equivalent tests, and are related by the following equality when there are no

ties in the data:

U + W =

m1m + 2n + 12

(13.6.3)

13.7 THE KOLMOGOROV–SMIRNOV

GOODNESS-OF-FIT TEST

When one wishes to know how well the distribution of sample data conforms to some

theoretical distribution, a test known as the Kolmogorov–Smirnov goodness-of-fit test

provides an alternative to the chi-square goodness-of-fit test discussed in Chapter 12. The

test gets its name from A. Kolmogorov and N. V. Smirnov, two Russian mathematicians

who introduced two closely related tests in the 1930s.

Kolmogorov’s work (6) is concerned with the one-sample case as discussed here.

Smirnov’s work (7) deals with the case involving two samples in which interest centers

on testing the hypothesis that the distributions of the two-parent populations are iden-

tical. The test for the first situation is frequently referred to as the Kolmogorov–Smirnov

one-sample test. The test for the two-sample case, commonly referred to as the

Kolmogorov–Smirnov two-sample test, will not be discussed here.

The Test Statistic In using the Kolmogorov–Smirnov goodness-of-fit test, a

comparison is made between some theoretical cumulative distribution function, FT 1x2,

and a sample cumulative distribution function, FS1x2. The sample is a random sample

from a population with unknown cumulative distribution function F1x2. It will be recalled

(Section 4.2) that a cumulative distribution function gives the probability that X is equal

to or less than a particular value, x. That is, by means of the sample cumulative distri-

bution function, FS1x2, we may estimate P1X … x2. If there is close agreement between

the theoretical and sample cumulative distributions, the hypothesis that the sample was

drawn from the population with the specified cumulative distribution function, FT1x2, is

supported. If, however, there is a discrepancy between the theoretical and observed cumu-

lative distribution functions too great to be attributed to chance alone, when H0 is true,

the hypothesis is rejected.

The difference between the theoretical cumulative distribution function, FT1x2, and

the sample cumulative distribution function, FS1x2, is measured by the statistic D, which

is the greatest vertical distance between FS1x2 and FT1x2. When a two-sided test is appro-

priate, that is, when the hypotheses are

the test statistic is

H0: F1x2 = FT1x2 for all x from - q to + q

HA: F1x2 Z FT1x2 for at least one x

D = sup ƒ FS1x2 - FT 1x2 ƒ

x

(13.7.1)

which is read, “D equals the supremum (greatest), over all x, of the absolute value of the

difference FS1X 2 minus FT1X 2.”

The null hypothesis is rejected at the a level of significance if the computed

value of D exceeds the value shown in Appendix Table M for 1 - a (two-sided) and

the sample size n.

Assumptions The assumptions underlying the Kolmogorov–Smirnov test include

the following:

1. The sample is a random sample.

2. The hypothesized distribution FT1x 2 is continuous.

When values of D are based on a discrete theoretical distribution, the test is con-

servative. When the test is used with discrete data, then, the investigator should bear in

mind that the true probability of committing a type I error is at most equal to a, the

stated level of significance. The test is also conservative if one or more parameters have

to be estimated from sample data.

EXAMPLE 13.7.1

Fasting blood glucose determinations made on 36 nonobese, apparently healthy, adult males

are shown in Table 13.7.1. We wish to know if we may conclude that these data are not

from a normally distributed population with a mean of 80 and a standard deviation of 6.

StatXact is often used for nonparametric statistical analysis. This particular soft-

ware program has a nonparametric module that contains nearly all of the commonly

used nonparametric tests, and many less common, but useful, procedures as well. Com-

puter analysis using StatXact for the data in Example 13.7.1 is shown in Figure 13.7.2.

Note that it provides the test statistic of D " 0.156 and the exact two-sided p value

of .3447.

A Precaution The reader should be aware that in determining the value of D, it

is not always sufficient to compute and choose from the possible values of

ƒ FS1x2 - FT1x2 ƒ . The largest vertical distance between FS1x2 and FT1x2 may not occur

at an observed value, x, but at some other value of X. Such a situation is illustrated in

Figure 13.7.3. We see that if only values of ƒ FS1x2 - FT 1x2 ƒ at the left endpoints of

the horizontal bars are considered, we would incorrectly compute D as ƒ .2 - .4 ƒ = .2.

One can see by examining the graph, however, that the largest vertical distance between

FS1x2 and FT1x2 occurs at the right endpoint of the horizontal bar originating at the

point corresponding to x = .4, and the correct value of D is ƒ .5 - .2 ƒ = .3.

One can determine the correct value of D algebraically by computing, in addition

to the differences ƒ FS1x2 - FT1x2 ƒ , the differences ƒ FS 1x i -12 - FT1x i2 ƒ for all values of

i = 1, 2, . . . , r + 1, where r = the number of different values of x and FS 1x 02 = 0.

The correct value of the test statistic will then be

D = maximum 5maximum3 ƒ FS1x i2 - FT1x i2 ƒ , ƒ FS1x i -12 - FT1x i2 ƒ 46 (13.7.2)

Advantages and Disadvantages The following are some important

points of comparison between the Kolmogorov–Smirnov and the chi-square goodness-

of-fit tests.

1. The Kolmogorov–Smirnov test does not require that the observations be grouped as

is the case with the chi-square test. The consequence of this difference is that the

Kolmogorov–Smirnov test makes use of all the information present in a set of data.

2. The Kolmogorov–Smirnov test can be used with any size sample. It will be recalled

that certain minimum sample sizes are required for the use of the chi-square test.

3. As has been noted, the Kolmogorov–Smirnov test is not applicable when parame-

ters have to be estimated from the sample. The chi-square test may be used in these

situations by reducing the degrees of freedom by 1 for each parameter estimated.

4. The problem of the assumption of a continuous theoretical distribution has already

been mentioned.

13.8 THE KRUSKAL–WALLIS ONE-WAY

ANALYSIS OF VARIANCE BY RANKS

In Chapter 8 we discuss how one-way analysis of variance may be used to test the null

hypothesis that several population means are equal. When the assumptions underlying

this technique are not met, that is, when the populations from which the samples are

drawn are not normally distributed with equal variances, or when the data for analysis

consist only of ranks, a nonparametric alternative to the one-way analysis of variance

may be used to test the hypothesis of equal location parameters. As was pointed out in

Section 13.5, the median test may be extended to accommodate the situation involving

more than two samples. A deficiency of this test, however, is the fact that it uses only a

small amount of the information available. The test uses only information as to whether

or not the observations are above or below a single number, the median of the combined

samples. The test does not directly use measurements of known quantity. Several

nonparametric analogs to analysis of variance are available that use more information by

taking into account the magnitude of each observation relative to the magnitude of every

other observation. Perhaps the best known of these procedures is the Kruskal–Wallis one-

way analysis of variance by ranks (8).

The Kruskal–Wallis Procedure The application of the test involves the

following steps.

1. The n 1, n 2, . . . , n k observations from the k samples are combined into a single

series of size n and arranged in order of magnitude from smallest to largest.

The observations are then replaced by ranks from 1, which is assigned to the small-

est observation, to n, which is assigned to the largest observation. When two or

more observations have the same value, each observation is given the mean of the

ranks for which it is tied.

2. The ranks assigned to observations in each of the k groups are added separately to

give k rank sums.

3. The test statistic

k R

j

12

H =

- 31n + 12

a

n

n1n + 12 j =1 j

2

(13.8.1)

is computed. In Equation 13.8.1,

k = the number of samples

n j = the number of observations in the jth sample

n = the number of observations in all samples combined

R j = the sum of the ranks in the jth sample

4. When there are three samples and five or fewer observations in each sample, the

significance of the computed H is determined by consulting Appendix Table N.

When there are more than five observations in one or more of the samples, H is

compared with tabulated values of x 2 with k - 1 degrees of freedom.

EXAMPLE 13.8.1

In a study of pulmonary effects on guinea pigs, Lacroix et al. (A-7) exposed ovalbu-

min (OA)-sensitized guinea pigs to regular air, benzaldehyde, or acetaldehyde. At the

end of exposure, the guinea pigs were anesthetized and allergic responses were

assessed in bronchoalveolar lavage (BAL). One of the outcome variables examined

was the count of eosinophil cells, a type of white blood cell that can increase with

allergies. Table 13.8.1 gives the eosinophil cell count 1\*1062 for the three treatment

groups.

Can we conclude that the three populations represented by the three samples dif-

fer with respect to eosinophil cell count? We can so conclude if we can reject the null

hypothesis that the three populations do not differ in eosinophil cell count.

Ties When ties occur among the observations, we may adjust the value of H by

dividing it by

1 -

gT

3

n - n

(13.8.2)

where T = t 3 - t. The letter t is used to designate the number of tied observations in a

group of tied values. In our example there are no groups of tied values but, in general,

there may be several groups of tied values resulting in several values of T.

The effect of the adjustment for ties is usually negligible. Note also that the effect

of the adjustment is to increase H, so that if the unadjusted H is significant at the cho-

sen level, there is no need to apply the adjustment.

More than Three Samples/Large Samples Now let us illustrate the

procedure when there are more than three samples and at least one of the n j is greater

than 5.

EXAMPLE 13.8.2

Table 13.8.3 shows the net book value of equipment capital per bed for a sample of

hospitals from each of five types of hospitals. We wish to determine, by means of the

Kruskal–Wallis test, if we can conclude that the average net book value of equipment

capital per bed differs among the five types of hospitals. The ranks of the 41 values,

along with the sum of ranks for each sample, are shown in the table.

13.9 THE FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE BY RANKS

Just as we may on occasion have need of a nonparametric analog to the parametric one-

way analysis of variance, we may also find it necessary to analyze the data in a two-way

classification by nonparametric methods analogous to the two-way analysis of variance.

Such a need may arise because the assumptions necessary for parametric analysis of

variance are not met, because the measurement scale employed is weak, or because results

are needed in a hurry. A test frequently employed under these circumstances is the Fried-

man two-way analysis of variance by ranks (9, 10). This test is appropriate whenever the

data are measured on, at least, an ordinal scale and can be meaningfully arranged in a

two-way classification as is given for the randomized block experiment discussed in Chap-

ter 8. The following example illustrates this procedure.

**EXAMPLE 13.9.1**

A physical therapist conducted a study to compare three models of low-volt electrical

stimulators. Nine other physical therapists were asked to rank the stimulators in order of

preference. A rank of 1 indicates first preference. The results are shown in Table 13.9.1.

We wish to know if we can conclude that the models are not preferred equally.

Ties When the original data consist of measurements on an interval or a ratio scale

instead of ranks, the measurements are assigned ranks based on their relative magni-

tudes within blocks. If ties occur, each value is assigned the mean of the ranks for which

it is tied.

Large Samples When the values of k and/or n exceed those given in Table O,

the critical value of x2r is obtained by consulting the x2 table (Table F) with the chosen

a and k - 1 degrees of freedom.

EXAMPLE 13.9.2

Table 13.9.2 shows the responses, in percent decrease in salivary flow, of 16 experimental

animals following different dose levels of atropine. The ranks (in parentheses) and the sum

of the ranks are also given in the table. We wish to see if we may conclude that the dif-

ferent dose levels produce different responses. That is, we wish to test the null hypothesis

of no difference in response among the four dose levels.

13.10 THE SPEARMAN RANK CORRELATION COEFFICIENT

Several nonparametric measures of correlation are available to the researcher. Of these

a frequently used procedure that is attractive because of the simplicity of the calculations involved is due to Spearman (11). The measure of correlation computed by this

method is called the Spearman rank correlation coefficient and is designated by rs . This

procedure makes use of the two sets of ranks that may be assigned to the sample values

of X and Y, the independent and continuous variables of a bivariate distribution.

Hypotheses The usually tested hypotheses and their alternatives are as follows:

(a) H0: X and Y are mutually independent.

HA: X and Y are not mutually independent.

(b) H0: X and Y are mutually independent.

HA: There is a tendency for large values of X and large values of Y to be paired

together.

(c) H0: X and Y are mutually independent.

HA: There is a tendency for large values of X to be paired with small values of Y.

The hypotheses specified in (a) lead to a two-sided test and are used when it is desired

to detect any departure from independence. The one-sided tests indicated by (b) and (c)

are used, respectively, when investigators wish to know if they can conclude that the vari-

ables are directly or inversely correlated.

**The Procedure** The hypothesis-testing procedure involves the following steps.

1. Rank the values of X from 1 to n (numbers of pairs of values of X and Y in the

sample). Rank the values of Y from 1 to n.

2. Compute di for each pair of observations by subtracting the rank of Yi from the

rank of X i .

3. Square each di and compute gd i2, the sum of the squared values.

4. Compute

rs = 1 -

6gd i2

n1n 2 - 12

(13.10.1)

5. If n is between 4 and 30, compare the computed value of rs with the critical values,

r \*,

s of Appendix Table P. For the two-sided test, H0 is rejected at the a significance

level if rs is greater than r \*s or less than -r \*,

s where r \*

s is at the intersection of the

column headed a>2 and the row corresponding to n. For the one-sided test with HA

specifying direct correlation, H0 is rejected at the a significance level if rs is greater

than r \*s for a and n. The null hypothesis is rejected at the a significance level in the

other one-sided test if rs is less than -r\*s for a and n.

6. If n is greater than 30, one may compute

z = rs 2n - 1

(13.10.2)

and use Appendix Table D to obtain critical values.

7. Tied observations present a problem. The use of Table P is strictly valid only when

the data do not contain any ties (unless some random procedure for breaking ties

is employed). In practice, however, the table is frequently used after some other

method for handling ties has been employed. If the number of ties is large, the fol-

lowing correction for ties may be employed:

T =

t3 - t

12

(13.10.3)

where t = the number of observations that are tied for some particular rank. When

this correction factor is used, rs is computed from

rs =

gx 2 + gy 2 - gd i2

22gx 2 gy 2

(13.10.4)

instead of from Equation 13.10.1.

In Equation 13.10.4,

g x2 =

n3 - n

- gTx

12

n3 - n

gy2 =

- g Ty

12

Tx = the sum of the values of T for the various tied ranks in X

Ty = the sum of the values of T for the various tied ranks in Y

Most authorities agree that unless the number of ties is excessive, the correction

makes very little difference in the value of rs. When the number of ties is small,

we can follow the usual procedure of assigning the tied observations the mean of

the ranks for which they are tied and proceed with steps 2 to 6.

EXAMPLE 13.10.1

In a study of the relationship between age and the EEG, data were collected on 20 sub-

jects between ages 20 and 60 years. Table 13.10.1 shows the age and a particular EEG

output value for each of the 20 subjects. The investigator wishes to know if it can be

concluded that this particular EEG output is inversely correlated with age.

Let us now illustrate the procedure for a sample with n > 30 and some tied

observations.

EXAMPLE 13.10.2

In Table 13.10.3 are shown the ages and concentrations (ppm) of a certain mineral in

the tissue of 35 subjects on whom autopsies were performed as part of a large research

project.

The ranks, di, di2, and sum(d i2) are shown in Table 13.10.4. Let us test, at the .05 level

of significance, the null hypothesis that X and Y are mutually independent against the

two-sided alternative that they are not mutually independent.

13.11 NONPARAMETRIC REGRESSION ANALYSIS

When the assumptions underlying simple linear regression analysis as discussed in Chap-

ter 9 are not met, we may employ nonparametric procedures. In this section we present

estimators of the slope and intercept that are easy-to-calculate alternatives to the least-

squares estimators described in Chapter 9.

Theil’s Slope Estimator

Theil (12) proposes a method for obtaining a point

estimate of the slope coefficient b. We assume that the data conform to the classic

regression model

yi = b 0 + b 1x 1 + Pi,

i = 1, . . . , n

where the xi are known constants, b0 and b1 are unknown parameters, and yi is an

observed value of the continuous random variable Y at i. For each value of xi, we assume

a subpopulation of Y values, and the ei are mutually independent. The x i are all distinct

(no ties), and we take x1< x2< . . . < xn.

The data consist of n pairs of sample observations, (x1, y1), (x2, y2), . . . , (xn , yn),

where the ith pair represents measurements taken on the ith unit of association.

To obtain Theil’s estimator of b1 we first form all possible sample slopes Sij =

(yj - yi)>(xj - xi), where i < j. There will be N = nC2 values of Sij . The estimator of

b1 which we designate by gbi is the median of Sij values. That is,

gb1 = median{Sij}

(13.11.1)

The following example illustrates the calculation of gb1.

EXAMPLE 13.11.1

In Table 13.11.1 are the plasma testosterone (ng/ml) levels (Y ) and seminal citric acid

(mg/ml) levels in a sample of eight adult males. We wish to compute the estimate of

the population regression slope coefficient by Theil’s method.

An Estimator of the Intercept Coefficient

Dietz (13) recommends

two intercept estimators. The first, designated (gbp)1,M is the median of the n terms

yi - gbi i in which gb1 is the Theil estimator. It is recommended when the researcher

is not willing to assume that the error terms are symmetric about 0. If the researcher is

willing to assume a symmetric distribution of error terms, Dietz recommends the estimator (gb0)2,M which is the median of the n(n + 1)/2 pairwise averages of the yi - gb1xi

terms. We illustrate the calculation of each in the following example.

EXAMPLE 13.11.2

Refer to Example 13.11.1. Let us compute aN 1,M and aN 2,M from the data on testosterone

and citric acid levels.