BIOSTATISTICS\_Daniel\_C13

CHAPTER13

NONPARAMETRIC AND

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.

Este capítulo explora una amplia variedad de técnicas que son útiles cuando se violan los supuestos subyacentes de las pruebas de hipótesis tradicionales o se desea realizar una prueba sin hacer suposiciones sobre la población muestreada.

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.

1. Comprender la transformación de rango y cómo se pueden utilizar procedimientos no paramétricos para escalas de medición débiles.

2. be able to calculate and interpret a wide variety of nonparametric tests commonly used in practice.

2. Ser capaz de calcular e interpretar una amplia variedad de pruebas no paramétricas comúnmente utilizadas en la práctica.

3. understand which nonparametric tests may be used in place of traditional parametric statistical tests when various test assumptions are violated.

3. comprender qué pruebas no paramétricas se pueden utilizar en lugar de las pruebas estadísticas paramétricas tradicionales cuando se violan varios supuestos de prueba.

**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.

La mayoría de los procedimientos de inferencia estadística que hemos analizado hasta ahora se clasifican como estadística paramétrica. Una excepción es nuestro uso de la chi-cuadrado, como prueba de bondad de ajuste y de independencia. Estos usos de la chi-cuadrado se clasifican como estadística no paramétrica.

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.

La pregunta obvia ahora es: "¿Cuál es la diferencia?". Para responder, recordemos la naturaleza de los procedimientos inferenciales que hemos categorizado como paramétricos. En cada caso, nuestro interés se centró en estimar o contrastar una hipótesis sobre uno o más parámetros poblacionales. Además, era fundamental para estos procedimientos conocer la forma funcional de la población de la que se extrajeron las muestras que sirvieron de base para la inferencia.

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.

Un ejemplo de prueba estadística paramétrica es la ampliamente utilizada prueba t. Los usos más comunes de esta prueba son para contrastar una hipótesis sobre una sola media poblacional o la diferencia entre dos medias poblacionales. Uno de los supuestos que sustentan el uso válido de esta prueba es que la población o poblaciones muestreadas tienen una distribución al menos aproximadamente normal.

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 non-parametric statistics. We will follow this convention.

Como aprenderemos, los procedimientos que analizamos en este capítulo no se refieren a parámetros poblacionales ni dependen del conocimiento de la población muestreada. En sentido estricto, solo los procedimientos que contrastan hipótesis que no son afirmaciones sobre parámetros poblacionales se clasifican como no paramétricos, mientras que aquellos que no hacen suposiciones sobre la población muestreada se denominan procedimientos de distribución libre. A pesar de esta distinción, se suele usar los términos no paramétrico y de distribución libre indistintamente, y se describen los diversos procedimientos de ambos tipos bajo el título de estadística no paramétrica. Seguiremos esta convención.

The above discussion implies the following four advantages of nonparametric statistics.

La discusión anterior implica las siguientes cuatro ventajas de las estadísticas no paramétricas.

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.

1. Permiten la comprobación de hipótesis que no son afirmaciones sobre los valores de los parámetros poblacionales. Algunas pruebas de chi-cuadrado de bondad de ajuste y las pruebas de independencia son ejemplos de pruebas que poseen esta ventaja.

2. Nonparametric tests may be used when the form of the sampled population is unknown.

2. Se pueden utilizar pruebas no paramétricas cuando se desconoce la forma de la población muestreada.

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.

3. Los procedimientos no paramétricos tienden a ser computacionalmente más sencillos y, en consecuencia, más rápidos de aplicar que los paramétricos. Esto puede ser una característica deseable en ciertos casos, pero cuando el tiempo no es un factor clave, no se prioriza como criterio para elegir una prueba no paramétrica. De hecho, la mayoría de los paquetes de software estadístico incluyen actualmente una amplia variedad de opciones de análisis no paramétrico, lo que hace innecesarias las consideraciones sobre la velocidad de cálculo.

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.

4. Se pueden aplicar procedimientos no paramétricos cuando los datos analizados consisten simplemente en clasificaciones. Es decir, es posible que los datos no se basen en una escala de medición lo suficientemente robusta como para permitir las operaciones aritméticas necesarias para llevar a cabo procedimientos paramétricos. El tema de las escalas de medición se aborda con más detalle en la siguiente sección.

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

Si bien las estadísticas no paramétricas presentan una serie de ventajas, también deben reconocerse sus desventajas.

1. The use of nonparametric procedures with data that can be handled with a parametric procedure results in a waste of data.

1. El uso de procedimientos no paramétricos con datos que pueden manejarse con un procedimiento paramétrico resulta en un desperdicio de datos.

2. The application of some of the nonparametric tests may be laborious for large samples.

2. La aplicación de algunas pruebas no paramétricas puede resultar laboriosa para muestras grandes.

**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.

Como se señaló en la sección anterior, una de las ventajas de los procedimientos estadísticos no paramétricos es que pueden utilizarse con datos basados ​​en una escala de medición débil. Para comprender plenamente esta afirmación, es necesario conocer y comprender el significado de la medición y las diversas escalas de medición más utilizadas. En este punto, el lector puede consultar la discusión sobre las escalas de medición en el capítulo 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.

Muchos expertos opinan que las diferentes pruebas estadísticas requieren distintas escalas de medición. Si bien esta idea parece seguirse en la práctica, existen perspectivas alternativas.

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.

Los datos basados ​​en rangos, como se analizará en este capítulo, son comunes en estadística. Por ejemplo, podemos simplemente registrar el orden en que una muestra de sujetos completa un evento en lugar del tiempo real empleado. Sin embargo, con mayor frecuencia, utilizamos una transformación de rangos en los datos reemplazando, antes del análisis, los datos originales por sus rangos. Aunque este procedimiento suele perder información (por ejemplo, la capacidad de calcular la media y la varianza), la escala de medición transformada permite el cálculo de la mayoría de los procedimientos estadísticos no paramétricos. De hecho, la mayoría de los procedimientos no paramétricos de uso común, incluyendo la mayoría de los presentados en este capítulo, pueden obtenerse aplicando primero la transformación de rangos y luego utilizando el procedimiento paramétrico estándar en los datos transformados en lugar de en los datos originales.

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 non-parametric test is not available in your available statistical software package.

Por ejemplo, si deseamos determinar si dos muestras independientes difieren, podemos emplear la prueba t para muestras independientes si los datos se distribuyen aproximadamente de forma normal. Si no podemos asumir distribuciones normales, podemos, como veremos en las secciones siguientes, emplear una prueba no paramétrica adecuada. En lugar de estos procedimientos, podríamos aplicar primero la transformación de rangos a los datos y luego usar la prueba t para muestras independientes en los rangos. Esto proporcionará una prueba equivalente a la prueba no paramétrica y es una herramienta útil si la prueba no paramétrica deseada no está disponible en su programa estadístico.

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.

Los lectores también deben tener en cuenta que otras transformaciones (por ejemplo, tomar el logaritmo de los datos originales) pueden normalizar suficientemente los datos como para que se puedan utilizar procedimientos paramétricos estándar en los datos transformados en lugar de utilizar métodos no paramétricos.

**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.

La conocida prueba t no es estrictamente válida para probar (1) la hipótesis nula de que la media poblacional es igual a un valor particular, ni (2) la hipótesis nula de que la media de una población de diferencias entre pares de mediciones es igual a cero, a menos que las poblaciones relevantes tengan una distribución al menos aproximadamente normal. El caso 2 se considerará una situación analizada mediante la prueba de comparaciones por pares en el capítulo 7. Cuando no se pueden realizar los supuestos de normalidad o cuando los datos disponibles son rangos en lugar de mediciones en una escala de intervalos o razones, el investigador puede optar por un procedimiento opcional. Aunque se sabe que la prueba t es bastante insensible a las violaciones del supuesto de normalidad, en ocasiones es conveniente una prueba alternativa.

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.

Una prueba no paramétrica de uso frecuente que no depende de los supuestos de la prueba t es la prueba de signos. Esta prueba se centra en la mediana, en lugar de la media, como medida de tendencia central o ubicación. La mediana y la media serán iguales en distribuciones simétricas. El único supuesto subyacente a la prueba es que la distribución de la variable de interés es continua. Este supuesto descarta el uso de datos nominales.

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.

La prueba del signo recibe su nombre porque los signos positivos y negativos, en lugar de los valores numéricos, proporcionan los datos brutos utilizados en los cálculos. Ilustramos el uso de la prueba del signo, primero con una sola muestra y luego con un ejemplo con muestras pareadas.

**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.

Los investigadores deseaban saber si la instrucción en cuidado y aseo personal mejoraría la apariencia de niñas con discapacidad intelectual. En una escuela para personas con discapacidad intelectual, 10 niñas seleccionadas al azar recibieron instrucción especial en cuidado y aseo personal. Dos semanas después de completar el curso, una enfermera y un trabajador social entrevistaron a cada una de ellas, quienes les asignaron una puntuación basada en su apariencia general. Los investigadores consideraron que las puntuaciones alcanzaban el nivel de una escala ordinal. Consideraron que, si bien una puntuación de, por ejemplo, 8 representaba una mejor apariencia que una de 6, no estaban dispuestos a afirmar que la diferencia entre las puntuaciones de 6 y 8 fuera igual a la diferencia entre, por ejemplo, las puntuaciones de 8 y 10; ni que la diferencia entre las puntuaciones de 6 y 8 representara el doble de mejora que la diferencia entre las puntuaciones de 5 y 6. Las puntuaciones se muestran en la Tabla 13.3.1. Deseamos saber si podemos concluir que la puntuación mediana de la población de la que suponemos que se extrajo esta muestra es diferente de 5.

**Solution:**

1. Data. See problem statement.

2. Assumptions. We assume that the measurements are taken on a con-

tinuous variable.

3. Hypotheses.

The population median is 5.

The population median is not 5.

Let a = .05.

HA:

H0:

**4. Test statistic.** The test statistic for the sign test is either the observed number of plus signs or the observed number of minus signs. The nature of the alternative hypothesis determines which of these test statistics is appropriate. In a given test, any one of the following alternative hypotheses is possible:

El estadístico de prueba para la prueba de signos es el número observado de signos positivos o negativos. La naturaleza de la hipótesis alternativa determina cuál de estos estadísticos de prueba es apropiado. En una prueba dada, cualquiera de las siguientes hipótesis alternativas es posible:

If the alternative hypothesis is

a sufficiently small number of minus signs causes rejection of The test statistic is the number of minus signs. Similarly, if the alternative hypothesis is

a sufficiently small number of plus signs causes rejection of The test statistic is the number of plus signs. If the alternative hypothesis is

either a sufficiently small number of plus signs or a sufficiently small number of minus signs causes rejection of the null hypothesis. We may take as the test statistic the less frequently occurring sign.

**5. Distribution of test statistic.** As a first step in determining the nature of the test statistic, let us examine the data in Table 13.3.1 to determine which scores lie above and which ones lie below the hypothesized median of 5. If we assign a plus sign to those scores that lie above the hypothesized median and a minus to those that fall below, we have the results shown in Table 13.3.2.

Como primer paso para determinar la naturaleza de la estadística de prueba, examinemos los datos de la Tabla 13.3.1 para determinar qué puntuaciones se encuentran por encima y cuáles por debajo de la mediana hipotética de 5. Si asignamos un signo más a las puntuaciones que se encuentran por encima de la mediana hipotética y un signo menos a las que se encuentran por debajo, tenemos los resultados que se muestran en la Tabla 13.3.2.

If the null hypothesis were true, that is, if the median were, in fact, 5, we would expect the numbers of scores falling above and below 5 to be

HA: P1+2 Z P1-2

H0.

HA: P1+2 6 P1-2

H0.

HA: P1+2 7 P1-2

HA: P1+2 Z P1-2 two-sided alternative

HA: P1+2 6 P1-2 one-sided alternative

HA: P1+2 7 P1-2 one-sided alternative

13.3 THE SIGN TEST 687

TABLE 13.3.1 General Appearance

Scores of 10 Mentally Retarded Girls

Girl Score Girl Score

1 4 66

2 5 7 10

3 8 87

4 8 96

5 9 10 6

approximately equal. This line of reasoning suggests an alternative way in which we could have stated the null hypothesis, namely, that the probability of a plus is equal to the probability of a minus, and these probabilities are equal to .5. Stated symbolically, the hypothesis would be

In other words, we would expect about the same number of plus signs as minus signs in Table 13.3.2 when Ho is true. A look at Table 13.3.2 reveals a preponderance of pluses; specifically, we observe eight pluses, one minus, and one zero, which was assigned to the score that fell exactly on the median. The usual procedure for handling zeros is to eliminate them from the analysis and reduce n, the sample size, accordingly. If we follow this procedure, our problem reduces to one consisting of nine observations of which eight are plus and one is minus.

En otras palabras, esperaríamos aproximadamente el mismo número de signos más que menos en la Tabla 13.3.2 cuando Ho es verdadera. Un vistazo a la Tabla 13.3.2 revela una preponderancia de signos más; específicamente, observamos ocho signos más, uno menos y un cero, que se asignó a la puntuación que se situó exactamente en la mediana. El procedimiento habitual para manejar los ceros es eliminarlos del análisis y reducir n, el tamaño de la muestra, en consecuencia. Si seguimos este procedimiento, nuestro problema se reduce a uno compuesto por nueve observaciones, de las cuales ocho son más y una es menos.

Since the number of pluses and minuses is not the same, we wonder if the distribution of signs is sufficiently disproportionate to cast doubt on our hypothesis. Stated another way, we wonder if this small a number of minuses could have come about by chance alone when the null hypothesis is true, or if the number is so small that something other than chance (that is, a false null hypothesis) is responsible for the results.

Dado que el número de signos positivos y negativos no es el mismo, nos preguntamos si la distribución de signos es lo suficientemente desproporcionada como para cuestionar nuestra hipótesis. Dicho de otro modo, nos preguntamos si este pequeño número de signos negativos podría haberse producido solo por casualidad cuando la hipótesis nula es verdadera, o si el número es tan pequeño que algo ajeno al azar (es decir, una hipótesis nula falsa) es responsable de los resultados.

Based on what we learned in Chapter 4, it seems reasonable to conclude that the observations in Table 13.3.2 constitute a set of n independent random variables from the Bernoulli population with parameter p. If we let the sampling distribution of k is the binomial probability distribution with parameter if the null hypothesis is true.

6. Decision rule. The decision rule depends on the alternative hypothesis.

For reject if, when is true, the probability of observing k or fewer minus signs is less than or equal to

For reject if the probability of observing, when is true, k or fewer plus signs is equal to or less than

For reject if (given that is true) the proba-

bility of obtaining a value of k as extreme as or more extreme than

was actually computed is equal to or less than

For this example the decision rule is: Reject if the p value for the

computed test statistic is less than or equal to .05.

H0

a>2.

HA H0 H0 : P1+2 Z P1-2,

H a. 0

HA H0 : P1+2 6 P1-2,

a.

HA H0 H0 : P1+2 7 P1-2,

p = .5

k = the test statistic,

H0

H0: P1+2 = P1-2 = .5

688 CHAPTER 13 NONPARAMETRIC AND DISTRIBUTION-FREE STATISTICS

TABLE 13.3.2 Scores Above and Below the Hypothesized Median Based

on Data of Example 13.3.1

Girl 1 2 3 4 5 6 7 8 9 10

Score relative to 0

hypothesized

median

- + + + + + + + +

(#) (!)

7. Calculation of test statistic. We may determine the probability of

observing x or fewer minus signs when given a sample of size n and

parameter p by evaluating the following expression:

(13.3.1)

For our example we would compute

8. Statistical decision. In Appendix Table B we find

With a two-sided test either a sufficiently small number of minuses or a sufficiently small number of pluses would cause rejection of the null hypothesis. Since, in our example, there are fewer minuses, we focus our attention on minuses rather than pluses. By setting equal to .05, we are saying that if the number of minuses is so small that the probability of observing this few or fewer is less than .025 (half of alfa), we will reject the null hypothesis. The probability we have computed, .0195, is less than .025. We, therefore, reject the null hypothesis.

Con una prueba bilateral, un número suficientemente pequeño de negativos o positivos provocaría el rechazo de la hipótesis nula. Dado que, en nuestro ejemplo, hay menos negativos, nos centramos en los negativos en lugar de los positivos. Al establecer 0,05, indicamos que si el número de negativos es tan pequeño que la probabilidad de observar estos pocos o menos es menor que 0,025 (la mitad de alfa), rechazamos la hipótesis nula. La probabilidad calculada, 0,0195, es menor que 0,025. Por lo tanto, rechazamos la hipótesis nula.

9. Conclusion. We conclude that the median score is not 5.

10. p value. The p value for this test is ■

**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

Cuando los datos a analizar consisten en observaciones en pares coincidentes y no se cumplen los supuestos subyacentes a la prueba t, o la escala de medición es débil, se puede emplear la prueba de signos para probar la hipótesis nula de que la diferencia mediana es 0. Una forma alternativa de enunciar la hipótesis nula es

One of the matched scores, say, is subtracted from the other score, If is less than the sign of the difference is and if is greater than 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

Por ejemplo, una de las puntuaciones coincidentes se resta de la otra. Si es menor que el signo de la diferencia y si es mayor que el signo de la diferencia. Si la diferencia mediana es 0, esperaríamos que un par elegido al azar tuviera la misma probabilidad de producir un valor de como de cuando se realiza la resta. Podemos enunciar la hipótesis nula, entonces, como

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.

En una muestra aleatoria de pares coincidentes, esperaríamos que el número de s y s fuera aproximadamente igual. Si hay más s o más s de los que se pueden explicar solo por azar cuando la hipótesis nula es verdadera, tendremos dudas sobre la veracidad de nuestra hipótesis nula. Mediante la prueba de signos, podemos determinar cuántos signos de un mismo signo constituyen más de los que se pueden explicar solo por azar.

+’ -

+’ -

H0: P1+2 = P1-2 = .5

+ -

-.

Xi Y , + i X , i

,

Xi Yi Y . i

,

P1Xi 7 Yi

2 = P1Xi 6 Yi

2 = .5

21.01952 = .0390.

a

a

P1k ... 1 ƒ 9, .52 = .0195

9C01.520

1.529-0 + 9C11.521

1.529-1 = .00195 + .01758 = .0195

P1k ... x ƒ n, p2 = a

x

k=0

nCkpk

qn-k

13.3 THE SIGN TEST 689

**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.

Un equipo de investigación dental deseaba saber si enseñar a las personas a cepillarse los dientes sería beneficioso. Se obtuvieron doce parejas de pacientes atendidos en una clínica dental mediante un cuidadoso emparejamiento en función de factores como edad, sexo, inteligencia y puntuaciones iniciales de higiene bucal. Un miembro de cada pareja recibió instrucciones sobre cómo cepillarse los dientes y otros aspectos de higiene bucal. Seis meses después, un higienista dental, que desconocía qué sujetos habían recibido la instrucción, examinó a los 24 sujetos y les asignó una puntuación de higiene bucal. Una puntuación baja indica un alto nivel de higiene bucal. Los resultados se muestran en la Tabla 13.3.3.

**Solution:**

1. Data. See problem statement.

2. Assumptions. We assume that the population of differences between

pairs of scores is a continuous variable.

**3. Hypotheses.** If the instruction produces a beneficial effect, this fact would be reflected in the scores assigned to the members of each pair. If we take the differences , we would expect to observe more ’s than s if instruction had been beneficial, since a low score indicates a higher level of oral hygiene. If, in fact, instruction is beneficial, the median of the hypothetical population of all such differences would be less than 0, that is, negative. If, on the other hand, instruction has no effect, the median of this population would be zero. The null and alternate hypotheses, then, are:

Si la instrucción produce un efecto beneficioso, este hecho se reflejaría en las puntuaciones asignadas a los miembros de cada pareja. Si consideramos las diferencias, esperaríamos observar más s que s si la instrucción hubiera sido beneficiosa, ya que una puntuación baja indica un mayor nivel de higiene bucal. Si, de hecho, la instrucción es beneficiosa, la mediana de la población hipotética de todas estas diferencias sería menor que 0, es decir, negativa. Si, por el contrario, la instrucción no tiene efecto, la mediana de esta población sería cero. Las hipótesis nula y alternativa son, entonces:

+’

Xi - Yi -

690 CHAPTER 13 NONPARAMETRIC AND DISTRIBUTION-FREE STATISTICS

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

(Yi (X ) i)

(Yi)

(Xi)

The median of the differences is zero

The median of the differences is negative

Let be .05.

4. Test statistic. The test statistic is the number of plus signs.

5. Distribution of test statistic. The sampling distribution of k is the bino-

mial distribution with parameters n and .5 if is true.

6. Decision rule. Reject if

**7. Calculation of test statistic.** As will be seen, the procedure here is identical to the single sample procedure once the score differences have been obtained for each pair. Performing the subtractions and observing signs yields the results shown in Table 13.3.4.

Como se observará, el procedimiento aquí es idéntico al de una sola muestra una vez obtenidas las diferencias de puntuación para cada par. Al realizar las restas y observar los signos, se obtienen los resultados que se muestran en la Tabla 13.3.4.

The nature of the hypothesis indicates a one-sided test so that all of is associated with the rejection region, which consists of all values of k (where k is equal to the number of # signs) for which the probability of obtaining that many or fewer pluses due to chance alone when is true is equal to or less than .05. We see in Table 13.3.4 that the experiment yielded one zero, two pluses, and nine minuses. When we eliminate the zero, the effective sample size is with two pluses and nine minuses. In other words, since a “small” number of plus signs will cause rejection of the null hypothesis, the value of our test statistic is k=2.

La naturaleza de la hipótesis indica una prueba unilateral, de modo que la totalidad de está asociada con la región de rechazo, que consiste en todos los valores de k (donde k es igual al número de signos #) para los cuales la probabilidad de obtener esa cantidad o menos de signos positivos, debido únicamente al azar, cuando es verdadera, es igual o menor que 0,05. En la Tabla 13.3.4, vemos que el experimento arrojó un cero, dos signos positivos y nueve negativos. Al eliminar el cero, el tamaño de muestra efectivo es de dos signos positivos y nueve negativos. En otras palabras, dado que un número pequeño de signos positivos provocará el rechazo de la hipótesis nula, el valor de nuestro estadístico de prueba es k = 2.

8. Statistical decision. We want to know the probability of obtaining no more than two pluses out of 11 tries when the null hypothesis is true.

As we have seen, the answer is obtained by evaluating the appropriate

binomial expression. In this example we find

By consulting Appendix Table B, we find this probability to be .0327.

Since .0327 is less than .05, we must reject

9. Conclusion. We conclude that the median difference is negative. That

is, we conclude that the instruction was beneficial.

10. p value. For this test, ■

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

p = .0327.

H0.

P1k ... 2 ƒ 11, .52 = a

2

k=0

11Ck1.52k

1.5211-k

k = 2.

n = 11

H0

a = .05

H P1k ... 2 ƒ 11, .52 ... .05. 0

H0

a

HA: 3P1+2 6 P1-24.

H 3P1+2 = P1-24. 0:

13.3 THE SIGN TEST 691

TABLE 13.3.4 Signs of Differences in Oral Hygiene Scores of 12 Subjects

Instructed and 12 Matched Subjects Not Instructed

Pair 1 2 3 4 5 6 7 8 9 10 11 12

Sign of score 0

differences

- - + - - - - - - + -

(Yi (X ) i)

(Xi ! Yi)

**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.

Como se ha demostrado, la prueba de signos puede utilizarse con una sola muestra o con dos muestras, donde cada miembro de una muestra se empareja con un miembro de la otra para formar una muestra de pares emparejados. También hemos visto que la hipótesis alternativa puede dar lugar a una prueba unilateral o bilateral. En ambos casos, nos centramos en el signo menos frecuente y calculamos la probabilidad de obtener esa cantidad o menos de ese signo.

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

Usamos el signo menos frecuente como nuestro estadístico de prueba porque las probabilidades binomiales en la Tabla B del Apéndice son probabilidades "menores o iguales a". Al usar el signo menos frecuente, podemos obtener la probabilidad que necesitamos directamente de la Tabla B sin tener que hacer ninguna resta. Si las probabilidades en la Tabla B fueran probabilidades "mayores o iguales a", que se encuentran a menudo en las tablas de la distribución binomial, usaríamos el signo más frecuente como nuestro estadístico de prueba para aprovechar la conveniencia de obtener la probabilidad deseada directamente de la tabla sin tener que hacer ninguna resta. De hecho, podríamos, en nuestros ejemplos presentes, usar el signo más frecuente como nuestro estadístico de prueba, pero como la Tabla B contiene probabilidades "menores o iguales a", tendríamos que realizar una operación de resta para obtener la probabilidad deseada. A modo de ilustración, considere el último ejemplo. Si usamos como nuestro estadístico de prueba el signo más frecuente, es 9, el número de menos. La probabilidad deseada, entonces, es la probabilidad de nueve o más menos, cuando y Es decir, queremos

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

En el capítulo 5 vimos que, cuando el tamaño de la muestra es grande y p se acerca a 0,5, la distribución binomial puede aproximarse mediante la distribución normal. La regla general utilizada fue que la aproximación normal es apropiada cuando tanto np como nq son mayores que 5. Cuando, como se hipotetizó en nuestros dos ejemplos, una muestra de tamaño 12 cumpliría la regla general. Siguiendo esta directriz, se podría utilizar la aproximación normal cuando se utiliza la prueba de signos para contrastar la hipótesis nula de que la mediana o la diferencia de medianas es 0 y n es igual o mayor que 12. Dado que el procedimiento implica aproximar una distribución continua mediante una distribución discreta, generalmente se utiliza la corrección de continuidad de 0,5. El estadístico de prueba es entonces

(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 is used when k - .5 k Ú n>2.

k + .5 k 6 n>2

z = 1k ; .52 - .5n

.52n

p = .5,

= .0327

= 1 - .9673

P1k Ú 9 ƒ 11, .52 = 1 - P1k ... 8 ƒ 11, .52

P1k = 9 ƒ 11, .52

n = 11 p = .5.

692 CHAPTER 13 NONPARAMETRIC AND DISTRIBUTION-FREE STATISTICS

Computer Analysis Many statistics software packages will perform the sign test.

For example, if we use MINITAB to perform the test for Example 13.3.1 in which the

data are stored in Column 1, the procedure and output would be as shown in Figure 13.3.1.

**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.

En ocasiones, deseamos contrastar una hipótesis nula sobre la media de una población, pero por alguna razón, ni z ni t son estadísticos de prueba adecuados. Si contamos con una muestra pequeña de una población con una distribución no normal, y el teorema del límite central no es aplicable, el estadístico z queda descartado. El estadístico t no es adecuado porque la población muestreada no se aproxima suficientemente a una distribución normal. Ante esta situación, solemos buscar un procedimiento estadístico no paramétrico adecuado. Como hemos visto, la prueba de los signos puede utilizarse cuando los datos consisten en una sola muestra o cuando tenemos datos pareados. Sin embargo, si los datos para el análisis se miden al menos en una escala de intervalos, la prueba de los signos puede resultar indeseable, ya que no aprovecharía al máximo la información contenida en los datos. Un procedimiento más adecuado podría ser la prueba de rangos con signo de Wilcoxon (1), que utiliza las magnitudes de las diferencias entre las mediciones y un parámetro de ubicación hipotético, en lugar de solo los signos de las diferencias.

**Assumptions** The Wilcoxon test for location is based on the following assump-

tions 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

(a) (b) (c)

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

1. Subtract the hypothesized mean from each observation to obtain

If any is equal to the mean, so that eliminate that from the calcula-

tions and reduce n accordingly.

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

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

ing 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 or 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 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 is equal to the expected value of We do not expect and computed from a given sample to be equal. However, when is true, we do not expect a large difference in their values. Consequently, a sufficiently small value of or a sufficiently small value of will cause rejection of Ho.

El estadístico de la prueba de Wilcoxon es o bien o bien dependiendo de la naturaleza de la hipótesis alternativa. Si la hipótesis nula es verdadera, es decir, si la media poblacional verdadera es igual a la media hipotética, y si se cumplen los supuestos, la probabilidad de observar una diferencia positiva de una magnitud dada es igual a la probabilidad de observar una diferencia negativa de la misma magnitud. Entonces, en un muestreo repetido, cuando la hipótesis nula es verdadera y se cumplen los supuestos, el valor esperado de es igual al valor esperado de No esperamos que y calculados a partir de una muestra dada sean iguales. Sin embargo, cuando es verdadera, no esperamos una gran diferencia en sus valores. En consecuencia, un valor suficientemente pequeño de o un valor suficientemente pequeño de provocará el rechazo de Ho.

When the alternative hypothesis is two-sided , either a sufficiently small

value of or a sufficiently small value of will cause us to reject The

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

smaller of the two T.

When is true, we expect our sample to yield a large value of There-

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

less than the hypothesized mean a sufficiently small value of will cause

rejection of and is the test statistic.

When is true, we expect our sample to yield a large value of There-

fore, for the one-sided alternative a sufficiently small value of will cause

rejection of and 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 through The following are the decision rules for the three possible alternative hypotheses:

Los valores críticos del estadístico de la prueba de Wilcoxon se presentan en la Tabla K del Apéndice. Se proporcionan los niveles de probabilidad exactos (P) con cuatro decimales para todos los posibles totales de rango (T) que arrojan un nivel de probabilidad diferente en el cuarto decimal, desde 0,0001 hasta 0,5000. Los totales de rango (T) se tabulan para todos los tamaños de muestra, desde [número] hasta [número]. A continuación, se presentan las reglas de decisión para las tres posibles hipótesis alternativas:

(a) Reject at the level of significance if the calculated T is smaller

than or equal to the tabulated T for n and preselected 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

(b) Reject at the level of significance if is less than or equal to

the tabulated T for n and preselected

(c) . Reject at the level of significance if is less than or equal to

the tabulated T for n and preselected

**EXAMPLE 13.4.1**

Cardiac output (liters/minute) was measured by thermodilution in a simple random sample of 15 postcardiac surgical patients in the left lateral position. The results were as follows:

El gasto cardíaco (litros/minuto) se midió mediante termodilución en una muestra aleatoria simple de 15 pacientes posoperados de cirugía cardíaca en decúbito lateral izquierdo. Los resultados fueron los siguientes:

4.91 4.10 6.74 7.27 7.42 7.50 6.56 4.64

5.98 3.14 3.23 5.80 6.17 5.39 5.77

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

is different from 5.05.

Solution:

1. Data. See statement of example.

2. Assumptions. We assume that the requirements for the application of

the Wilcoxon signed-ranks test are met.

3. Hypotheses.

4. Test statistic. The test statistic will be or whichever is smaller.

We will call the test statistic T.

5. Distribution of test statistic. Critical values of the test statistic are

given in Table K of the Appendix.

6. Decision rule. We will reject if the computed value of T is less than

or equal to 25, the critical value for and the clos-

est value to .0250 in Table K.

7. Calculation of test statistic. The calculation of the test statistic is

shown in Table 13.4.1.

8. Statistical decision. Since 34 is greater than 25, we are unable to reject

9. Conclusion. We conclude that the population mean may be 5.05.

10. p value. From Table K we see that p = 21.07572 = .1514.

**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 values, the difference between each of the n pairs of measurements. If we let we may follow the procedure described above to test any one of the following null hypotheses:

and

**Computer Analysis** Many statistics software packages will perform the

Wilcoxon signed-rank test. If, for example, the data of Example 13.4.1 are stored in

Column 1, we could use MINITAB to perform the test as shown in Figure 13.4.1.

**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).

Un procedimiento no paramétrico que puede utilizarse para comprobar la hipótesis nula de que se han extraído dos muestras independientes de poblaciones con medianas iguales es la prueba de la mediana. Esta prueba, atribuida principalmente a Mood (2) y Westenberg (3), también es analizada por Brown y 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?

¿Existen diferencias entre los estudiantes varones de escuelas secundarias urbanas y rurales en cuanto a su nivel de salud mental?

Solution:

1. Data. Members of a random sample of 12 male students from a rural junior high school and an independent random sample of 16 male

students from an urban junior high school were given a test to measure their level of mental health. The results are shown in Table 13.5.1.

To determine if we can conclude that there is a difference, we per- form a hypothesis test that makes use of the median test. Suppose we choose a .05 level of significance.

**2. Assumptions.** The assumptions underlying the test are (a) the samples are selected independently and at random from their respective populations; (b) the populations are of the same form, differing only in location; and (c) the variable of interest is continuous. The level of measurement must be, at least, ordinal. The two samples do not have to be of equal size.

Los supuestos que sustentan la prueba son: (a) las muestras se seleccionan de forma independiente y aleatoria de sus respectivas poblaciones; (b) las poblaciones tienen la misma forma, difiriendo únicamente en su ubicación; y (c) la variable de interés es continua. El nivel de medición debe ser, como mínimo, ordinal. Las dos muestras no tienen que tener el mismo tamaño.

3. Hypotheses.

is the median score of the sampled population of urban students,

and is the median score of the sampled population of rural students.

Let

4. Test statistic. As will be shown in the discussion that follows, the test

statistic is as computed, for example, by Equation 12.4.1 for a

contingency table.

5. Distribution of test statistic. When is true and the assumptions are

met, is distributed approximately as with 1 degree of freedom.

6. Decision rule. Reject if the computed value of is (since

).

7. Calculation of test statistic. The first step in calculating the test statis-

tic is to compute the common median of the two samples combined. This

is done by arranging the observations in ascending order and, because

the total number of observations is even, obtaining the mean of the two

middle numbers. For our example the median is

We now determine for each group the number of observations falling

above and below the common median. The resulting frequencies are arranged

in a table. For the present example we construct Table 13.5.2.

If the two samples are, in fact, from populations with the same median,

we would expect about one-half the scores in each sample to be above the

combined median and about one-half to be below. If the conditions relative

to sample size and expected frequencies for a contingency table as

discussed in Chapter 12 are met, the chi-square test with 1 degree of free-

dom may be used to test the null hypothesis of equal population medians.

For our examples we have, by Formula 12.4.1,

8. Statistical decision. Since the critical value of with

and 1 degree of freedom, we are unable to reject the null hypoth-

esis on the basis of these data.

9. Conclusion. We conclude that the two samples may have been drawn

from populations with equal medians.

10. p value. Since we have ■

**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 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 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.

A veces, uno o más valores observados serán exactamente iguales a la mediana común y, por lo tanto, no estarán ni por encima ni por debajo de ella. Cabe destacar que si es impar, al menos un valor siempre será exactamente igual a la mediana. Esto plantea la pregunta de qué hacer con observaciones de este tipo. Una solución es excluirlas del análisis si es grande y solo hay unos pocos valores que se encuentran en la mediana combinada. También podemos dicotomizar las puntuaciones entre las que superan la mediana y las que no, en cuyo caso las observaciones que igualan la mediana se contabilizarán en la segunda categoría.

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

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

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

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

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

compared with the critical with degrees of freedom.

Computer Analysis The median test calculations may be carried out using

MINITAB. To illustrate using the data of Example 13.5.1 we first store the mea-

surements in MINITAB Column 1. In MINITAB Column 2 we store codes that iden-

tify the observations as to whether they are for an urban (1) or rural (2) subject. The

MINITAB procedure and output are shown in Figure 13.5.1.

**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.

La prueba de la mediana, analizada en la sección anterior, no aprovecha al máximo toda la información presente en las dos muestras cuando la variable de interés se mide al menos en una escala ordinal. Reducir el contenido informativo de una observación a simplemente determinar si se encuentra por encima o por debajo de la mediana común es un desperdicio de información. Si, para contrastar la hipótesis deseada, se dispone de un procedimiento que utilice más información inherente a los datos, dicho procedimiento debería utilizarse si es posible. Un procedimiento no paramétrico de este tipo, que a menudo puede utilizarse en lugar de la prueba de la mediana, es la prueba de Mann-Whitney (5), a veces denominada prueba de Mann-Whitney-Wilcoxon. Dado que esta prueba se basa en los rangos de las observaciones, utiliza más información que la prueba de la mediana.

**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 population 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 populations 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.

Cuando se cumplen estos supuestos, podemos contrastar la hipótesis nula de que las dos poblaciones tienen medianas iguales frente a cualquiera de las tres alternativas posibles: (1) las poblaciones no tienen medianas iguales (prueba bilateral), (2) la mediana de la población 1 es mayor que la mediana de la población 2 (prueba unilateral), o (3) la mediana de la población 1 es menor que la mediana de la población 2 (prueba unilateral). Si las dos poblaciones son simétricas, de modo que dentro de cada población la media y la mediana son iguales, las conclusiones a las que llegamos con respecto a las dos medianas poblacionales también se aplicarán a las dos medias poblacionales. El siguiente ejemplo ilustra el uso de la prueba de Mann-Whitney.

EXAMPLE 13.6.1

A researcher designed an experiment to assess the effects of prolonged inhalation of cadmium 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.

Un investigador diseñó un experimento para evaluar los efectos de la inhalación prolongada de óxido de cadmio. Quince animales de laboratorio sirvieron como sujetos experimentales, mientras que diez animales similares sirvieron como controles. La variable de interés fue el nivel de hemoglobina después del experimento. Los resultados se muestran en la Tabla 13.6.1. Deseamos saber si podemos concluir que la inhalación prolongada de óxido de cadmio reduce el nivel de hemoglobina.

**Solution:**

1. Data. See Table 13.6.1.

2. Assumptions. We assume that the assumptions of the Mann– Whitney test

are met.

3. Hypotheses. The null and alternative hypotheses are as follows:

where is the median of a population of animals exposed to cadmium

oxide and is the median of a population of animals not exposed to

the substance. Suppose we let

4. Test statistic. To compute the test statistic we combine the two sam-

ples and rank all observations from smallest to largest while keeping

track of the sample to which each observation belongs. Tied observa-

tions are assigned a rank equal to the mean of the rank positions for

which they are tied. The results of this step are shown in Table 13.6.2.

The test statistic is

(13.6.1)

where n is the number of sample X observations and S is the sum of the ranks assigned to the sample observations from the population of X values. The choice of which sample’s values we label X is arbitrary.

donde n es el número de observaciones de muestra X y S es la suma de los rangos asignados a las observaciones de muestra de la población de valores X. La elección de qué valores de la muestra etiquetamos como X es arbitraria.

5. Distribution of test statistic. Critical values from the distribution of the

test statistic are given in Appendix Table L for various levels of

**6. Decision rule.** If the median of the X population is, in fact, smaller than the median of the Y population, as specified in the alternative hypothesis, we would expect (for equal sample sizes) the sum of the ranks assigned to the observations from the X population to be smaller than the sum of the ranks assigned to the observations from the Y population. The test statistic is based on this rationale in such a way that a sufficiently small value of T will cause rejection of In general, for one-sided tests of the type illustrated here the decision rule is:

Si la mediana de la población X es, de hecho, menor que la mediana de la población Y, como se especifica en la hipótesis alternativa, esperaríamos (para tamaños de muestra iguales) que la suma de los rangos asignados a las observaciones de la población X sea menor que la suma de los rangos asignados a las observaciones de la población Y. La estadística de prueba se basa en este razonamiento de tal manera que un valor suficientemente pequeño de T causará el rechazo de En general, para las pruebas unilaterales del tipo ilustrado aquí, la regla de decisión es:

Reject if the computed T is less than where is the criti-

cal value of T obtained by entering Appendix Table L with n, the number of

X observations; m, the number of Y observations; and the chosen level of

significance.

If we use the Mann–Whitney procedure to test

against

sufficiently large values of T will cause rejection so that the decision

rule is:

Reject if computed T is greater than where

For the two-sided test situation with

computed values of T that are either sufficiently large or sufficiently small

will cause rejection of The decision rule for this case, then, is:

Reject if the computed value of T is either less than or

greater than where is the critical value of T for n, m, and

given in Appendix Table L, and

For this example the decision rule is:

Reject if the computed value of T is smaller than 45, the critical value of

the test statistic for and found in Table L.

The rejection regions for each set of hypotheses are shown in

Figure 13.6.1.

7. Calculation of test statistic. For our present example we have, as

shown in Table 13.6.2, so that

T = 145 - 15115 + 12

2 = 25

8. Statistical decision. When we enter Table L with and

we find the critical value of to be 45. Since we

reject

9. Conclusion. We conclude that is smaller than This leads to the

conclusion that prolonged inhalation of cadmium oxide does reduce the

hemoglobin level.

10. p value. Since we have for this test

■

**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

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

ties in the data:

(13.6.3)

Computer Analysis Many statistics software packages will perform the

Mann–Whitney test. With the data of two samples stored in Columns 1 and 2, for

example, MINITAB will perform a one-sided or two-sided test. The MINITAB proce-

dure and output for Example 13.6.1 are shown in Figure 13.6.2.

The SPSS output for Example 13.6.1 is shown in Figure 13.6.3. As we see this out-

put provides the Mann–Whitney test, the Wilcoxon test, and large-sample z approximation.

**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. Thetest 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,

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

from a population with unknown cumulative distribution function 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, we may estimate 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, 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 is true,

the hypothesis is rejected.

The difference between the theoretical cumulative distribution function, and

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

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

priate, that is, when the hypotheses are

for all x from

for at least one x

the test statistic is

(13.7.1)

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

difference minus ”

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

value of D exceeds the value shown in Appendix Table M for (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 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 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.

Solution:

1. Data. See Table 13.7.1.

2. Assumptions. The sample available is a simple random sample from a

continuous population distribution.

3. Hypotheses. The appropriate hypotheses are

for all x from to

for at least one x

Let

4. Test statistic. See Equation 13.7.1.

5. Distribution of test statistic. Critical values of the test statistic for

selected values of are given in Appendix Table M.

6. Decision rule. Reject if the computed value of D exceeds .221, the

critical value of D for and

7. Calculation of test statistic. Our first step is to compute values of

F as shown in Table 13.7.2.

Each value of is obtained by dividing the corresponding

cumulative frequency by the sample size. For example, the first value of

We obtain values of by first converting each observed value

of x to a value of the standard normal variable, z. From Appendix Table

D we then find the area between and z. From these areas we are

able to compute values of The procedure, which is similar to that

used to obtain expected relative frequencies in the chi-square goodness-

of-fit test, is summarized in Table 13.7.3.

The test statistic D may be computed algebraically, or it may be

determined graphically by actually measuring the largest vertical distance

between the curves of and on a graph. The graphs of the

two distributions are shown in Figure 13.7.1.

Examination of the graphs of and reveals that

Now let us compute the value of D alge-

braically. The possible values of are shown in Table

13.7.4. This table shows that the exact value of D is .1547.

8. Statistical decision. Reference to Table M reveals that a computed D

of .1547 is not significant at any reasonable level. Therefore, we are not

willing to reject .

9. Conclusion. The sample may have come from the specified distribution.

10. p value. Since we have a two-sided test, and since . we

have . p 7 .20

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

The largest vertical distance between and 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 at the left endpoints of

the horizontal bars are considered, we would incorrectly compute D as

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

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

point corresponding to and the correct value of D is

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

to the differences the differences for all values of

where the number of different values of x and

The correct value of the test statistic will then be

D = maximum (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 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

(13.8.1)

is computed. In Equation 13.8.1,

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 with 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 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.

Solution:

1. Data. See Table 13.8.1.

2. Assumptions. The samples are independent random samples from

their respective populations. The measurement scale employed is at least

ordinal. The distributions of the values in the sampled populations are

identical except for the possibility that one or more of the populations

are composed of values that tend to be larger than those of the other

populations.

3. Hypotheses.

than at least one of the other populations.

Let

4. Test statistic. See Equation 13.8.1.

5. Distribution of test statistic. Critical values of H for various sample

sizes and levels are given in Appendix Table N.

6. Decision rule. The null hypothesis will be rejected if the computed

value of H is so large that the probability of obtaining a value that large

or larger when is true is equal to or less than the chosen significance

level,

7. Calculation of test statistic. When the three samples are combined into

a single series and ranked, the table of ranks shown in Table 13.8.2 may

be constructed.

The null hypothesis implies that the observations in the three sam-

ples constitute a single sample of size 15 from a single population. If

this is true, we would expect the ranks to be well distributed among the

three groups. Consequently, we would expect the total sum of ranks to

be divided among the three groups in proportion to group size. Departures

from these conditions are reflected in the magnitude of the test statis-

tics H.

From the data in Table 13.8.2 and Equation 13.8.1, we obtain

8. Statistical decision. Table N shows that when the are 5, 5, and 5,

the probability of obtaining a value of is less than .009. The

null hypothesis can be rejected at the .01 level of significance.

9. Conclusion. We conclude that there is a difference in the average

eosinophil cell count among the three populations.

10. p value. For this test, ■

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

dividing it by

(13.8.2)

where 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 is greater

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.

Solution: From the sums of the ranks we compute

Reference to Appendix Table F with degrees of freedom indi-

cates that the probability of obtaining a value of H as large as or larger

than 36.39, due to chance alone, when there is no difference among the

populations, is less than .005. We conclude, then, that there is a difference

among the five populations with respect to the average value of the vari-

able of interest. ■

Computer Analysis The MINITAB software package computes the Kruskal–

Wallis test statistic and provides additional information. After we enter the eosinophil

counts in Table 13.8.1 into Column 1 and the group codes into Column 2, the MINITAB

procedure and output are as shown in Figure 13.8.1.

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.

Solution:

1. Data. See Table 13.9.1.

2. Assumptions. The observations appearing in a given block are inde-

pendent of the observations appearing in each of the other blocks, and

within each block measurement on at least an ordinal scale is achieved.

3. Hypothesis. In general, the hypotheses are:

The treatments all have identical effects.

At least one treatment tends to yield larger observations than

at least one of the other treatments.

For our present example we state the hypotheses as follows:

The three models are equally preferred.

The three models are not equally preferred.

Let a = .05.

4. Test statistic. By means of the Friedman test we will be able to deter-

mine if it is reasonable to assume that the columns of ranks have been

drawn from the same population. If the null hypothesis is true we would

expect the observed distribution of ranks within any column to be the

result of chance factors and, hence, we would expect the numbers 1, 2,

and 3 to occur with approximately the same frequency in each column.

If, on the other hand, the null hypothesis is false (that is, the models are

not equally preferred), we would expect a preponderance of relatively

high (or low) ranks in at least one column. This condition would be

reflected in the sums of the ranks. The Friedman test will tell us whether

or not the observed sums of ranks are so discrepant that it is not likely

they are a result of chance when is true.

Since the data already consist of rankings within blocks (rows), our

first step is to sum the ranks within each column (treatment). These sums

are the shown in Table 13.9.1. A test statistic, denoted by Friedman

as is computed as follows:

(13.9.1)

where the number of rows (blocks) and the number of columns

(treatments).

5. Distribution of test statistic. Critical values for various values of n and

k are given in Appendix Table O.

6. Decision rule. Reject if the probability of obtaining (when is

true) a value of as large as or larger than actually computed is less

than or equal to

7. Calculation of test statistic. Using the data in Table 13.9.1 and Equa-

tions 13.9.1, we compute

8. Statistical decision. When we consult Appendix Table Oa, we find that

the probability of obtaining a value of as large as 8.222 due to chance

alone, when the null hypothesis is true, is .016. We are able, therefore,

to reject the null hypothesis.

9. Conclusion. We conclude that the three models of low-volt electrical

stimulator are not equally preferred.

10. p value. For this test, ■

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 is obtained by consulting the table (Table F) with the chosen

and 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.

Solution: From the data, we compute

Reference to Table F indicates that with degrees of free-

dom the probability of getting a value of as large as 30.32 due to chance

alone is, when is true, less than .005. We reject the null hypothesis and

conclude that the different dose levels do produce different responses.

Computer Analysis Many statistics software packages, including MINITAB, will

perform the Friedman test. To use MINITAB we form three columns of data. We may, for

example, set up the columns so that Column 1 contains numbers that indicate the treat-

ment to which the observations belong, Column 2 contains numbers indicating the blocks

to which the observations belong, and Column 3 contains the observations. If we do this

for Example 13.9.1, the MINITAB procedure and output are as shown in Figure 13.9.1.

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 calcula-

tions 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 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) and Y are mutually independent.

and Y are not mutually independent.

(b) and Y are mutually independent.

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

together.

(c) and Y are mutually independent.

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 for each pair of observations by subtracting the rank of from the

rank of

3. Square each and compute the sum of the squared values.

4. Compute

(13.10.1)

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

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

level if is greater than or less than where is at the intersection of the

column headed and the row corresponding to n. For the one-sided test with

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

than for and n. The null hypothesis is rejected at the significance level in the

other one-sided test if is less than for and n.

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

(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:

(13.10.3)

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

this correction factor is used, is computed from

(13.10.4)

instead of from Equation 13.10.1.

In Equation 13.10.4,

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

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 When the number of ties is small, rs.

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.

Solution:

1. Data. See Table 13.10.1.

2. Assumptions. We assume that the sample available for analysis is a

simple random sample and that both X and Y are measured on at least

the ordinal scale.

3. Hypotheses.

This EEG output and age are mutually independent.

There is a tendency for this EEG output to decrease with age.

Suppose we let a = .05.

4. Test statistic. See Equation 13.10.1.

5. Distribution of test statistic. Critical values of the test statistic are

given in Appendix Table P.

6. Decision rule. For the present test we will reject if the computed

value of is less than

7. Calculation of test statistic. When the X and Y values are ranked, we

have the results shown in Table 13.10.2. The and are shown

in the same table.

Substitution of the data from Table 13.10.2 into Equation 13.10.1

gives

8. Statistical decision. Since our computed is less than the crit-

ical we reject the null hypothesis.

9. Conclusion. We conclude that the two variables are inversely related.

10. p value. Since we have for this test -.76 6 -0.6586, p 6 .001.

Let us now illustrate the procedure for a sample with 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, and 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.

Solution: From the data in Table 13.10.4 we compute

To test the significance of we compute

z = .75235 - 1 = 4.37

Since 4.37 is greater than and we

reject and conclude that the two variables under study are not mutually

independent.

For comparative purposes let us correct for ties using Equation 13.10.3

and then compute by Equation 13.10.4.

In the rankings of X we had six groups of ties that were broken by

assigning the values 13.5, 17, 19.5, 21.5, 23.5, and 32.5. In five of the groups

two observations tied, and in one group three observations tied. We, there-

fore, compute five values of

and one value of

From these computations, we have so that

ax 2 = 352 - 35

12 - 4.5 = 3565.5

Since no ties occurred in the Y rankings, we have and

From Table 13.10.4 we have From these data we may now

compute by Equation 13.10.4

We see that in this case the correction for ties does not make any difference

in the value of ■

Computer Analysis We may use MINITAB, as well as many other statistical

software packages, to compute the Spearman correlation coefficient. To use MINITAB,

we must first have MINITAB rank the observations and store the ranks in separate

columns, one for the X ranks and one for the Y ranks. If we rank the X and Y values of

Example 13.10.1 and store them in Columns 3 and 4, we may obtain the Spearman rank

correlation coefficient with the procedure shown in Figure 13.10.1. Other software pack-

ages such as SAS® and SPSS, for example, automatically rank the measurements before

computing the coefficient, thereby eliminating an extra step in the procedure.

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 We assume that the data conform to the classic

regression model

where the are known constants, and are unknown parameters, and is an

observed value of the continuous random variable Y at For each value of we assume

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

(no ties), and we take

The data consist of n pairs of sample observations,

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

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

where There will be values of The estimator of

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

(13.11.1)

The following example illustrates the calculation of

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.

Solution: The ordered values of are shown in Table 13.11.2.

If we let the indicators of the first and second val-

ues of Y and X in Table 13.11.1, we may compute as follows:

When all the slopes are computed in a similar manner and ordered as in

Table 13.11.2, winds up as the tenth value in the ordered array.

The median of the values is .4878. Consequently, our estimate of

the population slope coefficient b1

N = .4878.

An Estimator of the Intercept Coefficient Dietz (13) recommends

two intercept estimators. The first, designated is the median of the n terms

in which 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 esti-

mator which is the median of the pairwise averages of the

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

EXAMPLE 13.11.2

Refer to Example 13.11.1. Let us compute and from the data on testosterone

and citric acid levels.

Solution: The ordered terms are: 13.5396, 24.6362, 39.3916, 48.1730,

54.8804, 59.1500, 91.7100, and 98.7810. The median, 51.5267, is the

estimator .

The ordered pairwise averages of the

are

The median of these averages, 53.1432, is the estimator The esti-

mating equation, then, is if we are willing to

assume that the distribution of error terms is symmetric about 0. If we are

not willing to make the assumption of symmetry, the estimating equation

is ■

EXERCISES

13.11.1 The following are the heart rates (HR: beats/minute) and oxygen consumption values (

cal/kg/24 h) for nine infants with chronic congestive heart failure:

HR(X): 163 164 156 151 152 167 165 153 155

53.9 57.4 41.0 40.0 42.0 64.4 59.1 49.9 43.2

Compute

13.11.2 The following are the body weights (grams) and total surface area of nine laboratory

animals:

Body weight (X): 660.2 706.0 924.0 936.0 992.1 888.9 999.4 890.3 841.2

Surface area (Y): 781.7 888.7 1038.1 1040.0 1120.0 1071.5 1134.5 965.3 925.0

Compute the slope estimator and two intercept estimators.

13.12 SUMMARY

This chapter is concerned with nonparametric statistical tests. These tests may be used

either when the assumptions underlying the parametric tests are not realized or when

the data to be analyzed are measured on a scale too weak for the arithmetic procedures

necessary for the parametric tests.

Nine nonparametric tests are described and illustrated. Except for the

Kolmogorov–Smirnov goodness-of-fit test, each test provides a nonparametric alternative

to a well-known parametric test. There are a number of other nonparametric tests avail-

able. The interested reader is referred to the many books devoted to nonparametric

methods, including those by Gibbons (14) and Pett (15).

1cm2

2

bN 1, 1bN 021,M, and 1bN 022, M

VO21Y 2:

VO2:

yi = 51.5267 + .4878xi