BIOSTATISTICS\_Daniel\_C08

CHAPTER OVERVIEW

The topic of this chapter, analysis of variance, provides a methodology for partitioning the total variance computed from a data set into components, each of which represents the amount of the total variance that can be attributed to a specific source of variation. The results of this partitioning can then be used to estimate and test hypotheses about population variances and means. In this chapter we focus our attention on hypothes

is testing of means. Specifically, we discuss the testing of differences among means when there is interest in more than two populations or two or more variables. The techniques discussed in this chapter are widely used in the health sciences.

El tema de este capítulo, el análisis de la varianza, proporciona una metodología para dividir la varianza total calculada a partir de un conjunto de datos en componentes, cada uno de los cuales representa la cantidad de la varianza total que se puede atribuir a una fuente específica de variación. Los resultados de esta partición pueden usarse luego para estimar y probar hipótesis sobre las varianzas y medias de la población. En este capítulo centramos nuestra atención en la prueba de hipótesis de medias. Específicamente, analizamos la prueba de diferencias entre medias cuando hay interés en más de dos poblaciones o dos o más variables. Las técnicas analizadas en este capítulo se utilizan ampliamente en las ciencias de la salud.

8.1 INTRODUCTION

8.2 THE COMPLETELY RANDOMIZED DESIGN

8.3 THE RANDOMIZED COMPLETE BLOCK DESIGN

8.4 THE REPEATED MEASURES DESIGN

8.5 THE FACTORIAL EXPERIMENT

8.6 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand how the total variation in a data set can be partitioned into different components.

2. be able to compare the means of more than two samples simultaneously.

3. understand multiple comparison tests and when their use is appropriate.

4. understand commonly used experimental designs.

8.1 INTRODUCTION

In the preceding chapters the basic concepts of statistics have been examined, and they provide a foundation for the present and succeeding chapters.

En los capítulos anteriores se han examinado los conceptos básicos de la estadística, que proporcionan una base para el presente capítulo y los siguientes.

This chapter is concerned with analysis of variance, which may be defined as a technique whereby the total variation present in a set of data is partitioned into two or more components. Associated with each of these components is a specific source of variation, so that in the analysis it is possible to ascertain the magnitude of the contributions of each of these sources to the total variation.

Este capítulo se ocupa del análisis de varianza, que puede definirse como una técnica mediante la cual la variación total presente en un conjunto de datos se divide en dos o más componentes. Asociada a cada uno de estos componentes hay una fuente específica de variación, de modo que en el análisis es posible determinar la magnitud de las contribuciones de cada una de estas fuentes a la variación total.

The development of analysis of variance (ANOVA) is due mainly to the work of R. A. Fisher (1), whose contributions to statistics, spanning the years 1912 to 1962, have had a tremendous influence on modern statistical thought (2, 3).

El desarrollo del análisis de varianza (ANOVA) se debe principalmente al trabajo de R. A. Fisher (1), cuyas contribuciones a la estadística, que abarcan los años 1912 a 1962, han tenido una tremenda influencia en el pensamiento estadístico moderno (2, 3).

Applications

Analysis of variance finds its widest application in the analysis of data derived from experiments. The principles of the design of experiments are well covered in many books, including those by Hinkelmann and Kempthorne (4), Montgomery (5), and Myers and Well (6). We do not study this topic in detail, since to do it justice would require a minimum of an additional chapter. Some of the important concepts in experimental design, however, will become apparent as we discuss analysis of variance.

El análisis de varianza encuentra su aplicación más amplia en el análisis de datos derivados de experimentos. Los principios del diseño de experimentos están bien tratados en muchos libros, incluidos los de Hinkelmann y Kempthorne (4), Montgomery (5) y Myers y Well (6). No estudiamos este tema en detalle, ya que para hacerle justicia requeriría como mínimo un capítulo adicional. Sin embargo, algunos de los conceptos importantes en el diseño experimental se harán evidentes cuando analicemos el análisis de la varianza.

Analysis of variance is used for two different purposes: (1) to estimate and test

hypotheses about population variances, and (2) to estimate and test hypotheses about population means. We are concerned here with the latter use. However, as we will see, our

conclusions regarding the means will depend on the magnitudes of the observed variances.

El análisis de varianza se utiliza para dos propósitos diferentes: (1) estimar y probar hipótesis sobre las varianzas poblacionales, y (2) estimar y probar hipótesis sobre las medias poblacionales. Nos ocupamos aquí del último uso. Sin embargo, como veremos, nuestras conclusiones sobre las medias dependerán de las magnitudes de las varianzas observadas.

As we shall see, the concepts and techniques that we cover under the heading of analysis of variance are extensions of the concepts and techniques covered in Chapter 7. In Chapter 7 we learned to test the null hypothesis that two means are equal. In this chapter we learn to test the null hypothesis that three or more means are equal. Whereas, for example, what we learned in Chapter 7 enables us to determine if we can conclude that two treatments differ in effectiveness, what we learn in this chapter enables us to determine if we can conclude that three or more treatments differ in effectiveness. The following example illustrates some basic ideas involved in the application of analysis of variance. These will be extended and elaborated on later in this chapter.

Como veremos, los conceptos y técnicas que cubrimos bajo el título de análisis de varianza son extensiones de los conceptos y técnicas tratados en el Capítulo 7. En el Capítulo 7 aprendimos a probar la hipótesis nula de que dos medias son iguales. En este capítulo aprenderemos a probar la hipótesis nula de que tres o más medias son iguales. Mientras que, por ejemplo, lo que aprendimos en el capítulo 7 nos permite determinar si podemos concluir que dos tratamientos difieren en efectividad, lo que aprendemos en este capítulo nos permite determinar si podemos concluir que tres o más tratamientos difieren en efectividad. El siguiente ejemplo ilustra algunas ideas básicas involucradas en la aplicación del análisis de varianza. Estos se ampliarán y desarrollarán más adelante en este capítulo.

EXAMPLE 8.1.1

Suppose we wish to know if three drugs differ in their effectiveness in lowering serum cholesterol in human subjects. Some subjects receive drug A, some drug B, and some drug C. After a specified period of time measurements are taken to determine the extent to which serum cholesterol was reduced in each subject. We find that the amount by which serum cholesterol was lowered is not the same in all subjects. In other words, there is variability among the measurements. Why, we ask ourselves, are the measurements not all the same?

Supongamos que deseamos saber si tres fármacos difieren en su eficacia para reducir el colesterol sérico en seres humanos. Algunos sujetos reciben el fármaco A, algunos fármacos B y algunos fármacos C. Después de un período de tiempo específico, se toman medidas para determinar en qué medida se redujo el colesterol sérico en cada sujeto. Encontramos que la cantidad en la que se redujo el colesterol sérico no es la misma en todos los sujetos. En otras palabras, existe variabilidad entre las mediciones. ¿Por qué, nos preguntamos, las medidas no son todas iguales?

Presumably, one reason they are not the same is that the subjects received different drugs. We now look at the measurements of those subjects who received drug A. We find that the amount by which serum cholesterol was lowered is not the same among these subjects. We find this to be the case when we look at the measurements for subjects who received drug B and those subjects who received drug C. We see that there is variability among the measurements within the treatment groups. Why, we ask ourselves again, are these measurements not the same? Among the reasons that come to mind are differences in the genetic makeup of the subjects and differences in their diets.

Presumiblemente, una de las razones por las que no son iguales es que los sujetos recibieron medicamentos diferentes. Ahora observamos las mediciones de aquellos sujetos que recibieron el fármaco A. Encontramos que la cantidad en la que se redujo el colesterol sérico no es la misma entre estos sujetos. Encontramos que este es el caso cuando observamos las mediciones de los sujetos que recibieron el fármaco B y los sujetos que recibieron el fármaco C. Vemos que hay variabilidad entre las mediciones dentro de los grupos de tratamiento. ¿Por qué, nos volvemos a preguntar, estas medidas no son iguales? Entre las razones que me vienen a la mente se encuentran las diferencias en la composición genética de los sujetos y las diferencias en sus dietas.

Through an analysis of the variability that we have observed, we will be able to reach a conclusion regarding the equality of the effectiveness of the three drugs. To do this we employ the techniques and concepts of analysis of variance.

Mediante un análisis de la variabilidad que hemos observado podremos llegar a una conclusión sobre la igualdad de eficacia de los tres fármacos. Para ello empleamos las técnicas y conceptos de análisis de varianza.

Variables

In our example we allude to three kinds of variables. We find these variables to be present in all situations in which the use of analysis of variance is appropriate. First we have the treatment variable, which in our example was “drug.” We had three “values” of this variable, drug A, drug B, and drug C. The second kind of variable we refer to is the response variable. In the example it is change in serum cholesterol. The response variable is the variable that we expect to exhibit different values when different “values” of the treatment variable are employed. Finally, we have the other variables that we mention— genetic composition and diet. These are called extraneous variables. These variables may have an effect on the response variable, but they are not the focus of our attention in the experiment. The treatment variable is the variable of primary concern, and the question to be answered is: Do the different “values” of the treatment variable result in differences, on the average, in the response variable?

En nuestro ejemplo aludimos a tres tipos de variables. Encontramos que estas variables están presentes en todas las situaciones en las que el uso del análisis de varianza es apropiado. Primero tenemos la variable de tratamiento, que en nuestro ejemplo era "fármaco". Teníamos tres “valores” de esta variable, fármaco A, fármaco B y fármaco C. El segundo tipo de variable al que nos referimos es la variable de respuesta. En el ejemplo se trata de un cambio en el colesterol sérico. La variable de respuesta es la variable que esperamos que exhiba diferentes valores cuando se emplean diferentes “valores” de la variable de tratamiento. Finalmente, tenemos las otras variables que mencionamos: composición genética y dieta. Éstas se llaman variables extrañas. Estas variables pueden tener un efecto sobre la variable respuesta, pero no son el foco de nuestra atención en el experimento. La variable de tratamiento es la variable de principal preocupación y la pregunta que hay que responder es: ¿Los diferentes “valores” de la variable de tratamiento dan como resultado diferencias, en promedio, en la variable de respuesta?

Assumptions

Underlying the valid use of analysis of variance as a tool of statistical inference are a set of fundamental assumptions. Although an experimenter must not expect to find all the assumptions met to perfection, it is important that the user of analysis of variance techniques be aware of the underlying assumptions and be able to recognize when they are substantially unsatisfied. Because experiments in which all the assumptions are perfectly met are rare, analysis of variance results should be considered as approximate rather than exact. These assumptions are pointed out at appropriate points in the following sections.

Detrás del uso válido del análisis de varianza como herramienta de inferencia estadística hay un conjunto de supuestos fundamentales. Aunque un experimentador no debe esperar encontrar que todos los supuestos se cumplan a la perfección, es importante que el usuario de técnicas de análisis de varianza sea consciente de los supuestos subyacentes y sea capaz de reconocer cuando están sustancialmente insatisfechos. Debido a que son raros los experimentos en los que se cumplen perfectamente todos los supuestos, los resultados del análisis de varianza deben considerarse aproximados más que exactos. Estos supuestos se señalan en los puntos apropiados de las siguientes secciones.

We discuss analysis of variance as it is used to analyze the results of two different experimental designs, the completely randomized and the randomized complete block designs. In addition to these, the concept of a factorial experiment is given through its use in a completely randomized design. These do not exhaust the possibilities. A discussion of additional designs may be found in the references (4–6).

Se analiza el análisis de varianza tal como se utiliza para analizar los resultados de dos diseños experimentales diferentes, el diseño completamente al azar y el diseño de bloques completos al azar. Además de estos, el concepto de experimento factorial se da a través de su uso en un diseño completamente al azar. Estos no agotan las posibilidades. Se puede encontrar una discusión sobre diseños adicionales en las referencias (4–6).

The ANOVA Procedure

In our presentation of the analysis of variance for the different designs, we follow the ten-step procedure presented in Chapter 7. The following is a restatement of the steps of the procedure, including some new concepts necessary for its adaptation to analysis of variance.

En nuestra presentación del análisis de varianza para los diferentes diseños, seguimos el procedimiento de diez pasos presentado en el Capítulo 7. A continuación se reformulan los pasos del procedimiento, incluidos algunos conceptos nuevos necesarios para su adaptación al análisis de varianza.

1. Description of data. In addition to describing the data in the usual way, we display the sample data in tabular form.

2. Assumptions. Along with the assumptions underlying the analysis, we present the model for each design we discuss. The model consists of a symbolic representation of a typical value from the data being analyzed.

Junto con los supuestos subyacentes al análisis, presentamos el modelo para cada diseño que analizamos. El modelo consta de una representación simbólica de un valor típico de los datos que se analizan.

3. Hypotheses.

4. Test statistic.

5. Distribution of test statistic.

6. Decision rule.

7. Calculation of test statistic. The results of the arithmetic calculations will be summarized in a table called the analysis of variance (ANOVA) table. The entries in the table make it easy to evaluate the results of the analysis.

8. Statistical decision.

9. Conclusion.

10. Determination of p value.

We discuss these steps in greater detail in Section 8.2.

The Use of Computers

The calculations required by analysis of variance are

lengthier and more complicated than those we have encountered in preceding chapters.

For this reason the computer assumes an important role in analysis of variance. All the

exercises appearing in this chapter are suitable for computer analysis and may be used

with the statistical packages mentioned in Chapter 1. The output of the statistical packages

may vary slightly from that presented in this chapter, but this should pose no major

problem to those who use a computer to analyze the data of the exercises. The basic

concepts of analysis of variance that we present here should provide the necessary background

for understanding the description of the programs and their output in any of the

statistical packages.

8.2 THE COMPLETELY RANDOMIZED DESIGN

We saw in Chapter 7 how it is possible to test the null hypothesis of no differencebetween two population means. It is not unusual for the investigator to be interested in testing the null hypothesis of no difference among several population means. The student first encountering this problem might be inclined to suggest that all possible pairs of sample means be tested separately by means of the Student t test. Suppose there are five populations involved. The number of possible pairs of sample means is 5C2 = 10.

Vimos en el capítulo 7 cómo es posible probar la hipótesis nula de que no hay diferencias entre dos medias poblacionales. No es inusual que el investigador esté interesado en probar la hipótesis nula de que no hay diferencias entre varias medias poblacionales. El estudiante que se encuentre por primera vez con este problema podría inclinarse a sugerir que todos los pares posibles de medias muestrales se prueben por separado mediante la prueba t de Student. Supongamos que hay cinco poblaciones involucradas. El número de posibles pares de medias muestrales es 5C2 = 10.

As the amount of work involved in carrying out this many t tests is substantial, it would be worthwhile if a more efficient alternative for analysis were available. A more important consequence of performing all possible t tests, however, is that it is very likely to lead to a false conclusion.

Como la cantidad de trabajo involucrada en la realización de tantas pruebas t es sustancial, valdría la pena si estuviera disponible una alternativa más eficiente para el análisis. Sin embargo, una consecuencia más importante de realizar todas las pruebas t posibles es que es muy probable que conduzca a una conclusión falsa.

Suppose we draw five samples from populations having equal means. As we have seen, there would be 10 tests if we were to do each of the possible tests separately. If we select a significance level of for each test, the probability of failing to reject a hypothesis of no difference in each case would be .95. By the multiplication rule of probability, if the tests were independent of one another, the probability of failing to reject a hypothesis of no difference in all 10 cases would be (.95)^10 = .5987. The probability of rejecting at least one hypothesis of no difference, then, would be 1-.5987 = .4013.

Supongamos que extraemos cinco muestras de poblaciones que tienen medias iguales. Como hemos visto, serían 10 pruebas si hiciéramos cada una de las posibles pruebas por separado. Si seleccionamos un nivel de significancia de para cada prueba, la probabilidad de no rechazar una hipótesis de no diferencia en cada caso sería .95. Según la regla de probabilidad de la multiplicación, si las pruebas fueran independientes entre sí, la probabilidad de no rechazar una hipótesis de no diferencia en los 10 casos sería (0,95)^10 = 0,5987. La probabilidad de rechazar al menos una hipótesis de no diferencia, entonces, sería 1-0,5987 = 0,4013.

Since we know that the null hypothesis is true in every case in this illustrative example, rejecting the null hypothesis constitutes the committing of a type I error. In the long run, then, in testing all possible pairs of means from five samples, we would commit a type I error 40 percent of the time. The problem becomes even more complicated in practice, since three or more t tests based on the same data would not be independent of one another. It becomes clear, then, that some other method for testing for a significant difference among several means is needed. Analysis of variance provides such a method.

Como sabemos que la hipótesis nula es verdadera en todos los casos de este ejemplo ilustrativo, rechazar la hipótesis nula constituye cometer un error de tipo I. Entonces, a largo plazo, al probar todos los pares posibles de medias de cinco muestras, cometeríamos un error de tipo I el 40 por ciento de las veces. El problema se vuelve aún más complicado en la práctica, ya que tres o más pruebas t basadas en los mismos datos no serían independientes entre sí. Resulta claro, entonces, que se necesita algún otro método para probar una diferencia significativa entre varias medias. El análisis de varianza proporciona dicho método.

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.

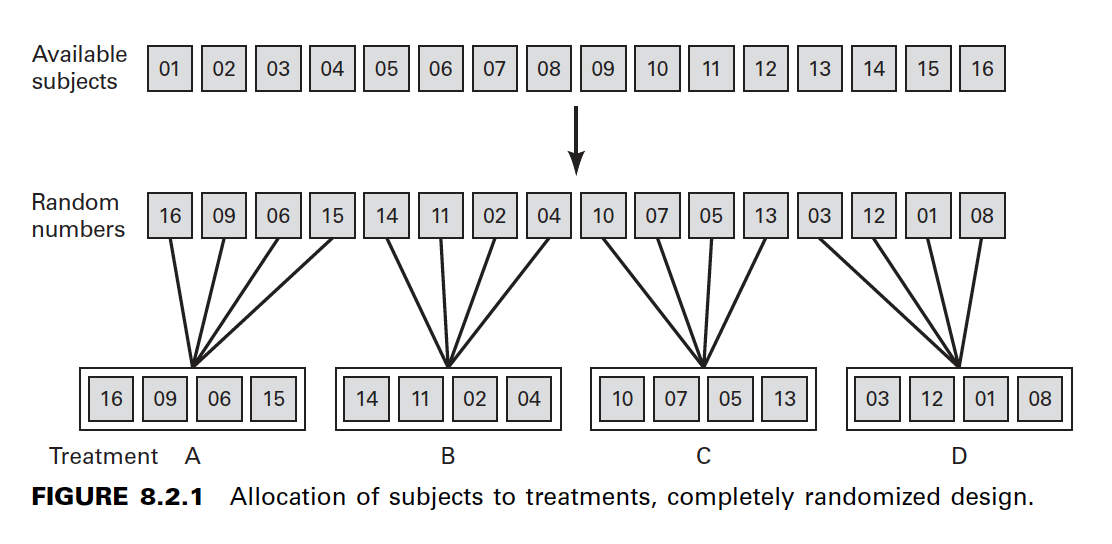
El tipo más simple de análisis de varianza es el conocido como análisis de varianza unidireccional, en el que solo se investiga una fuente de variación o factor. Es una extensión a tres o más muestras del procedimiento de prueba t (que se analiza en el Capítulo 7) para usar con dos muestras independientes.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

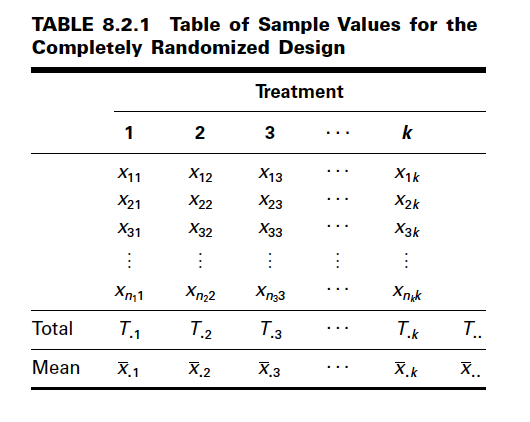
In a typical situation we want to use one-way analysis of variance to test the null hypothesis that three or more treatments are equally effective. The necessary experiment is designed in such a way that the treatments of interest are assigned completely at random to the subjects or objects on which 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.

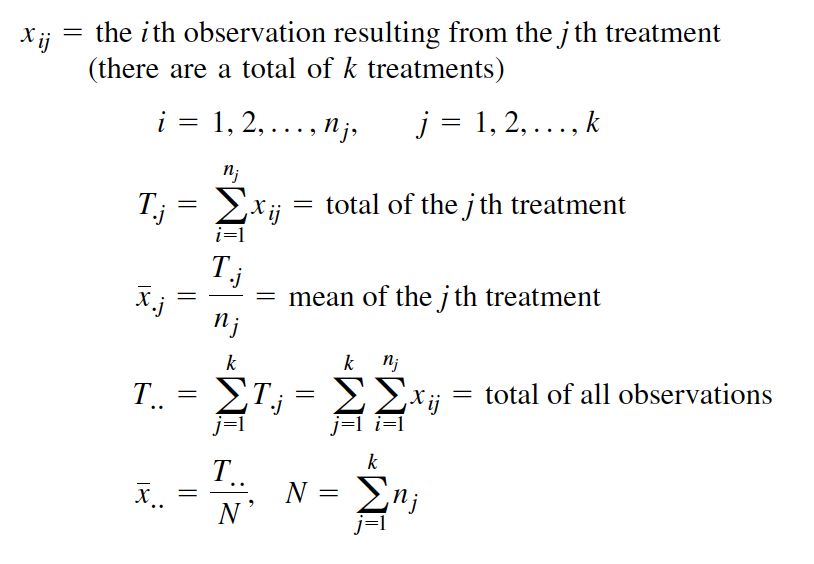




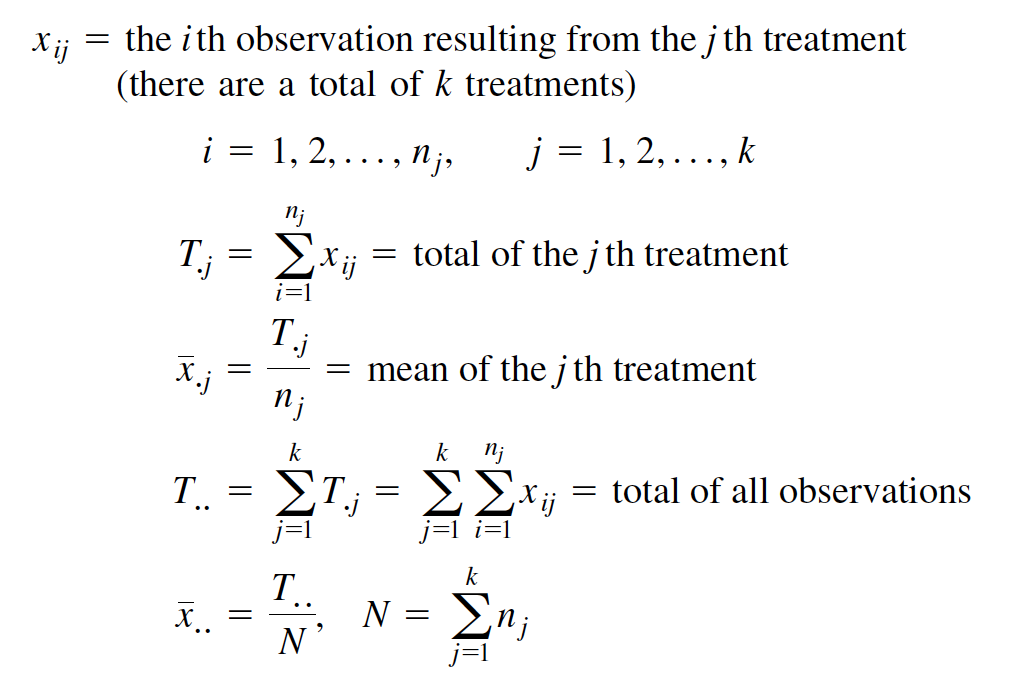
**Hypothesis Testing Steps**

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.

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:



(there are a total of k treatments)



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

experiment described here.

**The Model**

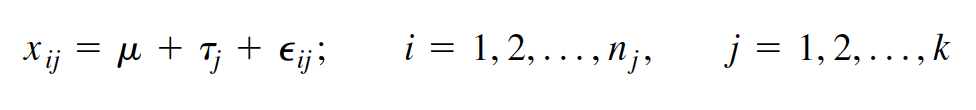
As already noted, a model is a symbolic representation of a typical value of a data set. To write down the model for the completely randomized experimental design, let us begin by identifying a typical value from the set of data represented by the sample displayed in Table 8.2.1. We use the symbol xij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

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 xij 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:

(8.2.1)



The terms in this model are defined as follows:

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

2. **taoj** represents the difference between the mean of the j th population and the grand

mean and is called the treatment effect.

3. **eij** 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**

The assumptions for the fixed-effects model are as follows:

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

respective populations.

(b) Each of the populations from which the samples come is normally distributed with

mean and variance

(c) Each of the populations has the same variance. That is,

the common variance.

(d) The are unknown constants and since the sum of all deviations of the

from their mean, is zero.

(e) The have a mean of 0, since the mean of is

(f) The have a variance equal to the variance of the since the and differ

only by a constant; that is, the error variance is equal to the common variance

specified in assumption c.

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

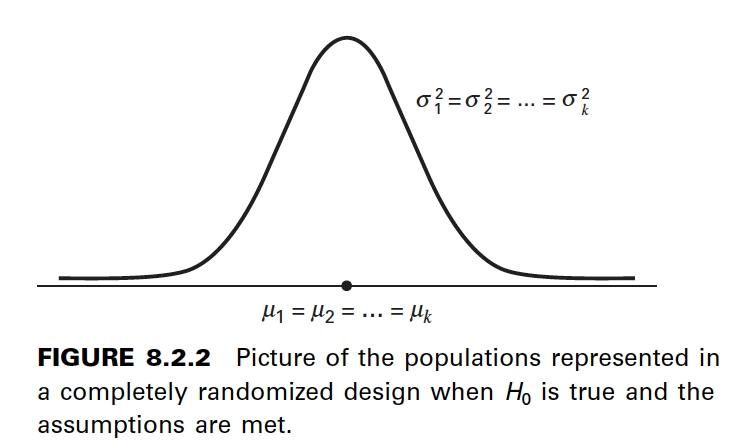
3. **Hypotheses.** We test the null hypothesis that all population or treatment means are

equal against the alternative that the members of at least one pair are not equal.

We may state the hypotheses formally as follows:

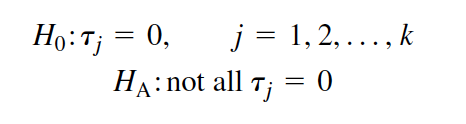
HA:not all mj are equal

H0 :m1 = m2 = . . . = mk



If the population means are equal, each treatment effect is equal to zero, so that, alternatively,

the hypotheses may be stated as



If Ho 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 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 Ho 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 es cierto, 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.

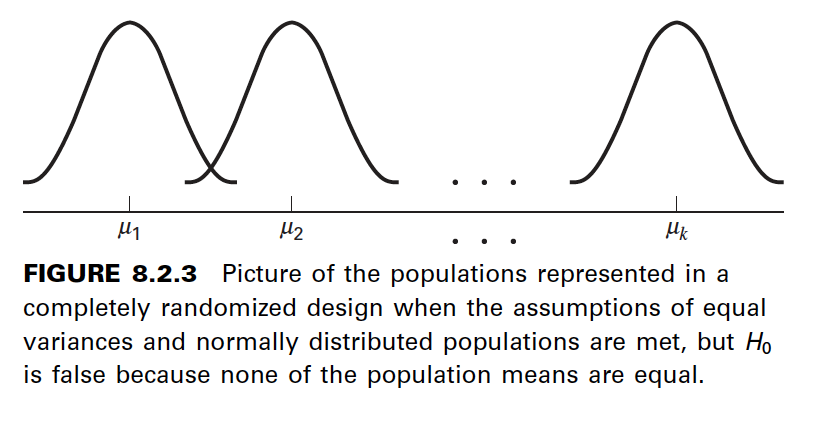
When Ho is false it may be false because one of the population means is different from the others, which are all equal. Or, perhaps, all the population means are different. These are only two of the possibilities when Ho is false. There are many other possible combinations of equal and unequal means.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

Cuando Ho es falso, puede ser falso 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 Ho 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 Ho 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 is true and the assumptions are met.

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 level.

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.

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.

Those who use a computer for calculations may wish to skip the following discussion

of the computations involved in obtaining the test statistic.

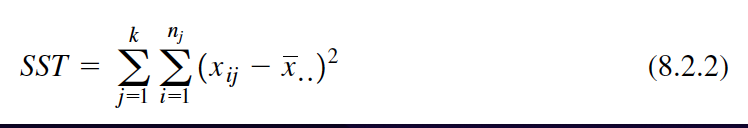
**The Total Sum of Squares**

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



(8.2.2)

where tells us to sum the squared deviations for each treatment group, and

tells us to add the k group totals obtained by applying 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.

**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

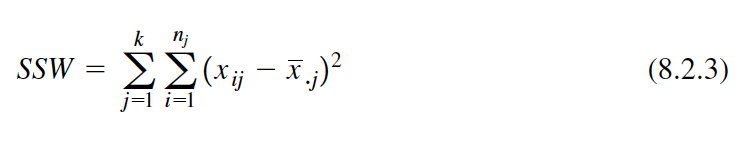
squared deviations of the individual observations from their mean. When these calculations

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:



SSW = a (8.2.3)

**The Among Groups Sum of Squares**

To obtain the second component

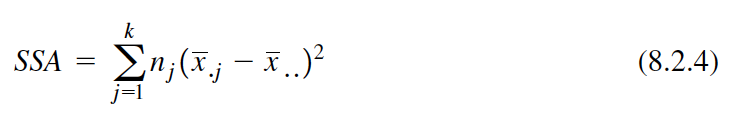
of the total sum of squares, we compute for each group the squared deviation of the

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:



(8.2.4)

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:

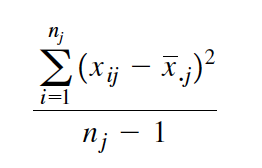
SST = SSA + SSW

From the sums of squares that we have now learned to compute, it is possible to obtain two estimates of the common population variance, sigma^2. 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 sigma^2.

A partir de las sumas de cuadrados que hemos aprendido a calcular, es posible obtener dos estimaciones de la varianza común de la población, sigma^2. 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 sigma^2.

**The First Estimate of sigma^2**

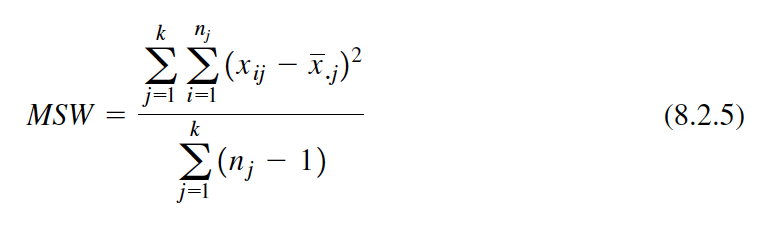
Within any sample,



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



(8.2.5)

This is our first estimate of sigma^2 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.

The within groups mean square is a valid estimate of sigma^2 only if the population

variances are equal. It is not necessary, however, for Ho to be true in order for the within

groups mean square to be a valid estimate of that is, the within groups mean square

estimates regardless of whether Ho is true or false, as long as the population variances

are equal.

**The Second Estimate of sigma^2**

The second estimate of may be obtained

from the familiar formula for the variance of sample means, If we solve this

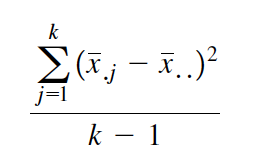
equation for the variance of the population from which the samples were drawn, we

have

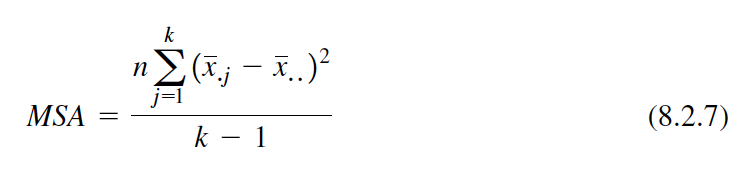


(8.2.6)

An unbiased estimate of computed from sample data is provided by



If we substitute this quantity into Equation 8.2.6, we obtain the desired estimate of sigma^2 ,



(8.2.7)

The reader will recognize the numerator of Equation 8.2.7 as the among groups

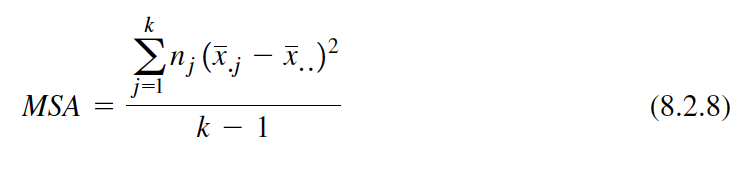
sum of squares for the special case when all sample sizes are equal. This sum of squares

when divided by the associated degrees of freedom is referred to as the among

groups mean square.

When the sample sizes are not all equal, an estimate of based on the variability

among sample means is provided by



(8.2.8)

If, indeed, the null hypothesis is true we would expect these two estimates of sigma^2

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.

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 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

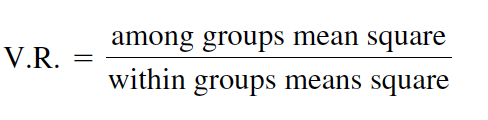
valid estimate of sigma^2.

**The Variance Ratio**

What we need to do now is to compare these two estimates

of and we do this by computing the following variance ratio, which is the desired

test statistic:



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.

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 is not the result of chance alone?

**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

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.

In Chapter 7 we learned that when the population variances are the same, they cancel

in the expression , leaving , 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

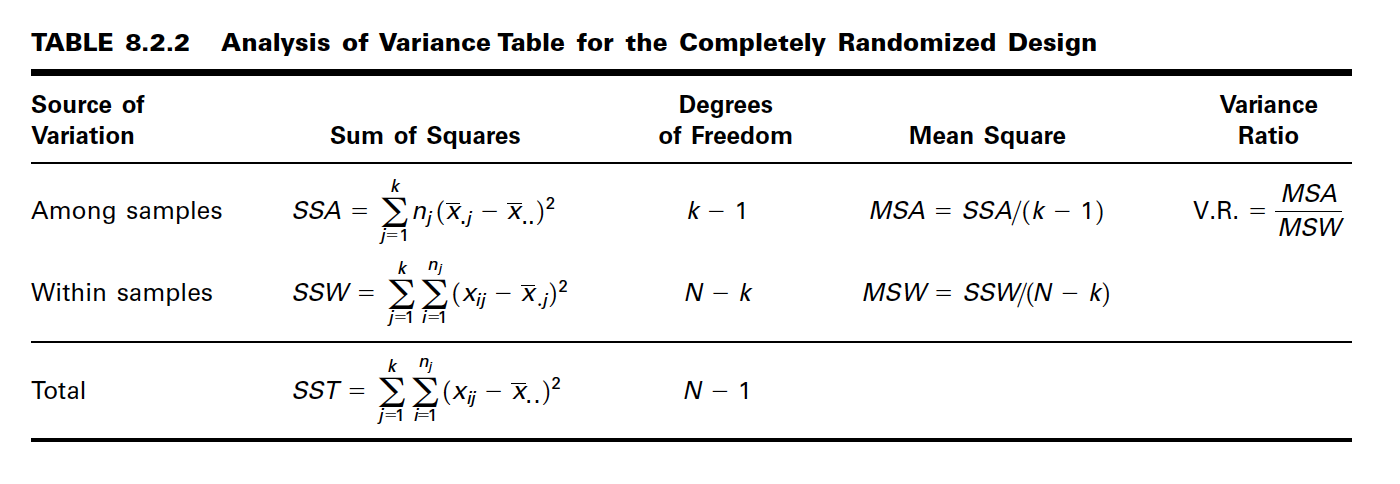
degrees of freedom).

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.

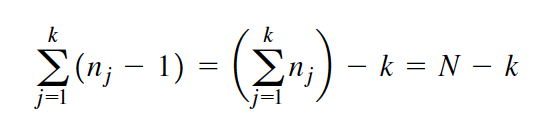


As we have seen, we compute V.R. in situations of this type by placing the among

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 the number of

groups minus 1, and the denominator degrees of freedom value is equal to



**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.

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

numerator degrees of freedom and denominator degrees of freedom.

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.

**Explaining a Rejected Null Hypothesis**

There are two possible explanations

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 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

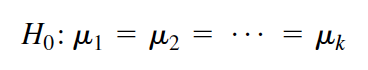
conclude to be a false null hypothesis.

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.

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



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.

9. Conclusion. When we reject we conclude that not all population means are

equal. When we fail to reject we conclude that the population means may all

be equal.

10. Determination of p value.

**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.

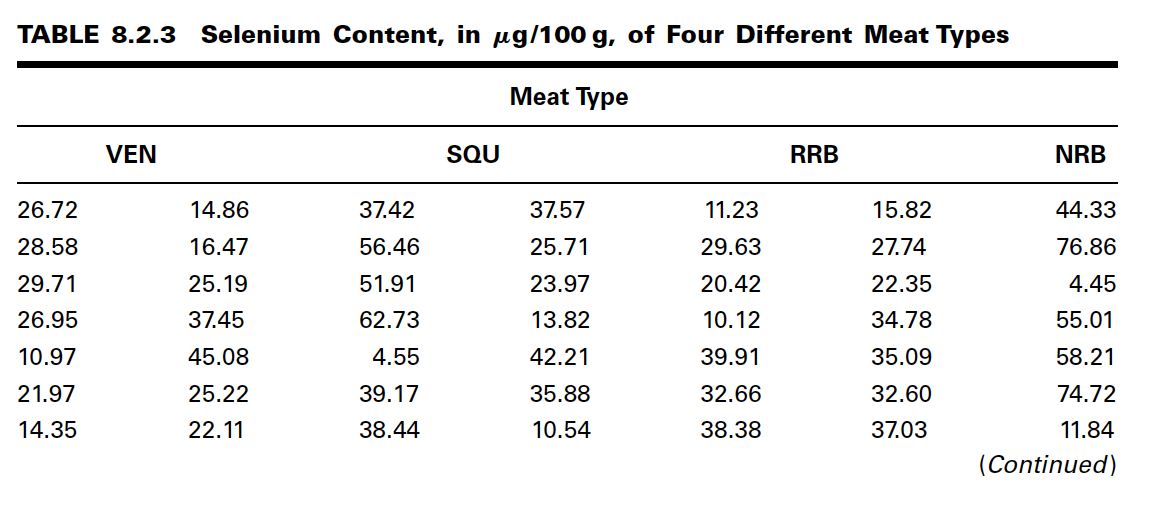
Solution:

1. Description of data. Selenium content of raw venison (VEN), squirrel

meat (SQU), region-raised beef (RRB), and nonregion-raised beef (NRB),

in 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

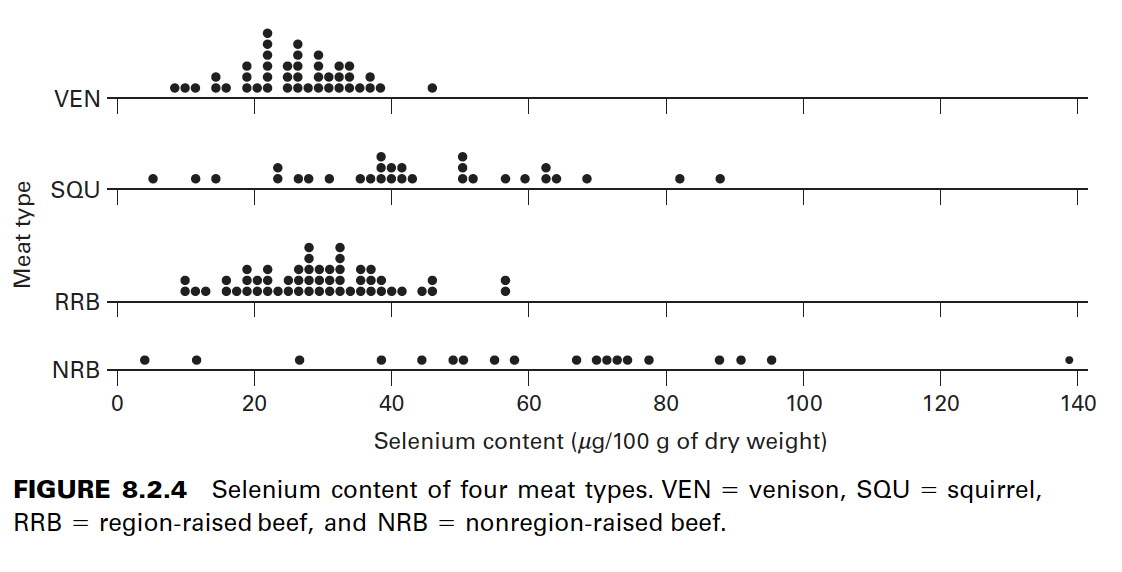
levels among the different meats.

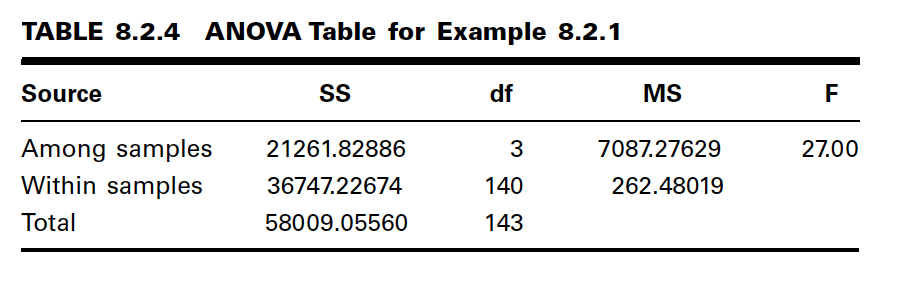
2. Assumptions. We assume that the four sets of data constitute independent

simple random samples from the four indicated populations. We

assume that the four populations of measurements are normally distributed

with equal variances.





3. Hypotheses.

(On average the four meats have the same

selenium content.)

Not all are equal (At least one meat yields an average selenium

content different from the average selenium content of at least one other

meat.)

4. Test statistic. The test statistic is

5. Distribution of test statistic. If is true and the assumptions are met,

the V.R. follows the F distribution with numerator degrees of

freedom and denominator degrees of freedom.

6. Decision rule. Suppose we let The critical value of F from

Appendix Table G is The decision rule, then, is reject if the

computed V.R. statistic is equal to or greater than 3.95.

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

By Equation 8.2.4 we compute

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

9. Conclusion. Since we reject 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.

10. p value. Since 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

the pooled standard deviation. This last quantity is equal to the square root of the error

mean square shown in the ANOVA table. Finally, the computer output gives graphic representations

of the 95 percent confidence intervals for the mean of each of the four populations

represented by the sample data.

Figure 8.2.6 contains a partial SAS® printout resulting from analysis of the data

of Example 8.2.1 through use of the SAS® statement PROC ANOVA. SAS® computes

some additional quantities as shown in the output. This quantity

tells us what proportion of the total variability present in the observations is

accounted for by differences in response to the treatments. (root MSE/selen

mean). Root MSE is the square root of MSW, and selen mean is the mean of the 18

observations.

Note that the test statistic V.R. is labeled differently by different statistical software

programs. MINITAB, for example, uses F rather than V.R. SAS® uses the label

F Value.

A useful device for displaying important characteristics of a set of data analyzed

by one-way analysis of variance is a graph consisting of side-by-side boxplots. For each

sample a boxplot is constructed using the method described in Chapter 2. Figure 8.2.7

shows the side-by-side boxplots for Example 8.2.1. Note that in Figure 8.2.7 the variable

of interest is represented by the vertical axis rather than the horizontal axis.

Alternatives If the data available for analysis do not meet the assumptions for oneway

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.

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

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,

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 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.

Tukey’s HSD Test Over the years several procedures for making multiple comparisons

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 The probability is then, that one or more of the null

hypotheses is false.

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

(8.2.9)

where 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.

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).

All possible differences between pairs of means are computed, and any difference

that yields an absolute value that exceeds HSD is declared significant.

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 where and 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

(8.2.10)

Any absolute value of the difference between two sample means that exceeds

HSD\* is declared significant.

Bonferroni’s Method Another very commonly used multiple comparison

test is based on a method developed by C. E. Bonferroni. As with Tukey’s method,

we desire to maintain an overall significance level of for the total of all pair-wise

tests. In the Bonferroni method, we simply divide the desired significance level by

the number of individual pairs that we are testing. That is, instead of testing at a significance

level of , we test at a significance level of where k is the number of

paired comparisons. The sum of all terms cannot, then, possibly exceed our stated

level of . For example, if one has three samples, A, B, and C, then there are

pair-wise comparisons. These are and If we choose a

significance level of then we would proceed with the comparisons and use

a Bonferroni-corrected significance level of Therefore, our p value must

be no greater then .017 in order to reject the null hypothesis and conclude that two

means differ.

Most computer packages compute values using the Bonferroni method and produce

an output similar to the Tukey’s HSD or other multiple comparison procedures. In

general, these outputs report the actual corrected p value using the Bonferroni method.

Given the basic relationship that then algebraically we can multiply both sides

of the equation by k to obtain In other words, the total is simply the sum of

all of the pk values, and the actual corrected p value is simply the calculated p value

multiplied by the number of tests that were performed.

EXAMPLE 8.2.2

Let us illustrate the use of the HSD test with the data from Example 8.2.1.

Solution: The first step is to prepare a table of all possible (ordered) differences

between means. The results of this step for the present example are displayed

in Table 8.2.5.

Suppose we let Entering Table H with and we

find that The actual value is which can be obtained from SAS®.

In Table 8.2.4 we have

The hypotheses that can be tested, the value of HSD\*, and the statistical decision

for each test are shown in Table 8.2.6.

SAS® uses Tukey’s procedure to test the hypothesis of no difference between

population means for all possible pairs of sample means. The output also contains

.------------------------------------------------

8.3 THE RANDOMIZED COMPLETE

BLOCK DESIGN

The randomized complete block design was developed about 1925 by R. A. Fisher, who was seeking methods of improving agricultural field experiments. The randomized complete block design is a design in which the units (called experimental units) to which the treatments are applied are subdivided into homogeneous groups called blocks, so that the number of experimental units in a block is equal to the number (or some multiple of the number) of treatments being studied. The treatments are then assigned at random to the experimental units within each block. It should be emphasized that each treatment appears in every block, and each block receives every treatment.

El diseño de bloques completos al azar fue desarrollado alrededor de 1925 por R. A. Fisher, quien buscaba métodos para mejorar los experimentos agrícolas de campo. El diseño de bloques completos al azar es un diseño en el que las unidades (llamadas unidades experimentales) a las que se aplican los tratamientos se subdividen en grupos homogéneos llamados bloques, de modo que el número de unidades experimentales en un bloque es igual al número (o algún múltiplo) del número) de tratamientos en estudio. Luego los tratamientos se asignan aleatoriamente a las unidades experimentales dentro de cada bloque. Cabe destacar que cada tratamiento aparece en cada bloque y cada bloque recibe todos los tratamientos.

**Objective**

The objective in using the randomized complete block design is to isolate and remove from the error term the variation attributable to the blocks, while assuring that treatment means will be free of block effects. The effectiveness of the design depends on the ability to achieve homogeneous blocks of experimental units. The ability to form homogeneous blocks depends on the researcher’s knowledge of the experimental material. When blocking is used effectively, the error mean square in the ANOVA table will be reduced, the V.R. will be increased, and the chance of rejecting the null hypothesis will be improved.

El objetivo al utilizar el diseño de bloques completos al azar es aislar y eliminar del término de error la variación atribuible a los bloques, asegurando al mismo tiempo que los medios de tratamiento estarán libres de efectos de bloque. La efectividad del diseño depende de la capacidad de lograr bloques homogéneos de unidades experimentales. La capacidad de formar bloques homogéneos depende del conocimiento que tenga el investigador del material experimental. Cuando el bloqueo se utiliza eficazmente, el error cuadrático medio en la tabla ANOVA se reducirá, el V.R. aumentará y mejorará la posibilidad de rechazar la hipótesis nula.

In animal experiments, the breed of animal may be used as a blocking factor. Litters may also be used as blocks, in which case an animal from each litter receives a treatment. In experiments involving human beings, if it is desired that differences resulting from age be eliminated, then subjects may be grouped according to age so that one person of each age receives each treatment. The randomized complete block design also may be employed effectively when an experiment must be carried out in more than one laboratory (block) or when several days (blocks) are required for completion.

En experimentos con animales, la raza del animal puede utilizarse como factor de bloqueo. Las camadas también se pueden utilizar como bloques, en cuyo caso un animal de cada camada recibe un tratamiento. En experimentos con seres humanos, si se desea eliminar las diferencias resultantes de la edad, entonces los sujetos pueden agruparse según la edad de modo que una persona de cada edad reciba cada tratamiento. El diseño de bloques completos al azar también puede emplearse eficazmente cuando un experimento debe realizarse en más de un laboratorio (bloques) o cuando se requieren varios días (bloques) para completarlo.

The random allocation of treatments to subjects is restricted in the randomized complete block design. That is, each treatment must be represented an equal number of times (one or more times) within each blocking unit. In practice this is generally accomplished by assigning a random permutation of the order of treatments to subjects within each block. For example, if there are four treatments representing three drugs and a placebo (drug A, drug B, drug C, and placebo [p]), then there are 4! " 24 possible permutations of the four treatments: (A, B, C, P) or (A, C, B, P) or (C, A, P, B), and so on. One permutation is then randomly assigned to each block.

La asignación aleatoria de tratamientos a los sujetos está restringida en el diseño de bloques completos al azar. Es decir, cada tratamiento debe representarse un número igual de veces (una o más veces) dentro de cada unidad de bloqueo. En la práctica, esto generalmente se logra asignando una permutación aleatoria del orden de los tratamientos a los sujetos dentro de cada bloque. Por ejemplo, si hay cuatro tratamientos que representan tres fármacos y un placebo (fármaco A, fármaco B, fármaco C y placebo [p]), ¡entonces hay 4! " 24 posibles permutaciones de los cuatro tratamientos: (A, B, C, P) o (A, C, B, P) o (C, A, P, B), y así sucesivamente. Luego se asigna aleatoriamente una permutación a cada bloque.

**Advantages**

One of the advantages of the randomized complete block design is that it is easily understood. Furthermore, certain complications that may arise in the course of an experiment are easily handled when this design is employed.

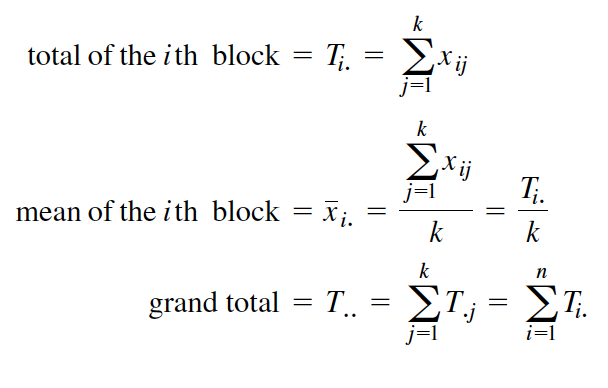
Una de las ventajas del diseño de bloques completos al azar es que es de fácil comprensión. Además, ciertas complicaciones que pueden surgir en el curso de un experimento se manejan fácilmente cuando se emplea este diseño.

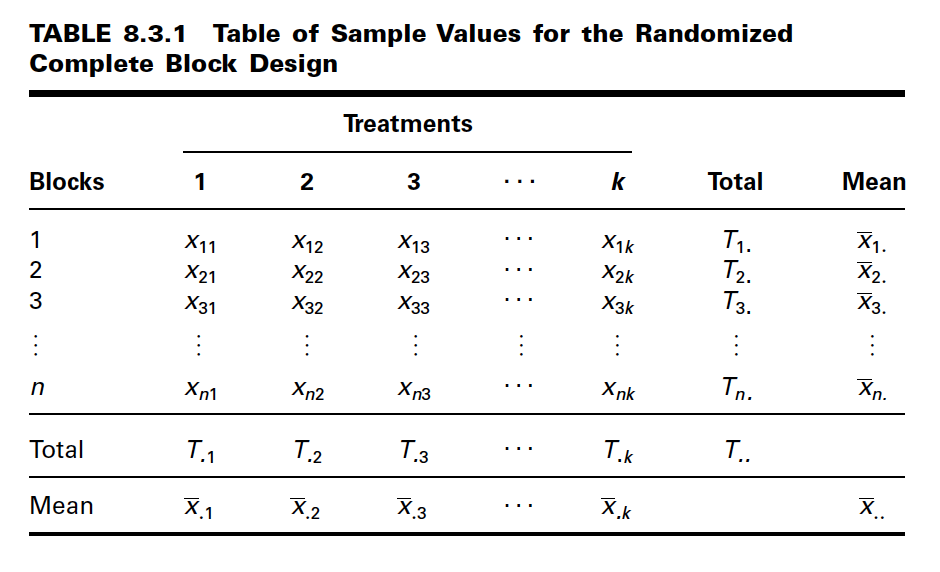
It is instructive here to point out that the paired comparisons analysis presented in Chapter 7 is a special case of the randomized complete block design. Example 7.4.1, for example, may be treated as a randomized complete block design in which the two points in time (Pre-op and Post-op) are the treatments and the individuals on whom the measurements were taken are the blocks.

Es instructivo señalar aquí que el análisis de comparaciones pareadas presentado en el Capítulo 7 es un caso especial del diseño de bloques completos al azar. El ejemplo 7.4.1, por ejemplo, puede tratarse como un diseño de bloques completos aleatorios en el que los dos momentos en el tiempo (preoperatorio y posoperatorio) son los tratamientos y los individuos en quienes se tomaron las mediciones son los bloques.

**Data Display**

In general, the data from an experiment utilizing the randomized complete block design may be displayed in a table such as Table 8.3.1. The following new notation in this table should be observed:





indicating that the grand total may be obtained either by adding row totals or by adding

column totals.

**Two-Way ANOVA**

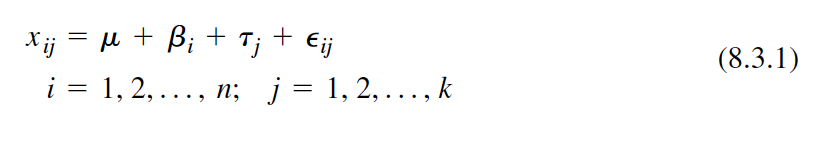
The technique for analyzing the data from a randomized complete block design is called two-way analysis of variance since an observation is categorized on the basis of two criteria—the block to which it belongs as well as the treatment group to which it belongs.

The steps for hypothesis testing when the randomized complete block design is used are as follows:

1. Data. After identifying the treatments, the blocks, and the experimental units, the data, for convenience, may be displayed as in Table 8.3.1.

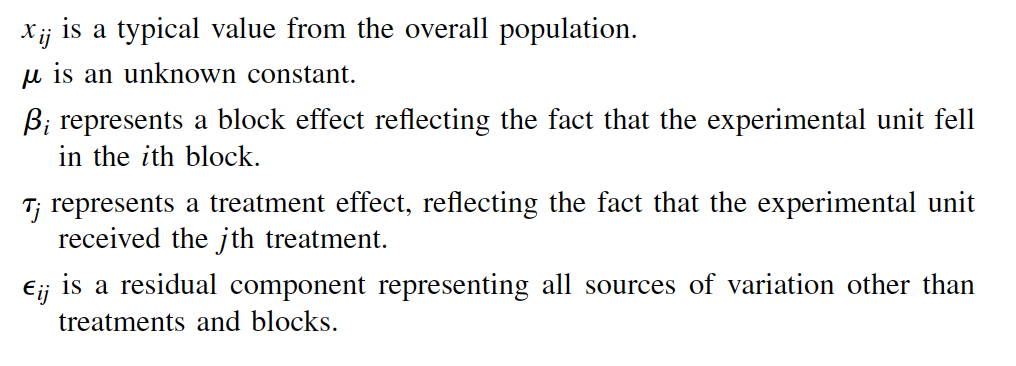
2. Assumptions. The model for the randomized complete block design and its underlying assumptions are as follows:

The Model



(8.3.1)

In this model



is a typical value from the overall population.

is an unknown constant.

represents a block effect reflecting the fact that the experimental unit fell

in the ith block.

represents a treatment effect, reflecting the fact that the experimental unit

received the j th treatment.

is a residual component representing all sources of variation other than

treatments and blocks.

Pij

tj

bi

m

xij

i = 1, 2, Á, n; j = 1, 2, Á, k

xij = m + bi + tj + Pij

336 CHAPTER 8 ANALYSIS OF VARIANCE

TABLE 8.3.1 Table of Sample Values for the Randomized

Complete Block Design

Treatments

Blocks 1 2 3 . . . k Total Mean

123

n

Total

Mean x.k x.. x.1 x.2 x.3 . . .

T.k T.. T.1 T.2 T.3 . . .

xnk Tn. xn. xn1 xn2 xn3 . . .

o o o o o o o o

x3k T3. x3. x31 x32 x33 . . .

x2k T2. x2. x21 x22 x23 . . .

x1k T1. x1. x11 x12 x13 . . .

**Assumptions of the Model**

(a) Each that is observed constitutes a random independent sample of size 1 from

one of the kn populations represented.

(b) Each of these kn populations is normally distributed with mean and the same

variance This implies that the are independently and normally distributed

with mean 0 and variance

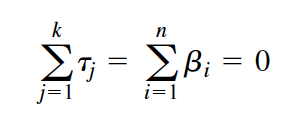
(c) The block and treatment effects are additive. This assumption may be interpreted

to mean that there is no interaction between treatments and blocks. In other words,

a particular block-treatment combination does not produce an effect that is greater

or less than the sum of their individual effects. It can be shown that when this

assumption is met,



The consequences of a violation of this assumption are misleading results. One need

not become concerned with the violation of the additivity assumption unless the

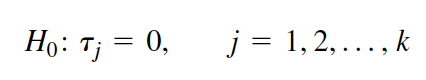
largest mean is more than 50 percent greater than the smallest.

When these assumptions hold true, the and are a set of fixed constants, and we

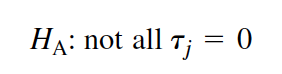
have a situation that fits the fixed-effects model.

3. Hypotheses.

We may test



against the alternative



A hypothesis test regarding block effects is not usually carried out under the

assumptions of the fixed-effects model for two reasons. First, the primary interest is in

treatment effects, the usual purpose of the blocks being to provide a means of eliminating

an extraneous source of variation. Second, although the experimental units are randomly

assigned to the treatments, the blocks are obtained in a nonrandom manner.

4. Test statistic. The test statistic is V.R.

5. Distribution of test statistic. When is true and the assumptions are met, V.R.

follows an F distribution.

6. Decision rule. Reject the null hypothesis if the computed value of the test statistic

V.R. is equal to or greater than the critical value of F.

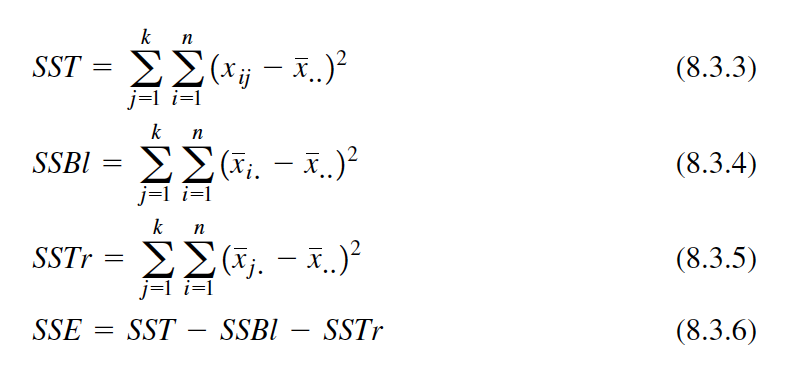
7. Calculation of test statistic. It can be shown that the total sum of squares for the

randomized complete block design can be partitioned into three components, one

each attributable to treatments (SSTr ), blocks (SSBl ), and error (SSE ). That is,

SST = SSBl + SSTr + SSE (8.3.2)

The formulas for the quantities in Equation 8.3.2 are as follows:



(8.3.3)

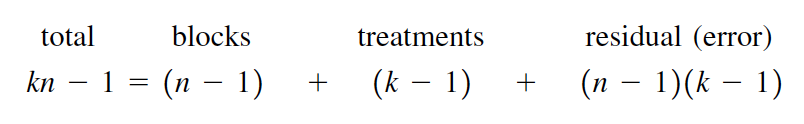
(8.3.4)

(8.3.5)

(8.3.6)

The appropriate degrees of freedom for each component of Equation 8.3.2

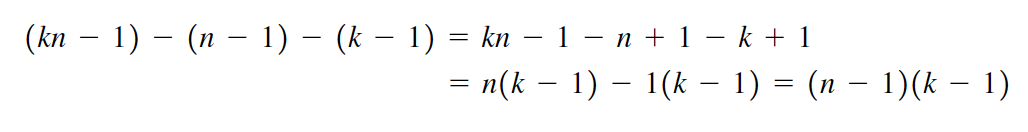
are



total blocks treatments residual (error)

The residual degrees of freedom, like the residual sum of squares, may be

obtained by subtraction as follows:



**The ANOVA Table**

The results of the calculations for the randomized complete

block design may be displayed in an ANOVA table such as Table 8.3.2.

8. Statistical decision. It can be shown that when the fixed-effects model applies and

the null hypothesis of no treatment effects is true, both the error, or

residual, mean square and the treatments mean square are estimates of the common

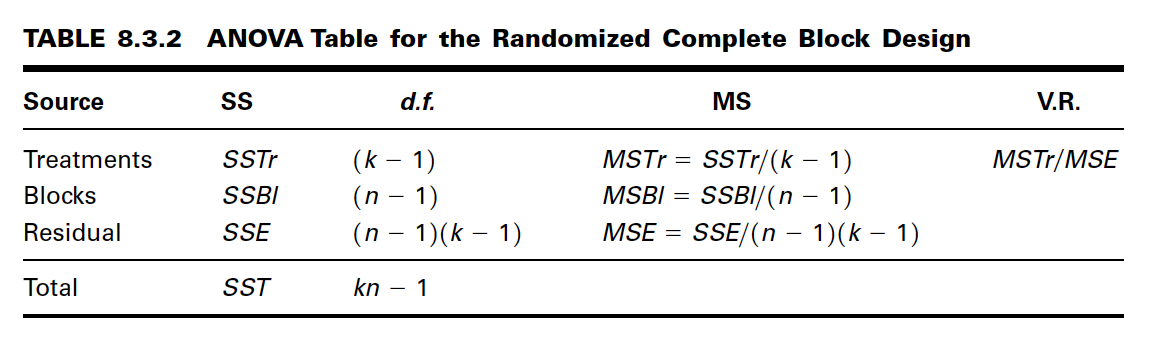
variance When the null hypothesis is true, therefore, the quantity

MSTr/MSE

is distributed as F with numerator degrees of freedom and

denominator degrees of freedom. The computed variance ratio, therefore,

is compared with the critical value of F.



9. Conclusion. If we reject we conclude that the alternative hypothesis is true.

If we fail to reject we conclude that may be true.

10. p value.

The following example illustrates the use of the randomized complete block

design.

**EXAMPLE 8.3.1**

A physical therapist wished to compare three methods for teaching patients to use a certain

prosthetic device. He felt that the rate of learning would be different for patients of

different ages and wished to design an experiment in which the influence of age could

be taken into account.

Solution: The randomized complete block design is the appropriate design for this

physical therapist.

1. Data. Three patients in each of five age groups were selected to participate

in the experiment, and one patient in each age group was randomly

assigned to each of the teaching methods. The methods of instruction constitute

our three treatments, and the five age groups are the blocks. The

data shown in Table 8.3.3 were obtained.

2. Assumptions. We assume that each of the 15 observations constitutes

a simple random of size 1 from one of the 15 populations defined by a

block-treatment combination. For example, we assume that the number

7 in the table constitutes a randomly selected response from a population

of responses that would result if a population of subjects under

the age of 20 received teaching method A. We assume that the responses

in the 15 represented populations are normally distributed with equal

variances.

H0, H0

H0,

8.3 THE RANDOMIZED COMPLETE BLOCK DESIGN 339

TABLE 8.3.3 Time (in Days) Required to Learn the Use

of a Certain Prosthetic Device

Teaching Method

Age Group A B C Total Mean

Under 20 7 9 10 26 8.67

20 to 29 8 9 10 27 9.00

30 to 39 9 9 12 30 10.00

40 to 49 10 9 12 31 10.33

50 and over 11 12 14 37 12.33

Total 45 48 58 151

Mean 9.0 9.6 11.6 10.07

3. Hypotheses.

Let

4. Test statistic. The test statistic is

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

met, V.R. follows an F distribution with 2 and 8 degrees of freedom.

6. Decision rule. Reject the null hypothesis if the computed V.R. is equal

to or greater than the critical F, which we find in Appendix Table G to

be 4.46.

7. Calculation of test statistic. We compute the following sums of squares:

The degrees of freedom are total blocks "

treatments and residual

The results of the calculations may be displayed in an ANOVA table

as in Table 8.3.4.

8. Statistical decision. Since our computed variance ratio, 20.91, is greater

than 4.46, we reject the null hypothesis of no treatment effects on the

assumption that such a large V.R. reflects the fact that the two sample

mean squares are not estimating the same quantity. The only other explanation

for this large V.R. would be that the null hypothesis is really true,

and we have just observed an unusual set of results. We rule out the second

explanation in favor of the first.

8.

5 - 1 = 4, = 3 - 1 = 2, = 15 - 1213 - 12 =

= 132152 - 1 = 14,

SSE = 46.9335 - 24.855 - 18.5335 = 3.545

SSTr = 5C19 - 10.0722 + 19.6 - 10.0722 + 111.6 - 10.0722 D = 18.5335

SSBl = 3C18.67 - 10.0722 + 19.00 - 10.0722 + Á + 112.33 - 10.0722 D = 24.855

SST = 17 - 10.0722 + 18 - 10.0722 + Á + 114 - 10.0722 = 46.9335

H0

V.R. = MSTr/MSE.

a = .05.

HA: not all tj = 0

H0: tj = 0 j = 1, 2, 3

340 CHAPTER 8 ANALYSIS OF VARIANCE

TABLE 8.3.4 ANOVA Table for Example 8.3.1

Source SS d.f. MS V.R.

Treatments 18.5335 2 9.26675 20.91

Blocks 24.855 4 6.21375

Residual 3.545 8 .443125

Total 46.9335 14

9. Conclusion. We conclude that not all treatment effects are equal to zero,

or equivalently, that not all treatment means are equal.

10. p value. For this test !

Computer Analysis Most statistics software packages will analyze data from a

randomized complete block design. We illustrate the input and output for MINITAB. We

use the data from the experiment to set up a MINITAB worksheet consisting of three

columns. Column 1 contains the observations, Column 2 contains numbers that identify

the block to which each observation belongs, and Column 3 contains numbers that identify

the treatment to which each observation belongs. Figure 8.3.1 shows the MINITAB

worksheet for Example 8.3.1. Figure 8.3.2 contains the MINITAB dialog box that initiates

the analysis and the resulting ANOVA table.

The ANOVA table from the SAS® output for the analysis of Example 8.3.1 is

shown in Figure 8.3.3. Note that in this output the model SS is equal to the sum of SSBl

and SSTr.

Alternatives When the data available for analysis do not meet the assumptions

of the randomized complete block design as discussed here, the Friedman procedure

discussed in Chapter 13 may prove to be a suitable nonparametric alternative.

p 6 .005.

8.3 THE RANDOMIZED COMPLETE BLOCK DESIGN 341

ROW C1 C2 C3

1 7 1 1

2 9 1 2

3 10 1 3

4 8 2 1

5 9 2 2

6 10 2 3

7 9 3 1

8 9 3 2

9 12 3 3

10 10 4 1

11 9 4 2

12 12 4 3

13 11 5 1

14 12 5 2

15 14 5 3

FIGURE 8.3.1 MINITAB worksheet for the data in Figure 8.3.2.

342 CHAPTER 8 ANALYSIS OF VARIANCE

Dialog box: Session command:

Stat " ANOVA " Twoway MTB > TWOWAY C1 C2 C3;

SUBC > MEANS C2 C3.

Type C1 in Response. Type C2 in Row factor and

check Display means. Type C3 in Column factor and

check Display means. Click OK.

Output:

Two-Way ANOVA: C1 versus C2, C3

Analysis of Variance for C1

Source DF SS MS F P

C2 4 24.933 6.233 14.38 0.001

C3 2 18.533 9.267 21.38 0.001

Error 8 3.467 0.433

Total 14 46.933

Individual 95% CI

C2 Mean ---+---------+----------+---------+--

1 8.67 (-----\*

-----)

2 9.00 (-----\*

-----)

3 10.00 (-----\*

-----)

4 10.33 (-----\*

-----)

5 12.33 (-----\*

-----)

---+---------+----------+---------+--

9.00 10.50 12.00 13.50

Individual 95% CI

C3 Mean ---+---------+----------+---------+--

1 9.00 (-----\*

-----)

2 9.60 (-----\*

-----)

3 11.60 (----\*

----)

---+---------+----------+---------+--

9.00 10.00 11.00 12.00

FIGURE 8.3.2 MINITAB dialog box and output for two-way analysis of variance,

Example 8.3.1.

EXERCISES

For Exercises 8.3.1 to 8.3.5 perform the ten-step hypothesis testing procedure for analysis of variance.

8.3.1 The objective of a study by Brooks et al. (A-11) was to evaluate the efficacy of using a virtual

kitchen for vocational training of people with learning disabilities. Twenty-four students participated

EXERCISES 343

in the study. Each participant performed four food preparation tasks and they were scored on the

quality of the preparation. Then each participant received regular vocational training in food preparation

(real training), virtual training using a TV and computer screen of a typical kitchen, workbook

training with specialized reading materials, and no training (to serve as a control). After each

of these trainings, the subjects were tested on food preparation. Improvement scores for each of the

four training methods are shown in the following table.

FIGURE 8.3.3 Partial SAS® output for analysis of Example 8.3.1.

The SAS System

Analysis of Variance Procedure

Dependent Variable: DAYS

Source DF Sum of Squares Mean Square F Value Pr > F

Model 6 43.46666667 7.24444444 16.72 0.0004

Error 8 3.46666667 0.43333333

Corrected Total 14 46.93333333

R-Square C.V. Root MSE DAYS Mean

0.926136 6.539211 0.65828059 10.06666667

Source DF Anova SS Mean Square F Value Pr > F

GROUP 2 18.53333333 9.26666667 21.38 0.0006

AGE 4 24.93333333 6.23333333 14.38 0.0010

Subject Real Virtual Workbook No

No. Training Training Training Training

1 2 10 2 &4

2 4 3 2 1

3 4 13 0 1

4 6 11 2 1

5 5 13 5 1

6 2 0 1 4

7 10 17 2 6

8 5 5 2 2

9 10 4 5 2

10 3 6 9 3

11 11 9 8 7

12 10 9 6 10

13 5 8 4 1

(Continued )

Subject Real Virtual Workbook No

No. Training Training Training Training

14 8 11 1 1

15 4 8 5 2

16 11 8 10 2

17 6 11 1 3

18 2 5 1 2

19 3 1 0 &3

20 7 5 0 &6

21 7 10 4 4

22 8 7 &2 8

23 4 9 3 0

24 9 6 3 5

Source: B. M. Brooks, Ph.D. Used with permission.

344 CHAPTER 8 ANALYSIS OF VARIANCE

After eliminating subject effects, can we conclude that the improvement scores differ among methods

of training? Let

8.3.2 McConville et al. (A-12) report the effects of chewing one piece of nicotine gum (containing 2 mg

nicotine) on tic frequency in patients whose Tourette’s disorder was inadequately controlled by

haloperidol. The following are the tic frequencies under four conditions:

a = .05.

Number of Tics During 30-Minute Period

After End of Chewing

Gum 0–30 30–60

Patient Baseline Chewing Minutes Minutes

1 249 108 93 59

2 1095 593 600 861

3 83 27 32 61

4 569 363 342 312

5 368 141 167 180

6 326 134 144 158

7 324 126 312 260

8 95 41 63 71

9 413 365 282 321

10 332 293 525 455

Source: Brian J. McConville, M. Harold Fogelson, Andrew B. Norman,

William M. Klykylo, Pat Z. Manderscheid, Karen W. Parker, and Paul R.

Sanberg. “Nicotine Potentiation of Haloperidol in Reducing Tic Frequency

in Tourette’s Disorder,” American Journal of Psychiatry, 148 (1991),

793–794. Copyright© 1991, American Psychiatric Association. Reprinted

by permission.

After eliminating patient effects, can we conclude that the mean number of tics differs among the

four conditions? Let a = .01.

Level of Remotivation Method

Initial

Motivation A B C D E

Nil 58 68 60 68 64

Very low 62 70 65 80 69

Low 67 78 68 81 70

Average 70 81 70 89 74

Do these data provide sufficient evidence to indicate a difference in mean scores among methods?

Let

8.3.4 The nursing supervisor in a local health department wished to study the influence of time of day

on length of home visits by the nursing staff. It was thought that individual differences among

nurses might be large, so the nurse was used as a blocking factor. The nursing supervisor collected

the following data:

Length of Home Visit by Time of Day

Early Late Early Late

Nurse Morning Morning Afternoon Afternoon

A 27 28 30 23

B 31 30 27 20

C 35 38 34 30

D 20 18 20 14

Do these data provide sufficient evidence to indicate a difference in length of home visit among

the different times of day? Let

8.3.5 Four subjects participated in an experiment to compare three methods of relieving stress. Each

subject was placed in a stressful situation on three different occasions. Each time a different method

for reducing stress was used with the subject. The response variable is the amount of decrease in

stress level as measured before and after treatment application. The results were as follows:

Treatment

Subject A B C

1 16 26 22

2 16 20 23

3 17 21 22

4 28 29 36

Can we conclude from these data that the three methods differ in effectiveness? Let a = .05.

a = .05.

a = .05.

EXERCISES 345

8.3.3 A remotivation team in a psychiatric hospital conducted an experiment to compare five methods for

remotivating patients. Patients were grouped according to level of initial motivation. Patients in each

group were randomly assigned to the five methods. At the end of the experimental period the patients

were evaluated by a team composed of a psychiatrist, a psychologist, a nurse, and a social worker,

none of whom was aware of the method to which patients had been assigned. The team assigned each

patient a composite score as a measure of his or her level of motivation. The results were as follows:

346 CHAPTER 8 ANALYSIS OF VARIANCE

8.3.6 In a study by Valencia et al. (A-13), the effects of environmental temperature and humidity on 24-

hour energy expenditure were measured using whole-body indirect calorimetry in eight normalweight

young men who wore standardized light clothing and followed a controlled activity regimen.

Temperature effects were assessed by measurements at 20, 23, 26, and 30 degrees Celsius

at ambient humidity and at 20 and 30 degrees Celsius with high humidity. What is the blocking

variable? The treatment variable? How many blocks are there? How many treatments? Construct

an ANOVA table in which you specify the sources of variability and the degrees of freedom for

each. What are the experimental units? What extraneous variables can you think of whose effects

would be included in the error term?

8.3.7 Hodgson et al. (A-14) conducted a study in which they induced gastric dilatation in six anesthetized

dogs maintained with constant-dose isoflurane in oxygen. Cardiopulmonary measurements

prior to stomach distension (baseline) were compared with measurements taken during .1, .5, 1.0,

1.5, 2.5, and 3.5 hours of stomach distension by analyzing the change from baseline. After distending

the stomach, cardiac index increased from 1.5 to 3.5 hours. Stroke volume did not change.

During inflation, increases were observed in systemic arterial, pulmonary arterial, and right atrial

pressure. Respiratory frequency was unchanged. PaO2 tended to decrease during gastric dilatation.

What are the experimental units? The blocks? Treatment variable? Response variable(s)? Can you

think of any extraneous variable whose effect would contribute to the error term? Construct an

ANOVA table for this study in which you identify the sources of variability and specify the degrees

of freedom.