**One-Way ANOVA**

The simplest type of analysis of variance is that known as one-way analysis of variance, in which only one source of variation, or factor, is investigated. It is an extension to three or more samples of the t test procedure (discussed in Chapter 7) for use with two independent samples. Stated another way, we can say that the t test for use with two independent samples is a special case of one-way analysis of variance.

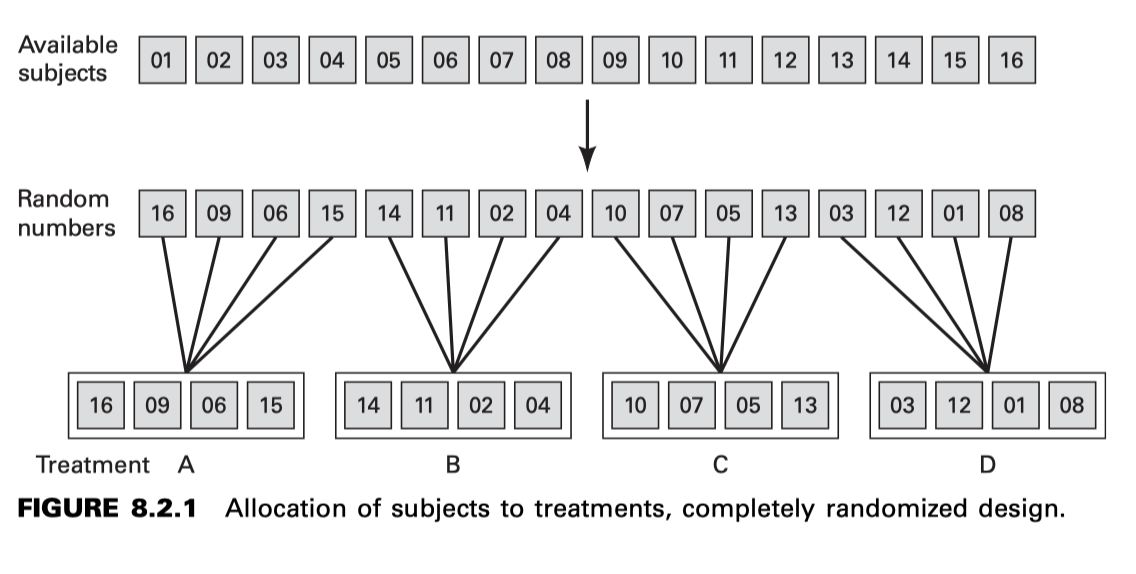
El tipo más simple de análisis de varianza es el conocido como **análisis de varianza unidireccional,** en el que solo se investiga una fuente de variación o **factor**. Es una extensión a tres o más muestras del procedimiento de prueba t (que se analiza en el Capítulo 7) para usar con **dos muestras independientes**. Dicho de otra manera, podemos decir que la prueba t para usar con dos muestras independientes es un caso especial de análisis de varianza unidireccional.

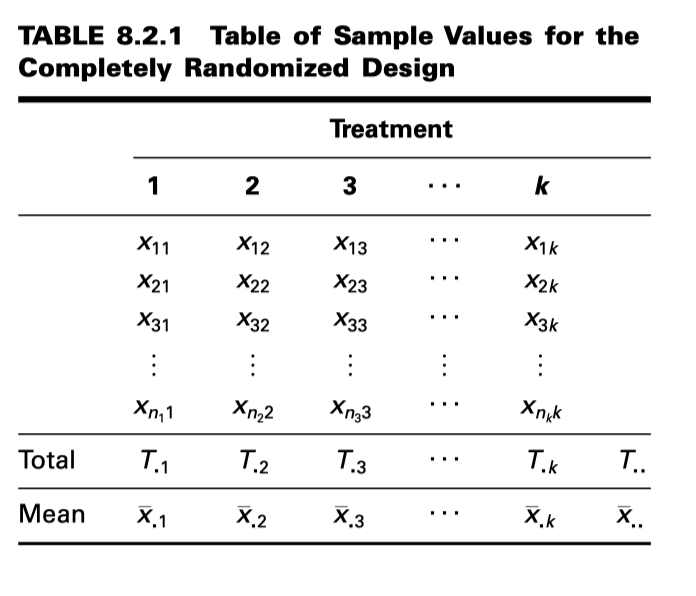
In a typical situation we want to use one-way analysis of variance to test the null hypothesis that three or more treatments are equally effective. The necessary experiment is designed in such a way that the treatments of interest are assigned completely at random to the subjects or objects on which the measurements to determine treatment effectiveness are to be made. For this reason the design is called the completely randomized experimental design.

En una situación típica queremos utilizar un análisis de varianza unidireccional para probar la hipótesis nula de que tres o más tratamientos **son igualmente efectivos**. El experimento necesario se diseña de tal manera que los tratamientos de interés se asignan completamente al azar a los sujetos u objetos sobre los cuales se van a realizar las mediciones para determinar la efectividad del tratamiento. Por esta razón el diseño se denomina **diseño experimental completamente al azar.**

We may randomly allocate subjects to treatments as follows. Suppose we have 16 subjects available to participate in an experiment in which we wish to compare four drugs. We number the subjects from 01 through 16. We then go to a table of random numbers and select 16 consecutive, unduplicated numbers between 01 and 16. To illustrate, let us use Appendix Table A and a random starting point that, say, is at the intersection of Row 4 and Columns 11 and 12. The two-digit number at this intersection is 98. The succeeding (moving downward) 16 consecutive two-digit numbers between 01 and 16 are 16, 09, 06, 15, 14, 11, 02, 04, 10, 07, 05, 13, 03, 12, 01, and 08. We allocate subjects 16, 09, 06, and 15 to drug A; subjects 14, 11, 02, and 04 to drug B; subjects 10, 07, 05, and 13 to drug C; and subjects 03, 12, 01, and 08 to drug D. We emphasize that the number of subjects in each treatment group does not have to be the same. Figure 8.2.1 illustrates the scheme of random allocation.

Podemos asignar sujetos aleatoriamente a los tratamientos de la siguiente manera. Supongamos que tenemos 16 sujetos disponibles para participar en un experimento en el que deseamos comparar cuatro fármacos. Numeramos los sujetos del 01 al 16. Luego vamos a una tabla de números aleatorios y seleccionamos 16 números consecutivos no duplicados entre 01 y 16. Para ilustrar, usemos la Tabla A del Apéndice y un punto de partida aleatorio que, digamos, está en la intersección de la fila 4 y las columnas 11 y 12. El número de dos dígitos en esta intersección es **98**. Los 16 números consecutivos de dos dígitos siguientes (moviéndose hacia abajo) entre 01 y 16 son 16, 09, 06, 15, 14, 11 , 02, 04, 10, 07, 05, 13, 03, 12, 01 y 08. Asignamos los sujetos 16, 09, 06 y 15 al fármaco A; los sujetos 14, 11, 02 y 04 al fármaco B; los sujetos 10, 07, 05 y 13 al fármaco C; y los sujetos 03, 12, 01 y 08 al fármaco D. Destacamos que el número de sujetos en cada grupo de tratamiento no tiene por qué ser el mismo. La Figura 8.2.1 ilustra el esquema de asignación aleatoria.





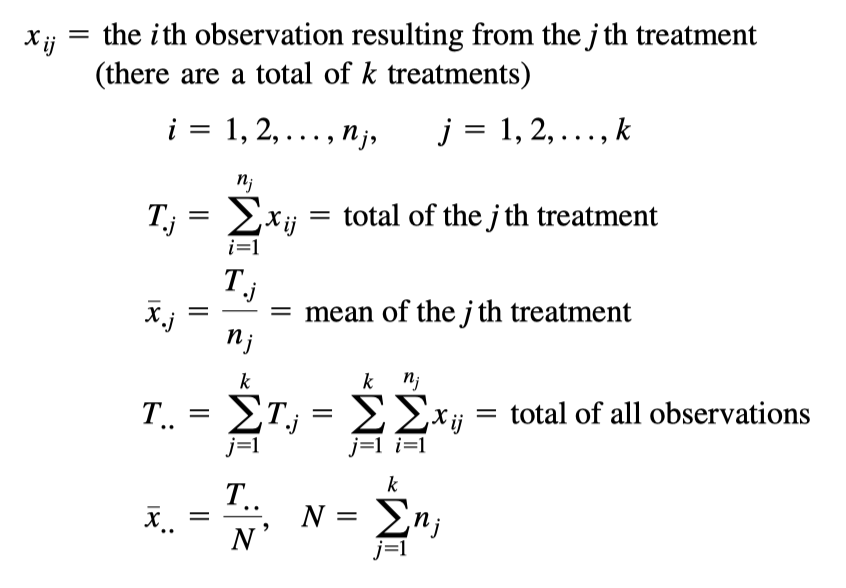
**Hypothesis Testing Steps**

Once we decide that the completely randomized design is the appropriate design, we may proceed with the hypothesis testing steps. We discuss these in detail ﬁrst, and follow with an example.

Una vez que decidimos que el diseño completamente aleatorio es el diseño apropiado, podemos continuar con los pasos de prueba de hipótesis. Primero los analizamos en detalle y seguimos con un ejemplo.

**1. Description of data.** The measurements (or observations) resulting from a completely randomized experimental design, along with the means and totals that can be computed from them, may be displayed for convenience as in Table 8.2.1. The symbols used in Table 8.2.1 are deﬁned as follows:

Las mediciones (u observaciones) resultantes de un diseño experimental completamente aleatorio, junto con las medias y los totales que pueden calcularse a partir de ellas, pueden presentarse por conveniencia como en la Tabla 8.2.1. Los símbolos utilizados en la Tabla 8.2.1 se definen de la siguiente manera:



**2. Assumptions.** Before stating the assumptions, let us specify the model for the experiment described here.

The Model As already noted, a model is a symbolic representation of a typical value of a data set. To write down the model for the completely randomized experimental design, let us begin by identifying a typical value from the set of data represented by the sample displayed in Table 8.2.1. We use the symbol x ij to represent this typical value.

El modelo Como ya se señaló, un modelo es una representación simbólica de un valor típico de un conjunto de datos. Para escribir el modelo para el diseño experimental completamente aleatorio, comencemos identificando un valor típico del conjunto de datos representado por la muestra que se muestra en la Tabla 8.2.1. Usamos el símbolo x ij para representar este valor típico.

The one-way analysis of variance model may be written as follows:

El modelo de análisis de varianza unidireccional se puede escribir de la siguiente manera:

x ij = m + t j + Pij ;

(8.2.1)

i = 1, 2, Á , n j ,

j = 1, 2, Á , k

The terms in this model are deﬁned as follows:

1. m represents the mean of all the k population means and is called the grand mean.

2. t j represents the difference between the mean of the jth population and the grand mean and is called the treatment effect.

3. P ij represents the amount by which an individual measurement differs from the mean of the population to which it belongs and is called the error term.

**Components of the Model**

By looking at our model we can see that a typical observation from the total set of data under study is composed of (1) the grand mean, (2) a treatment effect, and (3) an error term representing the deviation of the observation from its group mean.

Al observar nuestro modelo, podemos ver que una observación típica del conjunto total de datos bajo estudio se compone de (1) la media general, (2) un efecto de tratamiento y (3) un término de error que representa la desviación de la observación. de su media grupal.

In most situations we are interested only in the k treatments represented in our experiment. Any inferences that we make apply only to these treatments. We do not wish to extend our inference to any larger collection of treatments. When we place such a restriction on our inference goals, we refer to our model as the fixed-effects model, or model 1. The discussion in this book is limited to this model.

En la mayoría de las situaciones sólo nos interesan los k tratamientos representados en nuestro experimento. Cualquier inferencia que hagamos se aplica únicamente a estos tratamientos. No deseamos extender nuestra inferencia a una colección más amplia de tratamientos. Cuando imponemos tal restricción a nuestros objetivos de inferencia, nos referimos a nuestro modelo como modelo de efectos fijos, o modelo 1. La discusión en este libro se limita a este modelo.

Assumptions of the Model follows:

The assumptions for the ﬁxed-effects model are as

(a) The k sets of observed data constitute k independent random samples from the respective populations.

Los k conjuntos de datos observados constituyen k muestras aleatorias independientes de las respectivas poblaciones.

(b) Each of the populations from which the samples come is normally distributed with mean m j and variance sj 2 .

Cada una de las poblaciones de donde provienen las muestras se distribuye normalmente con media m j y varianza sj 2 .

(c) Each of the populations has the same variance. That is, s 1 2 = s 2 2 = . . . = s k 2 = s2 the common variance.

(d) The t j are unknown constants and gt j = 0 since the sum of all deviations of the m j from their mean, m, is zero.

(e) The P ij have a mean of 0, since the mean of x ij is m j .

(f) The P ij have a variance equal to the variance of the x ij , since the P ij and x ij differ only by a constant; that is, the error variance is equal to s2 , the common variance speciﬁed in assumption c.

(g) The P ij are normally (and independently) distributed.

3. Hypotheses. We test the null hypothesis that all population or treatment means are equal against the alternative that the members of at least one pair are not equal. We may state the hypotheses formally as follows:

Probamos la hipótesis nula de que todas las medias de población o tratamiento son iguales frente a la alternativa de que los miembros de al menos un par no son iguales. Podemos enunciar las hipótesis formalmente de la siguiente manera:

H 0 :m 1 = m 2 = . . . = mk

H A :not all m j are equal

If the population means are equal, each treatment effect is equal to zero, so that, alternatively, the hypotheses may be stated as

Si las medias poblacionales son iguales, cada efecto de tratamiento es igual a cero, de modo que, alternativamente, las hipótesis pueden expresarse como

j = 1, 2, Á , k

H0 : tj = 0,

HA : not all tj = 0

If H0 is true and the assumptions of equal variances and normally distributed populations are met, a picture of the populations will look like Figure 8.2.2. When H0 is true the population means are all equal, and the populations are centered at the same point (the common mean) on the horizontal axis. If the populations are all normally distributed with equal variances the distributions will be identical, so that in drawing their pictures each is superimposed on each of the others, and a single picture sufficiently represents them all.

Si H0 es verdadera y se cumplen los supuestos de varianzas iguales y poblaciones distribuidas normalmente, una imagen de las poblaciones se verá como en la Figura 8.2.2. Cuando H0 es verdadera, las medias poblacionales son todas iguales y las poblaciones están centradas en el mismo punto (la media común) en el eje horizontal. Si todas las poblaciones están distribuidas normalmente con varianzas iguales, las distribuciones serán idénticas, de modo que al dibujar sus imágenes, cada una se superpone a las demás, y una sola imagen las representa suficientemente a todas.

When H0 is false it may be false because one of the population means is different from the others, which are all equal. Or, perhaps, all the population means are different. These are only two of the possibilities when H0 is false. There are many other possible combinations of equal and unequal means. Figure 8.2.3 shows a picture of the populations when the assumptions are met, but H0 is false because no two population means are equal.

Cuando H0 es falsa, puede ser falsa porque una de las medias de la población es diferente de las demás, que son todas iguales. O, quizás, todas las medias poblacionales son diferentes. Éstas son sólo dos de las posibilidades cuando H0 es falsa. Hay muchas otras combinaciones posibles de medios iguales y desiguales. La Figura 8.2.3 muestra una imagen de las poblaciones cuando se cumplen los supuestos, pero H0 es falsa porque no hay dos medias poblacionales iguales.

**4. Test statistic.** The test statistic for one-way analysis of variance is a computed variance ratio, which we designate by V.R. as we did in Chapter 7. The two variances from which V.R. is calculated are themselves computed from the sample data. The methods by which they are calculated will be given in the discussion that follows.

El estadístico de prueba para el análisis de varianza unidireccional es una relación de varianza calculada, que denominamos V.R. como hicimos en el Capítulo 7. Las dos variaciones de las que V.R. se calcula se calculan a su vez a partir de los datos de la muestra. Los métodos mediante los cuales se calculan se darán en la discusión que sigue.

**5. Distribution of test statistic.** As discussed in Section 7.8, V.R. is distributed as the F distribution when H0 is true and the assumptions are met.

Como se discutió en la Sección 7.8, V.R. se distribuye como la distribución F cuando H0 es verdadera y se cumplen los supuestos.

**6. Decision rule.** In general, the decision rule is: reject the null hypothesis if the computed value of V.R. is equal to or greater than the critical value of F for the chosen a level.

En general, la regla de decisión es: rechazar la hipótesis nula si el valor calculado de V.R. es igual o mayor que el valor crítico de F para el nivel elegido.

**7. Calculation of test statistic.** We have defined analysis of variance as a process whereby the total variation present in a set of data is partitioned into components that are attributable to different sources. The term variation used in this context refers to the sum of squared deviations of observations from their mean, or sum of squares for short.

Hemos definido el análisis de varianza como un proceso mediante el cual la variación total presente en un conjunto de datos se divide en componentes que son atribuibles a diferentes fuentes. El término variación utilizado en este contexto se refiere a la suma de las desviaciones al cuadrado de las observaciones de su media, o suma de cuadrados para abreviar.

The initial computations performed in one-way ANOVA consist of the partitioning of the total variation present in the observed data into its basic components, each of which is attributable to an identifiable source.

Los cálculos iniciales realizados en ANOVA unidireccional consisten en dividir la variación total presente en los datos observados en sus componentes básicos, cada uno de los cuales es atribuible a una fuente identificable.

Those who use a computer for calculations may wish to skip the following discussion of the computations involved in obtaining the test statistic.

Quienes utilicen una computadora para realizar cálculos tal vez deseen omitir la siguiente discusión sobre los cálculos involucrados en la obtención de la estadística de prueba.

**The Total Sum of Squares** Before we can do any partitioning, we must first obtain the total sum of squares. The total sum of squares is the sum of the squares of the deviations of individual observations from the mean of all the observations taken together. This total sum of squares is defined as

Antes de poder hacer cualquier partición, primero debemos obtener la suma total de cuadrados. La suma total de cuadrados es la suma de los cuadrados de las desviaciones de las observaciones individuales de la media de todas las observaciones tomadas en conjunto. Esta suma total de cuadrados se define como

SST = a a 1x ij - x ..22

k

nj

(8.2.2)

j =1 i =1

j

where g i =1 tells us to sum the squared deviations for each treatment group, and g j =1 nj tells us to add the k group totals obtained by applying g i =1 . The reader will recognize Equation 8.2.2 as the numerator of the variance that may be computed from the complete set of observations taken together.

donde g i =1 nos indica sumar las desviaciones al cuadrado para cada grupo de tratamiento, y g j =1 nj nos indica sumar los k totales de los grupos obtenidos aplicando g i =1. El lector reconocerá la ecuación 8.2.2 como el numerador de la varianza que puede calcularse a partir del conjunto completo de observaciones tomadas en conjunto.

**The Within Groups Sum of Squares** Now let us show how to compute the first of the two components of the total sum of squares.

The first step in the computation calls for performing certain calculations within each group. These calculations involve computing within each group the sum of the squared deviations of the individual observations from their mean. When these calculations have been performed within each group, we obtain the sum of the individual group results. This component of variation is called the within groups sum of squares and may be designated SSW. This quantity is sometimes referred to as the residual or error sum of squares. The expression for these calculations is written as follows:

El primer paso en el cálculo requiere realizar ciertos cálculos dentro de cada grupo. Estos cálculos implican calcular dentro de cada grupo la suma de las desviaciones al cuadrado de las observaciones individuales de su media. Cuando estos cálculos se han realizado dentro de cada grupo, obtenemos la suma de los resultados de cada grupo. Este componente de variación se denomina suma de cuadrados dentro de los grupos y puede denominarse SSW. Esta cantidad a veces se denomina suma de cuadrados residual o de error. La expresión para estos cálculos se escribe de la siguiente manera:

SSW = a a 1x ij - x .j22

k

nj

j =1 i =1

(8.2.3)

**The Among Groups Sum of Squares** To obtain the second component of the total sum of squares, we compute for each group the squared deviation of the group mean from the grand mean and multiply the result by the size of the group. Finally, we add these results over all groups. This quantity is a measure of the variation among groups and is referred to as the sum of squares among groups or SSA. The formula for calculating this quantity is as follows:

Para obtener el segundo componente de la suma total de cuadrados, calculamos para cada grupo la desviación al cuadrado de la media del grupo de la media general y multiplicamos el resultado por el tamaño del grupo. Finalmente, sumamos estos resultados en todos los grupos. Esta cantidad es una medida de la variación entre grupos y se conoce como suma de cuadrados entre grupos o SSA. La fórmula para calcular esta cantidad es la siguiente:

SSA = a n j 1x .j - x ..22

k

(8.2.4)

j =1

In summary, then, we have found that the total sum of squares is equal to the sum of the among and the within sum of squares. We express this relationship as follows:

En resumen, entonces, hemos encontrado que la suma total de cuadrados es igual a la suma de la suma de cuadrados entre y dentro. Expresamos esta relación de la siguiente manera:

SST = SSA + SSW

From the sums of squares that we have now learned to compute, it is possible to obtain two estimates of the common population variance, s2. It can be shown that when the assumptions are met and the population means are all equal, both the among sum of squares and the within sum of squares, when divided by their respective degrees of freedom, yield independent and unbiased estimates of s2.

A partir de las sumas de cuadrados que ahora hemos aprendido a calcular, es posible obtener dos estimaciones de la varianza común de la población, s2. Se puede demostrar que cuando se cumplen los supuestos y las medias poblacionales son todas iguales, tanto la suma de cuadrados entre como la suma de cuadrados dentro, cuando se dividen por sus respectivos grados de libertad, producen estimaciones independientes e insesgadas de s2.

**The First Estimate of S2** Within any sample,

a 1x ij - x .j2

nj

2

i =1

nj - 1

provides an unbiased estimate of the true variance of the population from which the sample came. Under the assumption that the population variances are all equal, we may pool the k estimates to obtain

proporciona una estimación insesgada de la varianza real de la población de la que procede la muestra. Bajo el supuesto de que todas las varianzas poblacionales son iguales, podemos agrupar las k estimaciones para obtener

a a 1x ij - x .j2

k

MSW =

nj

j =1 i =1

a 1n j - 12

k

2

(8.2.5)

j =1

This is our first estimate of s2 and may be called the within groups variance, since it is the within groups sum of squares of Equation 8.2.3 divided by the appropriate degrees of freedom. The student will recognize this as an extension to k samples of the pooling of variances procedure encountered in Chapters 6 and 7 when the variances from two samples were pooled in order to use the t distribution. The quantity in Equation 8.2.5 is customarily referred to as the within groups mean square rather than the within groups variance.

Esta es nuestra primera estimación de s2 y puede denominarse varianza dentro de los grupos, ya que es la suma de cuadrados dentro de los grupos de la Ecuación 8.2.3 dividida por los grados de libertad apropiados. El estudiante reconocerá esto como una extensión a k muestras del procedimiento de combinación de varianzas encontrado en los Capítulos 6 y 7 cuando se combinaron las varianzas de dos muestras para usar la distribución t. La cantidad en la Ecuación 8.2.5 se conoce habitualmente como media cuadrática dentro de los grupos en lugar de varianza dentro de los grupos.

The within groups mean square is a valid estimate of s2 only if the population variances are equal. It is not necessary, however, for H0 to be true in order for the within groups mean square to be a valid estimate of s2; that is, the within groups mean square estimates s2 regardless of whether H0 is true or false, as long as the population variances are equal.

El cuadrado medio dentro de los grupos es una estimación válida de s2 sólo si las varianzas de la población son iguales. Sin embargo, no es necesario que H0 sea verdadera para que el cuadrado medio dentro de los grupos sea una estimación válida de s2; es decir, el cuadrado medio dentro de los grupos estima s2 independientemente de si H0 es verdadera o falsa, siempre que las varianzas poblacionales sean iguales.

**The Second Estimate of S2** The second estimate of s2 may be obtained from the familiar formula for the variance of sample means, s2x = s2>n. If we solve this equation for s2, the variance of the population from which the samples were drawn, we have

La segunda estimación de s2 puede obtenerse a partir de la conocida fórmula para la varianza de las medias muestrales, s2x = s2>n. Si resolvemos esta ecuación para s2, la varianza de la población de la que se extrajeron las muestras, tenemos

s2 = ns2x

(8.2.6)

An unbiased estimate of s2x computed from sample data is provided by

Una estimación insesgada de s2x calculada a partir de datos de muestra la proporciona

a 1x .j - x . .2

k

2

j =1

k - 1

If we substitute this quantity into Equation 8.2.6, we obtain the desired estimate of s2,

n a 1x.j - x.. 22

k

MSA =

j =1

k - 1

(8.2.7)

The reader will recognize the numerator of Equation 8.2.7 as the among groups sum of squares for the special case when all sample sizes are equal. This sum of squares when divided by the associated degrees of freedom k - 1 is referred to as the among groups mean square.

El lector reconocerá el numerador de la ecuación 8.2.7 como la suma de cuadrados entre grupos para el caso especial en el que todos los tamaños de muestra son iguales. Esta suma de cuadrados cuando se divide por los grados de libertad asociados k - 1 se denomina cuadrado medio entre grupos.

When the sample sizes are not all equal, an estimate of s2 based on the variability among sample means is provided by

Cuando los tamaños de muestra no son todos iguales, una estimación de s2 basada en la variabilidad entre las medias muestrales se obtiene mediante

a n j 1x.j - x..2

k

MSA =

2

j =1

k - 1

(8.2.8)

If, indeed, the null hypothesis is true we would expect these two estimates of s2 to be fairly close in magnitude. If the null hypothesis is false, that is, if all population means are not equal, we would expect the among groups mean square, which is computed by using the squared deviations of the sample means from the overall mean, to be larger than the within groups mean square.

Si, de hecho, la hipótesis nula es cierta, esperaríamos que estas dos estimaciones de s2 tuvieran una magnitud bastante similar. Si la hipótesis nula es falsa, es decir, si todas las medias de la población no son iguales, esperaríamos que la media cuadrática entre grupos, que se calcula utilizando las desviaciones al cuadrado de las medias muestrales de la media general, sea mayor que la media cuadrática interna. los grupos significan cuadrados.

In order to understand analysis of variance we must realize that the among groups mean square provides a valid estimate of s2 when the assumption of equal population variances is met and when H0 is true. Both conditions, a true null hypothesis and equal population variances, must be met in order for the among groups mean square to be a valid estimate of s2.

Para comprender el análisis de varianza debemos darnos cuenta de que el cuadrado medio entre grupos proporciona una estimación válida de s2 cuando se cumple el supuesto de varianzas poblacionales iguales y cuando H0 es verdadera. Se deben cumplir ambas condiciones, una hipótesis nula verdadera y varianzas poblacionales iguales, para que el cuadrado medio entre grupos sea una estimación válida de s2.

**The Variance Ratio** What we need to do now is to compare these two estimates of s2, and we do this by computing the following variance ratio, which is the desired test statistic:

Lo que tenemos que hacer ahora es comparar estas dos estimaciones de s2, y lo hacemos calculando la siguiente relación de varianza, que es el estadístico de prueba deseado:

V.R. =

among groups mean square

within groups means square

If the two estimates are about equal, V.R. will be close to 1. A ratio close to 1 tends to support the hypothesis of equal population means. If, on the other hand, the among groups mean square is considerably larger than the within groups mean square, V.R. will be considerably greater than 1. A value of V.R. sufficiently greater than 1 will cast doubt on the hypothesis of equal population means.

Si las dos estimaciones son aproximadamente iguales, V.R. será cercano a 1. Una proporción cercana a 1 tiende a apoyar la hipótesis de medias poblacionales iguales. Si, por otro lado, el cuadrado medio entre grupos es considerablemente mayor que el cuadrado medio dentro de los grupos, V.R. será considerablemente mayor que 1. Un valor de V.R. suficientemente mayor que 1 arrojará dudas sobre la hipótesis de medias poblacionales iguales.

We know that because of the vagaries of sampling, even when the null hypothesis is true, it is unlikely that the among and within groups mean squares will be equal. We must decide, then, how big the observed difference has to be before we can conclude that the difference is due to something other than sampling fluctuation. In other words, how large a value of V.R. is required for us to be willing to conclude that the observed difference between our two estimates of s2 is not the result of chance alone?

Sabemos que debido a los caprichos del muestreo, incluso cuando la hipótesis nula es verdadera, es poco probable que los cuadrados medios entre y dentro de los grupos sean iguales. Entonces, debemos decidir qué tan grande debe ser la diferencia observada antes de poder concluir que la diferencia se debe a algo distinto a la fluctuación muestral. En otras palabras, ¿qué tan grande es el valor de V.R. ¿Qué se requiere para que estemos dispuestos a concluir que la diferencia observada entre nuestras dos estimaciones de s2 no es resultado únicamente del azar?

**The F Test** To answer the question just posed, we must consider the sampling distribution of the ratio of two sample variances. In Chapter 6 we learned that the quantity 1s 12>s212>1s 22>s222 follows a distribution known as the F distribution when the sample variances are computed from random and independently drawn samples from normal populations. The F distribution, introduced by R. A. Fisher in the early 1920s, has become one of the most widely used distributions in modern statistics. We have already become acquainted with its use in constructing confidence intervals for, and testing hypotheses about, population variances. In this chapter, we will see that it is the distribution fundamental to analysis of variance. For this reason the ratio that we designate V.R. is frequently referred to as F, and the testing procedure is frequently called the F test. It is of interest to note that the F distribution is the ratio of two Chi-square distributions.

Para responder a la pregunta que acabamos de plantear, debemos considerar la distribución muestral de la razón de dos varianzas muestrales. En el capítulo 6 aprendimos que la cantidad 1s 12>s212>1s 22>s222 sigue una distribución conocida como distribución F cuando las varianzas muestrales se calculan a partir de muestras aleatorias e independientes de poblaciones normales. La distribución F, introducida por R. A. Fisher a principios de la década de 1920, se ha convertido en una de las distribuciones más utilizadas en la estadística moderna. Ya nos hemos familiarizado con su uso para construir intervalos de confianza y probar hipótesis sobre las varianzas poblacionales. En este capítulo veremos que la distribución es fundamental para el análisis de varianza. Por este motivo el ratio que denominamos V.R. Con frecuencia se hace referencia a ella como F, y el procedimiento de prueba frecuentemente se denomina prueba F. Es interesante observar que la distribución F es la relación de dos distribuciones Chi-cuadrado.

In Chapter 7 we learned that when the population variances are the same, they cancel in the expression 1s 21>s212>1s 22>s222, leaving s 21>s 22, which is itself distributed as F. The F distribution is really a family of distributions, and the particular F distribution we use in a given situation depends on the number of degrees of freedom associated with the sample variance in the numerator (numerator degrees of freedom) and the number of degrees of freedom associated with the sample variance in the denominator (denominator degrees of freedom).

En el capítulo 7 aprendimos que cuando las varianzas poblacionales son iguales, se cancelan en la expresión 1s 21>s212>1s 22>s222, dejando s 21>s 22, que a su vez se distribuye como F. ​​La distribución F es en realidad una familia. de distribuciones, y la distribución F particular que usamos en una situación dada depende del número de grados de libertad asociados con la varianza muestral en el numerador (grados de libertad del numerador) y del número de grados de libertad asociados con la varianza muestral en el denominador (grados de libertad del denominador).

Once the appropriate F distribution has been determined, the size of the observed V.R. that will cause rejection of the hypothesis of equal population variances depends on the significance level chosen. The significance level chosen determines the critical value of F, the value that separates the nonrejection region from the rejection region.

Una vez que se ha determinado la distribución F apropiada, el tamaño del V.R. Lo que provocará el rechazo de la hipótesis de varianzas poblacionales iguales depende del nivel de significancia elegido. El nivel de significancia elegido determina el valor crítico de F, el valor que separa la región de no rechazo de la región de rechazo.

As we have seen, we compute V.R. in situations of this type by placing the among groups mean square in the numerator and the within groups mean square in the denominator, so that the numerator degrees of freedom is equal to 1k - 12, the number of groups minus 1, and the denominator degrees of freedom value is equal to

Como hemos visto, calculamos V.R. en situaciones de este tipo colocando el cuadrado medio entre grupos en el numerador y el cuadrado medio dentro de grupos en el denominador, de modo que los grados de libertad del numerador sean iguales a 1k - 12, el número de grupos menos 1 y los grados del denominador del valor de la libertad es igual a

a 1n j - 12 = a a n j b - k = N - k

kk

j =1j =1

**The ANOVA Table** The calculations that we perform may be summarized and displayed in a table such as Table 8.2.2, which is called the ANOVA table.

Los cálculos que realizamos se pueden resumir y mostrar en una tabla como la Tabla 8.2.2, que se llama tabla ANOVA.

**8. Statistical decision.** To reach a decision we must compare our computed V.R. with the critical value of F, which we obtain by entering Appendix Table G with k - 1 numerator degrees of freedom and N - k denominator degrees of freedom.

Para tomar una decisión debemos comparar nuestro V.R. con el valor crítico de F, que obtenemos ingresando a la Tabla G del Apéndice con k - 1 grados de libertad del numerador y N - k grados de libertad del denominador.

If the computed V.R. is equal to or greater than the critical value of F, we reject the null hypothesis. If the computed value of V.R. is smaller than the critical value of F, we do not reject the null hypothesis.

Si el V.R. es igual o mayor que el valor crítico de F, rechazamos la hipótesis nula. Si el valor calculado de V.R. es menor que el valor crítico de F, no rechazamos la hipótesis nula.

**Explaining a Rejected Null Hypothesis**

There are two possible expla- nations for a rejected null hypothesis. If the null hypothesis is true, that is, if the two sample variances are estimates of a common variance, we know that the probability of getting a value of V.R. as large as or larger than the critical F is equal to our chosen level of significance. When we reject H0 we may, if we wish, conclude that the null hypothesis is true and assume that because of chance we got a set of data that gave rise to a rare event. On the other hand, we may prefer to take the position that our large computed V.R. value does not represent a rare event brought about by chance but, instead, reflects the fact that something other than chance is operative. This other something we conclude to be a false null hypothesis.

Hay dos explicaciones posibles para una hipótesis nula rechazada. Si la hipótesis nula es cierta, es decir, si las dos varianzas muestrales son estimaciones de una varianza común, sabemos que la probabilidad de obtener un valor de V.R. tan grande o mayor que el F crítico es igual a nuestro nivel de significancia elegido. Cuando rechazamos H0 podemos, si lo deseamos, concluir que la hipótesis nula es verdadera y suponer que debido al azar obtuvimos un conjunto de datos que dio lugar a un evento raro. Por otro lado, es posible que prefiramos adoptar la posición de que nuestro gran V.R. El valor no representa un acontecimiento raro provocado por el azar, sino que refleja el hecho de que algo distinto del azar está operativo. Este otro algo lo concluimos como una hipótesis nula falsa.

It is this latter explanation that we usually give for computed values of V.R. that exceed the critical value of F. In other words, if the computed value of V.R. is greater than the critical value of F, we reject the null hypothesis.

Esta última explicación es la que solemos dar para los valores calculados de V.R. que exceden el valor crítico de F. En otras palabras, si el valor calculado de V.R. es mayor que el valor crítico de F, rechazamos la hipótesis nula.

It will be recalled that the original hypothesis we set out to test was

Se recordará que la hipótesis original que nos propusimos probar era

H0 : m1 = m2 = Á = mk

Does rejection of the hypothesis about variances imply a rejection of the hypothesis of equal population means? The answer is yes. A large value of V.R. resulted from the fact that the among groups mean square was considerably larger than the within groups mean square. Since the among groups mean square is based on the dispersion of the sample means about their mean (called the grand mean), this quantity will be large when there is a large discrepancy among the sizes of the sample means. Because of this, then, a significant value of V.R. tells us to reject the null hypothesis that all population means are equal.

¿El rechazo de la hipótesis sobre las varianzas implica un rechazo de la hipótesis de igualdad de medias poblacionales? La respuesta es sí. Un gran valor de V.R. resultó del hecho de que el cuadrado medio entre grupos era considerablemente mayor que el cuadrado medio dentro de los grupos. Dado que el cuadrado medio entre grupos se basa en la dispersión de las medias muestrales respecto de su media (llamada gran media), esta cantidad será grande cuando haya una gran discrepancia entre los tamaños de las medias muestrales. Debido a esto, entonces, un valor significativo de V.R. nos dice que rechacemos la hipótesis nula de que todas las medias poblacionales son iguales.

**9. Conclusion.** When we reject H0, we conclude that not all population means are equal. When we fail to reject H0, we conclude that the population means may all be equal.

Cuando rechazamos H0, concluimos que no todas las medias poblacionales son iguales. Cuando no podemos rechazar H0, concluimos que todas las medias poblacionales pueden ser iguales.

10. Determination of p value.

Determinación del valor de p.

**EXAMPLE 8.2.1**

Game meats, including those from white-tailed deer and eastern gray squirrels, are used as food by families, hunters, and other individuals for health, cultural, or personal reasons. A study by David Holben (A-1) assessed the selenium content of meat from free-roaming white-tailed deer (venison) and gray squirrel (squirrel) obtained from a low selenium region of the United States. These selenium content values were also compared to those of beef produced within and outside the same region. We want to know if the selenium levels are different in the four meat groups.

Las carnes de caza, incluidas las del venado de cola blanca y las ardillas grises orientales, son utilizadas como alimento por familias, cazadores y otras personas por motivos de salud, culturales o personales. Un estudio realizado por David Holben (A-1) evaluó el contenido de selenio de la carne de venado de cola blanca (venado) y de ardilla gris (ardilla) en libertad obtenida de una región baja en selenio de los Estados Unidos. Estos valores de contenido de selenio también se compararon con los de la carne vacuna producida dentro y fuera de la misma región. Queremos saber si los niveles de selenio son diferentes en los cuatro grupos de carnes.

**Solution:**

**1. Description of data.**

Selenium content of raw venison (VEN), squirrel meat (SQU), region-raised beef (RRB), and nonregion-raised beef (NRB), in mg>100 g of dry weight, are shown in Table 8.2.3. A graph of the data in the form of a dotplot is shown in Figure 8.2.4. Such a graph highlights the main features of the data and brings into clear focus differences in selenium levels among the different meats.

El contenido de selenio del venado crudo (VEN), la carne de ardilla (SQU), la carne vacuna criada en la región (RRB) y la carne vacuna criada fuera de la región (NRB), en mg>100 g de peso seco, se muestra en la Tabla 8.2.3. En la Figura 8.2.4 se muestra una gráfica de los datos en forma de diagrama de puntos. Un gráfico de este tipo resalta las características principales de los datos y pone de relieve claramente las diferencias en los niveles de selenio entre las diferentes carnes.

**2. Assumptions.** We assume that the four sets of data constitute independent simple random samples from the four indicated populations. We assume that the four populations of measurements are normally distributed with equal variances.

Suponemos que los cuatro conjuntos de datos constituyen muestras aleatorias simples independientes de las cuatro poblaciones indicadas. Suponemos que las cuatro poblaciones de medidas se distribuyen normalmente con varianzas iguales.

**3. Hypotheses.**

H0 : m1 = m2 = m3 = m4 (On average the four meats have the same selenium content.)

En promedio, las cuatro carnes tienen el mismo contenido de selenio.

HA: Not all m’s are equal (At least one meat yields an average selenium content different from the average selenium content of at least one other meat.)

No todas las m son iguales (al menos una carne produce un contenido promedio de selenio diferente del contenido promedio de selenio de al menos otra carne).

**4. Test statistic.** The test statistic is V.R. = MSA>MSW.

**5. Distribution of test statistic.** If H0 is true and the assumptions are met, the V.R. follows the F distribution with 4 - 1 = 3 numerator degrees of freedom and 144 - 4 = 140 denominator degrees of freedom.

Si H0 es verdadera y se cumplen los supuestos, la V.R. sigue la distribución F con 4 - 1 = 3 grados de libertad del numerador y 144 - 4 = 140 grados de libertad del denominador.

**6. Decision rule.** Suppose we let a = .01. The critical value of F from Appendix Table G is 6 3.95. The decision rule, then, is reject H0 if the computed V.R. statistic is equal to or greater than 3.95.

Supongamos que dejamos a = .01. El valor crítico de F de la Tabla G del Apéndice es 6 3,95. La regla de decisión, entonces, es rechazar H0 si la V.R. estadística es igual o mayor que 3,95.

**7. Calculation of test statistic.** By Equation 8.2.2 we compute

SST = 58009.05560

By Equation 8.2.4 we compute

SSA = 21261.82886

SSW = 58009.05560 - 21261.82886 = 36747.22674

The results of our calculations are displayed in Table 8.2.4.

**8. Statistical decision.** Since our computed F of 27.00 is greater than 3.95 we reject H0.

Dado que nuestra F calculada de 27,00 es mayor que 3,95, rechazamos H0.

**9. Conclusion.** Since we reject H0, we conclude that the alternative hypothesis is true. That is, we conclude that the four meat types do not all have the same average selenium content.

Como rechazamos H0, concluimos que la hipótesis alternativa es verdadera. Es decir, concluimos que no todos los cuatro tipos de carne tienen el mismo contenido promedio de selenio.

**10. p value.** Since 27.00 7 3.95, p 6 .01 for this test.

**A Word of Caution**

The completely randomized design is simple and, therefore, widely used. It should be used, however, only when the units receiving the treatments are homogeneous. If the experimental units are not homogeneous, the researcher should consider an alternative design such as one of those to be discussed later in this chapter. In our illustrative example the treatments are treatments in the usual sense of the word. This is not always the case, however, as the term “treatment” as used in experimental design is quite general. We might, for example, wish to study the response to the same treatment (in the usual sense of the word) of several breeds of animals. We would, however, refer to the breed of animal as the “treatment.”

El diseño completamente aleatorizado es simple y, por tanto, ampliamente utilizado. Sin embargo, sólo debe utilizarse cuando las unidades que reciben los tratamientos sean homogéneas. Si las unidades experimentales no son homogéneas, el investigador debe considerar un diseño alternativo como uno de los que se analizarán más adelante en este capítulo. En nuestro ejemplo ilustrativo los tratamientos son tratamientos en el sentido habitual de la palabra. Sin embargo, este no es siempre el caso, ya que el término “tratamiento” tal como se utiliza en el diseño experimental es bastante general. Por ejemplo, podríamos desear estudiar la respuesta al mismo tratamiento (en el sentido habitual de la palabra) de varias razas de animales. Sin embargo, nos referiríamos a la raza del animal como el “tratamiento”.

We must also point out that, although the techniques of analysis of variance are more often applied to data resulting from controlled experiments, the techniques also may be used to analyze data collected by a survey, provided that the underlying assumptions are reasonably well met.

También debemos señalar que, aunque las técnicas de análisis de varianza se aplican más a menudo a datos resultantes de experimentos controlados, también pueden usarse para analizar datos recopilados mediante una encuesta, siempre que los supuestos subyacentes se cumplan razonablemente bien.

**Alternatives**

If the data available for analysis do not meet the assumptions for one- way analysis of variance as discussed here, one may wish to consider the use of the Kruskal-Wallis procedure, a nonparametric technique discussed in Chapter 13.

Si los datos disponibles para el análisis no cumplen con los supuestos para el análisis de varianza unidireccional como se analiza aquí, es posible que desee considerar el uso del procedimiento de Kruskal-Wallis, una técnica no paramétrica que se analiza en el Capítulo 13.

**Testing for Significant Differences Between Individual Pairs of Means**

When the analysis of variance leads to a rejection of the null hypothesis of no difference among population means, the question naturally arises regarding just which pairs of means are different. In fact, the desire, more often than not, is to carry out a significance test on each and every pair of treatment means. For instance, in Example 8.2.1, where there are four treatments, we may wish to know, after rejecting H0 : m1 = m2 = m3 = m4, which of the six possible individual hypotheses should be rejected. The experimenter, however, must exercise caution in testing for significant differences between individual means and must always make certain that the procedure is valid. The critical issue in the procedure is the level of significance. Although the probability, a, of rejecting a true null hypothesis for the test as a whole is made small, the probability of rejecting at least one true hypothesis when several pairs of means are tested is, as we have seen, greater than a. There are several multiple comparison procedures commonly used in practice. Below we illustrate two popular procedures, namely Tukey’s HSD test and Bonferroni’s method. The interested student is referred to the books by Hsu (7) and Westfall et al. (8) for additional techniques.

Cuando el análisis de varianza conduce a un rechazo de la hipótesis nula de que no hay diferencia entre las medias de las poblaciones, naturalmente surge la pregunta sobre qué pares de medias son diferentes. De hecho, lo que se desea, en la mayoría de los casos, es realizar una prueba de significancia de todos y cada uno de los pares de medios de tratamiento. Por ejemplo, en el ejemplo 8.2.1, donde hay cuatro tratamientos, es posible que deseemos saber, después de rechazar H0 : m1 = m2 = m3 = m4, cuál de las seis posibles hipótesis individuales debe rechazarse. Sin embargo, el experimentador debe tener cuidado al probar diferencias significativas entre medias individuales y siempre debe asegurarse de que el procedimiento sea válido. La cuestión crítica en el procedimiento es el nivel de significancia. Aunque la probabilidad, a, de rechazar una hipótesis nula verdadera para la prueba en su conjunto se hace pequeña, la probabilidad de rechazar al menos una hipótesis verdadera cuando se prueban varios pares de medias es, como hemos visto, mayor que a. Existen varios procedimientos de comparación múltiple comúnmente utilizados en la práctica. A continuación ilustramos dos procedimientos populares, a saber, la prueba HSD de Tukey y el método de Bonferroni. Se remite al estudiante interesado a los libros de Hsu (7) y Westfall et al. (8) para técnicas adicionales.

**Tukey’s HSD Test**

Over the years several procedures for making multiple com- parisons have been suggested. A multiple comparison procedure developed by Tukey (9) is frequently used for testing the null hypothesis that all possible pairs of treatment means are equal when the samples are all of the same size. When this test is employed we select an overall significance level of a. The probability is a, then, that one or more of the null hypotheses is false.

A lo largo de los años se han sugerido varios procedimientos para realizar comparaciones múltiples. Con frecuencia se utiliza un procedimiento de comparación múltiple desarrollado por Tukey (9) para probar la hipótesis nula de que todos los pares posibles de medias de tratamiento son iguales cuando todas las muestras son del mismo tamaño. Cuando se emplea esta prueba seleccionamos un nivel de significancia general de a. La probabilidad es, entonces, de que una o más de las hipótesis nulas sea falsa.

Tukey’s test, which is usually referred to as the HSD (honestly significant difference) test, makes use of a single value against which all differences are compared. This value, called the HSD, is given by

La prueba de Tukey, que generalmente se conoce como prueba HSD (diferencia honestamente significativa), utiliza un valor único con el que se comparan todas las diferencias. Este valor, llamado HSD, está dado por

HSD = qa,k,N -k

MSE

A n

(8.2.9)

where a is the chosen level of significance, k is the number of means in the experiment, N is the total number of observations in the experiment, n is the number of observations in a treatment, MSE is the error or within mean square from the ANOVA table, and q is obtained by entering Appendix Table H with a, k, and N - k.

donde a es el nivel de significancia elegido, k es el número de medias en el experimento, N es el número total de observaciones en el experimento, n es el número de observaciones en un tratamiento, MSE es el error o dentro del cuadrado medio del tabla ANOVA, y q se obtiene ingresando la Tabla H del Apéndice con a, k y N - k.

The statistic q, tabulated in Appendix Table H, is known as the studentized range statistic. It is defined as the difference between the largest and smallest treatment means from an ANOVA (that is, it is the range of the treatment means) divided by the error mean square over n, the number of observations in a treatment. The studentized range is discussed in detail by Winer (10).

El estadístico q, tabulado en la Tabla H del Apéndice, se conoce como estadístico de rango estudentizado. Se define como la diferencia entre las medias de tratamiento más grandes y más pequeñas de un ANOVA (es decir, es el rango de las medias de tratamiento) dividida por el cuadrado medio del error sobre n, el número de observaciones en un tratamiento. Winer (10) analiza en detalle el rango estudentizado.

All possible differences between pairs of means are computed, and any difference that yields an absolute value that exceeds HSD is declared significant.

Se calculan todas las diferencias posibles entre pares de medias y cualquier diferencia que produzca un valor absoluto que exceda HSD se declara significativa.

Tukey’s Test for Unequal Sample Sizes When the samples are not all the same size, as is the case in Example 8.2.1, Tukey’s HSD test given by Equation 8.2.9 is not applicable. Tukey himself (9) and Kramer (11), however, have extended the Tukey procedure to the case where the sample sizes are different. Their procedure, which is sometimes called the Tukey-Kramer method, consists of replacing MSE/n in Equation 8.2.9 with 1MSE >2211>n i + 1>n j2, where n i and n j are the sample sizes of the two groups to be compared. If we designate the new quantity by HSD\*, we have as the new test criterion

Prueba de Tukey para tamaños de muestra desiguales Cuando no todas las muestras son del mismo tamaño, como es el caso en el Ejemplo 8.2.1, la prueba HSD de Tukey dada por la Ecuación 8.2.9 no es aplicable. El propio Tukey (9) y Kramer (11), sin embargo, han extendido el procedimiento de Tukey al caso en que los tamaños de muestra son diferentes. Su procedimiento, que a veces se denomina método de Tukey-Kramer, consiste en reemplazar MSE/n en la ecuación 8.2.9 con 1MSE >2211>n i + 1>n j2, donde n i y n j son los tamaños de muestra de los dos grupos que se van a formar. comparado. Si designamos la nueva cantidad por HSD\*, tenemos como nuevo criterio de prueba

HSD\* = qa,k,N -k

MSE

A 2

a

1

1

+ b

ni

nj

(8.2.10)

Any absolute value of the difference between two sample means that exceeds HSD\* is declared significant.