BIOSTATISTICS\_Daniel\_C07

HYOTHESIS TESTING

CHAPTER OVERVIEW

This chapter covers hypothesis testing, the second of two general areas of statistical inference. Hypothesis testing is a topic with which you as a student are likely to have some familiarity. Interval estimation, discussed in the preceding chapter, and hypothesis testing are based on similar concepts. In fact, confidence intervals may be used to arrive at the same conclusions that are reached through the use of hypothesis tests. This chapter provides a format, followed throughout the remainder of this book, for conducting a hypothesis test.

Este capítulo cubre la prueba de hipótesis, la segunda de dos áreas generales de inferencia estadística. La prueba de hipótesis es un tema con el que usted, como estudiante, probablemente esté familiarizado. La estimación de intervalos, analizada en el capítulo anterior, y la prueba de hipótesis se basan en conceptos similares. De hecho, se pueden utilizar intervalos de confianza para llegar a las mismas conclusiones a las que se llega mediante el uso de pruebas de hipótesis. Este capítulo proporciona un formato, que se seguirá a lo largo del resto de este libro, para realizar una prueba de hipótesis.

TO P I C S

7.1 INTRODUCTION

7.2 HYPOTHESIS TESTING: A SINGLE POPULATION MEAN

7.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS

7.4 PAIRED COMPARISONS

7.5 HYPOTHESIS TESTING: A SINGLE POPULATION PROPORTION

7.6 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION

PROPORTIONS

7.7 HYPOTHESIS TESTING: A SINGLE POPULATION VARIANCE

7.8 HYPOTHESIS TESTING: THE RATIO OF TWO POPULATION VARIANCES

7.9 THE TYPE II ERROR AND THE POWER OF A TEST

7.10 DETERMINING SAMPLE SIZE TO CONTROL TYPE II ERRORS

7.11 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand how to correctly state a null and alternative hypothesis and carry out a structured hypothesis test.

Comprender cómo plantear correctamente una hipótesis nula y alternativa y realizar una prueba de hipótesis estructurada.

2. understand the concepts of type I error, type II error, and the power of a test.

comprender los conceptos de error tipo I, error tipo II y el poder de una prueba.

3. be able to calculate and interpret z, t, F, and chi-square test statistics for making statistical inferences.

Ser capaz de calcular e interpretar estadísticas de prueba z, t, F y chi-cuadrado para hacer inferencias estadísticas.

4. understand how to calculate and interpret p values.

7.1INTRODUCTION

One type of statistical inference, estimation, is discussed in the preceding chapter. The other type, hypothesis testing, is the subject of this chapter. As is true with estimation, the purpose of hypothesis testing is to aid the clinician, researcher, or administrator in reaching a conclusion concerning a population by examining a sample from that population. Estimation and hypothesis testing are not as different as they are made to appear by the fact that most textbooks devote a separate chapter to each. As we will explain later, one may use confidence intervals to arrive at the same conclusions that are reached by using the hypothesis testing procedures discussed in this chapter.

En el capítulo anterior se analiza un tipo de inferencia estadística, la estimación. El otro tipo, la prueba de hipótesis, es el tema de este capítulo. Como ocurre con la estimación, el propósito de la prueba de hipótesis es ayudar al médico, investigador o administrador a llegar a una conclusión sobre una población examinando una muestra de esa población. La estimación y la prueba de hipótesis no son tan diferentes como parece por el hecho de que la mayoría de los libros de texto dedican un capítulo separado a cada una. Como explicaremos más adelante, se pueden utilizar intervalos de confianza para llegar a las mismas conclusiones a las que se llega mediante los procedimientos de prueba de hipótesis analizados en este capítulo.

**Basic** **Concepts** In this section some of the basic concepts essential to an understanding of hypothesis testing are presented. The specific details of particular tests will be given in succeeding sections.

Conceptos básicos En esta sección se presentan algunos de los conceptos básicos esenciales para comprender la prueba de hipótesis. Los detalles específicos de pruebas particulares se darán en las secciones siguientes.

DEFINITION

A hypothesis may be defined simply as a statement about one or more populations.

Una hipótesis puede definirse simplemente como una afirmación sobre una o más poblaciones.

The hypothesis is frequently concerned with the parameters of the populations about which the statement is made. A hospital administrator may hypothesize that the average length of stay of patients admitted to the hospital is 5 days; a public health nurse may hypothesize that a particular educational program will result in improved communication between nurse and patient; a physician may hypothesize that a certain drug will be effective in 90 percent of the cases for which it is used. By means of hypothesis testing one determines whether or not such statements are compatible with the available data.

La hipótesis frecuentemente se relaciona con los parámetros de las poblaciones sobre las cuales se hace la afirmación. Un administrador de un hospital puede plantear la hipótesis de que la duración media de la estancia de los pacientes admitidos en el hospital es de cinco días; una enfermera de salud pública puede plantear la hipótesis de que un programa educativo particular dará como resultado una mejor comunicación entre la enfermera y el paciente; un médico puede plantear la hipótesis de que un determinado fármaco será eficaz en el 90 por ciento de los casos en los que se utiliza. Mediante pruebas de hipótesis se determina si dichas afirmaciones son compatibles o no con los datos disponibles.

**Types** **of** **Hypotheses** Researchers are concerned with two types of hypotheses—

research hypotheses and statistical hypotheses.

Tipos de hipótesis A los investigadores les interesan dos tipos de hipótesis: hipótesis de investigación e hipótesis estadísticas.

**DEFINITION**

The research hypothesis is the conjecture or supposition that motivates the research.

La hipótesis de investigación es la conjetura o suposición que motiva la investigación.

It may be the result of years of observation on the part of the researcher. A public health nurse, for example, may have noted that certain clients responded more readily to a particular type of health education program. A physician may recall numerous instances in which certain combinations of therapeutic measures were more effective than any one of them alone. Research projects often result from the desire of such health practitioners to7.1 determine whether or not their theories or suspicions can be supported when subjected to the rigors of scientific investigation.

Puede ser el resultado de años de observación por parte del investigador. Una enfermera de salud pública, por ejemplo, puede haber notado que ciertos clientes respondían más fácilmente a un tipo particular de programa de educación sanitaria. Un médico puede recordar numerosos casos en los que ciertas combinaciones de medidas terapéuticas fueron más efectivas que cualquiera de ellas por sí sola. Los proyectos de investigación a menudo surgen del deseo de dichos profesionales de la salud de determinar si sus teorías o sospechas pueden sustentarse o no cuando se las somete a los rigores de la investigación científica.

Research hypothses lead directly to statistical hypotheses.

DEFINITION

Statistical hypotheses are hypotheses that are stated in such a way that they may be evaluated by appropriate statistical techniques.

Las hipótesis estadísticas son hipótesis que se plantean de tal manera que pueden evaluarse mediante técnicas estadísticas apropiadas.

In this book the hypotheses that we will focus on are statistical hypotheses. We will assume that the research hypotheses for the examples and exercises have already been considered.

En este libro las hipótesis en las que nos centraremos son hipótesis estadísticas. Supondremos que ya se han considerado las hipótesis de investigación para los ejemplos y ejercicios.

**Hypothesis** **Testing** **Steps** For convenience, hypothesis testing will be presented as a ten-step procedure. There is nothing magical or sacred about this particular format. It merely breaks the process down into a logical sequence of actions and decisions.

Pasos de la prueba de hipótesis Por conveniencia, la prueba de hipótesis se presentará como un procedimiento de diez pasos. No hay nada mágico o sagrado en este formato en particular. Simplemente divide el proceso en una secuencia lógica de acciones y decisiones.

1. **Data**. The nature of the data that form the basis of the testing procedures must be understood, since this determines the particular test to be employed. Whether the data consist of counts or measurements, for example, must be determined.

1. Datos. Se debe comprender la naturaleza de los datos que forman la base de los procedimientos de prueba, ya que esto determina la prueba particular que se empleará. Es necesario determinar si los datos se componen, por ejemplo, de recuentos o de mediciones.

2. **Assumptions**. As we learned in the chapter on estimation, different assumptions lead to modifications of confidence intervals. The same is true in hypothesis testing: A general procedure is modified depending on the assumptions. In fact, the same assumptions that are of importance in estimation are important in hypothesis testing. We have seen that these include assumptions about the normality of the population distribution, equality of variances, and independence of samples.

2. Supuestos. Como aprendimos en el capítulo sobre estimación, diferentes supuestos conducen a modificaciones de los intervalos de confianza. Lo mismo ocurre en la prueba de hipótesis: se modifica un procedimiento general en función de los supuestos. De hecho, los mismos supuestos que son importantes en la estimación lo son en la prueba de hipótesis. Hemos visto que estos incluyen supuestos sobre la normalidad de la distribución de la población, la igualdad de varianzas y la independencia de las muestras.

3. **Hypotheses**. There are two statistical hypotheses involved in hypothesis testing, and these should be stated explicitly. The null hypothesis is the hypothesis to be tested. It is designated by the symbol H0. The null hypothesis is sometimes referred to as a hypothesis of no difference, since it is a statement of agreement with (or no difference from) conditions presumed to be true in the population of interest. In general, the null hypothesis is set up for the express purpose of being discredited. Consequently, the complement of the conclusion that the researcher is seeking to reach becomes the statement of the null hypothesis.

3. Hipótesis. Hay dos hipótesis estadísticas involucradas en la prueba de hipótesis, y estas deben establecerse explícitamente. La hipótesis nula es la hipótesis que se va a probar. Se designa con el símbolo H0. La hipótesis nula a veces se denomina hipótesis de no diferencia, ya que es una afirmación de acuerdo (o no diferencia) con condiciones que se presumen verdaderas en la población de interés. En general, la hipótesis nula se plantea con el expreso propósito de desacreditarla. En consecuencia, el complemento de la conclusión a la que el investigador busca llegar se convierte en el enunciado de la hipótesis nula.

In the testing process the null hypothesis either is rejected or is not rejected. If the null hypothesis is not rejected, we will say that the data on which the test is based do not provide sufficient evidence to cause rejection. If the testing procedure leads to rejection, we will say that the data at hand are not compatible with the null hypothesis, but are supportive of some other hypothesis. The alternative hypothesis is a statement of what we will believe is true if our sample data cause us to reject the null hypothesis. Usually the alternative hypothesis and the research hypothesis are the same, and in fact the two terms are used interchangeably. We shall designate the alternative hypothesis by the symbol HA.

En el proceso de prueba, la hipótesis nula se rechaza o no se rechaza. Si no se rechaza la hipótesis nula, diremos que los datos en los que se basa la prueba no aportan evidencia suficiente para provocar el rechazo. Si el procedimiento de prueba conduce al rechazo, diremos que los datos disponibles no son compatibles con la hipótesis nula, pero apoyan alguna otra hipótesis. La hipótesis alternativa es una afirmación de lo que creeremos que es verdadero si los datos de nuestra muestra nos hacen rechazar la hipótesis nula. Por lo general, la hipótesis alternativa y la hipótesis de investigación son las mismas y, de hecho, los dos términos se usan indistintamente. Designaremos la hipótesis alternativa con el símbolo HA.

**Rules for Stating Statistical Hypotheses** When hypotheses are of the type considered in this chapter an indication of equality (either =, …, or Ú ) must appear in the null hypothesis. Suppose, for example, that we want to answer the question: Can we conclude that a certain population mean is not 50? The null hypothesis is

Reglas para formular hipótesis estadísticas Cuando las hipótesis son del tipo considerado en este capítulo, debe aparecer una indicación de igualdad (ya sea = <=, >=) en la hipótesis nula. Supongamos, por ejemplo, que queremos responder la pregunta: ¿Podemos concluir que una determinada media poblacional no es 50? La hipótesis nula es

H0 : m = 50

and the alternative is

HA : m != 50

Suppose we want to know if we can conclude that the population mean is greater than 50. Our hypotheses are

H0: m <= 50

HA: m > 50

If we want to know if we can conclude that the population mean is less than 50, the hypotheses are

H0: m >= 50

HA: m < 50

In summary, we may state the following rules of thumb for deciding what statement goes in the null hypothesis and what statement goes in the alternative hypothesis:

En resumen, podemos establecer las siguientes reglas generales para decidir qué enunciado va en la hipótesis nula y qué enunciado va en la hipótesis alternativa:

(a) What you hope or expect to be able to conclude as a result of the test usually should be placed in the alternative hypothesis.

Lo que usted espera o espera poder concluir como resultado de la prueba generalmente debe ubicarse en la hipótesis alternativa.

(b) The null hypothesis should contain a statement of equality, either =, <=, or >=.

(c) The null hypothesis is the hypothesis that is tested.

La hipótesis nula es la hipótesis que se prueba.

(d) The null and alternative hypotheses are complementary. That is, the two together exhaust all possibilities regarding the value that the hypothesized parameter can assume.

Las hipótesis nula y alternativa son complementarias. Es decir, los dos juntos agotan todas las posibilidades respecto del valor que puede asumir el parámetro hipotético.

**A Precaution** It should be pointed out that neither hypothesis testing nor statistical inference, in general, leads to the proof of a hypothesis; it merely indicates whether the hypothesis is supported or is not supported by the available data. When we fail to reject a null hypothesis, therefore, we do not say that it is true, but that it may be true. When we speak of accepting a null hypothesis, we have this limitation in mind and do not wish to convey the idea that accepting implies proof.

Una precaución Cabe señalar que ni la prueba de hipótesis ni la inferencia estadística, en general, conducen a la prueba de una hipótesis; simplemente indica si la hipótesis está respaldada o no por los datos disponibles. Por lo tanto, cuando no podemos rechazar una hipótesis nula, no decimos que es verdadera, sino que puede ser cierta. Cuando hablamos de aceptar una hipótesis nula, tenemos esta limitación en mente y no queremos transmitir la idea de que aceptar implica prueba.

4. Test statistic. The test statistic is some statistic that may be computed from the data of the sample. As a rule, there are many possible values that the test statistic may assume, the particular value observed depending on the particular sample drawn. As we will see, the test statistic serves as a decision maker, since the decision to reject or not to reject the null hypothesis depends on the magnitude of the test statistic. An example of a test statistic is the quantity

4. Estadística de prueba. La estadística de prueba es una estadística que puede calcularse a partir de los datos de la muestra. Como regla general, hay muchos valores posibles que puede asumir la estadística de prueba, dependiendo el valor particular observado de la muestra particular extraída. Como veremos, el estadístico de prueba sirve para tomar decisiones, ya que la decisión de rechazar o no rechazar la hipótesis nula depende de la magnitud del estadístico de prueba. Un ejemplo de estadístico de prueba es la cantidad

z =

x - m0

s> 1n

(7.1.1)

where m0 is a hypothesized value of a population mean. This test statistic is related to the statistic

donde m0 es un valor hipotético de una media poblacional. Esta estadística de prueba está relacionada con la estadística

z =

with which we are already familiar.

con el que ya estamos familiarizados.

x - m

s> 1n

(7.1.2)

**General Formula for Test Statistic**

The following is a general formula for a test statistic that will be applicable in many of the hypothesis tests discussed in this book:

La siguiente es una fórmula general para un estadístico de prueba que será aplicable en muchas de las pruebas de hipótesis analizadas en este libro:

test statistic = relevant statistic - hypothesized parameter

standard error of the relevant statistic

In Equation 7.1.1, x is the relevant statistic, m0 is the hypothesized parameter, and s>1n

is the standard error of x, the relevant statistic.

En la ecuación 7.1.1, x es el estadístico relevante, m0 es el parámetro hipotético y s>1n es el error estándar de x, el estadístico relevante.

5. **Distribution of test statistic**. It has been pointed out that the key to statistical inference is the sampling distribution. We are reminded of this again when it becomes necessary to specify the probability distribution of the test statistic. The distribution of the test statistic

5. Distribución del estadístico de prueba. Se ha señalado que la clave de la inferencia estadística es la distribución muestral. Se nos recuerda esto nuevamente cuando se hace necesario especificar la distribución de probabilidad del estadístico de prueba. La distribución del estadístico de prueba.

z =

x - m0

s>1n

for example, follows the standard normal distribution if the null hypothesis is true and the assumptions are met.

6. **Decision rule.** All possible values that the test statistic can assume are points on the horizontal axis of the graph of the distribution of the test statistic and are divided into two groups; one group constitutes what is known as the rejection region and the other group makes up the nonrejection region. The values of the test statistic forming the rejection region are those values that are less likely to occur if the null hypothesis is true, while the values making up the acceptance region are more likely to occur if the null hypothesis is true. The decision rule tells us to reject the null hypothesis if the value of the test statistic that we compute from our sample is one of the values in the rejection region and to not reject the null hypothesis if the computed value of the test statistic is one of the values in the nonrejection region.

Regla de decisión. Todos los valores posibles que puede asumir el estadístico de prueba son puntos en el eje horizontal del gráfico de distribución del estadístico de prueba y se dividen en dos grupos; un grupo constituye lo que se conoce como región de rechazo y el otro grupo constituye la región de no rechazo. Los valores del estadístico de prueba que forman la región de rechazo son aquellos valores que tienen menos probabilidad de ocurrir si la hipótesis nula es verdadera, mientras que los valores que componen la región de aceptación tienen mayor probabilidad de ocurrir si la hipótesis nula es verdadera. La regla de decisión nos dice que rechacemos la hipótesis nula si el valor del estadístico de prueba que calculamos a partir de nuestra muestra es uno de los valores en la región de rechazo y que no rechacemos la hipótesis nula si el valor calculado del estadístico de prueba es uno de los valores en la región de no rechazo.

**Significance Level** The decision as to which values go into the rejection region and which ones go into the nonrejection region is made on the basis of the desired level of significance, designated by a. The term level of significance reflects the fact that hypothesis tests are sometimes called significance tests, and a computed value of the test statistic that falls in the rejection region is said to be significant. The level of significance, a, specifies the area under the curve of the distribution of the test statistic that is above the values on the horizontal axis constituting the rejection region.

Nivel de significancia La decisión sobre qué valores van a la región de rechazo y cuáles van a la región de no rechazo se toma en base al nivel de significancia deseado, designado por a. El término nivel de significancia refleja el hecho de que las pruebas de hipótesis a veces se denominan pruebas de significancia y se dice que un valor calculado del estadístico de prueba que cae en la región de rechazo es significativo. El nivel de significancia, a, especifica el área bajo la curva de distribución del estadístico de prueba que está por encima de los valores en el eje horizontal que constituye la región de rechazo.

**DEFINITION**

The level of significance A is a probability and, in fact, is the probability of rejecting a true null hypothesis.

Since to reject a true null hypothesis would constitute an error, it seems only reasonable that we should make the probability of rejecting a true null hypothesis small and, in fact, that is what is done. We select a small value of a in order to make the probability of rejecting a true null hypothesis small. The more frequently encountered values of a are .01, .05, and .10.

Dado que rechazar una hipótesis nula verdadera constituiría un error, parece razonable que hagamos pequeña la probabilidad de rechazar una hipótesis nula verdadera y, de hecho, eso es lo que se hace. Seleccionamos un valor pequeño de a para hacer pequeña la probabilidad de rechazar una hipótesis nula verdadera. Los valores de a que se encuentran con más frecuencia son 0,01, 0,05 y 0,10.

**Types of Errors** The error committed when a true null hypothesis is rejected is called the type I error. The type II error is the error committed when a false null hypothesis is not rejected. The probability of committing a type II error is designated by b .

Tipos de errores El error cometido cuando se rechaza una hipótesis nula verdadera se denomina error tipo I. El error tipo II es el error que se comete cuando no se rechaza una hipótesis nula falsa. La probabilidad de cometer un error tipo II se designa por b.

Whenever we reject a null hypothesis there is always the concomitant risk of committing a type I error, rejecting a true null hypothesis. Whenever we fail to reject a null hypothesis the risk of failing to reject a false null hypothesis is always present. We make a small, but we generally exercise no control over b , although we know that in most practical situations it is larger than a.

Siempre que rechazamos una hipótesis nula siempre existe el riesgo concomitante de cometer un error de tipo I, rechazando una hipótesis nula verdadera. Siempre que no logramos rechazar una hipótesis nula, siempre está presente el riesgo de no rechazar una hipótesis nula falsa. Hacemos a pequeño, pero generalmente no ejercemos control sobre b, aunque sabemos que en la mayoría de las situaciones prácticas es mayor que a.

We never know whether we have committed one of these errors when we reject or fail to reject a null hypothesis, since the true state of affairs is unknown. If the testing procedure leads to rejection of the null hypothesis, we can take comfort from the fact that we made a small and, therefore, the probability of committing a type I error was small. If we fail to reject the null hypothesis, we do not know the concurrent risk of committing a type II error, since b is usually unknown but, as has been pointed out, we do know that, in most practical situations, it is larger than a.

Nunca sabemos si hemos cometido uno de estos errores cuando rechazamos o no rechazamos una hipótesis nula, ya que se desconoce el verdadero estado de las cosas. Si el procedimiento de prueba conduce al rechazo de la hipótesis nula, podemos consolarnos con el hecho de que cometimos un error pequeño y, por lo tanto, la probabilidad de cometer un error de tipo I fue pequeña. Si no logramos rechazar la hipótesis nula, no conocemos el riesgo concurrente de cometer un error de tipo II, ya que b generalmente se desconoce pero, como se ha señalado, sí sabemos que, en la mayoría de las situaciones prácticas, es mayor que a.

Figure 7.1.1 shows for various conditions of a hypothesis test the possible actions that an investigator may take and the conditions under which each of the two types of error will be made. The table shown in this figure is an example of what is generally referred to as a confusion matrix.

La figura 7.1.1 muestra, para diversas condiciones de una prueba de hipótesis, las posibles acciones que puede realizar un investigador y las condiciones bajo las cuales se cometerá cada uno de los dos tipos de error. La tabla que se muestra en esta figura es un ejemplo de lo que generalmente se conoce como matriz de confusión.

7. **Calculation of test statistic.** From the data contained in the sample we compute a value of the test statistic and compare it with the rejection and nonrejection regions that have already been specified.

Condition of Null Hypothesis

Possible

Action

FIGURE 7.1.1

TrueFalse

Fail to

reject H0Correct actionType II error

Reject H0Type I errorCorrect action

Conditions under which type I and type II errors may be committed.7.1 INTRODUCTION

221

8. **Statistical decision.** The statistical decision consists of rejecting or of not rejecting the null hypothesis. It is rejected if the computed value of the test statistic falls in the rejection region, and it is not rejected if the computed value of the test statistic falls in the nonrejection region.

Decisión estadística. La decisión estadística consiste en rechazar o no rechazar la hipótesis nula. Se rechaza si el valor calculado del estadístico de prueba cae en la región de rechazo y no se rechaza si el valor calculado del estadístico de prueba cae en la región de no rechazo.

9. **Conclusion**. If H0 is rejected, we conclude that HA is true. If H0 is not rejected, we conclude that H0 may be true.

Conclusión. Si se rechaza H0, concluimos que HA es verdadera. Si no se rechaza H0, concluimos que H0 puede ser verdadera.

10. **p values.** The p value is a number that tells us how unusual our sample results are, given that the null hypothesis is true. A p value indicating that the sample results are not likely to have occurred, if the null hypothesis is true, provides justification for doubting the truth of the null hypothesis.

valores p. El valor p es un número que nos dice qué tan inusuales son los resultados de nuestra muestra, dado que la hipótesis nula es verdadera. Un valor de p que indica que es poco probable que los resultados de la muestra hayan ocurrido, si la hipótesis nula es verdadera, proporciona una justificación para dudar de la verdad de la hipótesis nula.

**DEFINITION**

A p value is the probability that the computed value of a test statistic is at least as extreme as a specified value of the test statistic when the null hypothesis is true. Thus, the p value is the smallest value of A for which we can reject a null hypothesis.

Un valor p es la probabilidad de que el valor calculado de un estadístico de prueba sea al menos tan extremo como un valor específico del estadístico de prueba cuando la hipótesis nula es verdadera. Por tanto, el valor p es el valor más pequeño de A para el cual podemos rechazar una hipótesis nula.

We emphasize that when the null hypothesis is not rejected one should not say that the null hypothesis is accepted. We should say that the null hypothesis is “not rejected.” We avoid using the word “accept” in this case because we may have committed a type II error. Since, frequently, the probability of committing a type II error can be quite high, we do not wish to commit ourselves to accepting the null hypothesis.

Destacamos que cuando no se rechaza la hipótesis nula no se debe decir que se acepta la hipótesis nula. Deberíamos decir que la hipótesis nula “no se rechaza”. Evitamos utilizar la palabra “aceptar” en este caso porque podemos haber cometido un error de tipo II. Dado que, frecuentemente, la probabilidad de cometer un error de tipo II puede ser bastante alta, no deseamos comprometernos a aceptar la hipótesis nula.

Figure 7.1.2 is a flowchart of the steps that we follow when we perform a hypothesis test.

**Purpose of Hypothesis Testing** The purpose of hypothesis testing is to assist administrators and clinicians in making decisions. The administrative or clinical decision usually depends on the statistical decision. If the null hypothesis is rejected, the administrative or clinical decision usually reflects this, in that the decision is compatible with the alternative hypothesis. The reverse is usually true if the null hypothesis is not rejected. The administrative or clinical decision, however, may take other forms, such as a decision to gather more data.

Propósito de la prueba de hipótesis El propósito de la prueba de hipótesis es ayudar a los administradores y médicos a tomar decisiones. La decisión administrativa o clínica suele depender de la decisión estadística. Si se rechaza la hipótesis nula, la decisión administrativa o clínica suele reflejarlo, en el sentido de que la decisión es compatible con la hipótesis alternativa. Lo contrario suele ser cierto si no se rechaza la hipótesis nula. Sin embargo, la decisión administrativa o clínica puede adoptar otras formas, como la decisión de recopilar más datos.

We also emphasize that the hypothesis testing procedures highlighted in the remainder of this chapter generally examine the case of normally distributed data or cases where the procedures are appropriate because the central limit theorem applies. In practice, it is not uncommon for samples to be small relative to the size of the population, or to have samples that are highly skewed, and hence the assumption of normality is violated. Methods to handle this situation, that is distribution-free or nonparametric methods, are examined in detail in Chapter 13. Most computer packages include an analytical procedure (for example, the Shapiro-Wilk or Anderson-Darling test) for testing normality.

También enfatizamos que los procedimientos de prueba de hipótesis destacados en el resto de este capítulo generalmente examinan el caso de datos distribuidos normalmente o casos en los que los procedimientos son apropiados porque se aplica el teorema del límite central. En la práctica, no es raro que las muestras sean pequeñas en relación con el tamaño de la población, o que tengan muestras muy sesgadas y, por tanto, se viole el supuesto de normalidad. Los métodos para manejar esta situación, es decir, métodos sin distribución o no paramétricos, se examinan en detalle en el capítulo 13. La mayoría de los paquetes de computadora incluyen un procedimiento analítico (por ejemplo, la prueba de Shapiro-Wilk o Anderson-Darling) para probar la normalidad.

It is important that such tests are carried out prior to analysis of data. Further, when testing two samples, there is an implicit assumption that the variances are equal. Tests for this assumption are provided in Section 7.8. Finally, it should be noted that hypothesis tests, just like confidence intervals, are relatively sensitive to the size of the samples being tested, and caution should be taken when interpreting results involving very small sample sizes.

Es importante que dichas pruebas se realicen antes del análisis de los datos. Además, cuando se prueban dos muestras, existe la suposición implícita de que las varianzas son iguales. Las pruebas para este supuesto se proporcionan en la Sección 7.8. Finalmente, cabe señalar que las pruebas de hipótesis, al igual que los intervalos de confianza, son relativamente sensibles al tamaño de las muestras que se prueban, y se debe tener precaución al interpretar resultados que involucran tamaños de muestra muy pequeños.

We must emphasize at this point, however, that the outcome of the statistical test is only one piece of evidence that influences the administrative or clinical decision. The statistical decision should not be interpreted as definitive but should be considered along with all the other relevant information available to the experimenter. With these general comments as background, we now discuss specific hypothesis tests.

Sin embargo, debemos enfatizar en este punto que el resultado de la prueba estadística es sólo una pieza de evidencia que influye en la decisión administrativa o clínica. La decisión estadística no debe interpretarse como definitiva sino que debe considerarse junto con toda la demás información relevante disponible para el experimentador. Con estos comentarios generales como antecedente, ahora analizamos las pruebas de hipótesis específicas.

7.2 HYPOTHESIS TESTING: A SINGLE POPULATION MEAN

In this section we consider the testing of a hypothesis about a population mean under three different conditions: (1) when sampling is from a normally distributed population of values with known variance; (2) when sampling is from a normally distributed population with unknown variance, and (3) when sampling is from a population that is not normally distributed. Although the theory for conditions 1 and 2 depends on normally distributed populations, it is common practice to make use of the theory when relevant populations are only approximately normally distributed. This is satisfactory as long as the departure from normality is not drastic. When sampling is from a normally distributed population and the population variance is known, the test statistic for testing H0: m = m0 is

En esta sección consideramos la prueba de una hipótesis sobre una media poblacional bajo tres condiciones diferentes: (1) cuando el muestreo se realiza a partir de una población de valores normalmente distribuida con varianza conocida; (2) cuando el muestreo se realiza a partir de una población distribuida normalmente con varianza desconocida, y (3) cuando el muestreo se realiza a partir de una población que no se distribuye normalmente. Aunque la teoría para las condiciones 1 y 2 depende de poblaciones distribuidas normalmente, es una práctica común hacer uso de la teoría cuando las poblaciones relevantes solo están distribuidas aproximadamente normalmente. Esto es satisfactorio siempre que la desviación de la normalidad no sea drástica. Cuando el muestreo se realiza a partir de una población distribuida normalmente y se conoce la varianza de la población, el estadístico de prueba para probar H0: m = m0 es

z =

x - m

s>1n

(7.2.1)

which, when H0 is true, is distributed as the standard normal. Examples 7.2.1 and 7.2.2 illustrate hypothesis testing under these conditions.

**Sampling from Normally Distributed Populations: Population Variances Known**

As we did in Chapter 6, we again emphasize that situations in which the variable of interest is normally distributed with a known variance are rare. The following example, however, will serve to illustrate the procedure.

Como hicimos en el capítulo 6, nuevamente enfatizamos que las situaciones en las que la variable de interés tiene una distribución normal con una varianza conocida son raras. Sin embargo, el siguiente ejemplo servirá para ilustrar el procedimiento.

EXAMPLE 7.2.1

Researchers are interested in the mean age of a certain population. Let us say that they are asking the following question: Can we conclude that the mean age of this population is different from 30 years?

Los investigadores están interesados en la edad media de una determinada población. Digamos que se hacen la siguiente pregunta: ¿Podemos concluir que la edad media de esta población es diferente de 30 años?

Solution:

Based on our knowledge of hypothesis testing, we reply that they can conclude that the mean age is different from 30 if they can reject the null hypothesis that the mean is equal to 30. Let us use the ten-step hypothesis testing procedure given in the previous section to help the researchers reach a conclusion.

Basándonos en nuestro conocimiento de las pruebas de hipótesis, respondemos que pueden concluir que la edad media es diferente de 30 si pueden rechazar la hipótesis nula de que la media es igual a 30. Usemos el procedimiento de prueba de hipótesis de diez pasos dado en el sección anterior para ayudar a los investigadores a llegar a una conclusión.

1. **Data**. The data available to the researchers are the ages of a simple random sample of 10 individuals drawn from the population of interest. From this sample a mean of x = 27 has been computed.

Los datos disponibles para los investigadores son las edades de una muestra aleatoria simple de 10 individuos extraídos de la población de interés. A partir de esta muestra se ha calculado una media de x = 27.

2. **Assumptions**. It is assumed that the sample comes from a population whose ages are approximately normally distributed. Let us also assume that the population has a known variance of s2 = 20.

Suposiciones. Se supone que la muestra proviene de una población cuyas edades se distribuyen aproximadamente normalmente. Supongamos también que la población tiene una varianza conocida de s2 = 20.

3. **Hypotheses**. The hypothesis to be tested, or null hypothesis, is that the mean age of the population is equal to 30. The alternative hypothesis is that the mean age of the population is not equal to 30. Note that we are identifying with the alternative hypothesis the conclusion the researchers wish to reach, so that if the data permit rejection of the null hypothesis, the researchers’ conclusion will carry more weight, since the accompanying probability of rejecting a true null hypothesis will be small. We will make sure of this by assigning a small value to a, the probability of committing a type I error. We may present the relevant hypotheses in compact form as follows:

Hipótesis. La hipótesis a probar, o hipótesis nula, es que la edad media de la población es igual a 30. La hipótesis alternativa es que la edad media de la población no es igual a 30. Tenga en cuenta que estamos identificando con la hipótesis alternativa la conclusión a la que los investigadores desean llegar, de modo que si los datos permiten el rechazo de la hipótesis nula, la conclusión de los investigadores tendrá más peso, ya que la probabilidad que la acompaña de rechazar una hipótesis nula verdadera será pequeña. De ello nos aseguraremos asignando un valor pequeño a a, la probabilidad de cometer un error tipo I. Podemos presentar las hipótesis relevantes en forma compacta de la siguiente manera:

H0: m = 30

HA: m Z 30

4. **Test statistic.** Since we are testing a hypothesis about a population mean, since we assume that the population is normally distributed, and since the population variance is known, our test statistic is given by Equation 7.2.1.

Estadística de prueba. Dado que estamos probando una hipótesis sobre una media poblacional, dado que asumimos que la población tiene una distribución normal y dado que se conoce la varianza de la población, nuestro estadístico de prueba viene dado por la Ecuación 7.2.1.

5. Distribution of test statistic. Based on our knowledge of sampling distributions and the normal distribution, we know that the test statistic is normally distributed with a mean of 0 and a variance of 1, if H0 is true. There are many possible values of the test statistic that the present situation can generate; one for every possible sample of size 10 that can be drawn from the population. Since we draw only one sample, we have only one of these possible values on which to base a decision.

Distribución del estadístico de prueba. Con base en nuestro conocimiento de las distribuciones muestrales y la distribución normal, sabemos que el estadístico de prueba tiene una distribución normal con una media de 0 y una varianza de 1, si H0 es verdadera. Hay muchos valores posibles del estadístico de prueba que la situación actual puede generar; uno por cada muestra posible de tamaño 10 que pueda extraerse de la población. Como sólo tomamos una muestra, sólo tenemos uno de estos valores posibles en el que basar una decisión.

6. **Decision rule.** The decision rule tells us to reject H0 if the computed value of the test statistic falls in the rejection region and to fail to reject H0 if it falls in the nonrejection region. We must now specify the rejection and nonrejection regions. We can begin by asking ourselves what magnitude of values of the test statistic will cause rejection of H0. If the null hypothesis is false, it may be so either because the population mean is less than 30 or because the population mean is greater than 30. Therefore, either sufficiently small values or sufficiently large values of the test statistic will cause rejection of the null hypothesis. We want these extreme values to constitute the rejection region. How extreme must a possible value of the test statistic be to qualify for the rejection region?

Regla de decisión. La regla de decisión nos dice que rechacemos H0 si el valor calculado del estadístico de prueba cae en la región de rechazo y que no rechacemos H0 si cae en la región de no rechazo. Ahora debemos especificar las regiones de rechazo y no rechazo. Podemos comenzar preguntándonos qué magnitud de los valores del estadístico de prueba provocarán el rechazo de H0. Si la hipótesis nula es falsa, puede serlo porque la media poblacional es menor que 30 o porque la media poblacional es mayor que 30. Por lo tanto, valores suficientemente pequeños o suficientemente grandes del estadístico de prueba provocarán el rechazo de la hipótesis nula. hipótesis. Queremos que estos valores extremos constituyan la región del rechazo. ¿Qué tan extremo debe ser un valor posible del estadístico de prueba para calificar para la región de rechazo?

The answer depends on the significance level we choose, that is, the size of the probability of committing a type I error. Let us say that we want the probability of rejecting a true null hypothesis to be a = .05. Since our rejection region is to consist of two parts, sufficiently small values and sufficiently large values of the test statistic, part of a will have to be associated with the large values and part with the small values. It seems reasonable that we should divide a equally and let a>2 = .025 be associated with small values and a>2 = .025 be associated with large values.

La respuesta depende del nivel de significancia que elijamos, es decir, del tamaño de la probabilidad de cometer un error de tipo I. Digamos que queremos que la probabilidad de rechazar una hipótesis nula verdadera sea a = 0,05. Dado que nuestra región de rechazo constará de dos partes, valores suficientemente pequeños y valores suficientemente grandes del estadístico de prueba, parte de a tendrá que asociarse con los valores grandes y parte con los valores pequeños. Parece razonable dividir a en partes iguales y dejar que a>2 = 0,025 se asocie con valores pequeños y a>2 = 0,025 se asocie con valores grandes.

Critical Value of Test Statistic

What value of the test statistic is so large that, when the null hypothesis is true, the probability of obtaining a value this large or larger is .025? In other words, what is the value of z to the right of which lies .025 of the area under the standard normal distribution? The value of z to the right of which lies .025 of the area is the same value that has .975 of the area between it and - q . We look in the body of Appendix Table D until we find .975 or its closest value and read the corresponding marginal entries to obtain our z value. In the present example the value of z is 1.96. Similar reasoning will lead us to find -1.96 as the value of the test statistic so small that when the null hypothesis is true, the probability of obtaining a value this small or smaller is .025.

¿Qué valor del estadístico de prueba es tan grande que, cuando la hipótesis nula es verdadera, la probabilidad de obtener un valor tan grande o mayor es .025? En otras palabras, ¿cuál es el valor de z a cuya derecha se encuentra 0,025 del área bajo la distribución normal estándar? El valor de z a la derecha del cual se encuentra .025 del área es el mismo valor que tiene .975 del área entre él y -q. Buscamos en el cuerpo de la Tabla D del Apéndice hasta encontrar .975 o su valor más cercano y leemos las entradas marginales correspondientes para obtener nuestro valor z. En el presente ejemplo, el valor de z es 1,96. Un razonamiento similar nos llevará a encontrar -1,96 como el valor del estadístico de prueba tan pequeño que cuando la hipótesis nula es verdadera, la probabilidad de obtener un valor tan pequeño o menor es 0,025.

Our rejection region, then, consists of all values of the test statistic equal to or greater than 1.96 and less than or equal to -1.96. The nonrejection region consists of all values in between. We may state the decision rule for this test as follows: reject H0 if the computed value of the test statistic is either Ú 1.96 or … -1.96. Otherwise, do not reject H0. The rejection and nonrejection regions are shown in Figure 7.2.1. The values of the test statistic that separate the rejection and nonrejection regions are called critical values of the test statistic, and the rejection region is sometimes referred to as the critical region.

Nuestra región de rechazo, entonces, consta de todos los valores del estadístico de prueba iguales o mayores que 1,96 y menores o iguales a -1,96. La región de no rechazo consta de todos los valores intermedios. Podemos enunciar la regla de decisión para esta prueba de la siguiente manera: rechazar H0 si el valor calculado del estadístico de prueba es Ú 1,96 o... -1,96. De lo contrario, no rechace H0. Las regiones de rechazo y no rechazo se muestran en la Figura 7.2.1. Los valores del estadístico de prueba que separan las regiones de rechazo y de no rechazo se denominan valores críticos del estadístico de prueba, y la región de rechazo a veces se denomina región crítica.

The decision rule tells us to compute a value of the test statistic from the data of our sample and to reject H0 if we get a value that is either equal to or greater than 1.96 or equal to or less than -1.96 and to fail to reject H0 if we get any other value. The value of a and, hence, the decision rule should be decided on before gathering the data. This prevents our being accused of allowing the sample results to influence our choice of a. This condition of objectivity is highly desirable and should be preserved in all tests.

La regla de decisión nos dice que calculemos un valor del estadístico de prueba a partir de los datos de nuestra muestra y rechacemos H0 si obtenemos un valor igual o mayor que 1,96 o igual o menor que -1,96 y que no rechacemos H0 si obtenemos cualquier otro valor. El valor de a y, por tanto, la regla de decisión deben decidirse antes de recopilar los datos. Esto evita que se nos acuse de permitir que los resultados de la muestra influyan en nuestra elección de a. Esta condición de objetividad es muy deseable y debe preservarse en todas las pruebas.

7. **Calculation of test statistic.** From our sample we compute

z =

27 - 30

-3

=

= -2.12

1.4142

120>10

8. **Statistical decision.** Abiding by the decision rule, we are able to reject the null hypothesis since -2.12 is in the rejection region. We can say that the computed value of the test statistic is significant at the .05 level.

Decisión estadística. Siguiendo la regla de decisión, podemos rechazar la hipótesis nula ya que -2,12 se encuentra en la región de rechazo. Podemos decir que el valor calculado del estadístico de prueba es significativo al nivel de 0,05.

9. **Conclusion.** We conclude that m is not equal to 30 and let our administrative or clinical actions be in accordance with this conclusion.

Conclusión. Concluimos que m no es igual a 30 y dejamos que nuestras acciones administrativas o clínicas estén de acuerdo con esta conclusión.

10. **p values.** Instead of saying that an observed value of the test statistic is significant or is not significant, most writers in the research literature prefer to report the exact probability of getting a value as extreme as or more extreme than that observed if the null hypothesis is true. In the present instance these writers would give the computed value of the test statistic along with the statement p = .0340. The statement p = .0340 means that the probability of getting a value as extreme as 2.12 in either direction, when the null hypothesis is true, is .0340. The value .0340 is obtained from Appendix Table D and is the probability of observing a z Ú 2.12 or a z … -2.12 when the null hypothesis is true. That is, when H0 is true, the probability of obtaining a value of z as large as or larger than 2.12 is .0170, and the probability of observing a value of z as small as or smaller than -2.12 is .0170. The probability of one or the other of these events occurring, when H0 is true, is equal to the sum of the two individual probabilities, and hence, in the present example, we say that p = .0170 + .0170 = .0340.

valores p. En lugar de decir que un valor observado del estadístico de prueba es significativo o no, la mayoría de los autores de la literatura de investigación prefieren informar la probabilidad exacta de obtener un valor tan extremo o más extremo que el observado si la hipótesis nula es verdadera. En el presente caso, estos autores darían el valor calculado del estadístico de prueba junto con la afirmación p = 0,0340. La afirmación p = 0,0340 significa que la probabilidad de obtener un valor tan extremo como 2,12 en cualquier dirección, cuando la hipótesis nula es verdadera, es 0,0340. El valor .0340 se obtiene de la Tabla D del Apéndice y es la probabilidad de observar a z Ú 2.12 o a z … -2.12 cuando la hipótesis nula es verdadera. Es decir, cuando H0 es verdadera, la probabilidad de obtener un valor de z tan grande o mayor que 2,12 es 0,0170, y la probabilidad de observar un valor de z tan pequeño o menor que -2,12 es 0,0170. La probabilidad de que ocurra uno u otro de estos eventos, cuando H0 es verdadera, es igual a la suma de las dos probabilidades individuales y, por tanto, en el presente ejemplo decimos que p = 0,0170 + 0,0170 = 0,0340.

Recall that the p value for a test may be defined also as the smallest value of a for which the null hypothesis can be rejected. Since, in Example 7.2.1, our p value is .0340, we know that we could have chosen an a value as small as .0340 and still have rejected the null hypothesis. If we had chosen an a smaller than .0340, we would not have been able to reject the null hypothesis. A general rule worth remembering, then, is this: if the p value is less than or equal to a, we reject the null hypothesis; if the p value is greater than a, we do not reject the null hypothesis.

Recuerde que el valor p de una prueba también se puede definir como el valor más pequeño de a para el cual se puede rechazar la hipótesis nula. Dado que, en el ejemplo 7.2.1, nuestro valor p es 0,0340, sabemos que podríamos haber elegido un valor de a tan pequeño como 0,0340 y aun así haber rechazado la hipótesis nula. Si hubiéramos elegido un valor menor que .0340, no habríamos podido rechazar la hipótesis nula. Entonces, una regla general que vale la pena recordar es la siguiente: si el valor de p es menor o igual a a, rechazamos la hipótesis nula; si el valor de p es mayor que a, no rechazamos la hipótesis nula.

The reporting of p values as part of the results of an investigation is more informative to the reader than such statements as “the null hypothesis is rejected at the .05 level of significance” or “the results were not significant at the .05 level.” Reporting the p value associated with a test lets the reader know just how common or how rare is the computed value of the test statistic given that H0 is true.

El informe de los valores p como parte de los resultados de una investigación es más informativo para el lector que afirmaciones como “la hipótesis nula se rechaza al nivel de significancia de 0,05” o “los resultados no fueron significativos al nivel de 0,05”. " Informar el valor p asociado con una prueba le permite al lector saber qué tan común o raro es el valor calculado del estadístico de prueba dado que H0 es verdadera.

**Testing H0 by Means of a Confidence Interval**

Earlier, we stated that one can use confidence intervals to test hypotheses. In Example 7.2.1 we used a hypothesis testing procedure to test H0: m = 30 against the alternative, HA: m Z 30. We were able to reject H0 because the computed value of the test statistic fell in the rejection region.

Anteriormente dijimos que se pueden utilizar intervalos de confianza para probar hipótesis. En el ejemplo 7.2.1 utilizamos un procedimiento de prueba de hipótesis para probar H0: m = 30 contra la alternativa, HA: m Z 30. Pudimos rechazar H0 porque el valor calculado del estadístico de prueba cayó en la región de rechazo.

Let us see how we might have arrived at this same conclusion by using a 100 11 - a2 percent confidence interval. The 95 percent confidence interval for m is

Veamos cómo podríamos haber llegado a la misma conclusión utilizando un intervalo de confianza de 100 11 - a2 por ciento. El intervalo de confianza del 95 por ciento para m es

27 ; 1.96 120>10

27 ; 1.96 11.4142

27 ; 2.7714

24.2286, 29.7714

Since this interval does not include 30, we say 30 is not a candidate for the mean we are estimating and, therefore, m is not equal to 30 and H0 is rejected. This is the same conclusion reached by means of the hypothesis testing procedure.

Como este intervalo no incluye 30, decimos que 30 no es candidato para la media que estamos estimando y, por lo tanto, m no es igual a 30 y se rechaza H0. Ésta es la misma conclusión a la que se llega mediante el procedimiento de prueba de hipótesis.

If the hypothesized parameter, 30, had been within the 95 percent confidence interval, we would have said that H0 is not rejected at the .05 level of significance. In general, when testing a null hypothesis by means of a two-sided confidence interval, we reject H0 at the a level of significance if the hypothesized parameter is not contained within the 100 11 - a2 percent confidence interval. If the hypothesized parameter is contained within the interval, H0 cannot be rejected at the a level of significance.

Si el parámetro hipotético, 30, hubiera estado dentro del intervalo de confianza del 95 por ciento, habríamos dicho que H0 no se rechaza en el nivel de significancia de 0,05. En general, cuando se prueba una hipótesis nula mediante un intervalo de confianza bilateral, rechazamos H0 en el nivel de significancia a si el parámetro hipotético no está contenido dentro del intervalo de confianza de 100 11 - a2 por ciento. Si el parámetro hipotético está contenido dentro del intervalo, H0 no puede rechazarse en un nivel de significancia.

**One-Sided Hypothesis Tests**

The hypothesis test illustrated by Example 7.2.1 is an example of a two-sided test, so called because the rejection region is split between the two sides or tails of the distribution of the test statistic. A hypothesis test may be one-sided, in which case all the rejection region is in one or the other tail of the distribution. Whether a one-sided or a two-sided test is used depends on the nature of the question being asked by the researcher.

La prueba de hipótesis ilustrada en el ejemplo 7.2.1 es un ejemplo de prueba bilateral, llamada así porque la región de rechazo se divide entre los dos lados o colas de la distribución del estadístico de prueba. Una prueba de hipótesis puede ser unilateral, en cuyo caso toda la región de rechazo está en una u otra cola de la distribución. El uso de una prueba unilateral o bilateral depende de la naturaleza de la pregunta que formula el investigador.

If both large and small values will cause rejection of the null hypothesis, a two-sided test is indicated. When either sufficiently “small” values only or sufficiently “large” values only will cause rejection of the null hypothesis, a one-sided test is indicated.

Si tanto los valores grandes como los pequeños provocarán el rechazo de la hipótesis nula, se indica una prueba bilateral. Cuando sólo valores suficientemente “pequeños” o valores suficientemente “grandes” provocan el rechazo de la hipótesis nula, está indicada una prueba unilateral.

EXAMPLE 7.2.2

Refer to Example 7.2.1. Suppose, instead of asking if they could conclude that m Z 30, the researchers had asked: Can we conclude that m 6 30? To this question we would reply that they can so conclude if they can reject the null hypothesis that m Ú 30.

Consulte el Ejemplo 7.2.1. Supongamos que, en lugar de preguntar si podían concluir que m Z 30, los investigadores hubieran preguntado: ¿Podemos concluir que m 6 30? A esta pregunta responderíamos que pueden concluir eso si pueden rechazar la hipótesis nula de que m Ú 30.

Solution:

Let us go through the ten-step procedure to reach a decision based on a one-sided test.

Repasemos el procedimiento de diez pasos para llegar a una decisión basada en una prueba unilateral.

1. Data. See the previous example.

2. Assumptions. See the previous example.

3. Hypotheses.

H0: m Ú 30

HA: m 6 30

The inequality in the null hypothesis implies that the null hypothesis consists of an infinite number of hypotheses. The test will be made only at the point of equality, since it can be shown that if H0 is rejected when the test is made at the point of equality it would be rejected if the test were done for any other value of m indicated in the null hypothesis.

La desigualdad en la hipótesis nula implica que la hipótesis nula consta de un número infinito de hipótesis. La prueba se realizará sólo en el punto de igualdad, ya que se puede demostrar que si se rechaza H0 cuando se realiza la prueba en el punto de igualdad, se rechazaría si la prueba se hiciera para cualquier otro valor de m indicado en el punto de igualdad. hipótesis nula.

4. Test statistic.

z =

x - m0

s>1n

5. Distribution of test statistic. See the previous example.

6. Decision rule. Let us again use a = .05. To determine where to place the rejection region, let us ask ourselves what magnitude of values would cause rejection of the null hypothesis. If we look at the hypotheses, we see that sufficiently small values would cause rejection and that large values would tend to reinforce the null hypothesis. We will want our rejection region to be where the small values are—at the lower tail of the distribution. This time, since we have a one-sided test, all of a will go in the one tail of the distribution. By consulting Appendix Table D, we find that the value of z to the left of which lies .05 of the area under the standard normal curve is -1.645 after interpolating. Our rejection and nonrejection regions are now specified and are shown in Figure 7.2.2.

Regla de decisión. Usemos nuevamente a = .05. Para determinar dónde ubicar la región de rechazo, preguntémonos qué magnitud de valores provocaría el rechazo de la hipótesis nula. Si analizamos las hipótesis, vemos que valores suficientemente pequeños provocarían rechazo y que valores grandes tenderían a reforzar la hipótesis nula. Querremos que nuestra región de rechazo esté donde están los valores pequeños: en la cola inferior de la distribución. Esta vez, como tenemos una prueba unilateral, todos los a irá en una cola de la distribución. Al consultar la tabla D del apéndice, encontramos que el valor de z a la izquierda del cual se encuentra 0,05 del área bajo la curva normal estándar es -1,645 después de interpolar. Nuestras regiones de rechazo y no rechazo ahora están especificadas y se muestran en la Figura 7.2.2.

Our decision rule tells us to reject H0 if the computed value of the test statistic is less than or equal to - 1.645.

7. Calculation of test statistic. From our data we compute

z =

27 - 30

= -2.12

120>10

8. Statistical decision. We are able to reject the null hypothesis since

-2.12 6 -1.645.

9. Conclusion. We conclude that the population mean is smaller than 30 and act accordingly.

Conclusión. Concluimos que la media poblacional es menor que 30 y actuamos en consecuencia.

10. p value. The p value for this test is .0170, since P1z … -2.122, when H0 is true, is .0170 as given by Appendix Table D when we determine the magnitude of the area to the left of -2.12 under the standard normal curve. One can test a one-sided null hypothesis by means of a one-sided confidence interval. However, we will not cover the construction and interpretation of this type of confidence interval in this book.

valor p. El valor p para esta prueba es 0,0170, ya que P1z... -2,122, cuando H0 es verdadera, es 0,0170 como se indica en la Tabla D del apéndice cuando determinamos la magnitud del área a la izquierda de -2,12 bajo la curva normal estándar. Se puede probar una hipótesis nula unilateral mediante un intervalo de confianza unilateral. Sin embargo, no cubriremos la construcción e interpretación de este tipo de intervalo de confianza en este libro.

If the researcher’s question had been, “Can we conclude that the mean is greater than 30?,” following the above ten-step procedure would have led to a one-sided test with all the rejection region at the upper tail of the distribution of the test statistic and a critical value of +1.645.

Si la pregunta del investigador hubiera sido: "¿Podemos concluir que la media es mayor que 30?", seguir el procedimiento de diez pasos anterior habría conducido a una prueba unilateral con toda la región de rechazo en la cola superior de la distribución de el estadístico de prueba y un valor crítico de +1,645.

**Sampling from a Normally Distributed Population: Population Variance Unknown**

As we have already noted, the population variance is usually unknown in actual situations involving statistical inference about a population mean. When sampling is from an approximately normal population with an unknown variance, the test statistic for testing H0: m = m0 is

Como ya hemos señalado, la varianza poblacional suele ser desconocida en situaciones reales que implican inferencia estadística sobre una media poblacional. Cuando el muestreo es de una población aproximadamente normal con una varianza desconocida, el estadístico de prueba para probar H0: m = m0 es

t =

x - m0

s>1n

(7.2.2)

which, when H0 is true, is distributed as Student’s t with n - 1 degrees of freedom. The following example illustrates the hypothesis testing procedure when the population is assumed to be normally distributed and its variance is unknown. This is the usual situation encountered in practice.

que, cuando H0 es cierta, se distribuye como la t de Student con n - 1 grados de libertad. El siguiente ejemplo ilustra el procedimiento de prueba de hipótesis cuando se supone que la población tiene una distribución normal y se desconoce su varianza. Esta es la situación habitual que se encuentra en la práctica.

EXAMPLE 7.2.3

Nakamura et al. (A-1) studied subjects with medial collateral ligament (MCL) and anterior cruciate ligament (ACL) tears. Between February 1995 and December 1997, 17 consecutive patients with combined acute ACL and grade III MCL injuries were treated by the same physician at the research center. One of the variables of interest was the length of time in days between the occurrence of the injury and the first magnetic resonance imaging (MRI). The data are shown in Table 7.2.1. We wish to know if we can conclude that the mean number of days between injury and initial MRI is not 15 days in a population presumed to be represented by these sample data.

Nakamura et al. (A-1) estudiaron sujetos con desgarros del ligamento colateral medial (LCM) y del ligamento cruzado anterior (LCA). Entre febrero de 1995 y diciembre de 1997, el mismo médico en el centro de investigación trató a 17 pacientes consecutivos con lesiones agudas combinadas del LCA y del LCM de grado III. Una de las variables de interés fue el tiempo en días transcurrido entre la aparición de la lesión y la primera resonancia magnética (RM). Los datos se muestran en la Tabla 7.2.1. Deseamos saber si podemos concluir que el número medio de días entre la lesión y la resonancia magnética inicial no es de 15 días en una población que se supone está representada por estos datos de muestra.

TABLE 7.2.1 Number of Days Until MRI for Subjects with MCL

and ACL Tears

SubjectDaysSubjectDaysSubjectDaysSubjectDays

1

214

96

70

1011

1228

2416

1714

9

3

418

268

94

813

1424

25121021153

Source: Norimasa Nakamura, Shuji Horibe, Yukyoshi Toritsuka, Tomoki Mitsuoka, Hideki Yoshikawa, and

Konsei Shino, “Acute Grade III Medial Collateral Ligament Injury of the Knee Associated with Anterior

Cruciate Ligament Tear,” American Journal of Sports Medicine, 31 (2003), 261–267.230

CHAPTER 7 HYPOTHESIS TESTING

**Solution:**

We will be able to conclude that the mean number of days for the population is not 15 if we can reject the null hypothesis that the population mean is equal to 15.

Podremos concluir que la media de días de la población no es 15 si podemos rechazar la hipótesis nula de que la media poblacional es igual a 15.

1. Data. The data consist of number of days until MRI on 17 subjects as previously described.

2. Assumptions. The 17 subjects constitute a simple random sample from a population of similar subjects. We assume that the number of days until MRI in this population is approximately normally distributed.

Suposiciones. Los 17 sujetos constituyen una muestra aleatoria simple de una población de sujetos similares. Suponemos que el número de días hasta la resonancia magnética en esta población se distribuye aproximadamente normalmente.

3. Hypotheses.

H0: m = 15

HA: m Z 15

4. Test statistic. Since the population variance is unknown, our test statistic is given by Equation 7.2.2.

5. Distribution of test statistic. Our test statistic is distributed as Stu- dent’s t with n - 1 = 17 - 1 = 16 degrees of freedom if H0 is true.

6. Decision rule. Let a = .05. Since we have a two-sided test, we put a>2 = .025 in each tail of the distribution of our test statistic. The t values to the right and left of which .025 of the area lies are 2.1199 and -2.1199. These values are obtained from Appendix Table E. The rejection and nonrejection regions are shown in Figure 7.2.3. The decision rule tells us to compute a value of the test statistic and reject H0 if the computed t is either greater than or equal to 2.1199 or less than or equal to - 2.1199.

Regla de decisión. Sea a = .05. Como tenemos una prueba bilateral, colocamos a>2 = 0,025 en cada cola de la distribución de nuestro estadístico de prueba. Los valores t a la derecha e izquierda de los cuales se encuentra 0,025 del área son 2,1199 y -2,1199. Estos valores se obtienen de la Tabla E del Apéndice. Las regiones de rechazo y no rechazo se muestran en la Figura 7.2.3. La regla de decisión nos dice que calculemos un valor del estadístico de prueba y rechacemos H0 si la t calculada es mayor o igual a 2.1199 o menor o igual a -2.1199.

7. Calculation of test statistic. From our sample data we compute a sample mean of 13.2941 and a sample standard deviation of 8.88654. Substituting these statistics into Equation 7.2.2 gives

t =

13.2941 - 15

-1.7059

= -.791

=

2.1553

8.88654> 117

.95

.025

.025

–2.1199

Rejection region

FIGURE 7.2.3

0

Nonrejection

region

2.1199

t

Rejection region

Rejection and nonrejection regions for Example 7.2.3.7.2 HYPOTHESIS TESTING: A SINGLE POPULATION MEAN

p/2 > .10

p/2 > .10

.10

.10

–1.337 –.791 0 .791 1.337

p > .20

FIGURE 7.2.4

231

t

Determination of p value for Example 7.2.3.

8. Statistical decision. Do not reject H0, since -.791 falls in the nonrejection region.

9. Conclusion. Our conclusion, based on these data, is that the mean of the population from which the sample came may be 15.

10. p value. The exact p value for this test cannot be obtained from Appendix Table E since it gives t values only for selected percentiles. The p value can be stated as an interval, however. We find that - .791 is less than -1.337, the value of t to the left of which lies .10 of the area under the t with 16 degrees of freedom. Consequently, when H0 is true, the probability of obtaining a value of t as small as or smaller than -.791 is greater than .10. That is P1t … -.7912 7 .10. Since the test was two-sided, we must allow for the possibility of a computed value of the test statistic as large in the opposite direction as that observed. Appendix Table E reveals that P1t Ú .7912 7 .10 also. The p value, then, is p 7 .20. Figure 7.2.4 shows the p value for this example.

valor p. El valor p exacto para esta prueba no se puede obtener de la Tabla E del Apéndice ya que proporciona valores t sólo para percentiles seleccionados. Sin embargo, el valor p se puede expresar como un intervalo. Encontramos que -0,791 es menor que -1,337, cuyo valor de t a la izquierda se encuentra a 0,10 del área bajo t con 16 grados de libertad. En consecuencia, cuando H0 es verdadera, la probabilidad de obtener un valor de t tan pequeño como o menor que -0,791 es mayor que 0,10. Eso es P1t… -.7912 7 .10. Dado que la prueba fue bilateral, debemos permitir la posibilidad de que un valor calculado del estadístico de prueba sea tan grande en la dirección opuesta a la observada. La Tabla E del Apéndice revela que P1t Ú .7912 7 .10 también. El valor p, entonces, es p 7 .20. La Figura 7.2.4 muestra el valor p para este ejemplo.

If in the previous example the hypotheses had been

H0: m Ú 15

HA: m 6 15

the testing procedure would have led to a one-sided test with all the rejection region at the lower tail of the distribution, and if the hypotheses had been

el procedimiento de prueba habría conducido a una prueba unilateral con toda la región de rechazo en la cola inferior de la distribución, y si las hipótesis hubieran sido

H0: m … 15

HA: m 7 15

we would have had a one-sided test with all the rejection region at the upper tail of the distribution.

Habríamos tenido una prueba unilateral con toda la región de rechazo en la cola superior de la distribución.

**Sampling from a Population That Is Not Normally Distributed**

If, as is frequently the case, the sample on which we base our hypothesis test about a population mean comes from a population that is not normally distributed, we may, if our sample is large (greater than or equal to 30), take advantage of the central limit theorem and use z = 1x - m02>1s>1n2 as the test statistic. If the population standard deviation is not known, the usual practice is to use the sample standard deviation as an estimate. The test statistic for testing H0: m = m0, then, is

Si, como suele ser el caso, la muestra en la que basamos nuestra prueba de hipótesis sobre una media poblacional proviene de una población que no está distribuida normalmente, podemos, si nuestra muestra es grande (mayor o igual a 30), aprovechar del teorema del límite central y utilice z = 1x - m02>1s>1n2 como estadístico de prueba. Si no se conoce la desviación estándar de la población, la práctica habitual es utilizar la desviación estándar de la muestra como estimación. El estadístico de prueba para probar H0: m = m0, entonces, es

z =

x - m0

s>1n

(7.2.3)

which, when H0 is true, is distributed approximately as the standard normal distribution if n is large. The rationale for using s to replace s is that the large sample, necessary for the central limit theorem to apply, will yield a sample standard deviation that closely approximates s.

que, cuando H0 es verdadera, se distribuye aproximadamente como la distribución normal estándar si n es grande. La razón para usar s para reemplazar s es que la muestra grande, necesaria para que se aplique el teorema del límite central, producirá una desviación estándar muestral que se aproxima mucho a s.

EXAMPLE 7.2.4

The goal of a study by Klingler et al. (A-2) was to determine how symptom recognition and perception influence clinical presentation as a function of race. They characterized symptoms and care-seeking behavior in African-American patients with chest pain seen in the emergency department. One of the presenting vital signs was systolic blood pressure. Among 157 African-American men, the mean systolic blood pressure was 146 mm Hg with a standard deviation of 27. We wish to know if, on the basis of these data, we may conclude that the mean systolic blood pressure for a population of African-American men is greater than 140.

El objetivo de un estudio de Klingler et al. (A-2) fue determinar cómo el reconocimiento y la percepción de los síntomas influyen en la presentación clínica en función de la raza. Caracterizaron los síntomas y el comportamiento de búsqueda de atención en pacientes afroamericanos con dolor en el pecho atendidos en el departamento de emergencias. Uno de los signos vitales de presentación fue la presión arterial sistólica. Entre 157 hombres afroamericanos, la presión arterial sistólica media fue de 146 mm Hg con una desviación estándar de 27. Deseamos saber si, sobre la base de estos datos, podemos concluir que la presión arterial sistólica media para una población de africanos -Los hombres estadounidenses son mayores a 140.

**Solution:**

We will say that the data do provide sufficient evidence to conclude that the population mean is greater than 140 if we can reject the null hypothesis that the mean is less than or equal to 140. The following test may be carried out:

Diremos que los datos proporcionan evidencia suficiente para concluir que la media poblacional es mayor que 140 si podemos rechazar la hipótesis nula de que la media es menor o igual a 140. Se puede realizar la siguiente prueba:

1. Data. The data consist of systolic blood pressure scores for 157 African-American men with x = 146 and s = 27.

Los datos consisten en puntuaciones de presión arterial sistólica de 157 hombres afroamericanos con x = 146 y s = 27.

2. Assumptions. The data constitute a simple random sample from a population of African-American men who report to an emergency department with symptoms similar to those in the sample. We are unwilling to assume that systolic blood pressure values are normally distributed in such a population.

Suposiciones. Los datos constituyen una muestra aleatoria simple de una población de hombres afroamericanos que acuden a un departamento de urgencias con síntomas similares a los de la muestra. No estamos dispuestos a suponer que los valores de presión arterial sistólica se distribuyan normalmente en dicha población.

3. Hypotheses.

H0: m … 140

HA: m 7 140

4. Test statistic. The test statistic is given by Equation 7.2.3, since s is unknown.

5. Distribution of test statistic. Because of the central limit theorem, the test statistic is at worst approximately normally distributed with m = 0 if H0 is true.

Distribución del estadístico de prueba. Debido al teorema del límite central, el estadístico de prueba, en el peor de los casos, tiene una distribución aproximadamente normal con m = 0 si H0 es verdadera.

6. Decision rule. Let a = .05. The critical value of the test statistic is 1.645. The rejection and nonrejection regions are shown in Figure 7.2.5. Reject H0 if computed z Ú 1.645.

Regla de decisión. Sea a = .05. El valor crítico del estadístico de prueba es 1,645. Las regiones de rechazo y no rechazo se muestran en la Figura 7.2.5. Rechace H0 si se calcula z Ú 1,645.

7. Calculation of test statistic.

z =

146 - 140

6

=

= 2.78

2.1548

27> 1157

8. Statistical decision. Reject H0 since 2.78 7 1.645.

9. Conclusion. Conclude that the mean systolic blood pressure for the sampled population is greater than 140.

Conclusión. Concluya que la presión arterial sistólica media de la población muestreada es superior a 140.

10. p value. The p value for this test is 1 - .9973 = .0027, since as shown in Appendix Table D, the area 1.00272 to the right of 2.78 is less than .05, the area to the right of 1.645.

**Procedures for Other Conditions**

If the population variance had been known, the procedure would have been identical to the above except that the known value of s, instead of the sample value s, would have been used in the denominator of the computed test statistic.

Si se hubiera conocido la varianza de la población, el procedimiento habría sido idéntico al anterior excepto que se habría utilizado el valor conocido de s, en lugar del valor muestral s, en el denominador del estadístico de prueba calculado.

Depending on what the investigators wished to conclude, either a two-sided test or a one-sided test, with the rejection region at the lower tail of the distribution, could have been made using the above data.

Dependiendo de lo que los investigadores quisieran concluir, se podría haber realizado una prueba bilateral o una prueba unilateral, con la región de rechazo en la cola inferior de la distribución, utilizando los datos anteriores.

When testing a hypothesis about a single population mean, we may use Figure 6.3.3 to decide quickly whether the test statistic is z or t.

Al probar una hipótesis sobre una media poblacional única, podemos usar la figura 6.3.3 para decidir rápidamente si el estadístico de prueba es z o t.

**Computer Analysis**

To illustrate the use of computers in testing hypotheses we

consider the following example.

**EXAMPLE 7.2.5**

The following are the head circumferences (centimeters) at birth of 15 infants:

Las siguientes son las circunferencias de la cabeza (centímetros) al nacer de 15 bebés:

33.38

34.34

33.46

32.15

33.95

34.13

33.99

33.85

34.45

34.10

34.23

34.19

We wish to test H0 : m = 34.5 against HA: m Z 34.5.

33.97

32.73

34.05234

CHAPTER 7 HYPOTHESIS TESTING

Dialog box:Session command:

Stat ➤ Basic Statistics ➤ 1-Sample tMTB > TTEST 34.5 C1

Type C1 in Samples in columns. Type 34.5

in the Test mean box. Click OK .

Output:

One-Sample T: C1

TEST OF MU " 34.5 VS NOT " 34.5

VARIABLE N

MEAN STDEV SE MEAN

95% CI

T

P

C1

15 33.7980 0.6303

0.1627 (33.4490, 34.1470) &4.31 0.001

FIGURE 7.2.6

**Solution:**

We assume that the assumptions for use of the t statistic are met. We enter the data into Column 1 and proceed as shown in Figure 7.2.6. To indicate that a test is one-sided when in Windows, click on the Options button and then choose “less than” or “greater than” as appropri-

ate in the Alternative box. If z is the appropriate test statistic, we choose

1-Sample z from the Basic Statistics menu. The remainder of the commands

are the same as for the t test.

We learn from the printout that the computed value of the test statis-

tic is -4.31 and the p value for the test is .0007. SAS® users may use the

output from PROC MEANS or PROC UNIVARIATE to perform hypothesis

tests.

When both the z statistic and the t statistic are inappropriate test

statistics for use with the available data, one may wish to use a nonpara-

metric technique to test a hypothesis about a single population measure

of central tendency. One such procedure, the sign test, is discussed in

Chapter 13.

**EXERCISES**

For each of the following exercises carry out the ten-step hypothesis testing procedure for the given

significance level. For each exercise, as appropriate, explain why you chose a one-sided test or a

two-sided test. Discuss how you think researchers and/or clinicians might use the results of your

hypothesis test. What clinical and/or research decisions and/or actions do you think would be

appropriate in light of the results of your test?

7.2.1

Escobar et al. (A-3) performed a study to validate a translated version of the Western Ontario and

McMaster Universities Osteoarthritis Index (WOMAC) questionnaire used with Spanish-speaking

patients with hip or knee osteoarthritis. For the 76 women classified with severe hip pain, theEXERCISES

235

WOMAC mean function score (on a scale from 0 to 100 with a higher number indicating less

function) was 70.7 with a standard deviation of 14.6. We wish to know if we may conclude that

the mean function score for a population of similar women subjects with severe hip pain is less

than 75. Let a = .01.

7.2.2

A study by Thienprasiddhi et al. (A-4) examined a sample of 16 subjects with open-angle glaucoma

and unilateral hemifield defects. The ages (years) of the subjects were:

62

57

62

41

68

62

48

50

51

53

60

34

51

62

57

61

Source: Phamornsak Thienprasiddhi, Vivienne

C. Greenstein, Candice S. Chen, Jeffrey M. Liebmann,

Robert Ritch, and Donald C. Hood, “Multifocal Visual

Evoked Potential Responses in Glaucoma Patients with

Unilateral Hemifield Defects,” American Journal of

Ophthalmology, 136 (2003), 34–40.

Can we conclude that the mean age of the population from which the sample may be presumed

to have been drawn is less than 60 years? Let a = .05.

7.2.3The purpose of a study by Lugliè et al. (A-5) was to investigate the oral status of a group of

patients diagnosed with thalassemia major (TM). One of the outcome measures was the decayed,

missing, and filled teeth index (DMFT). In a sample of 18 patients the mean DMFT index value

was 10.3 with a standard deviation of 7.3. Is this sufficient evidence to allow us to conclude that

the mean DMFT index is greater than 9.0 in a population of similar subjects? Let a = .10.

7.2.4A study was made of a sample of 25 records of patients seen at a chronic disease hospital on an

outpatient basis. The mean number of outpatient visits per patient was 4.8, and the sample stan-

dard deviation was 2. Can it be concluded from these data that the population mean is greater than

four visits per patient? Let the probability of committing a type I error be .05. What assumptions

are necessary?

7.2.5In a sample of 49 adolescents who served as the subjects in an immunologic study, one variable

of interest was the diameter of skin test reaction to an antigen. The sample mean and standard

deviation were 21 and 11 mm erythema, respectively. Can it be concluded from these data that the

population mean is less than 30? Let a = .05.

7.2.6Nine laboratory animals were infected with a certain bacterium and then immunosuppressed. The

mean number of organisms later recovered from tissue specimens was 6.5 (coded data) with a stan-

dard deviation of .6. Can one conclude from these data that the population mean is greater than

6? Let a = .05. What assumptions are necessary?

7.2.7A sample of 25 freshman nursing students made a mean score of 77 on a test designed to meas-

ure attitude toward the dying patient. The sample standard deviation was 10. Do these data pro-

vide sufficient evidence to indicate, at the .05 level of significance, that the population mean is

less than 80? What assumptions are necessary?

7.2.8We wish to know if we can conclude that the mean daily caloric intake in the adult rural popula-

tion of a developing country is less than 2000. A sample of 500 had a mean of 1985 and a stan-

dard deviation of 210. Let a = .05.

7.2.9A survey of 100 similar-sized hospitals revealed a mean daily census in the pediatrics service of

27 with a standard deviation of 6.5. Do these data provide sufficient evidence to indicate that the

population mean is greater than 25? Let a = .05.236

CHAPTER 7 HYPOTHESIS TESTING

7.2.10Following a week-long hospital supervisory training program, 16 assistant hospital administra-

tors made a mean score of 74 on a test administered as part of the evaluation of the training

program. The sample standard deviation was 12. Can it be concluded from these data that the

population mean is greater than 70? Let a = .05. What assumptions are necessary?

7.2.11A random sample of 16 emergency reports was selected from the files of an ambulance service.

The mean time (computed from the sample data) required for ambulances to reach their destina-

tions was 13 minutes. Assume that the population of times is normally distributed with a vari-

ance of 9. Can we conclude at the .05 level of significance that the population mean is greater

than 10 minutes?

7.2.12The following data are the oxygen uptakes (milliliters) during incubation of a random sample of

15 cell suspensions:

14.0, 14.1, 14.5, 13.2, 11.2, 14.0, 14.1, 12.2,

11.1, 13.7, 13.2, 16.0, 12.8, 14.4, 12.9

Do these data provide sufficient evidence at the .05 level of significance that the population mean

is not 12 ml? What assumptions are necessary?

7.2.13

Can we conclude that the mean maximum voluntary ventilation value for apparently healthy col-

lege seniors is not 110 liters per minute? A sample of 20 yielded the following values:

132, 33, 91, 108, 67, 169, 54, 203, 190, 133,

96, 30, 187, 21, 63, 166, 84, 110, 157, 138

Let a = .01. What assumptions are necessary?

7.2.14

The following are the systolic blood pressures (mm Hg) of 12 patients undergoing drug therapy

for hypertension:

183, 152, 178, 157, 194, 163, 144, 114, 178, 152, 118, 158

Can we conclude on the basis of these data that the population mean is less than 165? Let a = .05.

What assumptions are necessary?

7.2.15

Can we conclude that the mean age at death of patients with homozygous sickle-cell disease is

less than 30 years? A sample of 50 patients yielded the following ages in years:

15.5

27.6

6.9

20.7

23.7

2.0

45.0

13.5

30.9

4.8

45.1

1.0

1.9

36.6

33.2

1.7

66.4

31.2

1.1

27.1

.8

2.0

9.0

23.6

36.7

1.1

67.4

2.6

.9

3.2

18.2

2.5

29.7

7.6

38.0

9.7

61.7

13.5

23.5

3.5

28.1

16.2

2.6

6.3

21.8

18.2

31.7

14.4

40.2

2.4

Let a = .05. What assumptions are necessary?

7.2.16

The following are intraocular pressure (mm Hg) values recorded for a sample of 21 elderly subjects:

14.5

16.4

19.6

12.9

24.2

14.0

12.2

16.1

14.4

12.0

17.0

17.5

10.0

14.1

18.5

12.9

20.8

17.9

16.2

12.0

14.9

Can we conclude from these data that the mean of the population from which the sample was

drawn is greater than 14? Let a = .05. What assumptions are necessary?7.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS

237

7.2.17Suppose it is known that the IQ scores of a certain population of adults are approximately nor-

mally distributed with a standard deviation of 15. A simple random sample of 25 adults drawn

from this population had a mean IQ score of 105. On the basis of these data can we conclude that

the mean IQ score for the population is not 100? Let the probability of committing a type I error

be .05.

7.2.18A research team is willing to assume that systolic blood pressures in a certain population of males

are approximately normally distributed with a standard deviation of 16. A simple random sample

of 64 males from the population had a mean systolic blood pressure reading of 133. At the .05

level of significance, do these data provide sufficient evidence for us to conclude that the popula-

tion mean is greater than 130?

7.2.19A simple random sample of 16 adults drawn from a certain population of adults yielded a mean

weight of 63 kg. Assume that weights in the population are approximately normally distributed with

a variance of 49. Do the sample data provide sufficient evidence for us to conclude that the mean

weight for the population is less than 70 kg? Let the probability of committing a type I error be .01.

7.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS

Hypothesis testing involving the difference between two population means is most frequently employed to determine whether or not it is reasonable to conclude that the two population means are unequal. In such cases, one or the other of the following hypotheses may be formulated:

La prueba de hipótesis que implica la diferencia entre dos medias poblacionales se emplea con mayor frecuencia para determinar si es razonable o no concluir que las dos medias poblacionales son desiguales. En tales casos, se podrá formular una u otra de las siguientes hipótesis:

1. H0 : m1 - m2 = 0, HA : m1 - m2 Z 0

2. H0 : m1 - m2 Ú 0, HA : m1 - m2 6 0

3. H0 : m1 - m2 … 0, HA : m1 - m2 7 0

It is possible, however, to test the hypothesis that the difference is equal to, greater than or equal to, or less than or equal to some value other than zero.

Sin embargo, es posible probar la hipótesis de que la diferencia es igual, mayor o igual, o menor o igual a algún valor distinto de cero.

As was done in the previous section, hypothesis testing involving the difference between two population means will be discussed in three different contexts: (1) when sampling is from normally distributed populations with known population variances, (2) when sampling is from normally distributed populations with unknown population variances, and (3) when sampling is from populations that are not normally distributed.

Como se hizo en la sección anterior, las pruebas de hipótesis que involucran la diferencia entre dos medias poblacionales se discutirán en tres contextos diferentes: (1) cuando el muestreo se realiza a partir de poblaciones normalmente distribuidas con varianzas poblacionales conocidas, (2) cuando el muestreo se realiza a partir de poblaciones normalmente distribuidas con varianzas poblacionales desconocidas, y (3) cuando el muestreo se realiza a partir de poblaciones que no están distribuidas normalmente.

**Sampling from Normally Distributed Populations: Population Variances Known**

When each of two independent simple random samples has been drawn from a normally distributed population with a known variance, the test statistic for testing the null hypothesis of equal population means is

Cuando cada una de dos muestras aleatorias simples independientes se ha extraído de una población distribuida normalmente con una varianza conocida, el estadístico de prueba para probar la hipótesis nula de medias poblacionales iguales es

z =

1x 1 - x 22 - 1m1 - m220

s21

s22

+

n2

Cn1

(7.3.1)238

CHAPTER 7 HYPOTHESIS TESTING

where the subscript 0 indicates that the difference is a hypothesized parameter. When H0 is

true the test statistic of Equation 7.3.1 is distributed as the standard normal.

EXAMPLE 7.3.1

Researchers wish to know if the data they have collected provide sufficient evidence to indicate a difference in mean serum uric acid levels between normal individuals and individuals with Down’s syndrome. The data consist of serum uric acid readings on 12 individuals with Down’s syndrome and 15 normal individuals. The means are x1 = 4.5 mg/100 ml and x 2 = 3.4 mg/100 ml.

Los investigadores desean saber si los datos que han recopilado proporcionan evidencia suficiente para indicar una diferencia en los niveles medios de ácido úrico sérico entre individuos normales y personas con síndrome de Down. Los datos consisten en lecturas de ácido úrico sérico de 12 personas con síndrome de Down y 15 personas normales. Las medias son x1 = 4,5 mg/100 ml y x 2 = 3,4 mg/100 ml.

**Solution:**

We will say that the sample data do provide evidence that the population means are not equal if we can reject the null hypothesis that the population means are equal. Let us reach a conclusion by means of the ten-step hypothesis testing procedure.

Diremos que los datos muestrales proporcionan evidencia de que las medias poblacionales no son iguales si podemos rechazar la hipótesis nula de que las medias poblacionales son iguales. Lleguemos a una conclusión mediante el procedimiento de prueba de hipótesis de diez pasos.

1. Data. See problem statement.

2. Assumptions. The data constitute two independent simple random samples each drawn from a normally distributed population with a variance equal to 1 for the Down’s syndrome population and 1.5 for the normal population.

Suposiciones. Los datos constituyen dos muestras aleatorias simples independientes, cada una extraída de una población distribuida normalmente con una varianza igual a 1 para la población con síndrome de Down y 1,5 para la población normal.

3. Hypotheses.

H0 : m1 - m2 = 0

HA : m1 - m2 Z 0

An alternative way of stating the hypotheses is as follows:

H0 : m1 = m2

HA : m1 Z m2

4. Test statistic. The test statistic is given by Equation 7.3.1.

5. Distribution of test statistic. When the null hypothesis is true, the test statistic follows the standard normal distribution.

6. Decision rule. Let a = .05. The critical values of z are ; 1.96. Reject H0 unless -1.96 6 zcomputed 6 1.96. The rejection and nonrejection regions are shown in Figure 7.3.1.

7. Calculation of test statistic.

z =

14.5 - 3.42 - 0

11>12 + 1.5>15

=

1.1

= 2.57

.4282

8. Statistical decision. Reject H0, since 2.57 7 1.96.

9. Conclusion. Conclude that, on the basis of these data, there is an indi-

cation that the two population means are not equal.

10. p value. For this test, p = .0102.

A 95 Percent Confidence Interval for M 1 ! M 2

In the previous chapter the 95 percent confidence interval for m1 - m2, computed from the same data, was found to be .26 to 1.94. Since this interval does not include 0, we say that 0 is not a candidate for the difference between population means, and we conclude that the difference is not zero. Thus we arrive at the same conclusion by means of a confidence interval.

En el capítulo anterior se encontró que el intervalo de confianza del 95 por ciento para m1 - m2, calculado a partir de los mismos datos, era de 0,26 a 1,94. Como este intervalo no incluye a 0, decimos que 0 no es candidato para la diferencia entre medias poblacionales y concluimos que la diferencia no es cero. Así llegamos a la misma conclusión mediante un intervalo de confianza.

Sampling from Normally Distributed Populations: Population Variances Unknown

As we have learned, when the population variances are unknown, two possibilities exist. The two population variances may be equal or they may be unequal. We consider first the case where it is known, or it is reasonable to assume, that they are equal. A test of the hypothesis that two population variances are equal is described in Section 7.8.

Como hemos aprendido, cuando se desconocen las varianzas poblacionales, existen dos posibilidades. Las dos varianzas de la población pueden ser iguales o pueden ser desiguales. Consideramos primero el caso en el que se sabe, o es razonable suponer, que son iguales. En la sección 7.8 se describe una prueba de la hipótesis de que dos varianzas poblacionales son iguales.

Population Variances Equal

When the population variances are unknown, but assumed to be equal, we recall from Chapter 6 that it is appropriate to pool the sample variances by means of the following formula:

sp2 =

1n1 - 12s 21 + 1n 2 - 12s 22

n1 + n 2 - 2

When each of two independent simple random samples has been drawn from a normally distributed population and the two populations have equal but unknown variances, the test statistic for testing H0 : m1 = m2 is given by

Cuando cada una de dos muestras aleatorias simples independientes se ha extraído de una población distribuida normalmente y las dos poblaciones tienen varianzas iguales pero desconocidas, el estadístico de prueba para probar H0: m1 = m2 viene dado por

t =

1x 1 - x 22 - 1m1 - m220

s p2

Dn 1

+

s p2

(7.3.2)

n2

which, when H0 is true, is distributed as Student’s t with n1 + n2 - 2 degrees of freedom.

EXAMPLE 7.3.2

The purpose of a study by Tam et al. (A-6) was to investigate wheelchair maneuvering in individuals with lower-level spinal cord injury (SCI) and healthy controls (C). Subjects used a modified wheelchair to incorporate a rigid seat surface to facilitate the specified experimental measurements. Interface pressure measurement was recorded by using a high-resolution pressure-sensitive mat with a spatial resolution of four sensors per square centimeter taped on the rigid seat support. During static sitting conditions, average pressures were recorded under the ischial tuberosities (the bottom part of the pelvic bones). The data for measurements of the left ischial tuberosity (in mm Hg) for the SCI and control groups are shown in Table 7.3.1. We wish to know if we may conclude, on the basis of these data, that, in general, healthy subjects exhibit lower pressure than SCI subjects.

El propósito de un estudio de Tam et al. (A-6) fue investigar las maniobras de sillas de ruedas en personas con lesión de la médula espinal (LME) de nivel inferior y controles sanos (C). Los sujetos utilizaron una silla de ruedas modificada para incorporar una superficie de asiento rígida para facilitar las mediciones experimentales especificadas. La medición de la presión de la interfaz se registró utilizando una alfombra sensible a la presión de alta resolución con una resolución espacial de cuatro sensores por centímetro cuadrado pegada con cinta adhesiva al soporte rígido del asiento. Durante condiciones de sentado estático, se registraron presiones promedio debajo de las tuberosidades isquiáticas (la parte inferior de los huesos pélvicos). Los datos para las mediciones de la tuberosidad isquiática izquierda (en mm Hg) para los grupos SCI y control se muestran en la Tabla 7.3.1. Deseamos saber si podemos concluir, sobre la base de estos datos, que, en general, los sujetos sanos exhiben una presión más baja que los sujetos con LME.

Solution:

1. Data. See statement of problem.

2. Assumptions. The data constitute two independent simple random samples of pressure measurements, one sample from a population of control subjects and the other sample from a population with lower-level spinal cord injury. We shall assume that the pressure measurements in both populations are approximately normally distributed. The population variances are unknown but are assumed to be equal.

Suposiciones. Los datos constituyen dos muestras aleatorias simples independientes de mediciones de presión, una muestra de una población de sujetos de control y la otra muestra de una población con lesión de la médula espinal de nivel inferior. Supondremos que las mediciones de presión en ambas poblaciones se distribuyen aproximadamente normalmente. Las varianzas poblacionales se desconocen pero se supone que son iguales.

3. Hypotheses. H0: mC Ú mSCI, HA: mC 6 mSCI.

4. Test statistic. The test statistic is given by Equation 7.3.2.

5. Distribution of test statistic. When the null hypothesis is true, the test statistic follows Student’s t distribution with n 1 + n 2 - 2 degrees of freedom.

6. Decision rule. Let a = .05. The critical value of t is -1.7341. Reject

H0 unless t computed 7 -1.7341.

7. Calculation of test statistic. From the sample data we compute

x C = 126.1,

sC = 21.8,

x SCI = 133.1,

sSCI = 32.2

Next, we pool the sample variances to obtain

s 2p =

9121.822 + 9132.222

9 + 9

= 756.04

TABLE 7.3.1 Pressures (mm Hg) Under the Pelvis during

Static Conditions for Example 7.3.2

Control13111512413112211788114150169

SCI60150130180163130121119130148

Source: Eric W. Tam, Arthur F. Mak, Wai Nga Lam, John H. Evans, and York Y. Chow, “Pelvic Movement

and Interface Pressure Distribution During Manual Wheelchair Propulsion,” Archives of Physical Medicine

and Rehabilitation, 84 (2003), 1466–1472.7.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS

241

We now compute

t =

1126.1 - 133.12 - 0

756.04

756.04

+

A 10

10

= -.569

8. Statistical decision. We fail to reject H0, since -1.7341 6 -.569; that is, -.569 falls in the nonrejection region.

9. Conclusion. On the basis of these data, we cannot conclude that the population mean pressure is less for healthy subjects than for SCI subjects.

10. p value. For this test, p 7 .10 since -1.330 6 -.569.

Population Variances Unequal

When two independent simple random samples have been drawn from normally distributed populations with unknown and unequal variances, the test statistic for testing H0: m1 = m2 is

Cuando se extraen dos muestras aleatorias simples independientes de poblaciones distribuidas normalmente con varianzas desconocidas y desiguales, el estadístico de prueba para probar H0: m1 = m2 es

t¿ =

1x 1 - x 22 - 1m1 - m220

s 21

s 22

+

A n1 n2

(7.3.3)

The critical value of t¿ for an a level of significance and a two-sided test is approximately

El valor crítico de t¿ para un nivel de significancia y una prueba bilateral es aproximadamente

t 1¿ -1a>22 =

w1t 1 + w 2t 2

w1 + w2

(7.3.4)

where w1 = s 21 >n 1, w2 = s 22 >n 2, t 1 = t 1-1a>22 for n1 - 1 degrees of freedom, and t 2 = t 1-1a>22 for n 2 - 1 degrees of freedom. The critical value of t¿ for a one-sided test is found by computing t¿1-a by Equation 7.3.4, using t 1 = t 1-a for n1 - 1 degrees of freedom and t 2 = t 1-a for n 2 - 1 degrees of freedom.

For a two-sided test, reject H0 if the computed value of t¿ is either greater than or equal to the critical value given by Equation 7.3.4 or less than or equal to the negative of that value.

Para una prueba bilateral, rechace H0 si el valor calculado de t¿ es mayor o igual al valor crítico dado por la Ecuación 7.3.4 o menor o igual al negativo de ese valor.

For a one-sided test with the rejection region in the right tail of the sampling distribution, reject H0 if the computed t¿ is equal to or greater than the critical t¿. For a one-sided test with a left-tail rejection region, reject H0 if the computed value of t¿ is equal to or smaller than the negative of the critical t¿ computed by the indicated adaptation of Equation 7.3.4.

Para una prueba unilateral con la región de rechazo en la cola derecha de la distribución muestral, rechace H0 si el t¿ calculado es igual o mayor que el t¿ crítico. Para una prueba unilateral con una región de rechazo de cola izquierda, rechace H0 si el valor calculado de t¿ es igual o menor que el negativo del t¿ crítico calculado mediante la adaptación indicada de la Ecuación 7.3.4.

EXAMPLE 7.3.3

Dernellis and Panaretou (A-7) examined subjects with hypertension and healthy control subjects. One of the variables of interest was the aortic stiffness index. Measures of this variable were calculated from the aortic diameter evaluated by M-mode echocardiography and blood pressure measured by a sphygmomanometer. Generally, physicians wish to reduce aortic stiffness. In the 15 patients with hypertension (group 1), the mean aortic stiffness index was 19.16 with a standard deviation of 5.29. In the 30 control subjects (group 2), the mean aortic stiffness index was 9.53 with a standard deviation of 2.69. We wish to determine if the two populations represented by these samples differ with respect to mean aortic stiffness index.

Dernellis y Panaretou (A-7) examinaron sujetos con hipertensión y sujetos de control sanos. Una de las variables de interés fue el índice de rigidez aórtica. Las medidas de esta variable se calcularon a partir del diámetro aórtico evaluado mediante ecocardiografía en modo M y la presión arterial medida con un esfigmomanómetro. Generalmente, los médicos desean reducir la rigidez aórtica. En los 15 pacientes con hipertensión (grupo 1), el índice medio de rigidez aórtica fue de 19,16 con una desviación estándar de 5,29. En los 30 sujetos de control (grupo 2), el índice medio de rigidez aórtica fue de 9,53 con una desviación estándar de 2,69. Deseamos determinar si las dos poblaciones representadas por estas muestras difieren con respecto al índice medio de rigidez aórtica.

**Solution**:

1. Data. The sample sizes, means, and sample standard deviations are:

n 1 = 15, x 1 = 19.16, s1 = 5.29

n 2 = 30, x 2 = 9.53, s2 = 2.69

2. Assumptions. The data constitute two independent random samples, one from a population of subjects with hypertension and the other from a control population. We assume that aortic stiffness values are approximately normally distributed in both populations. The population variances are unknown and unequal.

Suposiciones. Los datos constituyen dos muestras aleatorias independientes, una de una población de sujetos con hipertensión y la otra de una población de control. Suponemos que los valores de rigidez aórtica se distribuyen aproximadamente normalmente en ambas poblaciones. Las variaciones poblacionales son desconocidas y desiguales.

3. Hypotheses.

H0: m1 - m2 = 0

HA: m1 - m2 Z 0

4. Test statistic. The test statistic is given by Equation 7.3.3.

5. Distribution of test statistic. The statistic given by Equation 7.3.3

does not follow Student’s t distribution. We, therefore, obtain its critical

values by Equation 7.3.4.

6. Decision rule. Let a = .05. Before computing t¿ we calculate

w1 = 15.2922>15 = 1.8656 and w 2 = 12.6922>30 = .2412. In Appen-

dix Table E we find that t 1 = 2.1448 and t 2 = 2.0452. By Equation

7.3.4 we compute

t¿ =

1.865612.14482 + .241212.04522

1.8656 + .2412

= 2.133

Our decision rule, then, is reject H0 if the computed t is either Ú 2.133

or … -2.133.

7. Calculation of test statistic. By Equation 7.3.3 we compute

t¿ =

119.16 - 9.532 - 0

2

B

15.292

15

2

+

12.692

30

=

9.63

= 6.63

1.45157.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS

243

8. Statistical decision. Since 6.63 7 2.133, we reject H0.

9. Conclusion. On the basis of these results we conclude that the two

population means are different.

10. p value. For this test p 6 .05.

Sampling from Populations That Are Not Normally Distributed

When sampling is from populations that are not normally distributed, the results of the central limit theorem may be employed if sample sizes are large (say, Ú302. This will allow the use of normal theory since the distribution of the difference between sample means will be approximately normal. When each of two large independent simple random samples has been drawn from a population that is not normally distributed, the test statistic for testing H0: m1 = m2 is

Cuando el muestreo se realiza a partir de poblaciones que no están distribuidas normalmente, se pueden emplear los resultados del teorema del límite central si los tamaños de muestra son grandes (por ejemplo, Ú302). Esto permitirá el uso de la teoría normal ya que la distribución de la diferencia entre las medias de las muestras será aproximadamente normal. Cuando cada una de dos muestras aleatorias simples independientes grandes se ha extraído de una población que no está distribuida normalmente, el estadístico de prueba para probar H0: m1 = m2 es

z =

1x 1 - x 22 - 1m1 - m220

(7.3.5)

s21

s22

+

n2

A n1

which, when H0 is true, follows the standard normal distribution. If the population variances are known, they are used; but if they are unknown, as is the usual case, the sample variances, which are necessarily based on large samples, are used as estimates. Sample variances are not pooled, since equality of population variances is not a necessary assumption when the z statistic is used.

que, cuando H0 es verdadera, sigue la distribución normal estándar. Si se conocen las varianzas poblacionales, se utilizan; pero si se desconocen, como es el caso habitual, se utilizan como estimaciones las varianzas muestrales, que necesariamente se basan en muestras grandes. Las varianzas muestrales no se agrupan, ya que la igualdad de las varianzas poblacionales no es un supuesto necesario cuando se utiliza el estadístico z.

EXAMPLE 7.3.4

The objective of a study by Sairam et al. (A-8) was to identify the role of various disease states and additional risk factors in the development of thrombosis. One focus of the study was to determine if there were differing levels of the anticardiolipin antibody IgG in subjects with and without thrombosis. Table 7.3.2 summarizes the researchers’ findings:

El objetivo de un estudio de Sairam et al. (A-8) fue identificar el papel de diversas enfermedades y factores de riesgo adicionales en el desarrollo de la trombosis. Uno de los objetivos del estudio fue determinar si había diferentes niveles del anticuerpo anticardiolipina IgG en sujetos con y sin trombosis. La Tabla 7.3.2 resume los hallazgos de los investigadores:

TABLE 7.3.2 IgG Levels for Subjects with and Without Thrombosis

for Example 7.3.4

Group

Thrombosis

No thrombosis

Mean IgG Level

(ml/unit)Sample SizeStandard Deviation

59.01

46.6153

5444.89

34.85

Source: S. Sairam, B. A. Baethge and T. McNearney, “Analysis of Risk Factors and

Comorbid Diseases in the Development of Thrombosis in Patients with Anticardiolipin

Antibodies,” Clinical Rheumatology, 22 (2003), 24–29.244

CHAPTER 7 HYPOTHESIS TESTING

We wish to know if we may conclude, on the basis of these results, that, in general, persons with thrombosis have, on the average, higher IgG levels than persons without thrombosis.

Deseamos saber si podemos concluir, sobre la base de estos resultados, que, en general, las personas con trombosis tienen, en promedio, niveles de IgG más altos que las personas sin trombosis.

Solution:

1. Data. See statement of example.

2. Assumptions. The statistics were computed from two independent samples that behave as simple random samples from a population of persons with thrombosis and a population of persons who do not have thrombosis. Since the population variances are unknown, we will use the sample variances in the calculation of the test statistic.

Suposiciones. Las estadísticas se calcularon a partir de dos muestras independientes que se comportan como muestras aleatorias simples de una población de personas con trombosis y una población de personas que no tienen trombosis. Dado que se desconocen las varianzas de la población, utilizaremos las varianzas de la muestra en el cálculo del estadístico de prueba.

3. Hypotheses.

H0: mT - mNT … 0

HA: mT - mNT 7 0

or, alternatively,

H0: mT … mNT

HA: mT 7 mNT

4. Test statistic. Since we have large samples, the central limit theorem allows us to use Equation 7.3.5 as the test statistic.

5. Distribution of test statistic. When the null hypothesis is true, the test statistic is distributed approximately as the standard normal.

6. Decision rule. Let a = .01. This is a one-sided test with a critical

value of z equal to 2.33. Reject H0 if z computed Ú 2.33.

7. Calculation of test statistic.

z =

59.01 - 46.61

44.892

34.852

+

A 53

54

= 1.59

8. Statistical decision. Fail to reject H0, since z = 1.59 is in the nonre-

jection region.

9. Conclusion. These data indicate that on the average, persons with

thrombosis and persons without thrombosis may not have differing IgG

levels.

10. p value. For this test, p = .0559. When testing a hypothesis about the

difference between two populations means, we may use Figure 6.4.1 to

decide quickly whether the test statistic should be z or t.

■

We may use MINITAB to perform two-sample t tests. To illustrate, let us

refer to the data in Table 7.3.1. We put the data for control subjects and spinal cord

injury subjects in Column 1 and Column 2, respectively, and proceed as shown in

Figure 7.3.2.

The SAS® statistical package performs the t test for equality of population means

under both assumptions regarding population variances: that they are equal and that they

are not equal. Note that SAS® designates the p value as Pr 7 ƒ t ƒ . The default output is

a p value for a two-sided test. The researcher using SAS® must divide this quantity in

half when the hypothesis test is one-sided. The SAS® package also tests for equality of

population variances as described in Section 7.8. Figure 7.3.3 shows the SAS® output

for Example 7.3.2.

**Alternatives to z and t**

Sometimes neither the z statistic nor the t statistic is an appropriate test statistic for use with the available data. When such is the case, one may wish to use a nonparametric technique for testing a hypothesis about the difference between two population measures of central tendency. The Mann-Whitney test statistic and the median test, discussed in Chapter 13, are frequently used alternatives to the z and t statistics.

A veces, ni el estadístico z ni el estadístico t son estadísticos de prueba apropiados para usar con los datos disponibles. Cuando tal sea el caso, es posible que deseemos utilizar una técnica no paramétrica para probar una hipótesis sobre la diferencia entre dos medidas poblacionales de tendencia central. El estadístico de prueba de Mann-Whitney y la prueba de la mediana, que se analizan en el capítulo 13, son alternativas utilizadas con frecuencia a los estadísticos z y t.

EXERCISES

In each of the following exercises, complete the ten-step hypothesis testing procedure. State

the assumptions that are necessary for your procedure to be valid. For each exercise, as appro-

priate, explain why you chose a one-sided test or a two-sided test. Discuss how you

think researchers or clinicians might use the results of your hypothesis test. What clinical or

research decisions or actions do you think would be appropriate in light of the results of your

test?

7.3.1Subjects in a study by Dabonneville et al. (A-9) included a sample of 40 men who claimed to engage

in a variety of sports activities (multisport). The mean body mass index (BMI) for these men was 22.41

with a standard deviation of 1.27. A sample of 24 male rugby players had a mean BMI of 27.75 with

a standard deviation of 2.64. Is there sufficient evidence for one to claim that, in general, rugby players

have a higher BMI than the multisport men? Let a = .01.

7.3.2The purpose of a study by Ingle and Eastell (A-10) was to examine the bone mineral density

(BMD) and ultrasound properties of women with ankle fractures. The investigators recruited 31

postmenopausal women with ankle fractures and 31 healthy postmenopausal women to serve as

controls. One of the baseline measurements was the stiffness index of the lunar Achilles. The

mean stiffness index for the ankle fracture group was 76.9 with a standard deviation of 12.6. In

the control group, the mean was 90.9 with a standard deviation of 12.5. Do these data provide

sufficient evidence to allow you to conclude that, in general, the mean stiffness index is higherEXERCISES

247

in healthy postmenopausal women than in postmenopausal women with ankle fractures? Let

a = .05.

7.3.3

Hoekema et al. (A-11) studied the craniofacial morphology of 26 male patients with obstructive

sleep apnea syndrome (OSAS) and 37 healthy male subjects (non–OSAS). One of the variables of

interest was the length from the most superoanterior point of the body of the hyoid bone to the

Frankfort horizontal (measured in millimeters).

Length (mm) Non–OSAS

96.80

100.70

94.55

99.65

109.15

102.75

97.70

92.10

91.90

89.50

97.00

97.70

97.00

94.55

106.45

94.55

94.05

89.45

89.85

98.20

101.00

88.25

92.60

98.25

90.85

95.25

88.80

101.40

90.55

109.80

Length (mm) OSAS

88.95

101.05

92.60

97.00

91.95

88.95

95.75

105.95

114.90

110.35

123.10

119.30

110.00

98.95

114.20

108.95

105.05

114.90

114.35

112.25

106.15

102.60

102.40

105.05

112.65

128.95

117.70

113.70

116.30

108.75

113.30

106.00

101.75

Source: A. Hoekema, D.D.S. Used with permission.

Do these data provide sufficient evidence to allow us to conclude that the two sampled popula-

tions differ with respect to length from the hyoid bone to the Frankfort horizontal? Let a = .01.

7.3.4

Can we conclude that patients with primary hypertension (PH), on the average, have higher total

cholesterol levels than normotensive (NT) patients? This was one of the inquiries of interest for Rossi

et al. (A-12). In the following table are total cholesterol measurements (mg/dl) for 133 PH patients

and 41 NT patients. Can we conclude that PH patients have, on average, higher total cholesterol

levels than NT patients? Let a = .05.

Total Cholesterol (mg/dl)

Primary Hypertensive Patients

207

172

191

221

203

241

208

199

185

235

214

134

226

221

223

181

217

208

202

218

216

168

168

214

203

280

212

260

210

265

206

198

210

211

274

223

175

203

168

220

214

215

206

247

221

199

196

239

199

244

214

236

Normotensive Patients

190

245

171

261

182

162

182

225

203

195

178

240

222

286

226

187

204

203

206

196

168

229

184

186

281

203

189

196

142

179

212

163

196

189

142

168

121

(Continued)248

CHAPTER 7 HYPOTHESIS TESTING

Total Cholesterol (mg/dl)

Primary Hypertensive Patients

222

213

272

185

181

238

141

203

222

221

180

276

226

224

206

203

225

227

239

265

228

226

236

195

284

183

266

258

214

260

178

217

200

226

207

232

182

215

239

210

207

224

251

212

201

249

212

259

189

235

239

239

210

203

188

237

231

222

174

219

Normotensive Patients

117

252

203

245

218

152

231

237

177

179

194

206

219

173

189

194

196

212

168

188

232

242

200

135

161

Source: Gian Paolo Rossi, M.D., F.A.C.C., F.A.H.A. Used with permission.

7.3.5

Garção and Cabrita (A-13) wanted to evaluate the community pharmacist’s capacity to positively

influence the results of antihypertensive drug therapy through a pharmaceutical care program in Por-

tugal. Eighty-two subjects with essential hypertension were randomly assigned to an intervention

or a control group. The intervention group received monthly monitoring by a research pharmacist

to monitor blood pressure, assess adherence to treatment, prevent, detect, and resolve drug-related

problems, and encourage nonpharmacologic measures for blood pressure control. The changes after

6 months in diastolic blood pressure 1pre - post, mm Hg) are given in the following table for

patients in each of the two groups.

Intervention Group

20

2

36

26

2

20

2

14

30

18

6

4

24

6

&2

8

8

16

14

8

20

12

6

24

42

20

14

&2

10

2

18

16

10

16

10

6

6

2

8

16

&12

Control Group

0

12

18

0

8

10

8

14

4

&2

&6

4

2

2

8

10

0

6

10

&2

2

12

2

0

0

&4

12

4

28

&18

12

0

8

10

14

8

0

2

&8

16

12

Source: José Garção, M.S.,

Pharm.D. Used with permission.

On the basis of these data, what should the researcher conclude? Let a = .05.

7.3.6

A test designed to measure mothers’ attitudes toward their labor and delivery experiences was

given to two groups of new mothers. Sample 1 (attenders) had attended prenatal classes held atEXERCISES

249

the local health department. Sample 2 (nonattenders) did not attend the classes. The sample sizes

and means and standard deviations of the test scores were as follows:

Samplenxs

1

215

224.75

3.001.0

1.5

Do these data provide sufficient evidence to indicate that attenders, on the average, score higher

than nonattenders? Let a = .05.

7.3.7

Cortisol level determinations were made on two samples of women at childbirth. Group 1 subjects

underwent emergency cesarean section following induced labor. Group 2 subjects delivered by

either cesarean section or the vaginal route following spontaneous labor. The sample sizes, mean

cortisol levels, and standard deviations were as follows:

Samplenxs

1

210

12435

64565

80

Do these data provide sufficient evidence to indicate a difference in the mean cortisol levels in the

populations represented? Let a = .05.

7.3.8

Protoporphyrin levels were measured in two samples of subjects. Sample 1 consisted of 50 adult

male alcoholics with ring sideroblasts in the bone marrow. Sample 2 consisted of 40 apparently

healthy adult nonalcoholic males. The mean protoporphyrin levels and standard deviations for the

two samples were as follows:

Samplexs

1

2340

45250

25

Can one conclude on the basis of these data that protoporphyrin levels are higher in the repre-

sented alcoholic population than in the nonalcoholic population? Let a = .01.

7.3.9

A researcher was interested in knowing if preterm infants with late metabolic acidosis and

preterm infants without the condition differ with respect to urine levels of a certain chemical.

The mean levels, standard deviations, and sample sizes for the two samples studied were as

follows:

Samplenxs

With condition

Without condition35

408.5

4.85.5

3.6

What should the researcher conclude on the basis of these results? Let a = .05.250

CHAPTER 7 HYPOTHESIS TESTING

7.3.10

Researchers wished to know if they could conclude that two populations of infants differ with respect

to mean age at which they walked alone. The following data (ages in months) were collected:

Sample from population A:9.5, 10.5, 9.0, 9.75, 10.0, 13.0,

10.0, 13.5, 10.0, 9.5, 10.0, 9.75

Sample from population B:12.5, 9.5, 13.5, 13.75, 12.0, 13.75,

12.5, 9.5, 12.0, 13.5, 12.0, 12.0

What should the researchers conclude? Let a = .05.

7.3.11

Does sensory deprivation have an effect on a person’s alpha-wave frequency? Twenty volunteer

subjects were randomly divided into two groups. Subjects in group A were subjected to a 10-day

period of sensory deprivation, while subjects in group B served as controls. At the end of the

experimental period, the alpha-wave frequency component of subjects’ electroencephalograms was

measured. The results were as follows:

Group A:10.2, 9.5, 10.1, 10.0, 9.8, 10.9, 11.4, 10.8, 9.7, 10.4

Group B:11.0, 11.2, 10.1, 11.4, 11.7, 11.2, 10.8, 11.6, 10.9, 10.9

Let a = .05.

7.3.12

Can we conclude that, on the average, lymphocytes and tumor cells differ in size? The following

are the cell diameters 1mm2 of 40 lymphocytes and 50 tumor cells obtained from biopsies of

tissue from patients with melanoma:

Lymphocytes

9.0

6.3

8.6

7.4

8.8

9.4

5.7

7.0

8.7

5.2

4.7

5.0

6.8

4.9

7.1

4.8

3.5

7.1

7.4

5.3

8.9

7.8

5.7

6.4

4.7

4.9

10.4

7.6

7.1

8.4

8.4

8.0

6.2

6.3

6.45.9

8.0

7.1

8.8

8.3

19.1

18.9

17.9

18.2

16.119.4

18.7

15.2

20.7

21.5

Tumor Cells

12.6

16.7

20.0

17.7

16.3

14.6

15.9

17.8

15.1

17.7

16.2

15.8

13.9

16.9

18.1

23.9

16.0

22.1

16.4

24.3

23.3

17.9

13.9

22.8

11.2

17.1

13.4

18.3

19.4

19.5

20.0

19.1

22.8

19.6

18.6

21.0

16.6

13.0

18.4

16.4

Let a = .05.

7.4 PAIRED COMPARISONS

In our previous discussion involving the difference between two population means, it was assumed that the samples were independent. A method frequently employed for assessing the effectiveness of a treatment or experimental procedure is one that makes use of related observations resulting from nonindependent samples. A hypothesis test based on this type of data is known as a paired comparisons test.

En nuestra discusión anterior sobre la diferencia entre dos medias poblacionales, se supuso que las muestras eran independientes. Un método frecuentemente empleado para evaluar la eficacia de un tratamiento o procedimiento experimental es aquel que utiliza observaciones relacionadas resultantes de muestras no independientes. Una prueba de hipótesis basada en este tipo de datos se conoce como prueba de comparaciones pareadas.

**Reasons for Pairing**

It frequently happens that true differences do not exist between two populations with respect to the variable of interest, but the presence of extraneous sources of variation may cause rejection of the null hypothesis of no difference. On the other hand, true differences also may be masked by the presence of extraneous factors.

Con frecuencia sucede que no existen diferencias verdaderas entre dos poblaciones con respecto a la variable de interés, pero la presencia de fuentes extrañas de variación puede provocar el rechazo de la hipótesis nula de no diferencia. Por otra parte, las verdaderas diferencias también pueden quedar enmascaradas por la presencia de factores extraños.

A better way to design the experiment would be to select just one simple random sample of subjects and let each member of the sample receive both sunscreens. We could, for example, randomly assign the sunscreens to the left or the right side of each subject’s back with each subject receiving both sunscreens. After a specified length of exposure to the sun, we would measure the amount of sun damage to each half of the back. If the half of the back receiving sunscreen A tended to be less damaged, we could more confidently attribute the result to the sunscreen, since in each instance both sunscreens were applied to equally pigmented skin.

Una mejor manera de diseñar el experimento sería seleccionar solo una muestra aleatoria simple de sujetos y dejar que cada miembro de la muestra reciba ambos protectores solares. Podríamos, por ejemplo, asignar aleatoriamente los protectores solares al lado izquierdo o derecho de la espalda de cada sujeto y cada sujeto recibiría ambos protectores solares. Después de un período específico de exposición al sol, mediríamos la cantidad de daño solar en cada mitad de la espalda. Si la mitad de la espalda que recibió el protector solar A tendiera a estar menos dañada, podríamos atribuir con mayor confianza el resultado al protector solar, ya que en cada caso ambos protectores solares se aplicaron a piel igualmente pigmentada.

The objective in paired comparisons tests is to eliminate a maximum number of sources of extraneous variation by making the pairs similar with respect to as many variables as possible.

El objetivo de las pruebas de comparaciones por pares es eliminar un número máximo de fuentes de variación extraña haciendo que los pares sean similares con respecto a tantas variables como sea posible.

Related or paired observations may be obtained in a number of ways. The same subjects may be measured before and after receiving some treatment. Litter mates of the same sex may be assigned randomly to receive either a treatment or a placebo. Pairs of twins or siblings may be assigned randomly to two treatments in such a way that members of a single pair receive different treatments. In comparing two methods of analysis, the material to be analyzed may be divided equally so that one-half is analyzed by one method and one-half is analyzed by the other. Or pairs may be formed by matching individuals on some characteristic, for example, digital dexterity, which is closely related to the measurement of interest, say, posttreatment scores on some test requiring digital manipulation.

Las observaciones relacionadas o emparejadas se pueden obtener de varias maneras. Los mismos sujetos podrán ser medidos antes y después de recibir algún tratamiento. Los compañeros de camada del mismo sexo pueden ser asignados aleatoriamente para recibir un tratamiento o un placebo. Se pueden asignar aleatoriamente parejas de gemelos o hermanos a dos tratamientos de tal manera que los miembros de una sola pareja reciban tratamientos diferentes. Al comparar dos métodos de análisis, el material a analizar se puede dividir en partes iguales de modo que la mitad se analice mediante un método y la otra mitad se analice mediante el otro. O se pueden formar parejas emparejando individuos en alguna característica, por ejemplo, destreza digital, que está estrechamente relacionada con la medición de interés, por ejemplo, puntuaciones postratamiento en alguna prueba que requiera manipulación digital.

Instead of performing the analysis with individual observations, we use di, the difference between pairs of observations, as the variable of interest.

En lugar de realizar el análisis con observaciones individuales, utilizamos di, la diferencia entre pares de observaciones, como variable de interés.

When the n sample differences computed from the n pairs of measurements constitute a simple random sample from a normally distributed population of differences, the test statistic for testing hypotheses about the population mean difference md is

Cuando las n diferencias muestrales calculadas a partir de los n pares de mediciones constituyen una muestra aleatoria simple de una población de diferencias distribuida normalmente, el estadístico de prueba para probar hipótesis sobre la diferencia de medias poblacional md es

t =

d - md0

sd

(7.4.1)252

CHAPTER 7 HYPOTHESIS TESTING

where d is the sample mean difference, md0 is the hypothesized population mean difference, sd = sd>1n, n is the number of sample differences, and sd is the standard deviation of the sample differences. When H0 is true, the test statistic is distributed as Student’s t with n - 1 degrees of freedom.

donde d es la diferencia de medias muestrales, md0 es la diferencia de medias poblacional hipotética, sd = sd>1n, n es el número de diferencias muestrales y sd es la desviación estándar de las diferencias muestrales. Cuando H0 es verdadera, el estadístico de prueba se distribuye como t de Student con n - 1 grados de libertad.

Although to begin with we have two samples—say, before levels and after levels— we do not have to worry about equality of variances, as with independent samples, since our variable is the difference between readings in the same individual, or matched individuals, and, hence, only one variable is involved. The arithmetic involved in performing a paired comparisons test, therefore, is the same as for performing a test involving a single sample as described in Section 7.2. The following example illustrates the procedures involved in a paired comparisons test.

Aunque para empezar tenemos dos muestras (digamos, antes de los niveles y después de los niveles), no tenemos que preocuparnos por la igualdad de varianzas, como ocurre con las muestras independientes, ya que nuestra variable es la diferencia entre lecturas en el mismo individuo o en individuos emparejados. y, por tanto, sólo interviene una variable. Por lo tanto, la aritmética involucrada en realizar una prueba de comparaciones pareadas es la misma que para realizar una prueba que involucra una sola muestra como se describe en la Sección 7.2. El siguiente ejemplo ilustra los procedimientos involucrados en una prueba de comparaciones pareadas.

EXAMPLE 7.4.1

John M. Morton et al. (A-14) examined gallbladder function before and after fundoplication—a surgery used to stop stomach contents from flowing back into the esophagus (reflux)—in patients with gastroesophageal reflux disease. The authors measured gallbladder functionality by calculating the gallbladder ejection fraction (GBEF) before and after fundoplication. The goal of fundoplication is to increase GBEF, which is measured as a percent. The data are shown in Table 7.4.1. We wish to know if these data provide sufficient evidence to allow us to conclude that fundoplication increases GBEF functioning.

John M. Morton et al. (A-14) examinaron la función de la vesícula biliar antes y después de la funduplicatura (una cirugía utilizada para impedir que el contenido del estómago regrese al esófago (reflujo)) en pacientes con enfermedad por reflujo gastroesofágico. Los autores midieron la funcionalidad de la vesícula biliar calculando la fracción de eyección de la vesícula biliar (FEGB) antes y después de la funduplicatura. El objetivo de la funduplicatura es aumentar el GBEF, que se mide como porcentaje. Los datos se muestran en la Tabla 7.4.1. Deseamos saber si estos datos proporcionan evidencia suficiente que nos permita concluir que la funduplicatura aumenta el funcionamiento del GBEF.

**Solution**:

We will say that sufficient evidence is provided for us to conclude that the fundoplication is effective if we can reject the null hypothesis that the population mean change md is different from zero in the appropriate direction. We may reach a conclusion by means of the ten-step hypothesis testing procedure.

Diremos que se proporciona evidencia suficiente para concluir que la funduplicatura es efectiva si podemos rechazar la hipótesis nula de que el cambio medio poblacional md es diferente de cero en la dirección apropiada. Podemos llegar a una conclusión mediante el procedimiento de prueba de hipótesis de diez pasos.

1. Data. The data consist of the GBEF for 12 individuals, before and after fundoplication. We shall perform the statistical analysis on the differences in preop and postop GBEF. We may obtain the differences in one of two ways: by subtracting the preop percents from the postop percents or by subtracting the postop percents from the preop percents. Let us obtain the differences by subtracting the preop

Datos. Los datos consisten en el GBEF de 12 individuos, antes y después de la funduplicatura. Realizaremos el análisis estadístico de las diferencias en GBEF preoperatorio y posoperatorio. Podemos obtener las diferencias de dos maneras: restando los porcentajes preoperatorios de los porcentajes posoperatorios o restando los porcentajes posoperatorios de los porcentajes preoperatorios. Obtenemos las diferencias restando el preoperatorio.

TABLE 7.4.1 Gallbladder Function in Patients with Presentations of

Gastroesophageal Reflux Disease Before and After Treatment

Preop (%)

Postop (%)

22

63.5

63.3

91.5

96

59

9.2

37.8

3.1

10.1

50

19.6

33

41

69

87.8

64

86

18.8

55

0

88

34

40

Source: John M. Morton, Steven P. Bowers, Tananchai A. Lucktong, Samer Mattar, W. Alan Bradshaw,

Kevin E. Behrns, Mark J. Koruda, Charles A. Herbst, William McCartney, Raghuveer K. Halkar, C. Daniel

Smith, and Timothy M. Farrell, “Gallbladder Function Before and After Fundoplication,” Journal of

Gastrointestinal Surgery, 6 (2002), 806–811.7.4 PAIRED COMPARISONS

253

percents from the postop percents. The di = postop - preop differ-

ences are:

41.5, 28.2, &37.0, 28.6, 7.0, &30.4, 8.0, 18.8, 22.0, 36.2, 88.0, 6.0

2. Assumptions. The observed differences constitute a simple random sample from a normally distributed population of differences that could be generated under the same circumstances.

Suposiciones. Las diferencias observadas constituyen una muestra aleatoria simple de una población de diferencias normalmente distribuidas que podrían generarse en las mismas circunstancias.

3. Hypotheses. The way we state our null and alternative hypotheses must be consistent with the way in which we subtract measurements to obtain the differences. In the present example, we want to know if we can conclude that the fundoplication is useful in increasing GBEF percentage. If it is effective in improving GBEF, we would expect the postop percents to tend to be higher than the preop percents. If, therefore, we subtract the preop percents from the postop percents (postop - preop), we would expect the differences to tend to be positive. Furthermore, we would expect the mean of a population of such differences to be positive. So, under these conditions, asking if we can conclude that the fundoplication is effective is the same as asking if we can conclude that the population mean difference is positive (greater than zero).

Hipótesis. La forma en que planteamos nuestras hipótesis nula y alternativa debe ser coherente con la forma en que restamos medidas para obtener las diferencias. En el presente ejemplo, queremos saber si podemos concluir que la funduplicatura es útil para aumentar el porcentaje de GBEF. Si es eficaz para mejorar la GBEF, esperaríamos que los porcentajes postoperatorios tiendan a ser más altos que los porcentajes preoperatorios. Por lo tanto, si restamos los porcentajes preoperatorios de los porcentajes posoperatorios (postoperatorio - preoperatorio), esperaríamos que las diferencias tiendan a ser positivas. Además, esperaríamos que la media de una población con tales diferencias fuera positiva. Entonces, bajo estas condiciones, preguntar si podemos concluir que la funduplicatura es efectiva es lo mismo que preguntar si podemos concluir que la diferencia de medias poblacionales es positiva (mayor que cero).

The null and alternative hypotheses are as follows:

H0: md … 0

HA: md 7 0

If we had obtained the differences by subtracting the postop percents from the preop weights (preop & postop), our hypotheses would have been

H0: md Ú 0

HA: md 6 0

If the question had been such that a two-sided test was indicated, the

hypotheses would have been

H0: md = 0

HA: md Z 0

regardless of the way we subtracted to obtain the differences.

4. Test statistic. The appropriate test statistic is given by Equation 7.4.1.

5. Distribution of test statistic. If the null hypothesis is true, the test

statistic is distributed as Student’s t with n - 1 degrees of freedom.

6. Decision rule. Let a = .05. The critical value of t is 1.7959. Reject

H0 if computed t is greater than or equal to the critical value. The rejec-

tion and nonrejection regions are shown in Figure 7.4.1.

254

CHAPTER 7 HYPOTHESIS TESTING

a = .05

0

Nonrejection region

FIGURE 7.4.1

Example 7.4.1.

t

1.7959

Rejection region

Rejection and nonrejection regions for

7. Calculation of test statistic. From the n = 12 differences di, we

compute the following descriptive measures:

d =

s 2d =

t =

141.52 + 128.22 + 1-37.02 + . . . + 16.02

g di

216.9

=

=

= 18.075

n

12

12

g1di - d 22

n - 1

n g d i - 1gd i22

2

=

n 1n - 12

=

12115669.492 - 1216.922

18.075 - 0

18.075

= 1.9159

=

9.4344

11068.0930>12

11221112

= 1068.0930

8. Statistical decision. Reject H0, since 1.9159 is in the rejection region.

9. Conclusion. We may conclude that the fundoplication procedure

increases GBEF functioning.

10. p value. For this test, .025 6 p 6 .05, since 1.7959 6 1.9159 6

2.2010.

A Confidence Interval for M d A 95 percent confidence interval for md may

be obtained as follows:

d ; t 1-1a>22sd

18.075 ; 2.2010 11068.0930>12

18.075 ; 20.765

-2.690, 38.840

The Use of z If, in the analysis of paired data, the population variance of the

differences is known, the appropriate test statistic is

z =

d - md

sd>1n

It is unlikely that sd will be known in practice.

(7.4.2)EXERCISES

255

Paired T-Test and CI: C2, C1

Paired T for C2 - C1

N

12

12

12

C2

C1

Difference

Mean

56.6083

38.5333

18.0750

StDev

27.8001

30.0587

32.6817

SE Mean

8.0252

8.6772

9.4344

95% lower bound for mean difference: 1.1319

T-Test of mean difference " 0 (vs $ 0): T-Value " 1.92 P-Value "

0.041

FIGURE 7.4.2 MINITAB procedure and output for paired comparisons test, Example 7.4.1

(data in Table 7.4.1).

If the assumption of normally distributed di’s cannot be made, the central limit

theorem may be employed if n is large. In such cases, the test statistic is Equation 7.4.2,

with sd used to estimate sd when, as is generally the case, the latter is unknown.

We may use MINITAB to perform a paired t-test. The output from this procedure

is given in Figure 7.4.2.

Disadvantages The use of the paired comparisons test is not without its problems.

If different subjects are used and randomly assigned to two treatments, considerable time

and expense may be involved in our trying to match individuals on one or more relevant

variables. A further price we pay for using paired comparisons is a loss of degrees of

freedom. If we do not use paired observations, we have 2n - 2 degrees of freedom avail-

able as compared to n - 1 when we use the paired comparisons procedure.

In general, in deciding whether or not to use the paired comparisons procedure,

one should be guided by the economics involved as well as by a consideration of the

gains to be realized in terms of controlling extraneous variation.

Alternatives If neither z nor t is an appropriate test statistic for use with available

data, one may wish to consider using some nonparametric technique to test a hypothe-

sis about a median difference. The sign test, discussed in Chapter 13, is a candidate for

use in such cases.

EXERCISES

In the following exercises, carry out the ten-step hypothesis testing procedure at the specified signif-

icance level. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided

test. Discuss how you think researchers or clinicians might use the results of your hypothesis test.256

CHAPTER 7 HYPOTHESIS TESTING

What clinical or research decisions or actions do you think would be appropriate in light of the results

of your test?

7.4.1 Ellen Davis Jones (A-15) studied the effects of reminiscence therapy for older women with depres-

sion. She studied 15 women 60 years or older residing for 3 months or longer in an assisted liv-

ing long-term care facility. For this study, depression was measured by the Geriatric Depression

Scale (GDS). Higher scores indicate more severe depression symptoms. The participants received

reminiscence therapy for long-term care, which uses family photographs, scrapbooks, and personal

memorabilia to stimulate memory and conversation among group members. Pre-treatment and post-

treatment depression scores are given in the following table. Can we conclude, based on these data,

that subjects who participate in reminiscence therapy experience, on average, a decline in GDS

depression scores? Let a = .01.

Pre–GDS: 12

Post–GDS: 11

10

10

16

11

2

3

12

9

18

13

11

8

16

14

16

16

10

10

14

12

21

22

9

9

19

16

20

18

Source: Ellen Davis Jones, N.D., R.N., FNP-C. Used with permission.

7.4.2

Beney et al. (A-16) evaluated the effect of telephone follow-up on the physical well-being dimen-

sion of health-related quality of life in patients with cancer. One of the main outcome variables was

measured by the physical well-being subscale of the Functional Assessment of Cancer Therapy

Scale–General (FACT-G). A higher score indicates higher physical well-being. The following table

shows the baseline FACT-G score and the follow-up score to evaluate the physical well-being dur-

ing the 7 days after discharge from hospital to home for 66 patients who received a phone call

48–72 hours after discharge that gave patients the opportunity to discuss medications, problems,

and advice. Is there sufficient evidence to indicate that quality of physical well-being significantly

decreases in the first week of discharge among patients who receive a phone call? Let a = .05.

Subject

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

Baseline

FACT-GFollow-up

FACT-GSubjectBaseline

FACT-GFollow-up

FACT-G

16

26

13

20

22

21

20

15

25

20

11

22

18

21

25

17

26

18

719

19

9

23

25

20

10

20

22

18

6

21

17

13

25

21

22

22

934

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

5225

21

14

23

19

19

18

20

18

22

7

23

19

17

22

19

5

22

1214

17

22

22

16

15

23

21

11

22

17

9

16

16

20

23

17

17

6

(Continued )EXERCISES

Subject

20

21

22

23

24

25

26

27

28

29

30

31

32

33

Baseline

FACT-GFollow-up

FACT-GSubjectBaseline

FACT-GFollow-up

FACT-G

25

22

15

19

23

19

21

24

21

28

18

25

25

2824

15

9

7

20

19

24

23

15

27

26

26

26

2853

54

55

56

57

58

59

60

61

62

63

64

65

6619

17

7

27

22

16

26

17

23

23

13

24

17

2219

20

6

10

16

14

24

19

22

23

3

22

21

21

257

Source: Johnny Beney, Ph.D. and E. Beth Devine, Pharm.D., M.B.A. et al. Used with

permission.

7.4.3

The purpose of an investigation by Morley et al. (A-17) was to evaluate the analgesic effective-

ness of a daily dose of oral methadone in patients with chronic neuropathic pain syndromes. The

researchers used a visual analogue scale (0–100 mm, higher number indicates higher pain) ratings

for maximum pain intensity over the course of the day. Each subject took either 20 mg of

methadone or a placebo each day for 5 days. Subjects did not know which treatment they were

taking. The following table gives the mean maximum pain intensity scores for the 5 days on

methadone and the 5 days on placebo. Do these data provide sufficient evidence, at the .05 level

of significance, to indicate that in general the maximum pain intensity is lower on days when

methadone is taken?

Subject

1

2

3

4

5

6

7

8

9

10

11

7.4.4

MethadonePlacebo

29.8

73.0

98.6

58.8

60.6

57.2

57.2

89.2

97.0

49.8

37.057.2

69.8

98.2

62.4

67.2

70.6

67.8

95.6

98.4

63.2

63.6

Source: John S. Morley, John Bridson, Tim P. Nash, John B.

Miles, Sarah White, and Matthew K. Makin, “Low-Dose

Methadone Has an Analgesic Effect in Neuropathic Pain:

A Double-Blind Randomized Controlled Crossover Trial,”

Palliative Medicine, 17 (2003), 576–587.

Woo and McKenna (A-18) investigated the effect of broadband ultraviolet B (UVB) therapy and top-

ical calcipotriol cream used together on areas of psoriasis. One of the outcome variables is the Pso-

riasis Area and Severity Index (PASI). The following table gives the PASI scores for 20 subjects

measured at baseline and after eight treatments. Do these data provide sufficient evidence, at the .01

level of significance, to indicate that the combination therapy reduces PASI scores?258

CHAPTER 7 HYPOTHESIS TESTING

Subject

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

7.4.5

BaselineAfter 8

Treatments

5.9

7.6

12.8

16.5

6.1

14.4

6.6

5.4

9.6

11.6

11.1

15.6

6.9

15.2

21.0

5.9

10.0

12.2

20.2

6.25.2

12.2

4.6

4.0

0.4

3.8

1.2

3.1

3.5

4.9

11.1

8.4

5.8

5.0

6.4

0.0

2.7

5.1

4.8

4.2

Source: W. K. Woo, M.D. Used with permission.

One of the purposes of an investigation by Porcellini et al. (A-19) was to investigate the effect on

CD4 T cell count of administration of intermittent interleukin (IL-2) in addition to highly active

antiretroviral therapy (HAART). The following table shows the CD4 T cell count at baseline and

then again after 12 months of HAART therapy with IL-2. Do the data show, at the .05 level, a

significant change in CD4 T cell count?

Subject1234567

CD4 T cell count at

entry (: 106/L)

CD4 T cell count at end

of follow-up (: 106/L)17358103181105301169

257108315362141549369

Source: Simona Procellini, Giuliana Vallanti, Silvia Nozza, Guido Poli, Adraino Lazzarin,

Guiseppe Tabussi, and Antonio Grassia, “Improved Thymopoietic Potential in Aviremic HIV-

Infected Individuals with HAART by Intermittent IL-2 Administration,” AIDS, 17 (2003),

1621–1630.

7.5 HYPOTHESIS TESTING: A SINGLE POPULATION PROPORTION

Testing hypotheses about population proportions is carried out in much the same way

as for means when the conditions necessary for using the normal curve are met. One-

sided or two-sided tests may be made, depending on the question being asked. When a

7.5 HYPOTHESIS TESTING: A SINGLE POPULATION PROPORTION

259

sample sufficiently large for application of the central limit theorem as discussed in Sec-

tion 5.5 is available for analysis, the test statistic is

z =

pN - p0

p0q0

A n

(7.5.1)

which, when H0 is true, is distributed approximately as the standard normal.

EXAMPLE 7.5.1

Wagenknecht et al. (A-20) collected data on a sample of 301 Hispanic women living in

San Antonio, Texas. One variable of interest was the percentage of subjects with impaired

fasting glucose (IFG). IFG refers to a metabolic stage intermediate between normal glu-

cose homeostasis and diabetes. In the study, 24 women were classified in the IFG stage.

The article cites population estimates for IFG among Hispanic women in Texas as 6.3

percent. Is there sufficient evidence to indicate that the population of Hispanic women

in San Antonio has a prevalence of IFG higher than 6.3 percent?

Solution:

1. Data. The data are obtained from the responses of 301 individuals

of which 24 possessed the characteristic of interest; that is, pN =

24>301 = .080.

2. Assumptions. The study subjects may be treated as a simple random

sample from a population of similar subjects, and the sampling distri-

bution of pN is approximately normally distributed in accordance with

the central limit theorem.

3. Hypotheses.

H0 : p … .063

HA : p 7 .063

We conduct the test at the point of equality. The conclusion we reach

will be the same as we would reach if we conducted the test using any

other hypothesized value of p greater than .063. If H0 is true, p = .063

and the standard error spN = 11.06321.9372>301. Note that we use the

hypothesized value of p in computing spN . We do this because the

entire test is based on the assumption that the null hypothesis is true.

To use the sample proportion, pN , in computing spN would not be consis-

tent with this concept.

4. Test statistic. The test statistic is given by Equation 7.5.1.

5. Distribution of test statistic. If the null hypothesis is true, the test sta-

tistic is approximately normally distributed with a mean of zero.

6. Decision rule. Let a = .05. The critical value of z is 1.645. Reject H0

if the computed z is Ú1.645.260

CHAPTER 7 HYPOTHESIS TESTING

7. Calculation of test statistic.

z =

.080 - .063

= 1.21

1.06321.9372

A

301

8. Statistical decision. Do not reject H0 since 1.21 6 1.645.

9. Conclusion. We cannot conclude that in the sampled population the

proportion who are IFG is higher than 6.3 percent.

10. p value. p = .1131 .

■

Tests involving a single proportion can be carried out using a variety

of computer programs. Outputs from MINITAB and NCSS, using the data

from Example 7.5.1, are shown in Figure 7.5.1. It should be noted that the

results will vary slightly, because of rounding errors, if calculations are done

by hand. It should also be noted that some programs, such as NCSS, use a

continuity correction in calculating the z-value, and therefore the test statis-

tic values and corresponding p values differ slightly from the MINITAB

output.

MINITAB Output

Test and CI for One Proportion

Test of p " 0.063 vs p $ 0.063

Sample

1

X

24

95% Lower

Sample p

Bound

0.079734

0.054053

N

301

Z-Value

1.19

P-Value

0.116

Using the normal approximation.

NCSS Output

Normal Approximation using (P0)

Alternative

HypothesisZ-ValueProb

LevelDecision

(5%)

P%$P0

P%P0

P$P01.0763

1.0763

1.07630.281780

0.859110

0.140890Accept H0

Accept H0

Accept H0

FIGURE 7.5.1

MINITAB and partial NCSS output for the data in Example 7.5.1EXERCISES

261

EXERCISES

For each of the following exercises, carry out the ten-step hypothesis testing procedure at the des-

ignated level of significance. For each exercise, as appropriate, explain why you chose a one-sided

test or a two-sided test. Discuss how you think researchers or clinicians might use the results of

your hypothesis test. What clinical or research decisions or actions do you think would be appro-

priate in light of the results of your test?

7.5.1Jacquemyn et al. (A-21) conducted a survey among gynecologists-obstetricians in the Flanders

region and obtained 295 responses. Of those responding, 90 indicated that they had performed at

least one cesarean section on demand every year. Does this study provide sufficient evidence for

us to conclude that less than 35 percent of the gynecologists-obstetricians in the Flanders region

perform at least one cesarean section on demand each year? Let a = .05.

7.5.2In an article in the journal Health and Place, Hui and Bell (A-22) found that among 2428 boys

ages 7 to 12 years, 461 were overweight or obese. On the basis of this study, can we conclude

that more than 15 percent of the boys ages 7 to 12 in the sampled population are obese or over-

weight? Let a = .05.

7.5.3Becker et al. (A-23) conducted a study using a sample of 50 ethnic Fijian women. The women com-

pleted a self-report questionnaire on dieting and attitudes toward body shape and change. The

researchers found that five of the respondents reported at least weekly episodes of binge eating dur-

ing the previous 6 months. Is this sufficient evidence to conclude that less than 20 percent of the

population of Fijian women engage in at least weekly episodes of binge eating? Let a = .05.

7.5.4The following questionnaire was completed by a simple random sample of 250 gynecologists. The

number checking each response is shown in the appropriate box.

1. When you have a choice, which procedure do you prefer for obtaining samples of endometrium?

(a) Dilation and curettage 175

(b) Vobra aspiration 75

2. Have you seen one or more pregnant women during the past year whom you knew to have ele-

vated blood lead levels?

(a) Yes 25

(b) No 225

3. Do you routinely acquaint your pregnant patients who smoke with the suspected hazards of

smoking to the fetus?

(a) Yes 238

(b) No 12

Can we conclude from these data that in the sampled population more than 60 percent prefer dila-

tion and curettage for obtaining samples of endometrium? Let a = .01.

7.5.5Refer to Exercise 7.5.4. Can we conclude from these data that in the sampled population fewer

than 15 percent have seen (during the past year) one or more pregnant women with elevated blood

lead levels? Let a = .05.

7.5.6Refer to Exercise 7.5.4. Can we conclude from these data that more than 90 percent acquaint

their pregnant patients who smoke with the suspected hazards of smoking to the fetus? Let

a = .05.262

CHAPTER 7 HYPOTHESIS TESTING

7.6 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

The most frequent test employed relative to the difference between two population

proportions is that their difference is zero. It is possible, however, to test that the

difference is equal to some other value. Both one-sided and two-sided tests may be

made.

When the null hypothesis to be tested is p1 - p2 = 0, we are hypothesizing that

the two population proportions are equal. We use this as justification for combining

the results of the two samples to come up with a pooled estimate of the hypothesized

common proportion. If this procedure is adopted, one computes

p =

x1 + x2

, and q = 1 - p

n1 + n2

where x 1 and x 2 are the numbers in the first and second samples, respectively, possess-

ing the characteristic of interest. This pooled estimate of p = p1 = p2 is used in com-

puting sN pN 1 - pN 2, the estimated standard error of the estimator, as follows:

sN pN1 - pN2 =

p11 - p2

A

n1

+

p11 - p2

n2

(7.6.1)

The test statistic becomes

z =

1pN 1 - pN 22 - 1p1 - p220

sN pN 1 - pN 2

(7.6.2)

which is distributed approximately as the standard normal if the null hypothesis is

true.

EXAMPLE 7.6.1

Noonan syndrome is a genetic condition that can affect the heart, growth, blood clot-

ting, and mental and physical development. Noonan et al. (A-24) examined the stature

of men and women with Noonan syndrome. The study contained 29 male and 44 female

adults. One of the cut-off values used to assess stature was the third percentile of adult

height. Eleven of the males fell below the third percentile of adult male height, while

24 of the females fell below the third percentile of female adult height. Does this study

provide sufficient evidence for us to conclude that among subjects with Noonan syn-

drome, females are more likely than males to fall below the respective third percentile

of adult height? Let a = .05.7.6 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

263

Solution:

1. Data. The data consist of information regarding the height status of Noo-

nan syndrome males and females as described in the statement of the

example.

2. Assumptions. We assume that the patients in the study constitute inde-

pendent simple random samples from populations of males and females

with Noonan syndrome.

3. Hypotheses.

H0 : p F … p M or p F - p M … 0

HA : p F 7 p M or p F - p M 7 0

where p F is the proportion of females below the third percentile of

female adult height and p M is the proportion of males below the third

percentile of male adult height.

4. Test statistic. The test statistic is given by Equation 7.6.2.

5. Distribution of test statistic. If the null hypothesis is true, the test sta-

tistic is distributed approximately as the standard normal.

6. Decision rule. Let a = .05. The critical value of z is 1.645. Reject H0

if computed z is greater than 1.645.

7. Calculation of test statistic. From the sample data we compute pN F =

24>44 = .545, pN M = 11>29 = .379, and p = 124 + 112>144 + 292 =

.479. The computed value of the test statistic, then, is

z =

A

1.545 - .3792

1.47921.5212

44

+

1.47921.5212

= 1.39

29

8. Statistical decision. Fail to reject H0 since 1.39 6 1.645.

9. Conclusion. In the general population of adults with Noonan syndrome

there may be no difference in the proportion of males and females who

have heights below the third percentile of adult height.

10. p value. For this test p = .0823.

■

Tests involving two proportions, using the data from Example 7.6.1,

can be carried out with a variety of computer programs. Outputs from

MINITAB and NCSS are shown in Figure 7.6.1. Again, it should be noted

that, because of rounding errors, the results will vary slightly if calculations

are done by hand.264

CHAPTER 7 HYPOTHESIS TESTING

MINITAB Output

Test and CI for Two Proportions

SampleXNSample p

124440.545455

211290.379310

Difference " p (1) & p (2)

Estimate for difference: 0.166144

95% lower bound for difference: &0.0267550

Test for difference " 0 (vs > 0):

Z " 1.39

P-Value " 0.082

NCSS Output

Test

NameTest

Statistic’s

DistributionTest

Statistic

ValueProb

LevelConclude H1

at 5%

Significance?

Z-TestNormal1.3900.0822No

FIGURE 7.6.1

MINITAB and partial NCSS output for the data in Example 7.6.1

EXERCISES

In each of the following exercises use the ten-step hypothesis testing procedure. For each exercise, as

appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think

researchers or clinicians might use the results of your hypothesis test. What clinical or research deci-

sions or actions do you think would be appropriate in light of the results of your test?

7.6.1Ho et al. (A-25) used telephone interviews of randomly selected respondents in Hong Kong to obtain

information regarding individuals’ perceptions of health and smoking history. Among 1222 current

male smokers, 72 reported that they had “poor” or “very poor” health, while 30 among 282 former

male smokers reported that they had “poor” or “very poor” health. Is this sufficient evidence to

allow one to conclude that among Hong Kong men there is a difference between current and for-

mer smokers with respect to the proportion who perceive themselves as having “poor” and “very

poor” health? Let a = .01.

7.6.2Landolt et al. (A-26) examined rates of posttraumatic stress disorder (PTSD) in mothers and

fathers. Parents were interviewed 5 to 6 weeks after an accident or a new diagnosis of cancer or

diabetes mellitus type I for their child. Twenty-eight of the 175 fathers interviewed and 43 of the7.7 HYPOTHESIS TESTING: A SINGLE POPULATION VARIANCE

265

180 mothers interviewed met the criteria for current PTSD. Is there sufficient evidence for us to

conclude that fathers are less likely to develop PTSD than mothers when a child is traumatized by

an accident, cancer diagnosis, or diabetes diagnosis? Let a = .05.

7.6.3In a Kidney International article, Avram et al. (A-27) reported on a study involving

529 hemodialysis patients and 326 peritoneal dialysis patients. They found that at baseline

249 subjects in the hemodialysis treatment group were diabetic, while at baseline 134 of the

subjects in the peritoneal dialysis group were diabetic. Is there a significant difference in dia-

betes prevalence at baseline between the two groups of this study? Let a = .05. What does your

finding regarding sample significance imply about the populations of subjects?

7.6.4In a study of obesity the following results were obtained from samples of males and females

between the ages of 20 and 75:

Males

Females

nNumber Overweight

150

20021

48

Can we conclude from these data that in the sampled populations there is a difference in the

proportions who are overweight? Let a = .05.

7.7 HYPOTHESIS TESTING: A SINGLE POPULATION VARIANCE

In Section 6.9 we examined how it is possible to construct a confidence interval for

the variance of a normally distributed population. The general principles presented in

that section may be employed to test a hypothesis about a population variance. When

the data available for analysis consist of a simple random sample drawn from a

normally distributed population, the test statistic for testing hypotheses about a

population variance is

x2 = 1n - 12s 2>s2

(7.7.1)

which, when H0 is true, is distributed as x2 with n - 1 degrees of freedom.

EXAMPLE 7.7.1

The purpose of a study by Wilkins et al. (A-28) was to measure the effectiveness of

recombinant human growth hormone (rhGH) on children with total body surface area

burns 7 40 percent. In this study, 16 subjects received daily injections at home of rhGH.

At baseline, the researchers wanted to know the current levels of insulin-like growth fac-

tor (IGF-I) prior to administration of rhGH. The sample variance of IGF-I levels (in

ng/ml) was 670.81. We wish to know if we may conclude from these data that the

population variance is not 600.

Solution:

1. Data. See statement in the example.

2. Assumptions. The study sample constitutes a simple random sample

from a population of similar children. The IGF-I levels are normally dis-

tributed.

3. Hypotheses.

H0 : s2 = 600

HA : s2 Z 600

4. Test statistic. The test statistic is given by Equation 7.7.1.

5. Distribution of test statistic. When the null hypothesis is true, the test

statistic is distributed as x2 with n - 1 degrees of freedom.

6. Decision rule. Let a = .05. Critical values of x2 are 6.262 and 27.488.

Reject H0 unless the computed value of the test statistic is between

6.262 and 27.488. The rejection and nonrejection regions are shown in

Figure 7.7.1.

7. Calculation of test statistic.

x2 =

151670.812

600

= 16.77

8. Statistical decision. Do not reject H0 since 6.262 6 16.77 6 27.488.

9. Conclusion. Based on these data we are unable to conclude that the

population variance is not 600.

10. p value. The determination of the p value for this test is complicated

by the fact that we have a two-sided test and an asymmetric sampling

distribution. When we have a two-sided test and a symmetric sam-

pling distribution such as the standard normal or t, we may, as we

have seen, double the one-sided p value. Problems arise when we

attempt to do this with an asymmetric sampling distribution such as the

chi-square distribution. In this situation the one-sided p value is

reported along with the direction of the observed departure from the

null hypothesis. In fact, this procedure may be followed in the case

of symmetric sampling distributions. Precedent, however, seems to

favor doubling the one-sided p value when the test is two-sided and

involves a symmetric sampling distribution.

For the present example, then, we may report the p value as

follows: p 7 .05 (two-sided test). A population variance greater than

600 is suggested by the sample data, but this hypothesis is not strongly

supported by the test.

If the problem is stated in terms of the population standard devi-

ation, one may square the sample standard deviation and perform the

test as indicated above.

One-Sided Tests

Although this was an example of a two-sided test, one-sided

tests may also be made by logical modification of the procedure given here.

For HA : s2 7 s20, reject H0 if computed x 2 Ú x 21-a

For HA : s2 6 s20, reject H0 if computed x2 … x 2a

Tests involving a single population variance can be carried out using MINITAB

software. Most other statistical computer programs lack procedures for carrying out these

tests directly. The output from MINITAB, using the data from Example 7.7.1, is shown

in Figure 7.7.2.

Test and CI for One Variance

Statistics

N

16

StDev

25.9

Variance

671

95% Confidence Intervals

Method

Standard

CI for

StDev

(19.1, 40.1)CI for

Variance

(366, 1607)

Chi-Square

16.77DF

15

Tests

Method

Standard

FIGURE 7.7.2

P-Value

0.666

MINITAB output for the data in Example 7.7.1.268

CHAPTER 7 HYPOTHESIS TESTING

EXERCISES

In each of the following exercises, carry out the ten-step testing procedure. For each exercise, as

appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think

researchers or clinicians might use the results of your hypothesis test. What clinical or research

decisions or actions do you think would be appropriate in light of the results of your test?

7.7.1

Recall Example 7.2.3, where Nakamura et al. (A-1) studied subjects with acute medial collateral

ligament injury (MCL) with anterior cruciate ligament tear (ACL). The ages of the 17 subjects were:

31, 26, 21, 15, 26, 16, 19, 21, 28, 27, 22, 20, 25, 31, 20, 25, 15

Use these data to determine if there is sufficient evidence for us to conclude that in a population

of similar subjects, the variance of the ages of the subjects is not 20 years. Let a = .01.

7.7.2

7.7.3

Robinson et al. (A-29) studied nine subjects who underwent baffle procedure for transposition of

the great arteries (TGA). At baseline, the systemic vascular resistance (SVR) 1measured in

WU \* m22 values at rest yielded a standard deviation of 28. Can we conclude from these data

that the SVR variance of a population of similar subjects with TGA is not 700? Let a = .10.

Vital capacity values were recorded for a sample of 10 patients with severe chronic airway obstruc-

tion. The variance of the 10 observations was .75. Test the null hypothesis that the population

variance is 1.00. Let a = .05.

7.7.4Hemoglobin (g percent) values were recorded for a sample of 20 children who were part of a study

of acute leukemia. The variance of the observations was 5. Do these data provide sufficient evi-

dence to indicate that the population variance is greater than 4? Let a = .05.

7.7.5A sample of 25 administrators of large hospitals participated in a study to investigate the nature and

extent of frustration and emotional tension associated with the job. Each participant was given a test

designed to measure the extent of emotional tension he or she experienced as a result of the duties

and responsibilities associated with the job. The variance of the scores was 30. Can it be concluded

from these data that the population variance is greater than 25? Let a = .05.

7.7.6In a study in which the subjects were 15 patients suffering from pulmonary sarcoid disease,

blood gas determinations were made. The variance of the PaO2 (mm Hg) values was 450. Test

the null hypothesis that the population variance is greater than 250. Let a = .05.

7.7.7Analysis of the amniotic fluid from a simple random sample of 15 pregnant women yielded the

following measurements on total protein (grams per 100 ml) present:

.69, 1.04, .39, .37, .64, .73, .69, 1.04,

.83, 1.00, .19, .61, .42, .20, .79

Do these data provide sufficient evidence to indicate that the population variance is greater than

.05? Let a = .05. What assumptions are necessary?

7.8 HYPOTHESIS TESTING: THE RATIO OF TWO POPULATION VARIANCES

As we have seen, the use of the t distribution in constructing confidence intervals and in

testing hypotheses for the difference between two population means assumes that the

population variances are equal. As a rule, the only hints available about the magnitudes

of the respective variances are the variances computed from samples taken from the

populations. We would like to know if the difference that, undoubtedly, will exist between

the sample variances is indicative of a real difference in population variances, or if the

difference is of such magnitude that it could have come about as a result of chance alone

when the population variances are equal.

Two methods of chemical analysis may give the same results on the average. It

may be, however, that the results produced by one method are more variable than the

results of the other. We would like some method of determining whether this is likely

to be true.

Variance Ratio Test Decisions regarding the comparability of two population

variances are usually based on the variance ratio test, which is a test of the null hypoth-

esis that two population variances are equal. When we test the hypothesis that two pop-

ulation variances are equal, we are, in effect, testing the hypothesis that their ratio is

equal to 1.

We learned in the preceding chapter that, when certain assumptions are met, the

quantity 1s 21>s212>1s 22> s222 is distributed as F with n 1 - 1 numerator degrees of freedom and n 2 - 1 denominator degrees of freedom. If we are hypothesizing that s21 = s22, we

assume that the hypothesis is true, and the two variances cancel out in the above expres-

sion leaving s 21>s 22, which follows the same F distribution. The ratio s 21>s 22 will be desig-

nated V.R. for variance ratio.

For a two-sided test, we follow the convention of placing the larger sample vari-

ance in the numerator and obtaining the critical value of F for a>2 and the appropriate

degrees of freedom. However, for a one-sided test, which of the two sample variances

is to be placed in the numerator is predetermined by the statement of the null hypothe-

sis. For example, for the null hypothesis that s21>s22, the appropriate test statistic is

V.R. = s 21>s 22. The critical value of F is obtained for a (not a>2) and the appropriate

degrees of freedom. In like manner, if the null hypothesis is that s21 Ú s22, the appropri-

ate test statistic is V.R. = s 22>s 21. In all cases, the decision rule is to reject the null hypoth-

esis if the computed V.R. is equal to or greater than the critical value of F.

EXAMPLE 7.8.1

Borden et al. (A-30) compared meniscal repair techniques using cadaveric knee speci-

mens. One of the variables of interest was the load at failure (in newtons) for knees fixed

with the FasT-FIX technique (group 1) and the vertical suture method (group 2). Each

technique was applied to six specimens. The standard deviation for the FasT-FIX method

was 30.62, and the standard deviation for the vertical suture method was 11.37. Can we

conclude that, in general, the variance of load at failure is higher for the FasT-FIX tech-

nique than the vertical suture method?

Solution:

1. Data. See the statement of the example.

2. Assumptions. Each sample constitutes a simple random sample of a

population of similar subjects. The samples are independent. We assume

the loads at failure in both populations are approximately normally

distributed.

270

CHAPTER 7 HYPOTHESIS TESTING

.05

0

F(5, 5)

5.05

Nonrejection region

FIGURE 7.8.1

Example 7.8.1.

Rejection region

Rejection and nonrejection regions,

3. Hypotheses.

H0 : s21 … s22

HA : s21 7 s22

4. Test statistic.

V.R. =

s 21

(7.8.1)

s 22

5. Distribution of test statistic. When the null hypothesis is true, the test

statistic is distributed as F with n 1 - 1 numerator and n 2 - 1 denom-

inator degrees of freedom.

6. Decision rule. Let a = .05. The critical value of F, from Appendix Table

G, is 5.05. Note that if Table G does not contain an entry for the given

numerator degrees of freedom, we use the column closest in value to the

given numerator degrees of freedom. Reject H0 if V.R. Ú 5.05.

The rejection and nonrejection regions are shown in Figure 7.8.1.

7. Calculation of test statistic.

V.R. =

130.6222

111.3722

= 7.25

8. Statistical decision. We reject H0, since 7.25 7 5.05; that is, the com-

puted ratio falls in the rejection region.

9. Conclusion. The failure load variability is higher when using the FasT-

FIX method than the vertical suture method.

10. p value. Because the computed V.R. of 7.25 is greater than 5.05, the p

value for this test is less than 0.05.

Several computer programs can be used to test the equality of two variances. Outputs

from these programs will differ depending on the test that is used. We saw in Figure 7.3.3,

for example, that the SAS system uses a folded F-test procedure. MINITAB uses two dif-

ferent tests. The first is an F-test under the assumption of normality, and the other is a mod-

ified Levene’s test (1) that is used when normality cannot be assumed. SPSS uses an unmod-

ified Levene’s test (2). Regardless of the options, these tests are generally considered superior

to the variance ratio test that is presented in Example 7.8.1. Discussion of the mathematics

behind these tests is beyond the scope of this book, but an example is given to illustrate

these procedures, since results from these tests are often provided automatically as outputs

when a computer program is used to carry out a t-test.

EXAMPLE 7.8.2

Using the data from Example 7.3.2, we are interested in testing whether the assumption

of the equality of variances can be assumed prior to performing a t-test. For ease of dis-

cussion, the data are reproduced below (Table 7.8.1):

TABLE 7.8.1 Pressures (mm Hg) Under the Pelvis During Static Conditions for

Example 7.3.2

Control13111512413112211788114150169

SCI60150130180163130121119130148

Partial outputs for MINITAB, SAS, and SPSS are shown in Figure 7.8.2. Regardless of

the test or program that is used, we fail to reject the null hypothesis of equal variances

1H0: s21 = s222 because all p values $ 0.05. We may now proceed with a t-test under

the assumption of equal variances.

EXERCISES

In the following exercises perform the ten-step test. For each exercise, as appropriate, explain why

you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians

might use the results of your hypothesis test. What clinical or research decisions or actions do you

think would be appropriate in light of the results of your test?

7.8.1Dora et al. (A-31) investigated spinal canal dimensions in 30 subjects symptomatic with disc hernia-

tion selected for a discectomy and 45 asymptomatic individuals. The researchers wanted to know if

spinal canal dimensions are a significant risk factor for the development of sciatica. Toward that end,

they measured the spinal canal dimension between vertebrae L3 and L4 and obtained a mean of 17.8

mm in the discectomy group with a standard deviation of 3.1. In the control group, the mean was

18.5 mm with a standard deviation of 2.8 mm. Is there sufficient evidence to indicate that in relevant

populations the variance for subjects symptomatic with disc herniation is larger than the variance for

control subjects? Let a = .05.

7.8.2Nagy et al. (A-32) studied 50 stable patients who were admitted for a gunshot wound that tra-

versed the mediastinum. Of these, eight were deemed to have a mediastinal injury and 42 did

not. The standard deviation for the ages of the eight subjects with mediastinal injury was 4.7

years, and the standard deviation of ages for the 42 without injury was 11.6 years. Can we con-

clude from these data that the variance of age is larger for a population of similar subjects with-

out injury compared to a population with mediastinal injury? Let a = .05.

7.8.3A test designed to measure level of anxiety was administered to a sample of male and a sample

of female patients just prior to undergoing the same surgical procedure. The sample sizes and the

variances computed from the scores were as follows:

Males:n = 16, s 2 = 150

Females:n = 21, s 2 = 275

Do these data provide sufficient evidence to indicate that in the represented populations the scores

made by females are more variable than those made by males? Let a = .05.

7.8.4In an experiment to assess the effects on rats of exposure to cigarette smoke, 11 animals were exposed

and 11 control animals were not exposed to smoke from unfiltered cigarettes. At the end of the exper-

iment, measurements were made of the frequency of the ciliary beat (beats/min at 20°C) in each ani-

mal. The variance for the exposed group was 3400 and 1200 for the unexposed group. Do these data

indicate that in the populations represented the variances are different? Let a = .05.

7.8.5Two pain-relieving drugs were compared for effectiveness on the basis of length of time elapsing

between administration of the drug and cessation of pain. Thirteen patients received drug 1, and

13 received drug 2. The sample variances were s 21 = 64 and s 22 = 16. Test the null hypothesis that

the two populations variances are equal. Let a = .05.

7.8.6Packed cell volume determinations were made on two groups of children with cyanotic congeni-

tal heart disease. The sample sizes and variances were as follows:

Groupns2

1

210

1640

847.9 THE TYPE II ERROR AND THE POWER OF A TEST

273

Do these data provide sufficient evidence to indicate that the variance of population 2 is larger

than the variance of population 1? Let a = .05.

7.8.7

Independent simple random samples from two strains of mice used in an experiment yielded the fol-

lowing measurements on plasma glucose levels following a traumatic experience:

54, 99, 105, 46, 70, 87, 55, 58, 139, 91

93, 91, 93, 150, 80, 104, 128, 83, 88, 95, 94, 97

Strain A:

Strain B:

Do these data provide sufficient evidence to indicate that the variance is larger in the population of

strain A mice than in the population of strain B mice? Let a = .05. What assumptions are necessary?

7.9 THE TYPE II ERROR AND THE POWER OF A TEST

In our discussion of hypothesis testing our focus has been on a, the probability of com-

mitting a type I error (rejecting a true null hypothesis). We have paid scant attention to

b , the probability of committing a type II error (failing to reject a false null hypothe-

sis). There is a reason for this difference in emphasis. For a given test, a is a single num-

ber assigned by the investigator in advance of performing the test. It is a measure of the

acceptable risk of rejecting a true null hypothesis. On the other hand, b may assume one

of many values. Suppose we wish to test the null hypothesis that some population param-

eter is equal to some specified value. If H0 is false and we fail to reject it, we commit

a type II error. If the hypothesized value of the parameter is not the true value, the value

of b (the probability of committing a type II error) depends on several factors: (1) the

true value of the parameter of interest, (2) the hypothesized value of the parameter,

(3) the value of a, and (4) the sample size, n. For fixed a and n, then, we may, before

performing a hypothesis test, compute many values of b by postulating many values for

the parameter of interest given that the hypothesized value is false.

For a given hypothesis test it is of interest to know how well the test controls type

II errors. If H0 is in fact false, we would like to know the probability that we will reject

it. The power of a test, designated 1 - b , provides this desired information. The quan-

tity 1 - b is the probability that we will reject a false null hypothesis; it may be com-

puted for any alternative value of the parameter about which we are testing a hypothesis.

Therefore, 1 - b is the probability that we will take the correct action when H0 is false

because the true parameter value is equal to the one for which we computed 1 - b . For

a given test we may specify any number of possible values of the parameter of interest

and for each compute the value of 1 - b . The result is called a power function. The

graph of a power function, called a power curve, is a helpful device for quickly assess-

ing the nature of the power of a given test. The following example illustrates the proce-

dures we use to analyze the power of a test.

EXAMPLE 7.9.1

Suppose we have a variable whose values yield a population standard deviation of 3.6.

From the population we select a simple random sample of size n = 100. We select a

value of a = .05 for the following hypotheses:

H0: m = 17.5,

HA: m Z 17.5274

CHAPTER 7 HYPOTHESIS TESTING

Solution:

When we study the power of a test, we locate the rejection and nonrejec-

tion regions on the x scale rather than the z scale. We find the critical val-

ues of x for a two-sided test using the following formulas:

x U = m0 + zs

1n

x L = m0 - zs

1n

and

(7.9.1)

(7.9.2)

where x U and x L are the upper and lower critical values, respectively, of x;

+z and -z are the critical values of z; and m0 is the hypothesized value of

m. For our example, we have

x U = 17.50 + 1.96

13.62

= 17.50 + 1.961.362

1102

= 17.50 + .7056 = 18.21

and

x L = 17.50 - 1.961.362 = 17.50 - .7056 = 16.79

Suppose that H0 is false, that is, that m is not equal to 17.5. In that case,

m is equal to some value other than 17.5. We do not know the actual value of

m. But if H0 is false, m is one of the many values that are greater than or

smaller than 17.5. Suppose that the true population mean is m1 = 16.5. Then

the sampling distribution of x 1 is also approximately normal, with

mx = m = 16.5. We call this sampling distribution f 1x 12, and we call the sam-

pling distribution under the null hypothesis f 1x 02.

b, the probability of the type II error of failing to reject a false null

hypothesis, is the area under the curve of f 1x 12 that overlaps the nonrejec-

tion region specified under H0. To determine the value of b, we find the

area under f 1x 12, above the x axis, and between x = 16.79 and x = 18.21.

The value of b is equal to P 116.79 … x … 18.212 when m = 16.5. This is

the same as

Pa

16.79 - 16.5

18.21 - 16.5

.29

1.71

… z …

b = Pa

… z …

b

.36

.36

.36

.36

= P1.81 … z … 4.752

L 1 - .7910 = .2090

Thus, the probability of taking an appropriate action (that is, rejecting

H0) when the null hypothesis states that m = 17.5, but in fact m = 16.5, is7.9 THE TYPE II ERROR AND THE POWER OF A TEST

FIGURE 7.9.1

275

Size of b for selected values for H1 for Example 7.9.1.

1 - .2090 = .7910. As we noted, m may be one of a large number of pos-

sible values when H0 is false. Figure 7.9.1 shows a graph of several such

possibilities. Table 7.9.1 shows the corresponding values of b and 1 - b

(which are approximate), along with the values of b for some additional

alternatives.

Note that in Figure 7.9.1 and Table 7.9.1 those values of m under the

alternative hypothesis that are closer to the value of m specified by H0 have

larger associated b values. For example, when m = 18 under the alterna-

tive hypothesis, b = .7190; and when m = 19.0 under HA, b = .0143. The

power of the test for these two alternatives, then, is 1 - .7190 = .2810 and

1 - .0143 = .9857, respectively. We show the power of the test graphically276

CHAPTER 7 HYPOTHESIS TESTING

TABLE 7.9.1 Values of B and 1 ! B for

Selected Alternative Values of M 1, Example

7.9.1

Possible Values of M Under

HA When H0 is False

16.0

16.5

17.0

18.0

18.5

19.0

B1!B

0.0143

0.2090

0.7190

0.7190

0.2090

0.01430.9857

0.7910

0.2810

0.2810

0.7910

0.9857

1–b

1.00

0.90

0.80

0.70

0.60

0.50

0.40

0.30

0.20

0.10

0

16.0

FIGURE 7.9.2

17.0

18.0

19.0

Alternative values of m

Power curve for Example 7.9.1.

in a power curve, as in Figure 7.9.2. Note that the higher the curve, the

greater the power.

■

Although only one value of a is associated with a given hypothesis test, there are many

values of b, one for each possible value of m if m0 is not the true value of m as hypoth-

esized. Unless alternative values of m are much larger or smaller than m0, b is relatively

large compared with a. Typically, we use hypothesis-testing procedures more often in

those cases in which, when H0 is false, the true value of the parameter is fairly close to

the hypothesized value. In most cases, b, the computed probability of failing to reject a

false null hypothesis, is larger than a, the probability of rejecting a true null hypothesis.

These facts are compatible with our statement that a decision based on a rejected null

hypothesis is more conclusive than a decision based on a null hypothesis that is not

rejected. The probability of being wrong in the latter case is generally larger than the

probability of being wrong in the former case.

Figure 7.9.2 shows the V-shaped appearance of a power curve for a two-sided test.

In general, a two-sided test that discriminates well between the value of the parameter

in H0 and values in H1 results in a narrow V-shaped power curve. A wide V-shaped curve7.9 THE TYPE II ERROR AND THE POWER OF A TEST

277

indicates that the test discriminates poorly over a relatively wide interval of alternative

values of the parameter.

Power Curves for One-Sided Tests The shape of a power curve for a

one-sided test with the rejection region in the upper tail is an elongated S. If the rejec-

tion region of a one-sided test is located in the lower tail of the distribution, the power

curve takes the form of a reverse elongated S. The following example shows the nature

of the power curve for a one-sided test.

EXAMPLE 7.9.2

The mean time laboratory employees now take to do a certain task on a machine is 65 sec-

onds, with a standard deviation of 15 seconds. The times are approximately normally distrib-

uted. The manufacturers of a new machine claim that their machine will reduce the mean

time required to perform the task. The quality-control supervisor designs a test to determine

whether or not she should believe the claim of the makers of the new machine. She chooses

a significance level of a = 0.01 and randomly selects 20 employees to perform the task on

the new machine. The hypotheses are

H0: m Ú 65,

HA: m 6 65

The quality-control supervisor also wishes to construct a power curve for the test.

Solution:

The quality-control supervisor computes, for example, the following value

of 1 - b for the alternative m = 55. The critical value of 1 - b for the

test is

65 - 2.33a

We find b as follows:

15

b = 57

120

b = P1x 7 57 ƒ m = 552 = Paz 7

= 1 - .7257 = .2743

57 - 55

b = P1z 7 .602

15> 120

Consequently, 1 - b = 1 - .2743 = .7257. Figure 7.9.3 shows the calcu-

lation of b. Similar calculations for other alternative values of m also yield

b = 0.2743

a = 0.01

55

FIGURE 7.9.3

57

65

b calculated for m = 55.

x–278

CHAPTER 7 HYPOTHESIS TESTING

1–b

1.00

0.90

0.80

0.70

0.60

0.50

0.40

0.30

0.20

0.10

51

FIGURE 7.9.4

53

55

57

59

Alternative values of m

61

63

65

Power curve for Example 7.9.2.

values of 1 - b. When plotted against the values of m, these give the power

curve shown in Figure 7.9.4.

■

Operating Characteristic Curves Another way of evaluating a test is to

look at its operating characteristic (OC) curve. To construct an OC curve, we plot val-

ues of b, rather than 1 - b, along the vertical axis. Thus, an OC curve is the comple-

ment of the corresponding power curve.

EXERCISES

Construct and graph the power function for each of the following situations.

7.9.1H0: m … 516, HA: m 7 516, n = 16, s = 32, a = 0.05.

7.9.2H0: m = 3, HA: m Z 3, n = 100, s = 1, a = 0.05.

7.9.3H0: m … 4.25, HA: m 7 4.25, n = 81, s = 1.8, a = 0.01.

7.10 DETERMINING SAMPLE SIZE

TO CONTROL TYPE II ERRORS

You learned in Chapter 6 how to find the sample sizes needed to construct confidence

intervals for population means and proportions for specified levels of confidence. You

learned in Chapter 7 that confidence intervals may be used to test hypotheses. The

method of determining sample size presented in Chapter 6 takes into account the prob-

ability of a type I error, but not a type II error since the level of confidence is deter-

mined by the confidence coefficient, 1 - a.7.10 DETERMINING SAMPLE SIZE TO CONTROL TYPE II ERRORS

279

In many statistical inference procedures, the investigator wishes to consider the

type II error as well as the type I error when determining the sample size. To illustrate

the procedure, we refer again to Example 7.9.2.

EXAMPLE 7.10.1

In Example 7.9.2, the hypotheses are

H0: m Ú 65,

HA: m 6 65

The population standard deviation is 15, and the probability of a type I error is set at

.01. Suppose that we want the probability of failing to reject H01b2 to be .05 if H0 is

false because the true mean is 55 rather than the hypothesized 65. How large a sample

do we need in order to realize, simultaneously, the desired levels of a and b?

Solution:

For a = .01 and n = 20, b is equal to .2743. The critical value is 57. Under

the new conditions, the critical value is unknown. Let us call this new crit-

ical value C. Let m0 be the hypothesized mean and m1 the mean under the

alternative hypothesis. We can transform each of the relevant sampling dis-

tributions of x, the one with a mean of m0 and the one with a mean of m1

to a z distribution. Therefore, we can convert C to a z value on the hori-

zontal scale of each of the two standard normal distributions. When we

transform the sampling distribution of x that has a mean of m0 to the stan-

dard normal distribution, we call the z that results z 0. When we transform

the sampling distribution x that has a mean of m1 to the standard normal

distribution, we call the z that results z 1. Figure 7.10.1 represents the situ-

ation described so far.

We can express the critical value C as a function of z 0 and m0 and

also as a function of z 1 and m1. This gives the following equations:

s

1n

s

C = m1 + z 1

1n

C = m0 - z 0

(7.10.1)

(7.10.2)

b

a

m1C

0z1

z0

m0

x–

z

0

z

FIGURE 7.10.1 Graphic representation of relationships in determination

of sample size to control both type I and type II errors.280

CHAPTER 7 HYPOTHESIS TESTING

We set the right-hand sides of these equations equal to each other and solve

for n, to obtain

n = c

1z 0 + z 12s 2

d

1m0 - m12

(7.10.3)

To find n for our illustrative example, we substitute appropriate quanti-

ties into Equation 7.10.3. We have m0 = 65, m1 = 55, and s = 15. From

Appendix Table D, the value of z that has .01 of the area to its left is -2.33.

The value of z that has .05 of the area to its right is 1.645. Both z 0 and z 1 are

taken as positive. We determine whether C lies above or below either m0 or

m1 when we substitute into Equations 7.10.1 and 7.10.2. Thus, we compute

n = c

12.33 + 1.64521152 2

d = 35.55

165 - 552

We would need a sample of size 36 to achieve the desired levels of a and

b when we choose m1 = 55 as the alternative value of m.

We now compute C, the critical value for the test, and state an appro-

priate decision rule. To find C, we may substitute known numerical values

into either Equation 7.10.1 or Equation 7.10.2. For illustrative purposes, we

solve both equations for C. First we have

C = 65 - 2.33 a15

b = 59.175

136

C = 55 - 1.645 a15

b = 59.1125

136

From Equation 7.10.2, we have

The difference between the two results is due to rounding error.

The decision rule, when we use the first value of C, is as follows:

Select a sample of size 36 and compute x , if x … 59.175 , reject H0. If

x 7 59.175, do not reject H0.

We have limited our discussion of the type II error and the power of

a test to the case involving a population mean. The concepts extend to cases

involving other parameters.

■

EXERCISES

7.10.1Given H0: m = 516, HA: m 7 516, n = 16, s = 32, a = .05. Let b = .10 and m1 = 520, and

find n and C. State the appropriate decision rule.

7.10.2Given H0: m … 4.500, HA: m 7 4.500, n = 16, s = .020, a = .01. Let b = .05 and m1 =

4.52, and find n and C. State the appropriate decision rule.

7.10.3Given H0: m … 4.25, HA: m 7 4.25, n = 81, s = 1.8, a = .01. Let b = .03 and m1 = 5.00,

and find n and C. State the appropriate decision ru