Capítulo

$\boxed{1}$

Introducción al Análisis de Varianza.

1.1. Introducción.

Hasta el momento se han realizado inferencias con respecto a un parámetro poblacional y a la comparación de un parámetro entre 2 poblaciones. Para ello hemos usado la distribución normal, t - student, Chi cuadrado y F. Por lo general, existen situaciones en las que deseamos comparar un parámetro entre tres o más poblaciones, como por ejemplo el salario promedio de los trabajadores en 5 estados de Venezuela. En principio el investigador podría pensar en resolver este problema haciendo comparaciones dos a dos y usar para ello la distribución normal o la t - student, según sea el caso. Dicho procedimiento es inadecuado por las siguientes razones:

1. El procedimiento es muy largo, ya que hay que realizar tantas pruebas como parejas de tratamientos existan. Por ejemplo, si se desea probar la igualdad de 5 medias poblacionales, usando comparaciones dos a dos, existen $\binom{5}{2} = 10$ combinaciones posibles, es decir se tendrían que realizar 10 pruebas de hipótesis, usando en cada uno de los casos la distribución normal o la t - student.

2. No se puede generalizar para todas las medias poblacionales, la conclusión se obtiene por parejas de medias poblacionales.

3. Existe una alta probabilidad de cometer error tipo I, debido a que cuando se compara una media poblacional con cada una de las otras medias poblacionales y se realiza una prueba para cada par de medias, es muy probable que se llegue a concluir que existen diferencias significativas para algunos pares cuando en realidad no existe diferencia entre ellas. Por ejemplo, si la probabilidad de no rechazar H_0 en cada prueba es $1-\alpha = 0{,}95$, entonces la probabilidad de aceptar H_0 en las 10 pruebas es $(0{,}95)^1 0 = 0{,}6$ si las pruebas son independientes, y la probabilidad de Rechazar H_0 en las 10 pruebas ES $1-(0{,}95)^{10} = 0{,}4$, el cuál es el error tipo I.

Por estas razones es necesario considerar un método que tome en consideración todas las medias al mismo tiempo. Una de las aplicaciones del análisis de varianza es precisamente resolver este problema. Aunque este enfoque es valido para introducir el análisis de varianza, sólo refleja una parte del interés de la técnica, pues la misma permite una mejor planificación de los experimentos. A continuación se desarrolla el análisis de varianza como una técnica en el diseño de experimentos, pues este enfoque abarca el enfoque inicialmente propuesto.

1.2. ¿Qué es el Análisis de Varianza?

Definición 1.1 (Análisis de Varianza) *Es una técnica estadística que divide y analiza la variabilidad total observada de una variable en porciones atribuibles a distintos factores de interés para el investigador.*

Para entender mejor la definición veamos el siguiente ejemplo:

Ejemplo 1.2.1 *Se desea estudiar el efecto que puedan tener 5 tipos de dietas en los tiempo de coagulación de la sangre extraída de 24 animales. El análisis de varianza supone que cualquier variación que existe entre los promedios del tiempo de coagulación de la sangre se atribuye a:*

1. *Variación de los tiempos de coagulación dentro de las dietas.*

2. *Variación debido a las dietas, esto es, debido a la composición de cada dieta.*

La variación dentro de cada dieta se debe, por supuesto, a diversas causas, tal vez al tipo de sangre, a la presión, o cualquier otro elemento no tomado en cuenta. De cualquier manera dicha variación es considerada como una variación al azar o aleatoria. En cambio, la variación debido a la dieta, es una variación que no depende de variables asociadas con el animal sino de la composición de la dieta. En este caso, el análisis de varianza busca identificar cuanto de la variación del tiempo de coagulación de la sangre se debe a la dieta y cuánto a otros elementos no tomados en cuenta .

Ejemplo 1.2.2 *Se desea conocer la efectividad que tienen cuatro metodologías para la enseñanza de la estadística en cierta universidad. Para ello se aplicaron las cuatro metodologías sobre un grupo de estudiantes pertenecientes a dicha universidad y se registro la nota final obtenida por cada alumno al final del semestre. En este caso la variabilidad de la nota entre los alumnos se puede explicar por:*

1. *La variabilidad aportada por las distintas metodologías y*

2. *la variabilidad aportada por los estudiantes*

Ejemplo 1.2.3 *Se desea estudiar la variabilidad que presenta una colección de calificaciones que provienen de tres asignaturas en cuatro cursos distintos a lo largo de los últimos anos, la varianza total se puede descomponer en cuatro sumandos:*

1. *Variación debido a las asignaturas,*
2. *variación debido a los cursos,*
3. *variación debido a los años y*
4. *variación aportada por los alumnos.*

1.2.1. El Análisis de Varianza en el Diseño de Experimentos.

Cuando se está realizando una investigación el investigador puede bien sea observar las características de los datos ya existentes (sin tener participación en su generación) o imponer deliberadamente una o más condiciones experimentales sobre los elementos en estudio. En el segundo caso caso, se dice que el experimento fue diseñado. El principal propósito del diseño de un experimento es reducir la variabilidad de las respuestas, pues previamente se establecen las variables que se piensan inciden sobre el fenómeno en estudio, así como sus posibles valores.
Algunos conceptos relacionados con el diseño de experimentos.

Definición 1.2 (Variable dependiente o respuesta) *Es la variable que nos interesa medir o respuesta que se va estudiar, para determinar el efecto que tiene sobre ella la o las variables independientes.*

Definición 1.3 (Variables independientes o factores) *Son las variables que pueden influenciar en la variabilidad de la variable respuesta. Estas son controladas completamente por el experimentador.*

Definición 1.4 (Nivel del Factor) *Es un valor de la variable independiente o factor.*

Definición 1.5 (Tratamiento) *Es un nivel o una combinación de dos o más niveles de un factor o factores.*

Definición 1.6 (Unidad Experimental) *Son los objetos sobre los cuales se aplican los tratamientos para obtener una respuesta.*

Definición 1.7 (Error Experimental) *Es la variación que no se puede atribuir a un cambio de tratamiento, es decir, a la que se produce por los factores extraños que pueden influir en la respuesta y que deben ser controlados o eliminados por el investigador.*

Definición 1.8 (Aleatorización) *Consiste en asignar en forma aleatoria los tratamientos a las unidades experimentales con el propósito de eliminar los sesgos que produce dicha asignación.*

Por lo general el diseño de un experimento comprende:

1. La selección de los factores que deben incluirse en el experimento y la especificación del o los parámetros de interés.

2. Decidir cuánta información se debe utilizar para estimar los parámetros.

3. Seleccionar los tratamientos que deben utilizarse en el experimento y el número de unidades experimentales que deben asignarse a cada uno.

4. Decidir como deben asignarse los tratamientos a las unidades experimentales.

Por lo tanto, dependiendo del número de factores, selección de los tratamientos y asignación de los tratamientos a las unidades experimentales existen distintos tipos de diseños de experimentos los cuales estudiaremos algunos de ellos más adelante.
Una vez diseñado y el experimento y recolectados los datos, interesa ver que factores de los que tomaron en cuenta influyen sobre la variable respuesta. Para ello se realiza el análisis de varianza, el cuál como se vio antes consiste en separar la variación total en cada uno de sus tratamientos y así determinar cuál de ellos afecta significativamente la respuesta.

1.2.2. Supuestos del Análisis de Varianza

Independientemente del diseño experimental usado para generar los datos, para que el análisis de varianza tenga validez, se deben cumplir los siguientes supuestos:

1. Cada tratamiento representa una población.

2. Normalidad: Las poblaciones de las que se extraen las muestras se distribuyen normal.

3. Homocedasticidad: Las varianzas poblacionales son iguales.

4. Los errores aleatorios son independientes y se distribuyen normal con media cero y varianza constante.

Cuando los tamaños de muestras son grandes e iguales, el análisis de varianza tiene la propiedad de ser robusta, es decir, la violación de los supuestos no afecta significativamente los resultados. Según Mendenhall, violar el supuesto de una varianza constante es más grave, en particular cuando los tamaños de las muestras no son casi iguales.

1.3. Diseño Completamente Aleatorizado (DCA)

1.3.1. Introducción

Denominado también diseño de una forma o vía de clasificación. Es un diseño útil para describir un experimento en el que se desean comparar k tratamientos (niveles de un factor), donde las unidades experimentales son homogéneas y los tratamientos son asignados en forma completamente aleatoria a estas unidades experimentales.

Supóngase que tenemos N unidades experimentales homogéneas y k tratamientos. Sean las N unidades experimentales particionadas aleatoriamente (con igual probabilidad) en k conjuntos de tamaño n_j . Sean los k tratamientos asignados a los k conjuntos de forma tal que el j-ésimo tratamiento es aplicado a cada una de las unidades experimentales en el j-ésimo conjunto. Este procedimiento define un **diseño completamente aleatorizado.**

Dentro de las ventajas del diseño completamente aleatorizado se encuentran:

1. Es completamente flexible. Puede usarse con cualquier número de tratamientos y de réplicas. El número de replicaciones puede variar de tratamiento a tratamiento, aunque esto no se debe hacer sin una buena razón, ya que si el diseño es balanceado (el mismo número de réplicas por tratamiento), la prueba estadística es relativamente insensible a pequeñas violaciones del supuesto de igualdad de varianzas y por otro lado, la potencia del test esta maximizado si las muestras son de igual tamaño.

2. El análisis estadístico es fácil de llevar a cabo aún si el diseño no es balanceado, si el error difiere de tratamiento a tratamiento y si los diversos tratamientos poseen varianzas distintas, lo cual se conoce como falta de homogeneidad (heterogenei-

dad) del error experimental. Bajo estas condiciones, las pruebas de hipótesis y la construcción del intervalo de confianza deben conducirse con un cuidado especial cuando hay heterogeneidad de la varianza.

3. La sencillez del análisis no se pierde si algunas unidades experimentales o tratamientos enteros faltan o se descartan. En este tipo de diseño, la información que se pierde debido a observaciones faltantes es mínima con relación a la sufrida por otros diseños. El número de grados de libertad para estimar el error experimental es máximo, lo que incide en un aumento en la precisión del experimento. Esto resulta significativamente importante en experimentos pequeños, es decir, en aquellos en los que se cuenta con pocos grados de libertad para el error experimental.

Como la aleatorización no tiene restricciones, el error experimental incluye toda la variación entre las unidades experimentales excepto, la debida a los tratamientos. Esto representa la principal desventaja del diseño completamente aleatorizado, lo cual se traduce en ineficiencia. En muchas situaciones es posible agrupar las unidades experimentales de modo que la variación entre las unidades de un mismo grupo sea menor que la variación entre las unidades de diferentes grupos. Ciertos diseños sacan ventaja de tal agrupamiento, ya que excluyen la variación del error experimental entre grupos y aumentan la precisión del experimento.

A pesar de lo expuesto anteriormente, la aleatorización completa resulta ser el procedimiento obvio en muchos tipos de experimentos de laboratorio, en los que una cantidad de material está completamente mezclada y luego se divide en porciones pequeñas para formar las unidades experimentales a los cuales se asignan los tratamientos en forma aleatoria o, en experimentos con animales y plantas con condiciones ambientales muy parecidas.

Ejemplo 1.9 *Supongamos que deseamos analizar el tiempo de coagulación para muestras de sangre tomadas de animales sometidos a cuatro diferentes drogas A, B, C y D. Las drogas fueron aplicadas aleatoriamente a los animales. Queremos entonces, medir el efecto de las drogas sobre el tiempo de coagulación.*

1.3.2. El Modelo

La respuesta observada para cada tratamiento, Y_{ij} es una variable aleatoria que puede ser expresada como la suma de dos componentes, a saber:

- Un componente que mide la media de tratamientos
- Un componente que representa al error aleatorio (termino de error aleatorio)

La media de los tratamientos muestra la influencia de los tratamientos sobre la variable respuesta y el error es una cantidad aleatoria que no puede predecirse con anticipación, pero cuyo valor esperado es igual a cero.

El modelo matemático apropiado para describir las observaciones, esta dada por:

$$Y_{ij} = \mu_j + \varepsilon_{ij} \qquad \begin{matrix} i = 1, 2, \cdots, n_j \\ j = 1, 2, \cdots, k \end{matrix} \tag{1.1}$$

donde:

Y_{ij} es la i-esima observación del j-ésimo tratamiento.

μ_j es la media del j-ésimo tratamiento

ε_{ij} es el error aleatorio, los cuales se suponen $N(0, \sigma^2)$ e independientes

El modelo estadístico propuesto en 1.1, describe dos situaciones diferentes con respecto al efecto de los tratamientos.

- Los k tratamientos pueden ser escogidos a criterio o conveniencia del investigador. En esta situación, se desea probar hipótesis sobre las medias de los tratamientos, y las conclusiones solamente pueden ser aplicadas a los niveles del factor (tratamientos) considerados en el análisis. Este modelo es llamado **modelo de efectos fijos**.

- Si los k tratamientos constituyen una muestra aleatoria de la población de tratamientos, las conclusiones pueden extenderse a la población de tratamientos. Aquí los μ_j son considerados variables aleatorias. En este caso, las hipótesis serán acerca de la variabilidad de los μ_j. Este modelo es llamado **modelo de efectos aleatorios** o **modelo de componentes de varianza**

En este curso sólo vamos a desarrollar el modelo de efectos aleatorios. Ahora bien, los datos observados de un diseño completamente aleatorizado pueden presentase como en la tabla 1.1

Tabla 1.1: Datos Muestrales de un DCA

	Tratamiento			
	1	2	$\cdots$	k
	Y_{11}	Y_{12}	$\cdots$	Y_{1k}
	Y_{21}	Y_{22}	$\cdots$	Y_{2k}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$	$\cdots$	$Y_{n_k k}$
Total	$Y_{,1}$	$Y_{,2}$	$\cdots$	$Y_{.k}$
n_j	n_1	n_2	$\cdots$	n_k
Media	$\bar{Y}_{,1}$	$\bar{Y}_{,2}$	$\cdots$	$\bar{Y}_{.K}$

La hipótesis a probar en este tipo de diseños es que la media de los tratamientos sean iguales, es decir,

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \mu_i \neq \mu_j \text{ para algún } i \neq j \end{aligned} \tag{1.2}$$

Otra manera de plantear el modelo de un diseño completamente aleatorizado, ecuación 1.1, se tiene al expresar la media del j-esimo tratamiento, μ_j como

$$\mu_j = \mu + \tau_j$$

donde

μ es la media general.

τ_j es el efecto del j-ésimo tratamiento

de esta manera, el modelo de un DCA se puede escribir como

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad \begin{array}{c} i = 1, 2, \cdots, n_j \\ j = 1, 2, \cdots, k \end{array} \tag{1.3}$$

En este caso, la hipótesis a probar se puede plantear como

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = \cdots = \tau_k = 0 \\ H_1 &: \tau_i \neq 0 \text{ para algún } i \end{aligned} \tag{1.4}$$

Para probar dicha hipótesis se realiza un análisis de varianza, cuyo desarrollo se verá a continuación.

1.3.3. Análisis de Varianza para el DCA

En el desarrollo analítico del Análisis de varianza (ANDEVA) se necesita calcular:

El gran total: $Y_{..} = \sum_{j=1}^{k} Y_{.j} = \sum_{j=1}^{k}\sum_{i=1}^{n_j} Y_{ij}$.

El total para el tratamiento j: $Y_{.j} = \sum_{i=1}^{n_j} Y_{ij}$. $\bar{Y}_{.j} = \frac{Y_{.j}}{n_j}$

El número de observaciones: $N = \sum_{j=1}^{k} n_j$.

La gran media: $\bar{Y}_{..} = \frac{Y_{..}}{N}$.

La media del tratamiento j: $\bar{Y}_{.j} = \frac{Y_{.j}}{n_j}$.

Como se dijo antes el análisis de varianza busca separar la variabilidad total en porciones significativas de variabilidad, en este caso, que solo hay un factor de interés además del error aleatorio, se busca separar la variabilidad de las observaciones con respecto a la media en 2 fuentes de variabilidad, una debida a los tratamientos y otra al error aleatorio.

Una medida de la desviación de las observaciones con respecto a la media está dada por $\sum_{j=1}^{k}\sum_{i=1}^{n}(Y_{ij} - \bar{Y}..)^2$, la cuál restándole y sumándole los promedios de los tratamientos y ordenándolo convenientemente se tiene que

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}..)^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{.j} + \bar{Y}_{.j} - \bar{Y}..)^2 \quad (1.5)$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}[(\bar{Y}_{.j} - \bar{Y}..) + (Y_{ij} - \bar{Y}_{.j})]^2 \quad (1.6)$$

Al desarrollar el segundo miembro de la ecuación 1.5, se obtiene

$$\begin{aligned}
\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{..})^2 &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}[(\bar{Y}_{.j}-\bar{Y}_{..})^2+2(\bar{Y}_{.j}-\bar{Y}_{..})(Y_{ij}-\bar{Y}_{.j})+(Y_{ij}-\bar{Y}_{.j})^2] \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-\bar{Y}_{..})^2+\sum_{j=1}^{k}\sum_{i=1}^{n_j}2(\bar{Y}_{.j}-\bar{Y}_{..})(Y_{ij}-\bar{Y}_{.j}) \\
&\quad +\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-\bar{Y}_{..})^2+\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2
\end{aligned}$$

Como las sumatorias que contienen productos cruzados son iguales a cero, se tiene que

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{..})^2=\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-\bar{Y}_{..})^2+\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 \tag{1.7}$$

La ecuación 1.7 representa la descomposición de la suma de cuadrados total. La cual se puede expresar de la siguiente manera

$$SC_T = SC_{Tr} + SC_E \tag{1.8}$$

Esta última ecuación es la ecuación fundamental del Análisis de Varianza.

Para efecto de cálculos, las formulas anteriores usualmente se desarrollan y se reescriben de la forma siguiente

$$\begin{aligned}
SC_T &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}Y_{ij}^2-\frac{Y_{..}^2}{N} \\
SC_{Tr} &= \sum_{j=1}^{k}\frac{Y_{.j}^2}{n_j}-\frac{Y_{..}^2}{N} \\
SC_E &= SC_T - SC_{Tr}
\end{aligned}$$

En base a estos estadísticos, se obtienen dos estadísticos adicionales, usualmente llamados Medias Cuadráticas o Cuadrados Medios y resultan de dividir cada suma de cuadrados por su correspondiente grados de libertad.
Cuadrado medio de tratamientos

$$CM_{Tr} = \frac{SC_{Tr}}{k-1}$$

y,
Cuadrado medio del error

$$CM_E = \frac{SC_E}{N-k}$$

Cuyos valores esperados están dados por:

$$\begin{aligned} E(CM_E) &= \sigma^2 \\ E(CM_{Tr}) &= \sigma^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{\tau_j^2}{k-1} \end{aligned}$$

Observemos que si $H_0 : \tau_j = 0 \forall j$, es verdadera, $E(CM_{Tr}) = \sigma^2$. Esto es, en este caso se tienen dos estimadores insesgados e independientes de σ^2, el CM_{Tr} y el CM_E.
Ahora bien, sabemos que $SC_T = SC_{Tr} + SC_E$ y Además, puede demostrarse que

$$\frac{SC_T}{\sigma^2} \sim \chi^2_{N-1} \tag{1.9}$$

Si H_0 es verdadera, y de acuerdo al teorema de Cochran es posible definir dos estadísticos chi-cuadrados independientes

$$\frac{SC_{Tr}}{\sigma^2} \sim \chi^2_{k-1} \tag{1.10}$$

$$\frac{SC_E}{\sigma^2} \sim \chi^2_{N-k} \tag{1.11}$$

Por lo tanto, el estadístico

$$F_0 = \frac{\frac{SC_{Tr}}{\sigma^2}/k-1}{\frac{SC_E}{\sigma^2}/N-k} = \frac{CM_{Tr}}{CM_E} \tag{1.12}$$

sigue una distribución F con $k-1$ y $N-k$ grados de libertad.

Estos resultados pueden ser resumidos bajo el formato general de la tabla de ANDEVA, como se muestra en la tabla 1.2.

Tabla 1.2: Tabla de Análisis de Varianza

Fuente de Variación	Suma de Cuadrados	Grado de Libertad	Cuadrado Medio	F
Tratamiento	SC_{Tr}	k-1	CM_{Tr}	F_0
Error	SC_E	N-k	CM_E	
Total	SC_T	N-1		

Rechazamos H_0 sí y solo sí: $F > F_{1-\alpha,k-1,N-k}$

Ejemplo 1.10 *Los datos que figuran en la tabla 1.3 son los resultados de un diseño completamente aleatorizado para el cual la respuesta son los kilowats hora, empleados por los sistemas de calentamiento (en cientos de kilowats hora) para casa muy similares en un lugar dado, como función de cinco aislamientos térmicos (en pulgadas). Con base en esta información, ¿Existe alguna razón para creer que por lo menos algunos consumos de energía promedio para los cinco niveles de aislamiento son diferentes?. Suponga un nivel de significación igual a 0.01. Se desea probar la hipótesis*

Tabla 1.3: Calor empleado para cinco niveles de aislamiento

Espesor del aislamiento del techo (pulgadas)				
4	6	8	10	12
14.4	14.5	13.8	13.0	13.1
14.8	14.1	14.1	13.4	12.8
15.2	14.6	13.7	13.2	12.9
14.3	14.2	13.6		13.2
14.6		14.0		13.3
				12.7

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_5 = \mu \\ H_1 &: \mu_i \neq \mu_j \textit{ para algún } i \neq j \end{aligned} \tag{1.13}$$

o de manera equivalente

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = \cdots = \tau_5 = 0 \\ H_1 &: \tau_j \neq 0 \textit{ para algún } j \end{aligned} \tag{1.14}$$

El número de observaciones y los totales se encuentran en la tabla 1.4. Por lo tanto,

Tabla 1.4: Cálculos del ejemplo 1.10

	Tratamiento				
	1	2	$\cdots$	k	
	14.4	14.5	13.8	13.0	13.1
	14.8	14.1	14.1	13.4	12.8
	15.2	14.6	13.7	13.2	12.9
	14.3	14.2	13.6		13.2
	14.6		14.0		13.3
					12.7
Total	73.3	57.4	69.2	39.6	78
n_j	$n_1 = 5$	$n_2 = 4$	$n_2 = 5$	$n_4 = 3$	$n_5 = 6$

las sumas de los cuadrados son las siguientes:

$$SC_T = 14{,}4^2 + 14{,}8^2 + ... + 12{,}7^2 - \frac{317{,}5^2}{23} = 11{,}05$$

$$SC_{Tr} = \frac{73{,}3^2}{5} + \frac{57{,}4^2}{4} + \frac{69{,}2^2}{5} + \frac{39{,}6^2}{3} + \frac{78^2}{6} - \frac{317{,}5^2}{23} = 9{,}836$$

$$SC_E = 11{,}05 - 9{,}836 = 1{,}214$$

La información se ha agrupado en una tabla de análisis de varianza que se muestra en la tabla 1.5 Dado que $F = 36{,}48 > F_{0{,}99{,}4{,}18} = 4{,}58$ *se rechaza la hipótesis nula de que*

Tabla 1.5: Tabla de Análisis de Varianza para el ejemplo 1.10

Fuente de Variación	Suma de Cuadrados	Grado de Libertad	Cuadrado Medio	F
Tratamiento	9.836	4	2.459	36.48
Error	1.214	18	0.0674	
Total	11.05	22		

no existe ningún efecto debido a los tratamientos. En relación con lo anterior, existe una razón para creer que parte de los consumos promedios de energía son diferentes para los cinco niveles de aislamiento.

1.4. Métodos A posteriori

En algunas investigaciones, sus objetivos o la naturaleza propia del problema indican que debe someterse a prueba la significación de determinados tratamientos o de una combinación de los mismos. Esto es, existen situaciones en las que los tratamiento bajo investigación tienen alguna relación lo cual incide en que unas comparaciones son de más interés que otras. A esto nos referimos como comparaciones a priori o, preplaneadas.

Ahora bien, si una vez realizado el experimento y analizada la información, rechazamos la hipótesis nula, significa que por lo menos una de las medias de los tratamientos es diferente del resto o, que al menos un efecto de tratamiento difiere significativamente de cero. Sin embargo, el rechazar la hipótesis nula no ofrece ninguna información que permita dar respuesta a la siguiente interrogante; ¿Cuál o cuales medias difieren?

En esta sección se van a desarrollar procedimientos que permiten probar la significación de algunas comparaciones entre los efectos de tratamientos o entre todas las posible parejas entre tratamientos, en el primer caso se dice que son comparaciones por contraste y en el segundo comparaciones múltiples. Veamos a continuación dichos procedimientos:

1.4.1. Comparaciones por Contrastes

Aunque por lo general, se está interesado en la comparación de los tratamientos agrupados por parejas, lo que se traduce en contrastar hipótesis del tipo $H_0 : \mu_1 = \mu_2$, o de manera equivalente $H_0 : \mu_1 - \mu_2 = 0$, dando como consecuencia un total de $\binom{k}{2}$ comparaciones, existen situaciones en las que es de interés comparar una combinación lineal de tratamientos, las cuales se traducen en contrastar hipótesis de la forma

$H_0 : \sum_{j=1}^{m} c_j\mu_j = 0$. La ecuación que se presenta en la hipótesis antes planteada se conoce como contraste.

Definición 1.11 (Contraste) *Un contraste* (L) *es una combinación lineal de las medias poblacionales de interés, es decir,*

$$L = \sum_{j=1}^{m} c_j\mu_j \tag{1.15}$$

donde

c_j *son números reales que cumplen con la condición* $\sum_{j=1}^{m} c_j = 0$

μ_j *es la media del* j*-esimo tratamiento.*

Por ejemplo, las hipótesis nulas del tipo $H_0 : \mu_i = \mu_j$, se pueden escribir como $H_0 : \mu_i - \mu_j = 0$, definen el contraste $L = c_1\mu_1 - c_2\mu_2$ donde $c_1 = 1$ y $c_2 = -1$. La hipótesis $H_0 : \frac{\mu_1+\mu_2}{2} = \mu_3$ define un contraste con $c_1 = \frac{1}{2}$, $c_2 = \frac{1}{2}$ y $c_3 = -1$.

Para probar dichas hipótesis, bajo el supuesto que la distribución de las poblaciones son $N(\mu_j, \sigma^2)$, se usa como estimador $\hat{L} = \sum_{j=1}^{m} c_j\hat{\mu}_j = \sum_{j=1}^{m} c_j\bar{Y}_j$, el cual se distribuye normal con parámetros

$$E[\hat{L}] = \sum_{j=1}^{m} c_j\mu_j \qquad \text{y} \qquad Var[\hat{L}] = \sigma^2 \sum_{j=1}^{m} \frac{c_j^2}{n_j}$$

$$L_0 = \frac{\hat{L} - \sum_{j=1}^{m} c_j\mu_j}{\sqrt{\sigma^2 \sum_{j=1}^{m} \frac{c_j^2}{n_j}}} \tag{1.16}$$

sigue una distribución normal con media igual a cero y varianza igual a 1.

Como por lo general σ^2 es desconocida, usamos CM_E como su estimador, de manera que,

$$L_0 = \frac{\hat{L} - \sum_{j=1}^{m} c_j \mu_j}{\sqrt{CM_E \sum_{j=1}^{m} \frac{c_j^2}{n_j}}} \tag{1.17}$$

el cual se distribuye t-student con $N-k$ grados de libertad. De esta forma la expresión

$$\hat{L} \pm t_{\alpha/2,N-k} \sqrt{CM_E \sum_{j=1}^{m} \frac{c_j^2}{n_j}} \tag{1.18}$$

constituye un intervalo de confianza del $100(1-\alpha)\,\%$ para L.

Si el intervalo contiene el cero, se concluye que L es estadísticamente igual a cero. Podemos indicar que rechazamos cuando $|L_0| > t_{\alpha/2,N-k}$.

Método de Scheffé

Es un método alternativo del t-student para probar contrastes. En este caso Scheffé propone el siguiente intervalo de confianza para el contraste L.

$$\hat{L} \pm A \sqrt{CM_E \sum_{j=1}^{m} \frac{c_j^2}{n_j}} \tag{1.19}$$

donde

$$A = \sqrt{(k-1)F_{\alpha,k-1,N-k}}$$

Nuevamente si el intervalo de confianza no contiene al cero, entonces decimos que la prueba es significativa, es decir que se rechaza la hipótesis de que el contraste sea igual a cero.

1.4.2. Comparaciones Múltiples

Método de la Diferencia Mínima Significativa (LSD)

Procedimiento propuesto por Fisher en el año 1.935 y que consiste en realizar todas las posibles comparaciones entre pares de medias, es decir, todos las $\binom{k}{2}$ pruebas de la forma:

$$\begin{aligned} H_0 &: \mu_i = \mu_j \\ H_1 &: \mu_i \neq \mu_j \forall i \neq j \end{aligned} \tag{1.20}$$

Para probar dicha hipótesis se usa como estadístico de prueba la diferencia entre los valores estimados de las medias (medias muestrales), es decir $\bar{Y}_{.j} - \bar{Y}_{.k}$, cuya distribución (suponiendo que las poblaciones son $N(\mu_j, \sigma^2)$) es $N\left[\mu_j - \mu_{j'}, \sigma^2(1/n_j + 1/n_{j'})\right]$. Por lo tanto, bajo la hipótesis nula cierta el estadístico

$$Z = \frac{\bar{Y}_{.j} - \bar{Y}_{.k}}{\sigma\sqrt{1/n_j + 1/n_k}} \tag{1.21}$$

se distribvuye normal estándar. Pro como σ^2 es desconocido, se usa el CM_E para estimarlo. asi, el estadístico

$$T = \frac{\bar{Y}_{.j} - \bar{Y}_{.k}}{\sqrt{CM_E(1/n_j + 1/n_k)}} \tag{1.22}$$

se distribuye t-student con $N-k$ grados de libertad. Por lo tanto, se rechaza H_0 si $|T| > t_{\alpha/2,N-k}$, lo cual es equivalente a rechazar H_0 si

$$|\bar{Y}_{.j} - \bar{Y}_{.k}| > t_{\alpha/2,N-k}\sqrt{CM_E(1/n_j + 1/n_k)}$$

Otra manera de contrastar la hipótesis es construyendo el intervalo de confianza para $\mu_j - \mu_k$ el cual es

$$|\bar{Y}_{.j} - \bar{Y}_{.k}| \pm t_{\alpha/2,N-k}\sqrt{CM_E(1/n_j + 1/n_k)}$$

Si el intervalo no contiene el cero rechazamos H_0.
Esta prueba es conocida como LSD, pues según Gutiérrez(2006), $t_{\alpha/2,N-k}\sqrt{CM_E(1/n_j + 1/n_k)}$ es la mínima diferencia que debe haber entre dos medias muestrales para poder considerar que los tratamientos correspondientes son significativamente diferentes.

Método de los Rangos Estudentizados o Método de Tukey

Procedimiento aplicado para probar hipótesis de la forma $H_0 : \mu_j - \mu_k = 0$, inicialmente en diseños balanceados. Este método hace uso de la Distribución del Rango Estudentizado, el cual se define a continuación.

Definición 1.12 *Sean $Z_1, ..., Z_m$ y U variables aleatorias independientes, tales que $Z_i \sim N(0;1)(i = 1,2,...m)$ y $U \sim \chi^2_m$. Sea ademas,*

$$q = \max_{i \neq j} \frac{|Z_i - Z_j|}{\sqrt{U/m}} \tag{1.23}$$

Decimos que q tiene una distribución de rango estudentizado , lo que se denota, $q \sim q_{k;m}$.

Para probar H_0 , se debe calcular:

$$T = q_{\alpha;k,N-k}\sqrt{CM_E/n} \tag{1.24}$$

donde $q_{\alpha;k,N-k}$ es el punto superior α de la distribución de rango estudentizado. Existen tablas de estos valores que pueden ser consultadas en libros de diseños de experimentos o modelos lineales.

Si $|\bar{Y}_{.j} - \bar{Y}_{.k}| > T$ concluimos que μ_j y μ_k son diferentes, en otro caso, se consideran iguales.

Para el caso no balanceado, Kramer (1.956) propone el siguiente cambio en 1.24

$$T = q_{\alpha;k,f_E}\sqrt{\frac{1}{2}\left(\frac{1}{n_j} + \frac{1}{n_k}\right) CM_E/n} \tag{1.25}$$

donde f_E son los grados de libertad para el error. Este método es referido como el método de Tukey-Kramer.

Método de los Rangos Múltiples de Duncan

Test diseñado para comparar todos los posibles pares de medias $[k(k-1)/2]$. A diferencia del test de Tukey, éste usa diferentes valores críticos, los cuales dependen del rango de $|\bar{Y}_{.j}$ y $\bar{Y}_{.k}$. Esto es, dependen del número de medias entre ellas, una vez que han sido ordenadas en forma ascendente.

Sean $\bar{Y}_{(,1)}, ..., \bar{Y}_{(.k)}$ las medias de tratamientos ordenadas en forma ascendente. Si entre $\bar{Y}_{(.j)}$ y $\bar{Y}_{(.k)}$ hay p medias, entonces un test rango estudentizado de tamano α, es conducido comparando $\bar{Y}_{(.j)} - \bar{Y}_{(.k)}$ con $D_p = r_\alpha(p, f_\varepsilon)\sqrt{\frac{CM_E}{n}}$, donde $r_\alpha(p, f_\varepsilon)$ es el rango significativo de la tabla de Duncan para el nivel α. Si $\bar{Y}_{(.j)} - \bar{Y}_{(.k)} > D_p$, entonces μ_j y μ_k son significativamente diferentes.

El procedimiento de Duncan se desarrolla de la siguiente manera:

1. Ordenar las medias en forma ascendente.

2. Obtener las diferencias entre cada par de medias de la siguiente manera:

$$\bar{Y}_{.(k)} - \bar{Y}_{.(1)}, \bar{Y}_{.(k)} - \bar{Y}_{.(2)}, ..., \bar{Y}_{.(k)} - \bar{Y}_{.(k-1)}, ..., \bar{Y}_{.(2)} - \bar{Y}_{.(1)}$$

3. Obtener $r_\alpha(p, f_\varepsilon)$ y comparar $\bar{Y}_{.(k)} - \bar{Y}_{.(1)}$ con D_k. Si esta diferencia no es significativa, debemos considerar las diferencias $\bar{Y}_{.(k)} - \bar{Y}_{.(2)}$ y $\bar{Y}_{.(k-1)} - \bar{Y}_{.(1)}$ y compararlas con D_{k-1} , y así sucesivamente hasta comparar las diferencias $\bar{Y}_{.(k)} - \bar{Y}_{.(k-1)}$, $\bar{Y}_{.(k-1)} - \bar{Y}_{.(k-2)}$ con D_2.

Si el modelo es no balanceado, n suele ser reemplazado por $n_h = \frac{k}{\sum_{j=1}^{k} \frac{1}{n_j}}$.ç

Método de Newman Keuls

Es un método restringido a la comparación entre pares de medias. Es similar en cuanto a su procedimiento, al Test de Rangos Múltiples de Duncan, no así en su eficiencia, ya que la prueba de Duncan es más eficaz. En este procedimiento las medias deben ser ordenadas en forma ascendente y se requiere del calculo de todas las posibles diferencias críticas entre las medias. Estas diferencias críticas están dadas por:

$$NK_p = q_{\alpha;p,f_E}\sqrt{\frac{CME}{n}} \qquad p = 2, 3, ..., k \tag{1.26}$$

donde q_α es el valor critico de la tabla de rango estudentizado. μ_j y μ_k, se consideran significativamente diferentes si y solo si $(\bar{Y}_{(.j)} - \bar{Y}_{(.k)}) > NK_p$.

Método de Dunnet

Existen situaciones en las que dentro del grupo de k tratamientos se tiene un tratamiento control, y el objetivo principal del experimento es comparar a los $(k-1)$ tratamientos restantes con éste. Esto es, si el tratamiento S es el control, entonces estamos interesados en probar la hipótesis:

$$H_0 : \mu_S = \mu_j \qquad\qquad j = 1, ..., k; \qquad j \neq S$$

Para hacer las $(k-1)$ comparaciones se usa el procedimiento desarrollado por Dunnett y el cual consiste en comparar $(\bar{Y}_{(.S)} - \bar{Y}_{(.j)})$ con el valor crítico

$$\begin{aligned}
D &= \left(\frac{2CM_E}{n}\right)^{-1/2} d_{k-1,\alpha,f} \quad si \quad H_1 : \mu_s > \mu_j \\
D' &= -\left(\frac{2CM_E}{n}\right)^{-1/2} d_{k-1,\alpha,f} \quad si \quad H_1 : \mu_s < \mu_j \\
D'' &= \left(\frac{2CM_E}{n}\right)^{-1/2} d_{k-1,\alpha/2,f} \quad si \quad H_1 : \mu_s \neq \mu_j
\end{aligned}$$

Luego, rechazamos H_0 sí y sólo sí:

$$\begin{aligned}
D \geq D \quad & si \quad H_1 : \mu_s > \mu_j \\
D \leq D' \quad & si \quad H_1 : \mu_s < \mu_j \\
|D| \neq D'' \quad & si \quad H_1 : \mu_s \neq \mu_j
\end{aligned}$$

1.5. Ejercicios

1. Demuestre que la suma de los productos cruzados en la descomposición de la suma de cuadrados es cero.

2. Obtenga las formulas usuales del análisis de varianza a partir de las formulas teóricas.

3. Demuestre que

$$\begin{aligned} E(CM_E) &= \sigma^2 \\ E(CM_{Tr}) &= \sigma^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{\tau_j^2}{k-1} \end{aligned}$$

4. Demuestre que

$$\frac{SC_T}{\sigma^2} \sim \chi^2_{N-1}$$

$$\frac{SC_{Tr}}{\sigma^2} \sim \chi^2_{k-1}$$

$$\frac{SC_E}{\sigma^2} \sim \chi^2_{N-k}$$

5. Demuestre que

$$F_0 = \frac{CM_{Tr}}{CM_E} \sim F_{k-1,N-k}.$$

Nota: Establezca los supuestos necesarios.

6. Los miembros de un equipo ciclista se dividen al azar en tres grupos que entrenan con métodos diferentes. El primer grupo realiza largos recorridos a ritmo pausado, el segundo grupo realiza series cortas de alta intensidad y el tercero trabaja en el gimnasio con pesas y se ejercita en el pedaleo de alta frecuencia. Después de

un mes de entrenamiento se realiza un test de rendimiento consistente en un recorrido cronometrado de 9 Km. Los tiempos empleados fueron los siguientes

Método I	Método II	Método III
15	14	13
16	13	12
14	15	11
15	16	14
17	14	11

A un nivel de confianza del 95 % ¿Puede considerarse que los tres métodos producen resultados equivalentes? O por el contrario ¿Hay algún método superior a los demás?

7. Una lista de palabras sin sentido se presenta en la pantalla del ordenador con cuatro procedimientos diferentes, asignados al azar a un grupo de sujetos. Posteriormente se les realiza una prueba de recuerdo de dichas palabras, obteniéndose los siguientes resultados:

Proc. I	Proc. II	Proc. III	Proc. IV
5	9	8	1
7	11	6	3
6	8	9	4
3	7	5	5
9	7	7	1
7		4	4
4		4	
2			

¿Qué conclusiones pueden sacarse acerca de las cuatro formas de presentación, con un nivel de significación del 5

8. Una egresada de contaduría tiene ofertas de trabajo de cuatro empresas. Para examinar un poco más las propuestas, solicitó a una muestra de personas de nuevo ingreso, decirle cuántos meses trabajaron cada una para su compañia, antes de recibir un aumento de sueldo. La información muestral es:

A	B	C	D
12	14	18	12
10	12	12	14
14	10	16	16
12	10		
9	7	7	1

Al nivel de significancia de 0,05; existe alguna diferencia entre las cuatro empresas, en el número medio de meses antes de recibir un aumento de sueldo?

9. Cierta ciudad está dividida en cuatro distritos. El jefe de policia quiere determinar si hay alguna diferencia en el número promedio de crímenes cometidos en cada distrito. Registró el número de crímenes reportados en cada distrito en una muestra de seis días. Al nivel de significancia 0,05; puede el funcionario concluir que hay diferencia en el número promedio de crímenes?

A	B	C	D
13	21	12	16
15	13	14	17
14	18	15	18
15	19	13	15
14	18	12	20
15	19	15	18

10. En una empresa electrónica se estudian cuatro tipos de circuitos para comparar la cantidad de ruido de fondo asociado a cada circuito. Se han obtenido los siguientes datos:

circuito	Ruido observado				
1	19	20	19	30	8
2	80	61	73	56	80
3	47	26	25	35	50
4	95	46	83	78	97

¿Es la cantidad media de ruido asociado a cada circuito la misma?, ¿qué circuito seleccionaría?