UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

MAESTRÍA EN ESTADÍSTICA
(SEGUNDA PROMOCIÓN)

# " DESARROLLO PROGRAMÁTICO DE LA ASIGNATURA DISEÑO DE EXPERIMENTOS "

PRESENTADO POR:

MARÍA DEL TRANSITO GUTIÉRREZ REYES

ALBA IDALIA CÓRDOVA CUÉLLAR

PARA OPTAR AL TÍTULO DE:

MASTER EN ESTADÍSTICA

CIUDAD UNIVERSITARIA, JULIO DE 2003

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

MAESTRÍA EN ESTADÍSTICA
(SEGUNDA PROMOCIÓN)

# " DESARROLLO PROGRAMÁTICO DE LA ASIGNATURA DISEÑO DE EXPERIMENTOS "

PRESENTADO POR:

MARÍA DEL TRANSITO GUTIÉRREZ REYES

ALBA IDALIA CÓRDOVA CUÉLLAR

PARA OPTAR AL TÍTULO DE:

MASTER EN ESTADÍSTICA

CIUDAD UNIVERSITARIA, JULIO DE 2003

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

MAESTRÍA EN ESTADÍSTICA
(SEGUNDA PROMOCIÓN)

# "DESARROLLO PROGRAMÁTICO DE LA ASIGNATURA DISEÑO DE EXPERIMENTOS"

PRESENTADO POR:

MARÍA DEL TRANSITO GUTIÉRREZ REYES

ALBA IDALIA CÓRDOVA CUÉLLAR

PARA OPTAR AL TÍTULO DE:

MASTER EN ESTADÍSTICA

ASESOR: Dr. JOSÉ NERYS FUNES TORRES

CIUDAD UNIVERSITARIA, JULIO DE 2003

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

MAESTRÍA EN ESTADÍSTICA
(SEGUNDA PROMOCIÓN)

# “DESARROLLO PROGRAMÁTICO DE LA ASIGNATURA DISEÑO DE EXPERIMENTOS”

PRESENTADO POR:

MARÍA DEL TRANSITO GUTIÉRREZ REYES

ALBA IDALIA CÓRDOVA CUÉLLAR

PARA OPTAR AL TÍTULO DE:

MASTER EN ESTADÍSTICA

ASESOR: Dr. JOSÉ NERYS FUNES TORRES ________________________

CIUDAD UNIVERSITARIA, JULIO DE 2003

UNIVERSIDAD DE EL SALVADOR

AUTORIDADES UNIVERSITARIAS

Dra. María Isabel Rodríguez
Rectora

Ing. José Francisco Marroquín
Vice-Rector Académico

Licda. Hortensia Dueñas de García
Vice-Rectora Administrativa

Licda. Lidia Margarita Muñoz Vela
Secretaria General

Lic. Pedro Rosalío Escobar
Fiscal

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

Licda. Leticia Noemí Paúl de Flores
Decana

Dr. Rafael Antonio Gómez Escoto
Vice-Decano

ESCUELA DE MATEMÁTICA

Lic. Alfredo Aguilar González
Director

# PRÓLOGO

Nadie ha tenido tanto impacto en el uso de los métodos estadísticos en el Diseño Experimental como el estadístico Sir Ronald Aylimer Fisher; nacido en Londres Inglaterra el 17 de Febrero de 1890 y murió en Adelaide Australia el 20 de Julio de 1962.

En octubre de 1919, Fisher fue contratado por la Estación Agrícola Experimental Rothamsted en Londres, Inglaterra para estar a cargo de la estadística y análisis de datos; por cerca de seis meses a un año, pero permaneció hasta el año de 1933, desarrollando y consolidando los principios básicos de Diseño y Análisis que hasta la fecha son prácticas necesarias para llegar a resultados de investigación válidos. Y fue Fisher el que uso por primera vez el análisis de varianza como herramienta primaria para el análisis estadístico en el Diseño Experimental. Por lo tanto, el Diseño experimental tuvo su inicio teórico a partir de 1935 por Sir Ronald Aylimer Fisher, quien sentó la base de la teoría del Diseño Experimental y que a la fecha se encuentra bastante desarrollada y ampliada.

Otros que han contribuido de manera significativa a las publicaciones sobre el Diseño Experimental, son: F. Yates, G.E.P. Box, R.C. Bose, O. Kempthorne y W. G. Cocran.

El Diseño Experimental es considerado dentro de la estadística como una rama; en la cual se desarrollan métodos para analizar procesos o sistemas.

Los métodos del Diseño Experimental juegan un papel importante en el desarrollo y depuración de procesos para mejorar el rendimiento. En muchos de los casos, el objetivo puede ser desarrollar un proceso consistente o robusto; esto significa un proceso que es afectado minimamente por fuentes de variabilidad externas que puedan intervenir en el proceso.

Los estadistas ven el Diseño Experimental como una herramienta muy eficaz y eficiente para obtener conclusiones válidas y objetivas de un proceso o sistema en particular.
Muchas de las primeras aplicaciones de los métodos del Diseño Experimental se dieron en el área de la agricultura y las ciencias biológicas, como resultado de ello, gran parte de la terminología usada proviene de estos antecedentes agrícolas.

En los años recientes en Estados Unidos hubo un interés muy grande por el Diseño Experimental, llegando así a utilizar experimentos diseñados por muchos años; y esto ha sido un factor importante en su éxito competitivo.

Los campos de aplicación de los métodos del Diseño Experimental son muy amplios y se utilizan en muchas disciplinas. Tal es así, que es posible considerar a la experimentación como parte del proceso científico y es una de las formas en las cuales aprendemos acerca de cómo funcionan los sistemas o procesos.

Las primeras aplicaciones se dieron en el área de la agricultura y ciencias biológicas; mientras que las primeras aplicaciones industriales se hicieron en la década de 1930, en la industria textil y de la lana británica.

Los métodos del Diseño Experimental también se introdujeron en las industrias químicas para el desarrollo de productos y procesos.

La industria de los semiconductores y la electrónica ha utilizado por muchos años y con considerable éxito los métodos Experimentales.

El experimento comparativo es el tipo de experimento que utilizan los investigadores en áreas como : **Biología, Medicina, Agricultura, Ingeniería, Psicología y otras Ciencias Experimentales**. En este experimento se establece más de un conjunto de circunstancias y se comparan entre si, las respuestas a las diferentes circunstancias.

En la ingeniería el Diseño Experimental es un medio de importancia crítica para mejorar el rendimiento del proceso de manufactura; también los métodos del Diseño Experimental son muy importantes, en los cuales se desarrollan nuevos productos y se mejoran otros ya existentes.

En estas áreas el uso del Diseño Experimental puede dar como resultado productos con mayor confiabilidad y mejor funcionamiento en el campo, menores costos y menor tiempo de diseño y desarrollo del producto.

Al aplicar las técnicas de un Diseño Experimental al inicio del desarrollo de un proceso, se pueden obtener los siguientes resultados:

1) Mejorar el rendimiento del proceso.
2) Menor variabilidad y mayor cumplimiento de los requerimientos u objetivos.
3) Menor tiempo de desarrollo.
4) Menores costos globales.

Considerando que el Diseño Experimental es una parte de la Estadística; la cual es aplicada en todas las Disciplinas Científicas Experimentales como Medicina, Agronomía, Química, Biología, Física, etc. Se pretende que el desarrollo de estos contenidos ayuden de mucho para llevar a cabo el Análisis y obtener conclusiones válidas de los diferentes Diseños Experimentales que se obtienen al llevar a cabo Experimentos en cada una de estas disciplinas. En el área experimental es muy importante la aplicación de los Diseños Experimentales; ya que ayudan a obtener conclusión válidas y además proporcionan una guía del principio al final del desarrollo del Experimento y dependiendo de lo que se quiera investigar; así será el Diseño a utilizar y el Análisis que se debe realizar.

También, con el desarrollo de estos contenidos se pretende proporcionar una guía sistematizada y un conocimiento más amplio de los Diseños Experimentales a los Estudiantes de la Carrera de Licenciatura en Estadística y Computación; ya que se observa una carencia de bibliografía en esta área de la Estadística para el desarrollo de los contenidos, a pesar de su gran importancia en la experimentación.

En el presente Desarrollo de estos contenidos se dará un enfoque teórico práctico de los Diseños Experimentales, ya que los investigadores realizan experimentos en todos los campos del saber, con el objetivo de descubrir algo acerca de un proceso o sistema en particular. También se pondrá en práctica la utilización de un software estadístico, con el objetivo de facilitar los cálculos matemáticos de cada uno de los diseños estudiados (que se describe al final de esta guía en detalle).

En esta guía programática se desarrollan cinco unidades. Iniciándose con un enfoque teórico de conceptos básicos que se utilizan en todo el desarrollo de los contenidos, se continua con el Análisis de los Experimentos más Simples que son los Diseños Unifactoriales llamados así; porque en ellos intervienen un sólo factor en el Experimento que es el de interés en el análisis. Luego se lleva a cabo el desarrollo del estudio de los Diseños por Bloques Completos y los Diseños por Bloques Incompletos; considerando en cada uno, otros Diseños que se estudian tomando en cuenta los mismos criterios de éstos.

En la siguiente Unidad Programática se desarrollan los contenidos de los Diseños Factoriales considerando los de dos factores, tres factores y una generalización del estudio de éstos (k factores). Por último, se da un enfoque del desarrollo de los Diseños Factoriales $2^k$ considerando los Diseños dos al cuadrado, dos al cubo y una generalización del estudio de lo que son los Diseños $2^k$.

## ÍNDICE GENERAL

# UNIDAD PROGRAMÁTICA I:

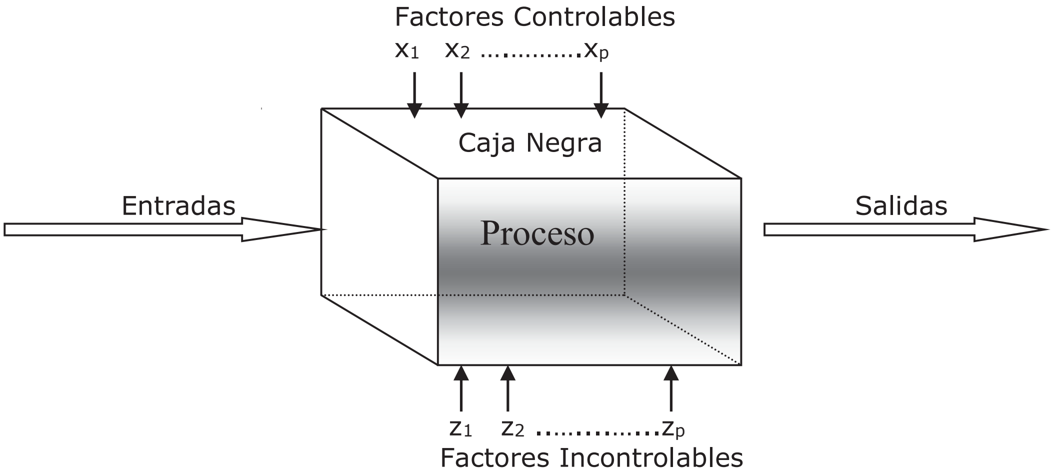
# "INTRODUCCIÓN A LOS DISEÑOS EXPERIMENTALES"

## 1. DEFINICIÓN DE UN DISEÑO EXPERIMENTAL.

**El diseño de un experimento** es la secuencia completa de los pasos que se deben tomar de antemano, para planear y asegurar la obtención de toda la información relevante y adecuada al problema bajo investigación, la cual será analizada estadísticamente para obtener conclusiones válidas y objetivas con respecto a los objetivos planteados.

**Un Diseño Experimental** es una prueba o serie de pruebas en las cuales existen cambios deliberados en las variables de entrada de un proceso o sistema, de tal manera que sea posible observar e identificar las causas de los cambios que se producen en la respuesta de salida.

Esquema de un proceso o sistema:




Un proceso suele visualizarse como una **Caja Negra** en donde existe una transformación de lo que entra al proceso, y que se observa en las salidas que produce.

Este proceso puede ser una combinación de máquinas, métodos, personas y otros recursos que transforman las entradas (a menudo un material) en las salidas que tienen una o más respuestas observables. Algunas de las variables del proceso digamos **$x_1$, $x_2$, ....,$x_p$ son controlables**, mientras que otras como **$z_1$, $z_2$, .....,$z_p$ son incontrolables** (no controlables). Cuando se realiza un diseño experimental es necesario tener en cuenta los siguientes **objetivos**:

1. Determinar cuales variables tienen mayor influencia en la respuesta o variable dependiente ( *y).*
2. Determinar el mejor valor de las (x) que influyen en (*y*), de modo que (*y)* tenga casi siempre un valor cercano al valor nominal deseado.

3. Determinar el mejor valor de las (x) que influyen en (*y*), de modo que la variabilidad de (*y*) sea pequeña.
4. Determinar el mejor valor de las (z) que influyen en (y), de modo que se minimicen los efectos de las variables incontrolables $z_1$, $z_2$,….,$z_p$.

## 2. PROPÓSITOS DE UN DISEÑO EXPERIMENTAL

El propósito de cualquier Diseño Experimental, es proporcionar una cantidad máxima de información pertinente al problema que se está investigando. Y ajustar el diseño que sea lo mas simple y efectivo; para ahorrar dinero, tiempo, personal y material experimental que se va ha utilizar. Es de acotar, que la mayoría de los diseños estadísticos simples, no sólo son fáciles de analizar, sino también son eficientes en el sentido económico y en el estadístico.

De lo anterior, se deduce que el diseño de un experimento es un proceso que explica tanto la **metodología estadística** como el **análisis económico**.

## 3. CONCEPTOS BÁSICOS

Los siguientes conceptos que se definen a continuación se utilizarán en el desarrollo de las unidades posteriores; los cuales fueron retomados de Douglas C. Montgomery , año 1991 y de Robert O. Kuehl , año 2001.

- **DISEÑO**: Consiste en planificar la forma de hacer el experimento, materiales y métodos a usar, etc.

- **EXPERIMENTO**: Conjunto de pruebas o ensayos cuyo objetivo es obtener información, que permita mejorar el producto o el proceso en estudio.

- **TRATAMIENTO**:
  - Es un conjunto particular de condiciones experimentales definidas por el investigador.
  - Son el conjunto de circunstancias creadas por el experimento, en respuesta a la hipótesis de investigación y son el centro de la misma.

- **FACTOR**: Es un grupo específico de tratamientos. (Ejemplo, Temperatura, humedad, tipos de suelos, etc.).

- **NIVELES DEL FACTOR**: Son diversas categorías de un factor. ( Por ejemplo, los niveles de temperatura son 20°C, 30°C, etc.). Un factor Cuantitativo tiene niveles asociados con puntos ordenados en alguna escala de medición, como temperatura; mientras que los niveles de un factor cualitativo representan distintas categorías o clasificaciones, como tipo de suelo, que no se puede acomodar conforme a alguna magnitud.

- **RÉPLICA**: Son las repeticiones que se realizan del experimento básico.

- **UNIDAD EXPERIMENTAL**:
  - Es el material experimental unitario que recibe la aplicación de un tratamiento.
  - Es la entidad física o el sujeto expuesto al tratamiento independientemente de las otras unidades. La unidad experimental una vez expuesta al tratamiento constituye una sola réplica del tratamiento.
  - Es el objeto o espacio al cual se aplica el tratamiento y donde se mide y analiza la variable que se investiga.
  - Es el elemento que se está estudiando.

- **UNIDAD MUESTRAL**: Es una fracción de la unidad experimental que se utiliza para medir el efecto de un tratamiento.

- **ERROR EXPERIMENTAL**: Es una medida de variación que existe entre dos o más unidades experimentales, que han recibido la aplicación de un mismo tratamiento de manera idéntica e independiente.

- **FACTORES CONTROLABLES**: Son aquellos parámetros o características del producto o proceso, para los cuales se prueban distintas variables o valores con el fin de estudiar cómo influyen sobre los resultados.

- **FACTORES INCONTROLABLES:** Son aquellos parámetros o características del producto o proceso, que es imposible de controlar al momento de desarrollar el experimento.

- **VARIABILIDAD NATURAL:** es la variación entre las unidades experimentales, que el experimentador no puede controlar ni eliminar.

- **VARIABLE DEPENDIENTE**: es la variable que se desea examinar o estudiar en un experimento. (Variable Respuesta).

- **HIPÓTESIS:**
  - Es una suposición o conjetura que se plantea el investigador de una realidad desconocida.
  - Es el supuesto que se hace sobre el valor de un parámetro (constante que caracteriza a una población) el cual puede ser validado mediante una prueba estadística.

## 4. TIPOS DE TRATAMIENTOS

A continuación se presentan ejemplos de tratamientos en algunas áreas, tales como:

1) **Experimentaciones Agrícolas, un tratamiento puede referirse a:**
   - Marca de Fertilizante.
   - Cantidad de Fertilizante.
   - Profundidad del Sembrado.
   - Variedad de Semilla.
   - Combinación de Cantidad de Fertilizante y Profundidad de Sembrado; esto es una combinación de tratamientos.
   - etc.

2) **Experimentaciones de Nutrición Animal, un tratamiento puede referirse a:**
   - Cría de Ganado Lanar
   - Sexo de los Animales
   - Padre del Animal Experimental
   - Tipo de Alimento
   - Ración Particular de Alimento de un Animal.
   - Raza del Animal
   - etc.

3) **Estudios Psicológicos y Sociológicos, un tratamiento puede referirse a:**
   - Edad
   - Sexo
   - Grado de Educación
   - Estatura
   - etc.

4) **En una investigación de los efectos de varios Factores en la eficiencia del lavado de ropa en casa, los tratamientos pueden ser varias combinaciones:**
   - Tipo de Ropa (dura y suave)
   - Temperatura del Agua
   - Tipo de Detergente
   - Duración del tiempo de Lavado
   - Tipo de Lavadora
   - Duración del Agente Limpiador
   - etc.

5) **En un Experimento para estudiar el Rendimiento de cierto Proceso químico, los tratamientos pueden ser todas las combinaciones de:**
   - La temperatura a la cual se ejecuta el Proceso
   - La cantidad de Catalizador Usada
   - etc.

6) **En un estudio de investigación y desarrollo concerniente a Baterías, los tratamientos podrían ser varias combinaciones:**
   - La cantidad de Electrolito
   - La Temperatura a la cual fue Activada la Batería
   - etc.

Es muy importante que cuando se elijan los tratamientos, éstos deben dar respuesta a una hipótesis de investigación. La hipótesis de investigación establece un conjunto de circunstancias y sus consecuencias. Los tratamientos deben ser una creación de las circunstancias para el experimento. Así, es necesario identificar los tratamientos con el papel que cada uno tiene en la evaluación de la hipótesis de investigación. Por lo tanto, el investigador debe asegurarse que los tratamientos elegidos concuerden con la hipótesis de investigación.

A continuación se presentan algunos experimentos reales; planteándose las hipótesis de investigación de cada uno de ellos y sus respectivos tratamientos, que dan respuesta a dicha hipótesis.

- **La hipótesis es:** La velocidad del tránsito depende del ancho de los carriles en las calles.

  **Para responder a esta hipótesis,** los tratamientos se deben definir seleccionando carriles con diferente anchura y se mide la velocidad de los automóviles en cada uno de ellos.

- **La hipótesis es:** La reproducción de los microbios del suelo depende de las condiciones de humedad.

  **Para responder a esta hipótesis**, se establecen tratamientos con distintos niveles de humedad para medir la reproducción de los microbios.

- **La hipótesis es:** El método para medir retrasos del tránsito depende del tipo de configuración usada en la señalización.

  **Para responder a esta hipótesis**, los tratamientos deben ser en relación a la evaluación de varios métodos para medir los retrasos del tránsito en intersecciones con diferentes tipos de configuraciones en los semáforos.

- **La hipótesis es**: Ciertas características demográficas familiares afectan de manera favorable el desarrollo de un niño.

  **Para responder a esta hipótesis,** los tratamientos deben ser en relación con el desarrollo de la adaptación social en niños pequeños según su relación con: 1) Educación de los padres, 2) Ingreso de los padres, 3) Estructura familiar y 4) Edad del niño.

- **La hipótesis es**: La energía requerida al reunir comida para la colonia de las abejas productoras de miel es independiente de la temperatura.

  **Para responder a esta hipótesis**, los tratamientos deben ser en relación al estudio de la cinética de bebida de las abejas productoras de miel a diferentes temperaturas ambientales.

- **La hipótesis es**: La temperatura ambiental en la cual las baterías son activadas altera su vida útil.

  **Para responder a esta hipótesis,** el tratamiento será temperatura y se debe probar un número determinado de baterías a diferentes niveles de temperatura.

## 5. TIPOS DE ERROR EXPERIMENTAL

Todos los experimentos están sujetos a posibles errores; los cuales se pueden disminuir pero no controlar totalmente y pueden ser de los siguientes tipos:

a) Errores de experimentación.
b) Errores de observación.
c) Errores de medición.
d) Variación natural en las unidades experimentales.
e) La interacción de los tratamientos con las unidades experimentales.
f) Factores extraños que pueden influir en las características de la investigación.

**Para reducir el Error Experimental se debe tomar en cuenta:**

a) Usar material experimental más homogéneo.
b) Usar información proporcionada por variables aleatorias relacionadas con la variable respuesta de interés.
c) Tener cuidado de dirigir el experimento.
d) Usar un número óptimo de replicaciones.
e) Utilizar un diseño experimental más eficiente.

La metodología estadística y un planeamiento adecuado del experimento permitirá reducir el error experimental, y esto hará posible que sean detectables las diferencias significativas entre los tratamientos que han sido usados. Si el investigador no hace el planeamiento del experimento en forma correcta, el error experimental del experimento va a ser grande y no se podrá detectar las diferencias significativas en el experimento. Por lo tanto, el investigador concluirá de manera equivocada que todos los tratamientos tienen el mismo efecto y que ningún tratamiento es ni mejor ni peor que los otros tratamientos.

## 6. PRINCIPIOS BÁSICOS

**Los tres principios básicos del Diseño de un Experimento son:**

1. **Replicación (Obtención de Réplicas).**
   Este principio se refiere al número de veces que se aplica un tratamiento a las unidades experimentales. El cual tiene dos propiedades importantes, la **primera** permite al experimentador obtener la estimación del error experimental; esta estimación se convierte en la unidad básica para determinar si las diferencias observadas en los datos

son estadísticamente significativas o para determinar la amplitud de un intervalo de confianza, y la **segunda** permite al experimentador calcular una estimación más precisa del efecto medio de cualquier factor en el experimento, si se usa la media de la muestra, como una estimación de dicho efecto. Lo que significa que la varianza de la media de la muestra se define como $\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$ ; donde $\sigma^2$ es la varianza de los datos y n el número de réplicas.

La implicación práctica de esto es que si el número de réplicas es pequeña (n=1) probablemente no se podría obtener inferencias satisfactorias con respecto al efecto del tratamiento; es decir, que la diferencia observada podría ser resultado, exclusivamente, del error experimental. El número de réplicas afecta la precisión de las estimaciones de las medias de tratamientos y la potencia de las pruebas estadísticas para detectar las diferencias entre las medias de los grupos en los tratamientos. Pero puede ser muy costoso económicamente la incorporación de una réplica en el Experimento.

**Las razones más notables para hacer réplicas en un Experimento son:**

1) Se demuestra que se puede reproducir los resultados, al menos bajo las condiciones experimentales actuales.
2) Proporciona cierto grado de seguridad contra resultados anormales en el experimento, debido a accidentes no previstos.
3) Proporciona las medias para estimar la varianza del error experimental. A un cuando la experimentación previa proporcione estimaciones de la varianza, la estimación a partir del experimento en curso puede ser más exacta porque refleja el comportamiento actual de las observaciones.
4) Proporciona la posibilidad de aumentar la precisión en la estimación de las medias de los tratamientos. Al incrementar las réplicas (n), disminuye la varianza muestral que esta definida como $s_{\bar{y}}^2 = \frac{s^2}{n}$, donde $s^2$ es la varianza de la muestra y n el número de réplicas, lo que aumenta la precisión de $\bar{y}$.

En conclusión se puede decir, que el número de réplicas dependerá que tan costoso sea económicamente el introducir una nueva réplica en el experimento; pero se recomienda que sea un número lo más razonablemente posible; para poder así obtener inferencias estadísticas satisfactorias con respecto al efecto de los tratamientos que se están estudiando. Entonces, el número de réplicas está determinado por las restricciones que se pueden asignar al problema.

2. **Aleatorización.**

La aleatorización garantiza el uso de los métodos estadísticos de Diseños de Experimentos, y consiste en el método por el cual las unidades experimentales reciben las aplicaciones de los tratamientos en forma aleatoria; es decir, que tanto la asignación del material experimental como el orden en que se realizan las pruebas individuales o ensayos se determinan aleatoriamente.

Al realizar la aleatorización adecuadamente en el experimento, se ayuda a "cancelar" los efectos de factores extraños que pudieran estar presentes; además simula el efecto de independencia y permite proceder como si las observaciones fueran independientes y con distribución normal; es decir, asegura el cumplimiento del supuesto de independencia del análisis de varianza. El método de aleatorización depende del Diseño de Experimento que será usado.

Para llevar a cabo la aleatorización de los tratamientos, se puede utilizar un programa de computadora o una tabla de números aleatorios; pero si no se cuenta con ello, se puede utilizar el siguiente método:

1) Asignar números a las unidades experimentales.
2) Elaboran tarjetas de papel con los mismos números de las unidades experimentales y colocarse en un recipiente.
3) Sacar al azar una por una las tarjetas del recipiente.
4) Los primeros $n$ números son las unidades experimentales asignadas al primer tratamiento.
5) Los segundos $n$ números corresponden a las unidades experimentales que se asignan al segundo tratamiento. Y así sucesivamente, hasta sacar todas las tarjetas de $n$ en $n$.

3. **Control Local:** Consiste en el uso de técnicas de bloqueo, balanceo y agrupamiento de las unidades experimentales para asegurar que el diseño usado sea eficiente; ya que los objetivos de la mayoría de los experimentos son las comparaciones claras y exactas entre los tratamientos a través de un conjunto apropiado de condiciones. Estos objetivos requieren estimaciones precisas de las medias y poderosas pruebas estadísticas, lo cual se puede obtener reduciendo la varianza del error experimental. El uso adecuado del control local describe las acciones que emplea un investigador para reducir o controlar la magnitud de la estimación del error experimental; incrementando la exactitud de las observaciones y estableciendo la base de la inferencia del estudio.

En este principio se hacen las comparaciones entre las condiciones de interés del experimento dentro de cada bloque.

**Las técnicas utilizadas en este principio se describen a continuación:**

- **El Bloqueo**

La bloquización proporciona control local del ambiente para reducir la variabilidad natural. Las unidades experimentales se distribuyen en grupos de unidades similares, con base en un factor o factores que se espera o se sabe que tienen alguna relación con la variable respuesta o con la medición que se supone responde de manera diferente a los diversos tratamientos. Es decir, que consiste en la distribución de las unidades experimentales en bloques de tal manera que las unidades dentro de un bloque sean relativamente homogéneas; ya que unidades experimentales heterogéneas producen valores grandes en la varianza del error experimental, es así que la mayor parte de la variación predecible entre las unidades queda confundida con el efecto de los bloques.

**Los cuatro criterios que se usan con más frecuencia para llevar a cabo el bloque en las unidades experimentales son:**

1) Proximidad (parcelas vecinas).
2) Características Físicas (edad o peso).
3) Tiempo (Tiempo de desarrollo).
4) Administración de tareas en el experimento.

- **Por Balanceo**

Es el bloqueo y la asignación de los tratamientos a las unidades experimentales de modo que resulte una configuración balanceada. La comparación precisa entre los tratamientos requiere la selección de unidades experimentales uniformes para reducir el error experimental. La naturaleza del experimento señala el equilibrio entre la variedad de las condiciones y la uniformidad de las unidades experimentales. Por ejemplo, si se trata de un experimento con vacas lecheras, la uniformidad de las unidades experimentales requiere elegir vacas de la misma cría, en la misma etapa de lactancia y con un número similar de lactancia.

- **Por Agrupamiento**

  Es la colocación de un conjunto de unidades experimentales homogéneas en grupos, de modo que los diferentes grupos puedan sujetarse a distintos tratamientos. Estos grupos pueden constar de diferente número de unidades experimentales.

En los tres principios analizados anteriormente el objetivo principal es disminuir en gran medida la Variabilidad Natural o error experimental, a continuación se presenta un ejemplo en el cual se evidencia este objetivo.

**Ejemplo**

Se hace una investigación sobre el efecto de administrar 10 mg. de vitamina B12 por libra de ración a cerdos en crecimiento, se tomaron ocho lotes de seis cerdos, cada uno tratados por pares. Los lotes se separaron por la administración de diferentes niveles de aureomicina. Se mide el aumento diario promedio del peso de tres cerdos (libras).

**Tratamientos** : Sin B12
Con B12
**Unidades experimentales**: Cerdos
**Variable Respuesta** : Aumento de peso.

Para llevar a cabo este experimento se deben agrupar los cerdos de la misma raza, edad y sexo, de forma aleatoria; ya que estas tres situaciones afectan significativamente en el peso de los cerdos. Y es así como se obtiene una muestra lo más homogénea posible, y que en el experimento sólo intervenga la variabilidad natural, reduciendo así el error experimental.

## 7. EJEMPLO DE UN DISEÑO EXPERIMENTAL.

Se llevó a cabo un experimento para determinar la eficacia de 6 fertilizantes de nitrógeno para una cierta variedad de maíz. Se contaba con 24 parcelas experimentales. Considerando que puede existir mucha variabilidad entre las parcelas experimentales, se decidió usar un diseño de experimento que pudiera tener la capacidad de controlar esta variabilidad. Cada uno de los seis fertilizantes fue aplicado a cuatro parcelas experimentales, siguiendo el método de aleatorización del diseño utilizado y cada parcela experimental tenía cinco surcos de plantas de maíz. Luego se obtuvo la cosecha de plantas, de cada una de las parcelas se tomaron solamente tres surcos y fueron los centrales. Las plantas cosechadas se llevaron al laboratorio para determinar el rendimiento por medio del peso de las semillas, haciendo esto separadamente para cada una de las parcelas.

## COMENTARIOS DEL EJEMPLO

**En el ejemplo planteado se puede observar:**

- Hay seis tratamientos que son los seis fertilizantes.
- Hay veinticuatro unidades experimentales, que son las parcelas experimentales.
- La unidad muestral no es la totalidad de la unidad experimental sino una parte de ella (las 4 parcelas).
- El investigador toma la decisión de cosechar tres surcos centrales en cada unidad experimental; ya que considera que de esta manera se puede evitar cualquier efecto del fertilizante que se aplica a una parcela y que pueda influir el resultado de las parcelas vecinas.
- El número de replicaciones (réplicas) es igual a cuatro por cada tratamiento.
- Existe un control local ya que el investigador habrá usado un diseño (por ejemplo: El Diseño de Bloques Aleatorios; el cual se estudiara en el detalle en las siguientes unidades programáticas), que controla la variabilidad entre las parcelas en el campo experimental.
- La variable respuesta en este experimento es el rendimiento.

## 8. FORMA DE ALEATORIZAR ESTE EXPERIMENTO

1) Asignar números a cada una de las parcelas experimentales de 1 a 24.
2) Elaboran tarjetas de papel con los mismos números de las parcelas (1 a 24) y luego colocarlos en un recipiente.
3) Sacar al azar una por una las tarjetas del recipiente.
4) Existen 6 tratamientos, que son los seis fertilizantes y como el número de réplicas es igual a cuatro entonces los primeros cuatro números sacados que corresponden a las primeras cuatro parcelas serán asignadas al fertilizante número uno.
5) Los segundos cuatro números sacados que corresponden a las segundas cuatro parcelas serán asignadas al fertilizante número dos.
6) Y así sucesivamente hasta obtener los últimos cuatro números que corresponden a las cuatro parcelas que serán asignadas al fertilizante número seis.

En general, un experimento de este tipo puede tener simultáneamente otras variables respuestas como por ejemplo: altura de plantas, grosor de las plantas, determinación del contenido de humedad de los granos, etc. Pero en el análisis del experimento el rendimiento es la variable de interés para el investigador.

A pesar de haber tomado todas las precauciones necesarias en la conducción del experimento, se podrá decir que siempre existirá el Error Experimental en cualquier experimento, no importa que tan bien sea planteado y conducido el experimento. Basta con observar y comparar los valores del rendimiento para dos ó más parcelas que han recibido la aplicación de un mismo fertilizante. Estos valores no serán iguales y por lo tanto el error experimental no es nulo y existe.

Algunas de las razones por las cuales puede surgir el **Error Experimental** en este experimento son las siguientes:

1) Las parcelas experimentales en el campo deben tener variación en la fertilidad del suelo, textura del suelo, Ph del suelo, pendiente, la cantidad de luz solar que puede recibir cada planta, etc.
2) El número total de plantas por cada parcela podría no ser igual. Esto puede ocurrir por defectos en la calidad de las semillas y el método de siembra utilizado.
3) Puede existir pérdida del material experimental cosechado que se lleva al laboratorio para determinar el peso.
4) Puede existir limitación y defectos en la máquina que se usa para determinar el peso del material que se ha cosechado.
5) Puede existir variación de criterios y técnicas que usan diferentes personas que han trabajado en la conducción del experimento.

## 9. DIRECTRICES O PROCEDIMIENTO PARA EL DISEÑO EXPERIMENTAL.

Se considera necesario que todos los participantes que realizan un enfoque estadístico en el cual se diseña y analiza un experimento tengan de antemano una idea clara de lo qué es exactamente lo que se va a estudiar; es decir, como se van ha recopilar los datos, y como se van a analizar.

**A continuación se presenta una guía del procedimiento a seguir; descrita en el libro de Diseño y Análisis de Experimentos de Douglas C. Mongomery:**

**Primero: Reconocimiento y Planeamiento del problema.**

Es necesario desarrollar todas las ideas sobre los objetivos del experimento. Una clara comprensión y planteamiento del problema con frecuencia contribuye sustancialmente a un mayor entendimiento del fenómeno y a la solución final del problema.

Suele ser importante solicitar la opinión de todas las partes implicadas; ya que normalmente saben mucho del problema.

**Segundo: Elección de Factores y Niveles.**

Se deben seleccionar los factores que van a ser investigados en el experimento, los intervalos de variación y los niveles específicos a los cuales se hará el experimento. Además debe considerarse la forma en que se controlarán estos factores para mantenerlos en los valores deseados y como se les medirá.

**Tercero: Selección de la Variable Respuesta.**

El experimentador debe seleccionar la variable respuesta o variable dependiente de tal forma que esté seguro que la respuesta, que se va a medir, realmente proporcione información útil a cerca del problema en estudio.

Las respuestas en un problema pueden  ser múltiples y además con mayor frecuencia, el promedio o la desviación estándar (o ambas) de la característica medida serán la variable respuesta. Por ejemplo, en el experimento de eficacia de seis fertilizantes de nitrógeno para cierta variedad de maíz (analizado anteriormente), la variable respuesta es el rendimiento. En general, un experimento de este tipo puede tener simultáneamente otras variables respuestas; como por ejemplo altura de plantas, determinación del contenido de humedad de las semillas, color de las semillas, etc. Pero en este momento se habla solamente del rendimiento, el cual se considera que es la variable de interés mas importante para el investigador.

**Cuarto: Elección del Diseño Experimental.**

Para definir o determinar el diseño experimental a utilizar se debe considerar el tamaño muestral (número de repeticiones), seleccionar un orden adecuado para los ensayos experimentales y determinar si hay implicaciones de bloque u otras restricciones de aleatorización. Además tener presente los objetivos experimentales.

El investigador debe decidir qué constituye una unidad experimental, cuántas réplicas de las unidades experimentales exige cada tratamiento y qué tratamiento asignar a cada una de ellas. También, debe determinar si agrupará por bloques las unidades experimentales en grupos homogéneos para controlar el error experimental.

Con base al Diseño Experimental debe proponerse un Modelo Matemático adecuado.

**Quinto: Ejecución del Experimento.**

En esta fase se lleva a cabo la recolección de los datos. Se debe observar cuidadosamente el proceso para asegurar que todo se realice conforme lo planteado; ya que los errores en el procedimiento suelen anular la validez del experimento.

**Sexto: Análisis de los Datos.**

Se deben utilizar métodos estadísticos para analizar los datos, para que los resultados y conclusiones sean objetivos más que apreciativos. Si todo se ha realizado correctamente los métodos estadísticos que se necesitan no son complicados.

El análisis de residuos y la verificación de la idoneidad del modelo son también técnicas de análisis de gran utilidad.

**Séptimo: Conclusiones y Recomendaciones.**

Consiste en la interpretación de las inferencias estadísticas. Y para llevar a cabo la presentación de los resultados son muy útiles los métodos gráficos, en especial cuando se presentan a otros personas. Es bien importante también realizar corridas de seguimiento y pruebas de confirmación para llevar a cabo la validación de las conclusiones del experimento.

**El procedimiento a seguir para el Diseño y Análisis de un Experimento descrito por otros autores es el siguiente: Descrito en Unidad Didáctica/Diseño de Experimentos y Teoría de Muestras, preparado por Miguel Martín Dávila (Universidad Nacional de Educación a Distancia) y otros libros de Diseño de Experimentos.**

**a) Definición del Experimento.**

1) Definir el problema.
2) Seleccionar las variables aleatorias dependientes e independientes.
3) Selección de los factores.
4) Elección de los niveles de estos factores. Estos niveles pueden ser:
    i) Cuantitativos o Cualitativos.
    ii) Fijos o Aleatorios.
5) Determinar la forma de combinar los niveles de dichos factores.

**b) Diseño.**

1) Número de observaciones a tomar.
2) Orden del experimento.
3) Método de aleatorización a utilizar.
4) Obtener un modelo matemático que represente el experimento a realizarse.

**c) Análisis.**

1) Colección y reducción de la información.
2) Cálculo de pruebas estadísticas.
3) Interpretación de resultados por el experimentador.

## 10. UTILIZACIÓN DE LOS MÉTODOS ESTADÍSTICOS EN LA EXPERIMENTACIÓN.

La mayoría de las investigaciones que se realizan en el campo de la ingeniería, ciencia e industria es empírica y utiliza mucho la experimentación.

El uso de los métodos estadísticos puede incrementar la eficiencia de los experimentos y, ayudar a justificar las conclusiones que se obtienen.

La utilización de las técnicas estadísticas en la experimentación requiere que el investigador considere los siguientes puntos:

**a) Uso del conocimiento no estadístico del problema.**

Se debe tomar en cuenta que los investigadores conocen a fondo su campo de especialidad; ya sea porque tienen una considerable experiencia práctica o una formación académica. Muchas veces se puede utilizar una gran cantidad de teoría para explicar las relaciones que hay entre los factores y la variable respuesta.

Este tipo de conocimiento no estadístico se debe tomar en cuenta para elegir los factores y las respuestas, también al decidir el número de réplicas que se quieren realizar, al analizar los datos, etc. Es por tanto que la estadística no puede sustituir el hecho de reflexionar sobre el problema.

**b) Mantener el Diseño y el Análisis tan simple como sea posible.**

Casi siempre, lo más adecuado son los métodos de diseño y análisis estadístico más simples. Por lo tanto, es recomendable el uso de técnicas estadísticas poco complejas y muy refinadas. Si se realiza el diseño cuidadosamente y correctamente, el análisis se espera que sea relativamente sencillo. Sin embargo, es poco probable que aun la estadística más compleja y elegante corrija la situación si se ha actuado indebidamente en la elaboración del diseño.

**c) Reconocer la diferencia entre la significación práctica y estadística.**

No hay seguridad de que una diferencia sea suficientemente grande, desde el punto de vista práctico, por el sólo hecho de que dos condiciones experimentales producen respuestas medias, estadísticamente diferentes. Por ejemplo, un ingeniero puede determinar que una modificación en el sistema de inyección de gasolina de un automóvil mejora el rendimiento medio en un 0.1mi/gal. Éste es un resultado estadísticamente significativo. Sin embargo, esta diferencia es demasiado pequeña desde el punto de vista práctico si el costo de la modificación es de 1,000 dólares .

**d) Usualmente los experimentos son iterativos.**

En las primeras etapas de un estudio no es conveniente diseñar experimentos demasiado extensos; ya que sólo se requiere que se conozcan los factores importantes, los intervalos en que estos factores van a ser investigados, el número apropiado de niveles para cada factor y las unidades de medición adecuadas a cada factor y la respuesta. Por lo general, al principio de un experimento no se está en capacidad de definir estos aspectos, pero es posible conocerlos a medida que se avanza la experimentación. Esto favorece al empleo del enfoque iterativo o secuencial; pero por regla general, la mayoría de los experimentos son iterativos.

## 11. IMPORTANCIA DEL ANÁLISIS DE VARIANZA.

En el caso que nos encontremos con experimentos en donde hay que realizar varias pruebas de hipótesis a la vez, y se trabaje con el mismo nivel de confianza ($\alpha$); es decir, aquellos experimentos en los cuales es necesario hacer la comparación de más de dos tratamientos simultáneos, podría utilizarse **PRUEBA DE HIPÓTESIS MÚLTIPLES** (Comparación por pares), pero es recomendable aplicar el **ANÁLISIS DE VARIANZA;** que es la técnica estadística que sirve para analizar la variación total de los resultados experimentales de un diseño en particular, descomponiéndolo en fuentes de variación independientes atribuibles a cada uno de los efectos en que se constituye el diseño experimental**.** Esta técnica tiene como objetivo identificar la importancia de los diferentes factores o tratamientos en estudio y determinar como interactúan entre sí.

Al llevar a cabo la prueba de hipótesis pueden cometerse dos tipos de errores, que son:

a) **Error tipo I**: Se da cuando la hipótesis nula ($H_0$) es rechazada siendo verdadera.
b) **Error tipo II**: Se comete cuando la hipótesis nula ($H_0$) no es rechazada siendo falsa.

Las probabilidades de cometer estos tipos de errores generalmente se denotan por:

$\alpha$ **=** P(Error tipo I)

$\beta$ = P(Error tipo II)

En el siguiente cuadro se presentan las diferentes situaciones que se pueden dar con la hipótesis nula ($H_0$).

| Decisión | $H_0$ es cierta | $H_0$ es Falsa |
|---|---|---|
| Aceptar $H_0$ | Decisión Correcta | Error tipo II ($\beta$) |
| Rechazar $H_0$ | Error tipo I ($\alpha$) | Decisión Correcta |

La utilización del análisis de varianza justifica la disminución de la **Probabilidad de Cometer el Error Tipo I** en el experimento.

**Por ejemplo**: Supongamos que se desea probar la igualdad de cinco medias usando la prueba de hipótesis múltiple.

**Las hipótesis a probar son:**

**$H_0$**: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

**$H_1$**: $\mu_1 \# \mu_2$ $\mu_2 \# \mu_3$ $\mu_3 \# \mu_5$

$\mu_1 \# \mu_3$ $\mu_2 \# \mu_4$ $\mu_4 \# \mu_5$

$\mu_1 \# \mu_4$ $\mu_2 \# \mu_5$

$\mu_1 \# \mu_5$ $\mu_3 \# \mu_4$

Como se puede observar el número de comparaciones de $H_1$, es: $\binom{5}{2} = \frac{5!}{3!2!} = 10$.

Es decir existen 10 posibles pares de medias, y si la probabilidad de aceptar correctamente la hipótesis nula ($H_0$) en cada una de las **Pruebas Individuales** es $1 - \alpha = 0.95$ (5 % de probabilidad de rechazar $H_0$), entonces la probabilidad de aceptar correctamente la hipótesis nula en las **10 pruebas** es $(0.95)^{10} = 0.6$ (40% de probabilidad para rechazar $H_0$) si estas son independientes. Es así como se produce un incremento sustancial del **error tipo I**, al utilizar la prueba de Hipótesis múltiple.

Por lo tanto, el procedimiento apropiado para probar la igualdad de varias medias es el **Análisis de Varianza.** Probablemente esta es la técnica más útil en el campo de la inferencia estadística.

# UNIDAD PROGRAMÁTICA II:

# “ DISEÑOS UNIFACTORIALES ”

# DISEÑO UNIFACTORIAL

## 1. DESCRIPCIÓN

En el análisis de los resultados de los experimentos se pueden observar diferentes aplicaciones de los Diseños Experimentales. Hay experimentos muy útiles en los cuales existe un sólo factor de interés; el cual se analiza por medio de la comparación de dos condiciones que intervienen en el Experimento (a menudo llamadas tratamientos o niveles del factor); a este tipo de experimentos se le denomina **Experimentos de Comparación Simple**, estudiados en los cursos de Estadística básica. El análisis de los datos de este tipo de Experimentos resulta ser sencillo, ya que se utilizan técnicas de la Inferencia Estadística, llamada **Prueba de Hipótesis** (o pruebas de significancia) que son las que ayudan al experimentador a comparar estas condiciones.

Si en el tipo de Diseño Experimental planteado anteriormente se requiere más de dos niveles del factor que se analiza, éstos son considerados como **"Diseños Unifactoriales"**. Teniendo en cuenta que para el análisis de éstos se utiliza el Análisis de Varianza, ya que se requiere probar la igualdad de varias medias, la cual se explicará posteriormente.

En los experimentos de los Diseños Unifactoriales, el número de observaciones recolectadas en cada tratamiento pueden ser iguales o diferentes. Cuando el número de observaciones sea diferente se dice que el Diseño está **Desequilibrado o Desbalanceado;** en caso contrario el Diseño está **Equilibrado o Balanceado**.

**Ejemplo 1**

Se sospecha que la temperatura ambiental en la cual las baterías son activadas, altera su vida útil. Treinta baterías homogéneas fueron probadas, seis en cada una de 5 temperaturas. La vida útil de las baterías se mide en segundos.

**Interpretación del Ejemplo 1**

El ejemplo anterior es considerado como un Diseño de Experimentos de un sólo factor, con 5 niveles del factor. El factor en estudio es la temperatura, en el cual intervienen 5 niveles del factor que son los diferentes valores de la temperatura en que son probadas las treinta baterías; esto significa que para cada valor de la temperatura se deberán probar seis baterías que es el número de observaciones o réplicas que se aplican a cada tratamiento. Se puede observar que para cada temperatura se aplica el mismo número de réplicas (seis); por lo tanto, es un Diseño Balanceado.

El análisis de la información obtenida en este tipo de Diseño de Experimentos se debe realizar utilizando el **" Análisis de Varianza"**.

Si en este ejemplo sólo probamos dos valores de temperaturas, manteniendo el número de réplicas para cada uno de los valores de la temperatura entonces es considerado un "Diseño de Comparación Simple", ya que existe un sólo factor y dos niveles del factor; por tanto, el análisis de la información en este tipo de Diseño de Experimentos se debe realizar utilizando la prueba de hipótesis.

## 2. REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

Bajo el supuesto que se tienen "*a*" niveles o "*a*" tratamientos de un único factor, la respuesta que se observa en cada uno de los "*a*" tratamientos es una variable aleatoria. La representación típica de los datos para un experimento Unifactorial, se presenta a continuación:

| Tratamiento (Nivel) | Observaciones | | | Totales ($y_{i.}$) | Promedios ($\bar{y}_{i.}$) |
|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ ………….. | $y_{1n}$ | $y_{1.}$ | $\bar{y}_{1.}$ |
| 2 | $y_{21}$ | $y_{22}$ ………….. | $y_{2n}$ | $y_{2.}$ | $\bar{y}_{2.}$ |
| 3 | $y_{31}$ | $y_{32}$ ……… | $y_{3n}$ | $y_{3.}$ | $\bar{y}_{3.}$ |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| a | $y_{a1}$ | $y_{a2}$ ……… | $y_{an}$ | $y_{a.}$ | $\bar{y}_{a.}$ |
| | | | | $y_{..}$ | $\bar{y}_{..}$ |

Un dato en la tabla, por ejemplo $y_{ij}$ , representa la $j$-ésima observación tomada bajo el tratamiento $i$. En general habrán, $n$ observaciones en el tratamiento $i$. En la tabla se encuentra considerado el caso en que hay un número igual de observaciones, n en cada tratamiento.

Sea $y_{i.}$ la representación del total de las observaciones bajo el tratamiento $i$-ésimo y $\bar{y}_{i.}$ la representación del promedio de las observaciones bajo el tratamiento i-ésimo.

De manera similar considérese que $y_{..}$ representa el gran total de todas la observaciones y, $\bar{y}_{..}$ la gran media de todas las observaciones. Matemáticamente se expresan como sigue:

**Caso Balanceado.**

$$y_{i.} = \sum_{j=1}^{n} y_{ij} \quad , \quad y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} \quad \text{ó} \quad y_{..} = \sum_{i=1}^{a} y_{i.} \quad , \quad \bar{y}_{i.} = \frac{y_{i.}}{n} \quad , \quad \bar{y}_{..} = \frac{y_{..}}{N}$$

$i$ = 1,2,.....,$a$

$j$ = 1,2,.....,$n$

siendo $a$ : número de tratamientos

$n$ : número de observaciones

$N = an$, número total de observaciones.

**Caso Desbalanceado.**

$$y_{i.} = \sum_{j=1}^{n_i} y_{ij} \quad , \quad y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij} \quad \text{ó} \quad y_{..} = \sum_{i=1}^{a} y_{i.} \quad , \quad \bar{y}_{i.} = \frac{y_{i.}}{n_i} \quad , \quad \bar{y}_{..} = \frac{y_{..}}{N}$$

i = 1,2,.....,a

j = 1,2,.....,$n_i$

siendo $n_i$ : número de observación del tratamiento i,

$$N = \sum_{i=1}^{a} n_i$$

En las expresiones anteriores la notación del subíndice "punto", representa la sumatoria del subíndice que él reemplaza.

## 3. MODELO ESTADÍSTICO

Sea "y" la variable que se va a medir en las distintas unidades experimentales y $y_{ij}$ el valor de la j-ésima observación del tratamiento i.

Se pueden describir las observaciones de la tabla anterior por medio del siguiente Modelo Estadístico Lineal:

$$y_{ij} = \mu + \iota_i + \varepsilon_{ij} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$

donde:

$y_{ij}$ : Es la observación ij-ésima

$\mu$ : Es un parámetro común para todos los tratamientos, llamado media general.

$\iota_i$ : Es un parámetro asociado con el tratamiento i-ésimo denominado efecto del tratamiento i-ésimo.

$\varepsilon_{ij}$ : Es la componente del error aleatorio (variabilidad natural).

i : Variando de 1 hasta el número de Tratamientos (a).

j : Variando de 1 hasta el número de réplicas de cada tratamiento ($n$).

El objetivo será probar hipótesis adecuadas con respecto a los efectos del tratamiento y hacer estimaciones de ellos.

Para llevar a cabo esta prueba de hipótesis, se debe suponer que los errores del modelo son variables aleatorias independientes con distribución normal, con media cero y varianza $\sigma^2$. Se supone que $\sigma^2$ es constante para todos los niveles del factor.

El Modelo Estadístico recibe el nombre de "Análisis de Varianza de Clasificación Unidireccional", debido a que se investiga un sólo factor.

Se debe realizar el experimento en orden aleatorio, es decir, que tanto la asignación del material experimental como el orden en que se realizan las pruebas individuales se determinan aleatoriamente, con el objetivo de que el medio ambiente en el que se usan las unidades experimentales sean lo más uniformemente posible; por lo anterior este diseño es un diseño completamente aleatorizado.

Este modelo describe dos situaciones con respecto al efecto de los tratamientos, que son:

- Si el experimentador selecciona específicamente los "$a$" tratamientos que intervienen en el experimento, entonces este Modelo se denomina **"Modelo de Efectos Fijos".** En este tipo de Modelos se desea probar hipótesis en relación a las medias de los tratamientos y las conclusiones sólo se aplicarán a los niveles del factor considerados en el análisis. Las conclusiones no pueden extenderse a tratamientos similares que no se consideraron.

- La selección de los "$a$" tratamientos para el experimento pueden hacerse utilizando una muestra aleatoria de una población de tratamientos, este modelo se denomina **"Modelo de Efectos Aleatorios"**. En este caso es conveniente generalizar las conclusiones (basadas en la muestra de tratamientos), a todos los tratamientos de la población, sin que importe que se hayan o no considerado dentro del análisis. Se supone que la población de niveles del factor es infinita o lo suficientemente grande para ser considerada infinita. Los casos en que la población de niveles del factor es suficientemente pequeña, para emplear un enfoque de población finita no se encuentra muy seguido. Los efectos de los tratamientos $(\iota_i)$ son variables aleatorias, y se considera que no es necesario conocer sus valores particulares para los tratamientos a investigar. En vez de ello se debe probar hipótesis acerca de la variabilidad de los $\iota_i$ y se tratará de estimar esta variabilidad.

## 4. SUMAS Y MEDIAS DE CUADRADOS

El término "Análisis de Varianza" consiste en la descomposición de la variabilidad total de los datos, en sus partes que la forman.

Sea:

$SS_T$ : Suma Total de Cuadrados Corregida.

$SS_{Tratamientos}$ : Suma de Cuadrados debida a los Tratamientos (entre Tratamientos).

$SS_E$ : Suma de Cuadrados debida al Error (dentro de los tratamientos).

La suma total de cuadrados corregida, que es considerada como una medida de variabilidad total de los datos, puede ser escrita de la siguiente forma:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

Al descomponer esta sumatoria queda:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$ (**Ver Apéndice (1)**)

Esto indica que la variabilidad total de los datos, medida por la suma total de cuadrados corregida ($SS_T$), puede descomponerse en la suma de los cuadrados de las diferencias entre las medias de los tratamientos y el promedio general, y en la suma de cuadrados de las diferencias entre las observaciones dentro del tratamiento y el promedio del mismo.

La diferencia entre las medias de tratamientos observadas y el promedio general, miden la diferencia entre las medias de tratamientos, y las causas de las diferencias de las observaciones dentro de los tratamientos, con respecto al promedio del tratamiento, puede ser solamente el error aleatorio.

Por lo tanto:

$$SS_T = SS_{Tratamientos} + SS_E$$

donde:

$SS_T$ : Tiene $N$-1 grados de libertad porque existe un total de $N = an$ observaciones, y un sólo parámetro a estimar que es $\mu$ .

$SS_{Tratamientos}$ : Tiene $a$ - 1 grados de libertad porque existen "$a$" niveles del factor (y "$a$" medias de tratamientos) y sólo un parámetro a estimar que es $\iota_i$.

$SS_E$ : Tiene $N$ - $a$ grados de libertad porque existe "$n$" réplicas dentro de cada tratamiento, las cuales proporcional n-1 grados de libertad para estimar el error experimental. Como hay "$a$" tratamientos, se tiene $a$(n-1) = $a$n- $a$ = N- $a$ .

Las fórmulas matemáticas para obtener las sumas de cuadrados son:

**Caso Balanceado.**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \quad \text{ó} \quad SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$$ **(Ver Apéndice (2))**

$$SS_{Tratamientos} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \quad \text{ó} \quad SS_{Tratamientos} = n\sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$$ **(Ver Apéndice(3))**

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$$

**Caso Desbalanceado.**

Sea $N = \sum_{i=1}^{a} n_i$

donde $n_i$ es el número de observaciones realizadas del tratamiento *i*.

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \quad \text{ó} \quad SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$$

$$SS_{Tratamientos} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \quad \text{ó} \quad SS_{Tratamientos} = \sum_{i=1}^{a} n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$$

En ambos casos otra forma de obtener la Suma de Cuadrados del Error es por diferencia, es decir:

$$SS_E = SS_T - SS_{Tratamientos}$$

Las Medias de Cuadrados no son más que la suma de cuadrados divididos por sus respectivos grados de libertad.

Matemáticamente las medias de cuadrados se definen de la siguiente manera:

$$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} \quad , \quad MS_E = \frac{SS_E}{N-a}$$

donde :

$MS_{Tratamientos}$ : Suma de Cuadrados Medios entre Tratamientos

$MS_E$ : Suma de Cuadrados Medios del Error

En el libro de Diseño y Análisis de Experimentos de Douglas C. Montgomery, Pág. 52 se, llega a determinar que los valores esperados de las medias de cuadrados son:

$$E(MS_{Tratamientos}) = \sigma^2 + \frac{n\sum_{i=1}^{a} \iota_i}{a-1}$$

$$E(MS_E) = \sigma^2$$

## 5. ANÁLISIS ESTADÍSTICO

En este apartado se plantearán las bases estadísticas para lograr la fiabilidad en las conclusiones que se obtengan en el experimento. Considerando separadamente los modelos antes planteados (Modelo de Efectos Fijos y Modelo Efectos Aleatorio), ya que poseen ciertas diferencias en su análisis.

### Modelo de Efectos Fijos.

En este tipo de Modelo interesa probar la igualdad de las medias de los "$a$" tratamientos; o sea las hipótesis a probar son:

$H_o : \mu_1 = \mu_2 = ...... = \mu_a$

$H_1 : \mu_i \neq \mu_j$ , para al menos un par (i,j)

Cuando $H_0$ es verdadera significa que todos los tratamientos tienen la media común "$\mu$".

También se pueden expresar las hipótesis anteriores en función de los efectos de los tratamientos $\iota_i$ , de la siguiente forma:

$H_o : \iota_1 = \iota_2 = ........= \iota_a = 0$

$H_1 : \iota_i \neq 0$ , para al menos un i

Esto significa que es posible hablar de probar la igualdad de las medias de los tratamientos, o bien de probar que los efectos de los tratamientos son iguales a cero.

Para poder llegar a probar la igualdad del nivel medio de "$a$" tratamientos debemos utilizar el Análisis de Varianza.

Como los errores $\varepsilon_{ij}$ son NID($0,\sigma^2$) (NID: variable aleatoria Normal e Independientemente Distribuida), las observaciones $y_{ij}$ son NID ($\mu+\tau_i$ ,$\sigma^2$); se puede demostrar que $\frac{SS_T}{\sigma^2}$ se distribuye como ji-cuadrada con N-1 grados de libertad, ya que $SS_T$ es una suma de cuadrados de variables aleatorias normalmente distribuidas. Si $H_o$ es verdadera, y $\frac{SS_E}{\sigma^2}$ tiene una distribución ji-cuadrada con N-a grados de libertad, entonces $\frac{SS_{Tratamientos}}{\sigma^2}$ tiene una distribución ji-cuadrada con $a$-1 grados de libertad.

Uno de los teoremas útiles para realizar el Análisis de la Varianza es el **Teorema de Cochran.**

**Teorema de Cochran.**

Sean $Z_i$ variables aleatorias NID(0,1) para i=1,2,…,v y $\sum_{i=1}^{v} Z_i^2 = Q_1 + Q_2 + \ldots + Q_s$ donde $s \leq v$ y $Q_i$ tiene $v_i$ grados de libertad (i=1,2,…..,s). Entonces $Q_1$ ,$Q_2$ ,……..,$Q_s$ son variables aleatorias independientes con distribución ji-cuadrada con $v_1$, $v_2$, …., $v_s$ grados de libertad, respectivamente, si y sólo si $v = v_1 + v_2 + \ldots + v_s$.

Aplicando las hipótesis y conclusiones del teorema enunciado anteriormente a las sumas de cuadrados se tiene:

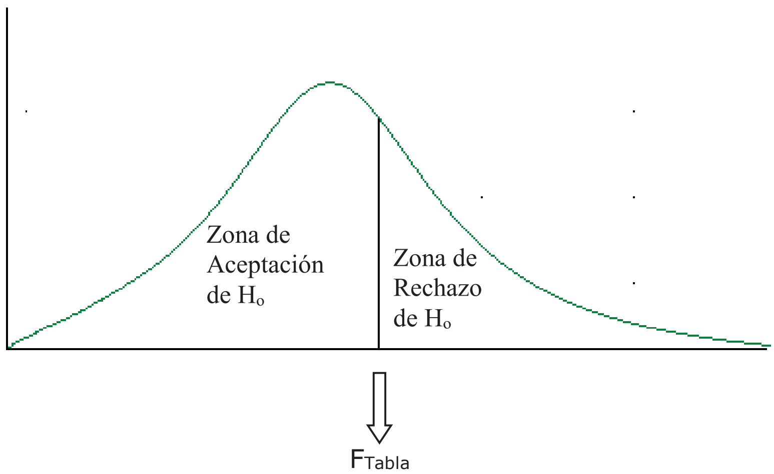
Como $SS_T = SS_{Tratamientos} + SS_E$ y los grados de libertad de $SS_{Tratamientos}$ es $a$ – 1, los grados de libertad de $SS_E$ son $N$ – a entonces la suma de los grados de libertad de $SS_{Tratamientos}$ y de $SS_E$ es igual a $N$ – 1; que es el total de los grados de libertad, por lo tanto por el **Teorema de Cochran** $\frac{SS_{Tratamientos}}{\sigma^2}$ y $\frac{SS_E}{\sigma^2}$ son variables aleatorias independientes con distribución ji-cuadrada.

Por lo tanto, si la hipótesis nula de igualdad de medias de los tratamientos es verdadera, se tiene la siguiente estadística:

$$F_o = \frac{\frac{SS_{Tratamientos}}{a-1}}{\frac{SS_E}{N-a}} = \frac{MS_{Tratamientos}}{MS_E}$$

, la cual tiene una distribución F con $a$-1 y $N$-$a$ grados de libertad.

En general se observa que el valor esperado de la media de cuadrados $MS_E$ es un estimador insesgado de $\sigma^2$; mientras que si la hipótesis nula es verdadera ($H_o$: $\iota_1 = \iota_2 = \ldots\ldots = \iota_a = 0$), $MS_{Tratamientos}$ resulta ser un estimador insesgado de $\sigma^2$; de lo contrario, el valor esperado de $MS_{Tratamientos}$ es mayor que $\sigma^2$. Por lo tanto, si la hipótesis alternativa es verdadera ($H_1$: $\iota_i \neq 0$ , para al menos un $i$) el valor esperado del numerador de la estadística, es mayor que el valor esperado del denominador, entonces debe rechazarse $H_o$ si el valor de la estadística es demasiado grande. Esto, significa una región crítica unilateral superior.




De tal modo la hipótesis nula ($H_o$) se rechazará si:

$F_o > F_{\alpha,(a-1),(N-a)}$

Donde $F_o$ se obtiene a través del Análisis de Varianza y $F_{\alpha,(a-1),(N-a)}$ se obtiene a través de la tabla F; con $a - 1$ grados de libertad en el numerador y $N - a$ grados de libertad en el denominador.

En la siguiente tabla se resume el Análisis de Varianza de un Diseño Unifactorial.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Entre Tratamientos | $SS_{Tratamientos}$ | $a-1$ | $MS_{Tratamientos}$ | $F_o = \dfrac{MS_{Tratamientos}}{MS_E}$ |
| Error (Dentro de Tratamientos) | $SS_E$ | $N-a$ | $MS_E$ | |
| Total | $SS_T$ | $N-1$ | | |

**Ejemplo 2**

El Ministerio de Educación esta interesado en implementar tres programas de estudio; con el objetivo de medir la habilidad de lectura en los alumnos. Para ello, se eligen alumnos del sexto grado de un Colegio de San Salvador, de los cuales fueron asignados al azar 27 alumnos, a cada uno de los tres grupos. Se utilizó un programa diferente en cada grupo, se llevó a cabo un examen al inicio y al final de la implementación de los programas, los valores obtenidos representan la diferencia que hay entre la nota del examen que se hizo al inicio y al final de la implementación del programa, obteniéndose los siguientes datos, en base 100:

| Tratamiento (nivel) | Observaciones | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Programa 1 | 20 | 18 | 18 | 23 | 22 | 17 | 15 | 13 | 21 |
| Programa 2 | 15 | 20 | 13 | 12 | 16 | 17 | 21 | 15 | 13 |
| Programa 3 | 12 | 15 | 18 | 20 | 18 | 17 | 10 | 24 | 16 |

**Solución.**

Antes de realizar los cálculos matemáticos, se definirá la variable de estudio y las hipótesis que se desean probar.

**Variable de estudio**: Habilidad de Lectura

$H_o$ : $\mu_1 = \mu_2 = \mu_3$ (No existe diferencia entre los grupos)

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3$ (Existe diferencia entre los grupos)

**El significado verbal de las hipótesis es:**

$H_o$ : Con la implementación de los tres programas de estudio, no existe diferencia significativa en la habilidad de lectura entre los grupos de alumnos del sexto grado.

$H_1$ : Con la implementación de los tres programas de estudio, existe diferencia significativa en la habilidad de lectura entre los grupos de alumnos del sexto grado.

**Datos**

Tratamientos: $a$ =3

Número de observaciones por grupo :$n$ = 9

Número total de observaciones: $N = an$ = 3x9 = 27

$i$ = 1,2,3 $j$ = 1,2,….,9

## Cálculos Matemáticos

**Totales**

$y_{i.} = \sum_{j=1}^{9} y_{ij}$

$y_{1.} = \sum_{j=1}^{9} y_{ij}$ = 20+18+…..+21 = 167 $\quad$ $y_{2.} = \sum_{j=1}^{9} y_{ij}$ = 15+20+…..+13 = 142

$y_{3.} = \sum_{j=1}^{9} y_{ij}$ = 12+15+…..+16 = 150 $\quad$ $y_{..} = \sum_{i=1}^{3}\sum_{j=1}^{9} y_{ij}$ = 20+18+18+….+24+16 = 459

**Medias de los Tratamientos**

$\bar{y}_{i.} = \frac{y_{i.}}{n}$

$\bar{y}_{1.} = \frac{y_{1.}}{9} = \frac{167}{9} = 18.55$ $\quad$ $\bar{y}_{2.} = \frac{y_{2.}}{9} = \frac{142}{9} = 15.77$

$\bar{y}_{3.} = \frac{y_{3.}}{9} = \frac{150}{9} = 16.66$ $\quad$ $\bar{y}_{..} = \frac{y_{..}}{27} = \frac{459}{27} = 17$

**Sumas de Cuadrados**

$SS_T = \sum_{i=1}^{3}\sum_{j=1}^{9} y_{ij}^2 - \frac{y_{..}^2}{N}$ = [ $(20)^2+(18)^2+(18)^2+.....+(24)^2+(16)^2$] - $\frac{(459)^2}{27}$

= 8141 - 7803

$SS_T$ = 338

$SS_{Tratamientos} = \frac{\sum_{i=1}^{3} y_{i.}^2}{n} - \frac{y_{..}^2}{N} = \frac{(167)^2 + (142)^2 + (150)^2}{9} - \frac{(459)^2}{27}$

= 7839.22 - 7803

$SS_{Tratamientos}$ = 36.22

$SS_E = SS_T - SS_{Tratamientos}$

$SS_E$ = 338 - 36.22

$SS_E$ = 301.78

**Medias de Cuadrados**

$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} = \frac{36.22}{3-1} = \frac{36.22}{2} = 18.11$

$MS_E = \frac{SS_E}{N-a} = \frac{301.78}{27-3} = \frac{301.78}{24} = 12.57$

**Estadística**

$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{18.11}{12.57} = 1.44$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Programas de Estudios | 36.22 | 2 | 18.11 | 1.44 |
| Error (Dentro de Tratamientos) | 301.78 | 24 | 12.57 | |
| Total | 338.00 | 26 | | |

Utilizando un nivel de significancia del 5% ($\alpha = 0.05$), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 2 grados de libertad ($a$-1) en el numerador y 24 grados de libertad ($N$-$a$) en el denominador.

$$F_{\alpha,a-1,N-a} = F_{0.05,2,24} = 3.40$$

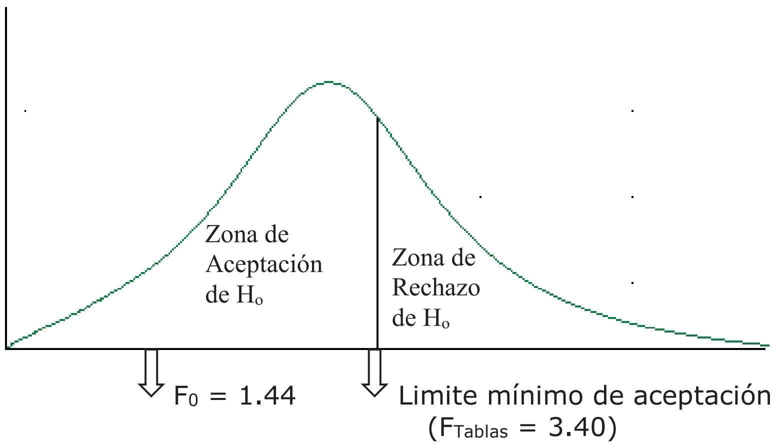
Comparando el $F_0$ calculado en el análisis de varianza y el $F_{Tablas}$ , se puede observar que:

$$F_0 < F_{Tablas}$$

$$1.44 < 3.40$$

Por tanto, se acepta la hipótesis nula ($H_0$) y se rechaza la hipótesis alternativa ($H_1$).

También, se puede observar gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de aceptación de $H_0$.

**Conclusión**

Con la implementación de los tres programas de estudio, no existe diferencia significativa en la habilidad de lectura entre los grupos de alumnos del sexto grado.

## 6. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Usando el método de mínimos cuadrados se pueden obtener los estimadores de $\mu$ y $\tau_i$ del modelo: $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$.

Se debe tomar en cuenta la restricción que $\sum_{i=1}^{a} \hat{\tau}_i = 0$; ya que los $\tau_i$ son desviaciones de la media general; por lo tanto, las estimaciones son: (Ver Douglas C. Montgomery, año 1991, Páginas 57 y 58).

$$\hat{\mu} = \bar{y}_{..} \qquad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \quad i = 1,2,\ldots,a$$

Esto significa que la media general puede ser estimada usando el promedio total de las observaciones, y que cualquiera de los efectos de los tratamientos son sólo la diferencia entre el promedio del tratamiento y el promedio total.

En algunas ocasiones es importante determinar un intervalo de confianza para la media del tratamiento i-ésimo. Por lo tanto, la media del tratamiento i-ésimo viene dada por:

$\mu_i = \mu + \tau_i \qquad i= 1,2,\ldots,a$

Un estimador puntual de $\mu_i$ es: $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$. Suponiendo que los errores están normalmente distribuidos, y $\bar{y}_{i.}$ son NID($\mu_i$, $\frac{\sigma^2}{n}$), entonces si se conociera $\sigma^2$, se utiliza una distribución normal para construir un intervalo de confianza. Si se utiliza $MS_E$ como un estimador de $\sigma^2$, el intervalo de confianza se basaría en una distribución t. Entonces un intervalo de confianza del 100%(1-$\alpha$) para la media del i-ésimo tratamiento, $\mu$ está determinado por:

$$\bar{y}_{i.} \pm t_{\frac{\alpha}{2},N-a}\sqrt{\frac{MS_E}{n}}$$ **Caso Balanceado**

Un intervalo de confianza del 100%(1-$\alpha$) para la diferencia de las medias de dos tratamientos cualesquiera, digamos $\mu_i$ - $\mu_j$, seria:

$$\bar{y}_{i.} - \bar{y}_{j.} \pm t_{\frac{\alpha}{2},N-a}\sqrt{\frac{2MS_E}{n}}$$ **Caso Balanceado**

En el **caso desbalanceado** utilizar la misma fórmula, solamente tomar n = $\frac{a}{\sum_{i=1}^{a}\frac{1}{n_i}}$

**Ejemplo 3**

Al usar los datos del ejemplo 2, para encontrar las estimaciones de la media general y los efectos de los programas de estudio se tiene:

$\hat{\mu} = \bar{y}_{..}$ = 17 (Media General)

Significa que si se implementan los tres programas de estudio en toda la población de sexto grado de El Salvador se espera que la media general sea igual a 17 puntos.

$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$ , i = 1,2,3

$\hat{\tau}_1 = \bar{y}_{1.} - \bar{y}_{..}$ = 18.55 - 17 = 1.55 (Efecto del programa 1)

Significa que la habilidad de lectura en los alumnos de sexto grado aumenta un 1.55 puntos con la implementación del programa 1.

$\hat{\tau}_2 = \bar{y}_{2.} - \bar{y}_{..}$ = 15.77 - 17 = -1.23 (Efecto del programa 2)

Significa que la habilidad de lectura en los alumnos de sexto grado disminuirá en 1.23 puntos con la implementación del programa 2.

$\hat{\tau}_3 = \bar{y}_{3.} - \bar{y}_{..}$ = 16.66 - 17 = -0.34 (Efecto del programa 3)

Significa que la habilidad de lectura en los alumnos de sexto grado disminuirá en 0.34 puntos con la implementación del programa 3.

Un intervalo de confianza del 95% para el efecto medio del programa 2, en el grupo 2, se obtiene de la siguiente manera:

**Datos**

$MS_E$ = 12.57 , $N$ = 27 , $\bar{y}_{3.}$ =16.66 , $n$ = 9 , $a$ = 3 , $\alpha$ = 0.05 , $\bar{y}_{2.}$ =15.77

$$\bar{y}_{i.} \pm t_{\frac{\alpha}{2},N-a}\sqrt{\frac{MS_E}{n}}$$

$$\bar{y}_{2.} \pm t_{\frac{0.05}{2},27-3}\sqrt{\frac{MS_E}{9}}$$

$$15.77 \pm t_{0.025,24}\sqrt{\frac{12.57}{9}}$$

$$15.77 \pm (2.064)\ (1.1818)$$

$$15.77 \pm 2.439$$

Por tanto, el intervalo buscado es: $13.33 \leq \mu_2 \leq 18.21$; es decir, que se tiene el 95% de confianza que en 100 muestras del mismo tamaño tomado anteriormente (n = 9); 95 de esas muestras la media del programa 2 caerá dentro del intervalo de [13.33 , 18.21].

Para la diferencia del efecto medio de los programas de estudio 1 y 3, en los grupos 1 y 3 respectivamente, el intervalo de confianza del 95% es:

$$\bar{y}_{i.} - \bar{y}_{j.} \pm t_{\frac{\alpha}{2},N-a}\sqrt{\frac{2MS_E}{n}}$$

$$\bar{y}_{1.} - \bar{y}_{3.} \pm t_{\frac{0.05}{2},27-3}\sqrt{\frac{2MS_E}{9}}$$

$$18.55 - 16.66 \pm t_{0.025,24}\sqrt{\frac{2(12.57)}{9}}$$

$$18.55 - 16.66 \pm (2.064)\,(1.67)$$

$$1.89 \pm 3.45$$

Por lo tanto, el intervalo buscado es: $-1.56 \leq \mu_1 - \mu_3 \leq 5.34$; es decir, que se tiene el 95% de confianza que en 100 muestras del mismo tamaño (n=9), en 95 de ellas la diferencia de la media poblacional del programa 1 y programa 3 caerá dentro del intervalo de [-1.56 , 5.34].

## 7. COMPARACIÓN ENTRE TRATAMIENTOS

Si al efectuar el Análisis de Varianza para un Modelo de Efectos Fijos, la hipótesis nula es rechazada. Se llega a la conclusión que existe diferencia entre las medias o que hay diferencia entre los tratamientos. En muchas situaciones en la industria, este resultado es de poco interés; ya que no se especifica exactamente cuales tratamientos son diferentes y el experimentador espera hallar diferencias, y está más interesado en investigar que tratamientos difieren entre si, o dicho de otra manera, en investigar contrastes entre los tratamientos.

Cuando se da esta situación puede ser útil realizar comparaciones adicionales entre grupos de medias de los tratamientos. Las comparaciones entre medias de tratamientos se realizan en términos de los totales de tratamientos $y_{i.}$ o de los promedios de tratamientos $\bar{y}_{i.}$. Los procedimientos para efectuar esta comparación se conocen como **Métodos de Comparación Múltiple o Pruebas a Posteriori.**

## 7.1 COMPARACIÓN DE MEDIAS DE TRATAMIENTOS INDIVIDUALES

### 7.1.1 Contrastes Ortogonales

Los contrastes se utilizan mucho en métodos de comparación múltiple; el cual se estudiará a continuación.

Supongamos que el factor que se esta estudiando tiene cinco niveles (tratamientos) y al llevar a cabo el Análisis de Varianza se rechaza $H_o$, con base a esta información es posible suponer que el tratamiento uno y dos producen la misma diferencia.

Esto implica que es necesario probar las siguientes hipótesis:

$H_o : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Estas hipótesis pueden ser probadas investigando una combinación lineal apropiada de los totales de los tratamientos, por ejemplo:

$y_{1.} - y_{2.} = 0$

Por otro lado, si se suponen que el promedio de los tratamientos 1 y 3 no difieren del promedio de los tratamientos 4 y 5, las hipótesis que deben probarse son:

$H_o : \mu_1 + \mu_3 = \mu_4 + \mu_5$

$H_1 : \mu_1 + \mu_3 \neq \mu_4 + \mu_5$

y esto implica la combinación lineal:

$y_{1.} + y_{3.} - y_{4.} - y_{5.} = 0$

De acuerdo a este análisis se puede generalizar que la comparación de medias de tratamientos conlleva a una combinación lineal de totales de tratamientos de la forma:

$c = \sum_{i=1}^{a} c_i y_{i.}$ tomando en cuenta que al formarse esta combinación lineal, los $c_i$ pueden tomar cualquier valor; dependiendo de las comparaciones de medias que se están investigando, siempre y cuando se tome en cuenta la restricción $\sum_{i=1}^{a} c_i = 0$

**Por ejemplo**, si deseo comparar tres veces la media del tratamiento dos con la suma de las medias de los tratamientos uno, tres y cuatro

**La hipótesis a probar será:** **Contraste**

$H_o : 3\mu_2 = \mu_1 + \mu_3 + \mu_4$ $c = y_{1.} + y_{3.} + y_{4.} - 3y_{2.}$

$H_1 : 3\mu_2 \neq \mu_1 + \mu_3 + \mu_4$

Tales combinaciones lineales se conocen como **Contrastes**. Las suma de cuadrados de un contraste viene dada por:

$$SS_c = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{n\sum_{i=1}^{a} c_i^2}$$ y tiene un sólo grado de libertad.

Si el diseño es desbalanceado, la comparación de las medias de tratamientos requiere que $\sum_{i=1}^{a} n_i c_i = 0$, en este caso la suma de cuadrados de un contraste se transforma en:

$$SS_c = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{\sum_{i=1}^{a} n_i c_i^2}$$

Para probar un contraste se debe comparar su suma de cuadrados con la media de cuadrados del error.

La estadística que resulta tiene una distribución F con 1 y $N - a$ grados de libertad, la hipótesis nula se rechazará si:

$$F_o > F_{\alpha,1,N-a}$$

Donde $F_o$ se obtiene a través del Análisis de Varianza de los contrastes y $F_{\alpha,1,N-a}$ se obtiene por medio de la tabla $F$ con 1 grado de libertad en el numerador y $N - a$ grados de libertad en el denominador.

Un caso especial de lo planteado anteriormente son los Contrastes Ortogonales.

Se dice que dos contrastes con coeficientes $\{c_i\}$ y $\{d_i\}$ son ortogonales si,

$\sum_{i=1}^{a} c_i d_i = 0$ en el caso de un **Diseño Balanceado**,

ó

$\sum_{i=1}^{a} n_i c_i d_i = 0$ en caso de un **Diseño Desbalanceado**.

El conjunto de $a$ - 1 contrastes ortogonales de "$a$" tratamientos descomponen la suma de cuadrados debido a los tratamientos en $a$-1 componentes independientes de un sólo grado de libertad. Por lo tanto, las pruebas realizadas sobre los contrastes ortogonales son independientes.

Hay muchas formas de elegir los coeficientes de los contrastes ortogonales para un conjunto dado de tratamientos. Por lo general, la naturaleza del experimento debe sugerir las comparaciones que resulten de interés.

Los coeficientes de los contrastes deben ser elegidos antes de realizar el experimento y analizar los datos. La razón de ello es que si las comparaciones son seleccionadas después de analizar los datos, la mayoría de los investigadores construirían pruebas que corresponderían a grandes diferencias observadas en los promedios y el error tipo I tiende a incrementarse.

El procedimiento para llevar a cabo la comparación de medias de tratamientos individuales por medio de contraste es el siguiente:

- Se asumen ($a$-1) hipótesis nulas.
- En base a dichas hipótesis, se establecen los contrastes.
- Se calculan los contrastes.
- Se calcula la suma de cuadrados para los contrastes.
- Se realiza el Análisis de Varianza incluyendo los contrastes.
- Se analizan los resultados de los contrastes.

**Ejemplo 4**

Se supone que la cantidad de carbón usada en la producción de acero tiene un efecto en su resistencia a la tensión. En la tabla se muestran los valores de la resistencia a la tensión del acero para cada uno de los 4 diferentes porcentajes de carbón. Con estos datos efectúe el análisis apropiado e interprete sus resultados.

| % de Carbón | Observaciones | | | |
|---|---|---|---|---|
| 0.10 | 23 | 28 | 28 | 30 |
| 0.20 | 31 | 29 | 36 | 38 |
| 0.30 | 36 | 40 | 42 | 44 |
| 0.40 | 48 | 45 | 40 | 40 |

**Solución.**

**Variable Respuesta: Resistencia a la tensión.**

**Las hipótesis que se desean probar son:**

$H_o$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (Las medias son iguales)

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ (Las medias son diferentes)

**El significado verbal es:**

$H_o$ : La cantidad de carbón usada en la producción de acero no tiene efecto significativo en la resistencia a la tensión.

$H_1$ : La cantidad de carbón usada en la producción de acero tiene efecto significativo en la resistencia a la tensión.

**Datos**

$a = 4$ , $n = 4$ , $N = 16$ , $i = 1,2,3,4$ , $j = 1,2,3,4$

**Cálculos Matemáticos**

**Totales**

$$y_{i.} = \sum_{j=1}^{4} y_{ij}$$

$$y_{1.} = \sum_{j=1}^{4} y_{ij} = 23+28+28+30 = 109 \qquad y_{2.} = \sum_{j=1}^{4} y_{ij} = 31+29+36+38 = 134$$

$$y_{3.} = \sum_{j=1}^{4} y_{ij} = 36+40+42+44 = 162 \qquad y_{4.} = \sum_{j=1}^{4} y_{ij} = 48+45+40+40 = 173$$

$$y_{..} = \sum_{i=1}^{4}\sum_{j=1}^{4} y_{ij} = 23+28+28+....+40+40 = 578 \text{ ó } y_{..} = \sum_{j=1}^{4} y_{i.} = 109+134+162+173 = 578$$

**Medias de los Tratamientos.**

$$\bar{y}_{i.} = \frac{y_{i.}}{n}$$

$$\bar{y}_{1.} = \frac{y_{1.}}{4} = \frac{109}{4} = 27.25 \qquad \bar{y}_{2.} = \frac{y_{2.}}{4} = \frac{134}{4} = 33.50$$

$$\bar{y}_{3.} = \frac{y_{3.}}{4} = \frac{162}{4} = 40.50 \qquad \bar{y}_{4.} = \frac{y_{4.}}{4} = \frac{173}{4} = 43.25$$

$$\bar{y}_{..} = \frac{y_{..}}{N} = \frac{578}{16} = 36.125$$

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{4} y_{ij}^2 - \frac{y_{..}^2}{N} = [\,(23)^2 + (28)^2 + (28)^2 + ..... + (40)^2 + (40)^2\,] - \frac{(578)^2}{16}$$

$= 21664 - 20880.25$

$SS_T = 783.75$

$$SS_{Tratamientos} = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} = \frac{(109)^2 + (134)^2 + (162)^2 + (173)^2}{4} - \frac{(578)^2}{16}$$

$= 21502.50 - 20880.25$

$SS_{Tratamientos} = 622.25$

$SS_E = SS_T - SS_{Tratamientos}$

$SS_E = 783.75 - 622.25$

$SS_E = 161.50$

**Medias de Cuadrados**

$$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} = \frac{622.25}{4-1} = \frac{622.25}{3} = 207.417$$

$$MS_E = \frac{SS_E}{N-a} = \frac{161.50}{16-4} = \frac{161.50}{12} = 13.458$$

**Estadística**

$$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{207.417}{13.458} = 15.41$$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| % de Carbón | 622.25 | 3 | 207.417 | 15.41 |
| Error (Dentro de Tratamientos) | 161.50 | 12 | 13.458 | |
| Total | 783.75 | 15 | | |

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 3 grados de libertad ($a$-1) en el numerador y 12 grados de libertad ($N$-$a$) en el denominador.

$F_{\alpha,a-1,N-a} = F_{0.05,3,12} = 3.49$

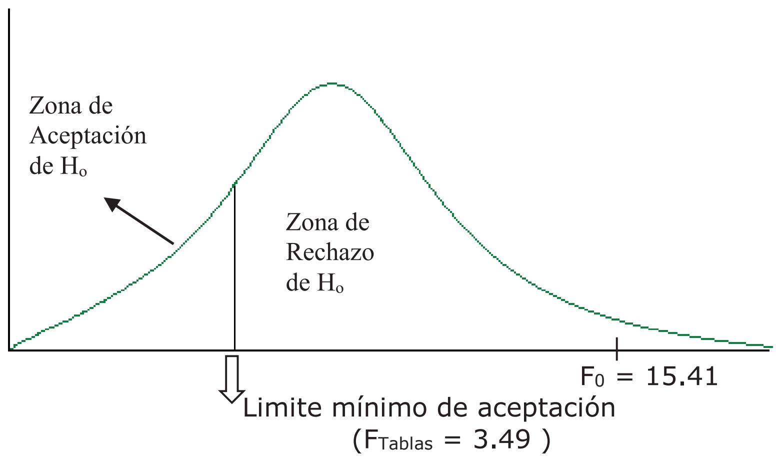
Comparando el $F_0$ calculado en el análisis de varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

15.41 > 3.49

Por tanto, se Rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$); es decir, que las medias de los tratamientos difieren.

También se puede observar gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

Se concluye que la cantidad de carbón usada en la producción de acero tiene efectos significativos en la resistencia a la tensión.

Como se ha rechazado $H_0$, existe diferencia entre las medias de los tratamientos, pero no se especifica entre que medias de tratamientos existen las diferencias.

Se podría estar interesado en querer saber entre que medias de tratamientos existe diferencia, para ello se utilizará el método de contrastes ortogonales para contestar esta inquietud.

1. **Se definen las 4 – 1 = 3 hipótesis.**

$H_0 : \mu_1 = \mu_2$ $\quad$ $H_0 : \mu_3 = \mu_4$ $\quad$ $H_0: \mu_1 + \mu_2 = \mu_3 + \mu_4$

$H_1 : \mu_1 \neq \mu_2$ $\quad$ $H_1 : \mu_3 \neq \mu_4$ $\quad$ $H_1: \mu_1 + \mu_2 \neq \mu_3 + \mu_4$

2. **Un conjunto de comparaciones entre medias y los contrastes ortogonales son:**

| Hipótesis | Contrastes (c = $\sum_{i=1}^{a} c_i y_{i.}$ ) |
|---|---|
| $H_0: \mu_1 = \mu_2$ | $C_1 = y_{1.} - y_{2.}$ |
| $H_0: \mu_3 = \mu_4$ | $C_2 = y_{3.} - y_{4.}$ |
| $H_0: \mu_1 + \mu_2 = \mu_3 + \mu_4$ | $C_3 = y_{1.} + y_{2.} - y_{3.} - y_{4.}$ |

3. **Cálculo de los contrastes.**

$C_1 = y_{1.} - y_{2.} = 1(109) - 1(134) = -25$

$C_2 = y_{3.} - y_{4.} = 1(162) - 1(173) = -11$

$C_3 = y_{1.} + y_{2.} - y_{3.} - y_{4.} = 1(109) + 1(134) -1(162) - 1(173) = -92$

4. **Cálculo de la Suma de cuadrados de los contrastes.**

$$SS_c = \frac{(\sum_{i=1}^{a} c_i y_{i.})^2}{n\sum_{i=1}^{a} c_i^2} \quad \text{( Por ser balanceado)}$$

$$SS_{c_1} = \frac{(-25)^2}{4(1^2 + (-1)^2)} = 78.125$$

$$SS_{c_2} = \frac{(-11)^2}{4(1^2 + (-1)^2)} = 15.125$$

$$SS_{c_3} = \frac{(-92)^2}{4(1^2 + 1^2 + (-1)^2 + (-1)^2)} = 529.000$$

**5. Tabla de Análisis de Varianza con los contrastes.**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| % de carbón | 622.250 | 3 | 207.417 | 15.41 |
| Contrastes ortogonales | | | | |
| $C_1 = y_{1.} - y_{2.}$ | 78.125 | 1 | 78.125 | 5.81 |
| $C_2 = y_{3.} - y_{4.}$ | 15.125 | 1 | 15.125 | 1.12 |
| $C_3 = y_{1.} + y_{2.} - y_{3.} - y_{4.}$ | 529.000 | 1 | 529.000 | 39.31 |
| Error | 161.5 | 12 | 13.458 | |
| Total | 783.75 | 15 | | |

**6. Análisis de los resultados de los contrastes.**

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 1 grado de libertad en el numerador y 12 grados de libertad ($N$-$a$) en el denominador.

$$F_{\alpha,1,N-a} = F_{0.05,1,12} = 4.75$$

Al comparar el $F_0$ obtenido en cada uno de los contrastes del análisis de varianza con el $F_{Tablas}$; se llega a las siguientes conclusiones:

a) El valor del $F_0$ del contraste 1 es mayor que el $F_{Tablas}$ (5.81 >4.75) entonces se Rechaza $H_0$, y por lo tanto, hay diferencia significativa entre los porcentajes de carbón uno y dos.

b) Como el $F_0$ del contraste tres es menor que $F_{Tablas}$ (1.12 < 4.75) se acepta $H_0$; y por lo tanto, no hay diferencia significativa entre los porcentajes de carbón tres y cuatro.

c) Al comparar el valor de $F_0$ del contraste dos con el $F_{Tablas}$ se observa que el $F_0$ es mayor (39.31 > 4.75) entonces se Rechaza $H_0$; y por lo tanto, se dice que el promedio de los porcentajes de carbón uno y dos difieren significativamente del promedio de los porcentajes de carbón tres y cuatro.

### 7.1.2 Método de Scheffé para comparar todos los contrastes.

Existen situaciones en que el investigador no sabe de antemano los contrastes que desea comparar, o le interesa llevar a cabo más de a-1 posibles comparaciones. Estas comparaciones de interés en muchos experimentos son descubiertas sólo hasta después de hacer un examen preliminar de los datos.

Scheffé hizo una propuesta de un método para comparar cualquier contraste, o los posibles contrastes entre medias de tratamientos. Con este método en cualquiera de las posibles comparaciones el error tipo I es cuando mucho igual a $\alpha$.

Supongamos que existe un conjunto de m contrastes de interés de las medias de tratamientos.

$W_k = C_{1k}\mu_1 + C_{2k}\mu_2 + C_{3k}\mu_3 + \ldots\ldots\ldots + C_{ak}\mu_a$, con $k = 1,2,\ldots,m$, estos contrastes usando los promedios de tratamientos $\bar{y}_{i.}$ son:

$C_k = C_{1k}\bar{y}_{1.} + C_{2k}\bar{y}_{2.} + C_{3k}\bar{y}_{3.} + \ldots\ldots\ldots + C_{ak}\bar{y}_{a.}$ con $k = 1,2,\ldots,m$

El error estándar de estos contraste viene dado por:

$$S_{c_k} = \sqrt{\frac{MS_E}{n}\sum_{i=1}^{a} c_{ik}^2}$$ **Caso Balanceado.**

$$S_{c_k} = \sqrt{MS_E \sum_{i=1}^{a} \frac{C_{ik}^2}{n_i}}$$ **Caso Desbalanceado**, donde $n_i$ es el número de observaciones del i-ésimo tratamiento.

El valor crítico con el que $C_k$ debe ser comparado es:

$$S_{\alpha,k} = S_{c_k}\sqrt{(a-1)F_{\alpha,a-1,N-a}}$$

Para llegar a probar la hipótesis de que el contraste $C_k$ difiere significativamente de cero, es necesario comparar $C_k$ con el valor crítico. Si $|C_k| > S_{\alpha,k}$, la hipótesis nula de que el contraste $W_k$ es igual a cero debe rechazarse.

Este método de Scheffé se puede utilizar para construir intervalos de confianza para todos los posibles contrastes de las medias de tratamientos; los cuales pueden ser construidos como $C_k - S_{\alpha,k} \leq W_k \leq C_k + S_{\alpha,k}$, estos son intervalos de confianza simultáneos; en el sentido de que la probabilidad de que todos ellos sean simultáneamente verdaderos es al menos $1 - \alpha$.

El procedimiento para llevar a cabo la comparación de medias por medio del método de Scheffé es el siguiente:

- Definir los contrastes de interés.
- Calcular los valores numéricos de los contrastes.
- Calcular el error estándar para cada contraste.
- Encontrar los valores críticos.
- Realizar las conclusiones.

**Ejemplo 5**

Considerando los datos del ejemplo del porcentaje de carbón (Ejemplo 4), ya que se rechazó Ho.

Se aplicará el Método anterior:

**Datos**

$\bar{y}_{1.} = 27.25$ , $\bar{y}_{2.} = 33.50$ , $\bar{y}_{3.} = 40.50$ , $\bar{y}_{4.} = 43.25$

$MS_E = 13.458$ , $n = 4$ , $a = 4$

- **Definición de los contrastes.**

  Si suponemos que los contrastes de interés el, dos y el tres son los mismos que se definieron en el Método de los contrastes ortogonales, tenemos:

  $W_1 = 2\mu_1 + \mu_2 - \mu_3 - 2\mu_4$

  $W_2 = \mu_3 - \mu_4$

  $W_3 = \mu_1 + \mu_2 - \mu_3 - \mu_4$

- **El cálculo de los valores numéricos de estos contrastes son:**

  $C_1 = 2\bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.} - 2\bar{y}_{4.} = 2(27.25) + 33.50 - 40.50 - 2(43.25) = -39$

  $C_2 = \bar{y}_{3.} - \bar{y}_{4.} = 40.50 - 43.25 = -2.75$

  $C_3 = \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.} - \bar{y}_{4.} = 27.25 + 33.50 - 40.5 - 43.25 = -23$

- **Cálculo de los errores estándares de los contrastes.**

$$S_{c_1} = \sqrt{\frac{MS_E}{4}\sum_{i=1}^{4} c_{i1}^2} = \sqrt{\frac{13.458}{4}((2)^2 + (1)^2 + (-1)^2 + (-2)^2)} = 5.80$$

$$S_{c_2} = \sqrt{\frac{MS_E}{4}\sum_{i=1}^{4} c_{i2}^2} = \sqrt{\frac{13.458}{4}(1^2 + (-1)^2)} = 2.59$$

$$S_{c_3} = \sqrt{\frac{MS_E}{4}\sum_{i=1}^{4} c_{i3}^2} = \sqrt{\frac{13.458}{4}(1^2 + 1^2 + (-1)^2 + (-1)^2)} = 3.67$$

- **Utilizando el 5% y encontrando los valores críticos tenemos:**

  $S_{0.05,1} = S_{c1}\sqrt{(a-1)F_{0.05,a-1,N-a}} = 5.80\sqrt{3F_{0.05,3,12}} = 5.80\sqrt{(3)(3.49)} = 18.77$

  $S_{0.05,2} = S_{c2}\sqrt{(a-1)F_{0.05,a-1,N-a}} = 2.59\sqrt{3F_{0.05,3,12}} = 2.59\sqrt{(3)(3.49)} = 8.38$

  $S_{0.05,3} = S_{c3}\sqrt{(a-1)F_{0.05,a-1,N-a}} = 3.67\sqrt{3F_{0.05,3,12}} = 3.67\sqrt{(3)(3.49)} = 11.87$

- **Conclusiones**

  a) Como $|C_1| > S_{0.05,1}$ ($|39| > 18.77$) entonces se rechaza $H_o$ y se concluye que $W_1 = 2\mu_1 + \mu_2 - \mu_3 - 2\mu_4$ es diferente de cero; es decir, que el doble del promedio del porcentaje de carbón uno y el promedio del porcentaje de carbón dos, difieren del promedio del porcentaje de carbón tres y el doble del promedio del porcentaje de carbón cuatro.

b) Como $|C_2| < S_{0.05,2}$ ($|-2.75| < 8.38$) entonces se acepta $H_o$ y se concluye que $W_2 = \mu_3 - \mu_4$ es igual a cero; es decir, que no existe diferencia significativa entre los porcentajes de carbón tres y cuatro.

c) Como $|C_3| > S_{0.05,3}$ ($|-23| > 11.87$) entonces se rechaza $H_o$ y se concluye que $W_3 = \mu_1 + \mu_2 - \mu_3 - \mu_4$ es diferente de cero; es decir, que el promedio de los porcentajes de carbón uno y dos difiere del promedio de los porcentajes de carbón tres y cuatro.

## 7.2 COMPARACIÓN DE PAREJAS DE MEDIAS DE TRATAMIENTOS.

En un experimento el investigador puede estar interesado en comparar todas las parejas de "$a$" medias de tratamiento y que la hipótesis nula que se desea probar es $H_o$: $\mu_i = \mu_j$ y la alternativa $H_1$: $\mu_i \neq \mu_j$ para toda $i \neq j$, para llevar acabo estas comparaciones existen muchos métodos, a continuación se estudiarán los más utilizados.

### 7.2.1 Método de la Mínima Diferencia Significativa (LSD)

Supongamos que se desea probar la hipótesis nula $H_o$: $\mu_i = \mu_j$ para toda $i \neq j$. Para llegar a probar esta hipótesis se debe utilizar la estadística t :

$$t_o = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

, si se supone una hipótesis nula bilateral, la pareja de medias $\mu_i$ y $\mu_j$ se consideran diferentes si $|\bar{y}_{i.} - \bar{y}_{j.}| >$ LSD; donde LSD, se encuentra de la siguiente manera:

**Caso Desbalanceado**

$$\text{LSD} = t_{\frac{\alpha}{2}, N-a}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

**Caso Balanceado**

$$\text{LSD} = t_{\frac{\alpha}{2}, N-a}\sqrt{\frac{2MS_E}{n}}$$

Se deben comparar las diferencias observadas entre cada par de promedios con el valor correspondiente de la LSD. Si $|\bar{y}_{i.} - \bar{y}_{j.}| >$ LSD se concluye que las medias poblacionales $\mu_i$ y $\mu_j$ son diferentes.

**Procedimiento**

- Encontrar el valor del LSD.
- Calcular el valor absoluto de las diferencias de los promedios y hacer las comparaciones con el LSD.
- Hacer las conclusiones **.**

**Ejemplo 6**

Si en el ejemplo 4 del porcentaje de carbón se desea saber cuales son las parejas de medias que difieren, se hace de la siguiente manera:

**Datos**

$\alpha$ = 0.05 , $MS_E$ = 13.458 , $N$ = 16 , n = 4 , $a$ = 4

$\bar{y}_{1.}$ = 27.25 , $\bar{y}_{2.}$ = 33.50 , $\bar{y}_{3.}$ = 40.5 , $\bar{y}_{4.}$ = 43.25

- **Encontrando el valor del LSD, con la fórmula establecida.**

$$LSD = t_{\frac{0.05}{2},16-4}\sqrt{\frac{2(13.458)}{4}} = t_{0.025,12}\sqrt{6.729} = (2.179)(2.594)$$

LSD = 5.65

- **Calculando la diferencia de los promedios.**

1 vrs 2 : $|\bar{y}_{1.} - \bar{y}_{2.}|$ = |27.25 – 33.50| = |-6.25| > LSD *

1 vrs 3 : $|\bar{y}_{1.} - \bar{y}_{3.}|$ = |27.25 – 40.50| = |-13.25| > LSD *

1 vrs 4 : $|\bar{y}_{1.} - \bar{y}_{4.}|$ = |27.25 – 43.25| = |-16.00| > LSD *

2 vrs 3 : $|\bar{y}_{2.} - \bar{y}_{3.}|$ = |33.50 – 40.50| = |-7.00| > LSD *

2 vrs 4 : $|\bar{y}_{2.} - \bar{y}_{4.}|$ = |33.50 – 43.25| = | -9.75| > LSD *

3 vrs 4 : $|\bar{y}_{3.} - \bar{y}_{4.}|$ = |40.50 – 43.25| = |-2.75| < LSD

Se dice que una pareja de medias difieren significativamente si el valor absoluto de las diferencias de los promedios de los tratamientos correspondientes es mayor que LSD = 5.65.

- **Conclusiones.**

a) Se observa que la pareja de medias que no difieren significativamente son: la media tres y la media cuatro, ya que |–2.75 |< 5.65, por lo tanto, no existe diferencia significativa entre el porcentaje de carbón tres y cuatro.

b) En las demás parejas de medias (*) el valor absoluto de las diferencia de los promedios a resultado ser mayor que el valor encontrado del LSD; por lo tanto, los demás porcentajes tomados como parejas difieren significativamente.

### 7.2.2 Prueba de Intervalos Múltiples de Duncan

Con este método se prueban las verdaderas diferencias que existen entre los pares de medias. Este método es muy eficiente para detectar diferencias entre medias cuando estas diferencias en realidad existen; es por esta razón que es muy utilizado.

Primero se colocan los "*a*" promedios de tratamiento en orden ascendente y se determina el error estándar de cada promedio.

$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{n}}$ para el **caso balanceado** y en **caso desbalanceado** se debe reemplazar "*n*" por la media armónica $n_h$ de los $\{n_i\}$, en donde $n_h = \frac{a}{\sum_{i=1}^{a}\left(\frac{1}{n_i}\right)}$; es decir, que $S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E \sum_{i=1}^{a} \frac{1}{n_i}}{a}}$

Luego se calculan los intervalos significantes $r_\alpha$ (p,f), para p=2,3,…,a, para ello existen tablas (Ver Apéndice:Tabla de Intervalos Significativos para la Prueba de Intervalos Múltiples de Duncan, Douglas C. Montgomery, año 1991); en donde $\alpha$ es el nivel de significancía y f es el número de grados de libertad del error.

Estos intervalos se convierten en un conjunto de ($a$-1) mínimos intervalos significativos ($R_p$) para $p$= 2,3,…,$a$, y se calculan de la forma siguiente: $R_p = r_\alpha(p,f)\, S_{\bar{y}_{i.}}$

Las diferencias observadas entre las medias se prueban, comenzando por el valor más alto contra el más pequeño, esta diferencia deberá ser comparada con el intervalo mínimo significativo $R_a$. Después se calcula la diferencia entre el valor más alto y el segundo más pequeño y se compara con el intervalo mínimo significativo $R_{a-1}$. Este proceso se continua hasta que han sido consideradas las diferencias entre todas las posibles $\frac{a(a-1)}{2}$ pares de medias. Se concluye que el par de medias en estudio es significativamente diferente si la diferencia observada es mayor que el intervalo mínimo significativo correspondiente.

**Procedimiento**

- Se colocan las medias en orden ascendente.
- Se calcula el error estándar.
- Se obtienen los intervalos significativos.
- Se obtienen los mínimos intervalos significativos.
- Calcular la diferencia de las medias y realizar las comparaciones con los mínimos intervalos significativos correspondientes.
- Hacer las conclusiones.

**Ejemplo 7**

Considerando la información obtenida en el ejemplo del porcentaje de carbón, deseamos saber que parejas de medias son significativamente diferentes.

**Datos**

$\alpha$ = 0.05 , $MS_E$ = 13.458 , $a$ = 4 , $N$ = 16 , $n$ = 4 , grados de libertad del error = 12

$\bar{y}_{1.}$ = 27.25 , $\bar{y}_{2.}$ = 33.50 , $\bar{y}_{3.}$ = 40.50 , $\bar{y}_{4.}$ = 43.25

**Solución**

- **Medias de tratamientos ordenadas ascendentemente.**

  $\bar{y}_{1.}$ = 27.25 , $\bar{y}_{2.}$ = 33.50 , $\bar{y}_{3.}$ = 40.5 , $\bar{y}_{4.}$ = 43.25

- **Obtención del error estándar de cada media.**

  $$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{n}} = \sqrt{\frac{13.458}{4}} = 1.834$$

- Tomando un $\alpha$ = 0.05 y 12 grados de libertad (f), de las tablas de intervalos significativos de Duncan, se obtienen los siguientes valores de los intervalos significativos, para p=2,3,4.

  $r_{0.05}(2,12)$ = 3.08 $r_{0.05}(3,12)$ = 3.23 $r_{0.05}(4,12)$ = 3.33

- **Calculando los mínimos intervalos significativos para p = 2,3,4.**

  $R_2 = r_{0.05}(2,12)\, S_{\bar{y}_{i.}}$ = (3.08)(1.834) = 5.65

  $R_3 = r_{0.05}(3,12)\, S_{\bar{y}_{i.}}$ = (3.23)(1.834) = 5.92

  $R_4 = r_{0.05}(4,12)\, S_{\bar{y}_{i.}}$ = (3.33)(1.834) = 6.11

- **Realizando las comparaciones y las diferencias de las medias (vrs:versus).**

  4 vrs 1 : $\bar{y}_{4.} - \bar{y}_{1.}$ = 43.25 – 27.25 = 16.00 > 6.11 ($R_4$) *

  4 vrs 2 : $\bar{y}_{4.} - \bar{y}_{2.}$ = 43.25 – 33.50 = 9.75 > 5.92 ($R_3$) *

  4 vrs 3 : $\bar{y}_{4.} - \bar{y}_{3.}$ = 43.25 – 40.50 = 2.75 < 5.65 ($R_2$)

  3 vrs 1 : $\bar{y}_{3.} - \bar{y}_{1.}$ = 40.50 – 27.25 = 13.25 > 5.92 ($R_3$) *

  3 vrs 2 : $\bar{y}_{3.} - \bar{y}_{2.}$ = 40.50 – 33.50 = 7.00 > 5.65 ($R_2$) *

  2 vrs 1 : $\bar{y}_{2.} - \bar{y}_{1.}$ = 33.50 – 27.25 = 6.25 > 5.65 ($R_2$) *

Un par de medias es significativamente diferente si la diferencia observada es mayor que el intervalo mínimo significativo correspondiente.

- **Conclusiones**

a) Como el valor de las diferencias de las medias tres y cuatro resultó ser menor que el mínimo intervalo significativo correspondiente; entonces se dice que no existe diferencias significativas entre los porcentajes de carbón tres y cuatro.

b) Por el contrario las demás diferencias de medias (*) resultaron ser mayores que el mínimo intervalo significativo correspondiente; entonces se dice que existe diferencias significativas entre sus medias.

### 7.2.3 Prueba de Tukey

Este método esta basado en el de intervalos. El procedimiento consiste en el uso de $q_\alpha(a,f)$ para encontrar el valor crítico de todas las comparaciones por pares, sin importar cuántas medias estén en el grupo.

Por lo tanto, se declaran dos medias significativamente diferentes si:

$|\bar{y}_{i.} - \bar{y}_{j.}| > T_\alpha$ , donde $T_\alpha = q_\alpha(a,f)\, S_{\bar{y}_{i.}}$ con $S_{\bar{y}_{i.}} = \sqrt{\dfrac{MS_E}{n}}$ .

Se debe observar que en todas las comparaciones sólo se usa un valor crítico.

Para calcular el valor crítico $q_\alpha(a,f)$, existen tablas (Ver Apéndice: Tabla de puntos Porcentuales de la Estadística de Amplitud Studentizada, Douglas C. Montgomery, año 1991); en donde $\alpha$ es el nivel de significancia para valores de 0.01 y 0.05, "$a$" es el número de tratamientos y $f$ es el número de grados de libertad del error.

**Procedimiento**

- Se calcula el valor crítico de todas las comparaciones por pares.
- Se obtiene el error estándar de cada promedio.
- Obtener el $T_\alpha$.
- Calcular la diferencia de las medias y realizar las comparaciones con el valor crítico.
- Hacer las conclusiones.

**Ejemplo 8**

Considerando la información obtenida en el ejemplo del porcentaje de carbón, se desea saber que parejas de medias son significativamente diferentes.

**Datos**

$\alpha$ = 0.05 , $MS_E$ = 13.458 , $a$ = 4 , $N$ = 16 , $n$ = 4, grados de libertad del error = 12

$\bar{y}_{1.}$ = 27.25 , $\bar{y}_{2.}$ = 33.50 , $\bar{y}_{3.}$ = 40.50 , $\bar{y}_{4.}$ = 43.25

**Solución**

- **Se calcula el valor crítico de todas las comparaciones por pares.**

  $q_\alpha(a,f) = q_{0.05}(4,12) = 4.20$

- **Se obtiene el error estándar de cada promedio.**

  $$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{n}} = \sqrt{\frac{13.458}{4}} = 1.834$$

- **Obtener el $T_\alpha$.**

  $T_\alpha = q_\alpha(a,f)\, S_{\bar{y}_{i.}} = (4.20)(1.834) = 7.70$

- **Calcular la diferencia de las medias y realizar las comparaciones con el valor crítico.**

  1 vrs 2 : $|\bar{y}_{1.} - \bar{y}_{2.}| = |\,27.25 - 33.50\,| = |\,-6.25\,| = 6.25 < 7.70\ (T_\alpha)$

  1 vrs 3 : $|\bar{y}_{1.} - \bar{y}_{3.}| = |\,27.25 - 40.50\,| = |\,-13.25\,| = 13.25 > 7.70\ (T_\alpha)$ *

  1 vrs 4 : $|\bar{y}_{1.} - \bar{y}_{4.}| = |\,27.25 - 43.25\,| = |\,-16.00\,| = 16 > 7.70\ (T_\alpha)$ *

  2 vrs 3 : $|\bar{y}_{2.} - \bar{y}_{3.}| = |\,33.50 - 40.50\,| = |\,-7.00\,| = 7 < 7.70\ (T_\alpha)$

  2 vrs 4 : $|\bar{y}_{2.} - \bar{y}_{4.}| = |\,33.50 - 43.25\,| = |\,-9.75\,| = 9.75 > 7.70\ (T_\alpha)$ *

  3 vrs 4 : $|\bar{y}_{3.} - \bar{y}_{4.}| = |\,40.50 - 43.25\,| = |-2.75\,| = 2.75 < 7.70\ (T_\alpha)$

  Dos medias son significativamente diferentes si $|\bar{y}_{i.} - \bar{y}_{j.}| > T_\alpha$ .

- **Conclusiones**

a) Como el valor absoluto de las diferencias de las medias uno y tres, uno y cuatro, dos y cuatro resultó ser mayor que el $T_\alpha$ (7.70) (*); entonces se dice que existe diferencias significativas entre las medias uno con la tres y la cuatro , la dos con la cuatro.

b) Por el contrario, el valor absoluto de las demás diferencias resultaron ser menor a $T_\alpha$, entonces se dice que no existe diferencias significativas entre las medias uno y dos, dos y tres, tres y cuatro.

## 7.3 COMPARACIÓN DE TRATAMIENTOS CON UN CONTROL.

Si en un experimento el analista o experimentador le interesa comparar una media específica con las $a$-1 medias del experimento, esta media específica se llama control y por lo tanto, sólo debe realizarse $a$-1 comparaciones. Existe un procedimiento para llevar a cabo esta comparación y fue desarrollado por Dunnett.

Si suponemos que el tratamiento "$a$" es el control, entonces las hipótesis que se desean probar son:

$H_o$: $\mu_i = \mu_a$

$H_1$: $\mu_i \neq \mu_a$ para $i = 1,2,\ldots, a-1$

Este procedimiento es una modificación de la prueba t. Para cada hipótesis se deben calcular las diferencias que se observan en las medias muéstrales; de la siguiente manera: $|\bar{y}_{i.} - \bar{y}_{a.}|$ $\quad i = 1,2,\ldots,a-1$.

La hipótesis nula $H_o: \mu_i = \mu_a$ se rechaza con un nivel $\alpha$, si:

$$|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a-1,f)\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_a}\right)}$$ **Caso Desbalanceado**

$$|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a-1,f)\sqrt{\frac{2MS_E}{n}}$$ **Caso Balanceado**

donde $d_\alpha(a-1,f)$ se encuentra en la tabla de Dunnett (Ver Apéndice: Tabla de Valores Críticos para la Prueba de Dunnett de Comparación de Tratamientos con un Control, Douglas C. Montgomery, año 1991), con $\alpha$ que constituye el nivel de significancia conjunta asociado a las $a$-1 pruebas y $f$ los grados de libertad del error.

**Procedimiento**

- Encontrar el factor de comparación en la tabla Dunnett.
- Calcular las diferencias de las medias de tratamiento con el tratamiento control y realizar las comparaciones con el factor de comparación.
- Hacer conclusiones.

**Ejemplo 9**

Supongamos que nos interesa comparar la media del tratamiento cuatro (control) con las demás medias del ejemplo del porcentaje de carbón.

**Datos**

$\alpha$ = 0.05 , $MS_E$ = 13.458 , $N$ = 16 , $a$ = 4 , $n$ = 4 , $f$ = 12 , $a$ - 1 = 3

$\bar{y}_{1.}$ = 27.25 , $\bar{y}_{2.}$ = 33.50 , $\bar{y}_{3.}$ = 40.5 , $\bar{y}_{4.}$ = 43.25

**Solución**

- **Tomando un $\alpha$ = 0.05 y $f$ = 12 se obtiene el factor de comparación de la tabla de Dunnett.**

  $d_\alpha(a\text{-}1,f) = d_{0.05}(4\text{-}1,12) = d_{0.05}(3,12) =$ 2.68 entonces

  $d_\alpha(a\text{-}1,f)\sqrt{\dfrac{2MS_E}{n}} = 2.68\sqrt{\dfrac{2(13.458)}{4}} = 2.68(2.59) = 6.95$

- **Calculando la diferencia de las medias con la media del tratamiento control y realizando las comparaciones con el factor de comparación.**

  1 vrs 4 : $|\bar{y}_{1.} - \bar{y}_{4.}| = |27.25 - 43.25| = |-16.00| > 6.95$ *

  2 vrs 4 : $|\bar{y}_{2.} - \bar{y}_{4.}| = |33.50 - 43.25| = |-9.75| > 6.95$ *

  3 vrs 4 : $|\bar{y}_{3.} - \bar{y}_{4.}| = |40.50 - 43.25| = |-2.75| < 6.95$

Dos medias se consideran significativamente diferentes si: $|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a\text{-}1,f)\sqrt{\dfrac{2MS_E}{n}}$

- **Conclusiones**
  a) Como el valor absoluto de la diferencia de las medias uno con cuatro y dos con cuatro (tratamiento cuatro es el tratamiento control) son mayores que el factor de comparación, entonces se concluye que existe diferencia significativa entre el porcentaje de carbón uno y cuatro, y entre el dos y cuatro; es decir, $\mu_1 \neq \mu_4$ y $\mu_2 \neq \mu_4$.
  b) Además, el valor absoluto de la diferencia de la media tres y cuatro resultó ser menor que el factor de comparación, entonces se concluye que no existe diferencia significativa entre el porcentaje de carbón tres y cuatro; es decir $\mu_3 = \mu_4$.

## 8. MODELO DE EFECTOS ALEATORIOS

El modelo estadístico lineal es:

$$y_{ij} = \mu + \iota_i + \varepsilon_{ij} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$ donde $\iota_i$ y $\varepsilon_{ij}$ son variables aleatorias.

En este modelo carece de sentido probar las hipótesis que están relacionadas con los efectos de los tratamientos individuales; por lo tanto, se deben probar las siguientes hipótesis.

$H_0 : \sigma^2_\iota = 0$

$H_1 : \sigma^2_\iota > 0$

Esto quiere decir, que si $H_0$ es verdadera todos los tratamientos son idénticos y que si $H_1$ es verdadera, existe variabilidad entre los tratamientos.

Para probar estas hipótesis se requiere que las $\{\varepsilon_{ij}\}$ sean NID(0,$\sigma^2$), que las $\{\iota_i\}$ sean NID(0, $\sigma^2_\tau$) y además $\iota_i$ , $\varepsilon_{ij}$ sean variables aleatorias independientes.

La suma total de cuadrados aún es válida $SS_T = SS_{Tratamientos} + SS_E$ ; es decir, que la variabilidad total en las observaciones se descompone en la variación entre los tratamientos ($SS_{Tratamientos}$) y en la variación dentro de los tratamientos ($SS_E$).

Si $\iota_i$ tiene una varianza $\sigma^2_\tau$ y es independiente de $\varepsilon_{ij}$ , la varianza de cualquier observación viene dada por:

$V(y_{ij}) = \sigma^2_\tau + \sigma^2$, donde $\sigma^2_\tau$ y $\sigma^2$ se conocen como componentes de varianza y por lo tanto, el modelo lineal descrito anteriormente recibe el nombre de **Modelo de Efectos Aleatorios.**

Recordando que $\frac{SS_E}{\sigma^2}$ tiene una distribución ji-cuadrada con $N$-$a$ grados de libertad y si la hipótesis nula es verdadera $\frac{SS_{Tratamientos}}{\sigma^2}$ tiene una distribución ji-cuadrada con $a$-1 grados de libertad; además que ambas variables son aleatorias independientes. Y si la hipótesis nula $\sigma^2_\tau = 0$ es verdadera entonces la razón:

$$F_o = \frac{\dfrac{SS_{Tratamientos}}{a-1}}{\dfrac{SS_E}{N-a}} = \frac{MS_{Tratamientos}}{MS_E}$$

tiene una distribución F con $a$-1 y $N$-$a$ grados de libertad.

Otro aspecto de interés es conocer los valores esperados de las medias de cuadrados para poder describir completamente el procedimiento de prueba. Los cuales están dados por:

$E(MS_{Tratamientos}) = \sigma^2 + n\sigma_\tau^2$

$E(MS_E) \quad = \sigma^2$

(Ver Duglas C. Montgomery, año 1991, Página 73).

Al analizar el valor esperado de las medias de cuadrados, se observa que si: Ho es verdadera, tanto el numerador como el denominador de la estadística de prueba ($F_o$) son estimadores insesgados de $\sigma^2$ y si $H_1$ es verdadera, el valor esperado del numerador es mayor que el del denominador.

Por lo tanto, se debe rechazar $H_o$ para valores grandes de $F_o$. En consecuencia hay que rechazar $H_o$, si $F_o > F_{\alpha,a-1,N-a}$.

El procedimiento para realizar los cálculos matemáticos y la Tabla de Análisis de Varianza es idéntico al Modelo de Efectos Fijos. Sin embargo, existe diferencia al momento de plantear las hipótesis y al llevar a cabo las conclusiones; ya que éstas son aplicadas a toda la población de tratamientos.

Se considera importante estimar las componentes de varianza ($\sigma_\tau^2$ y $\sigma^2$), y el procedimiento utilizado se conoce como "Método de Análisis de Varianza", ya que utiliza los renglones de la tabla de Análisis de Varianza; y consiste en igualar los valores esperados de las medias de cuadrados con su correspondiente valor observado en la Tabla de Análisis de Varianza, para luego despejar las componentes de varianza; por lo tanto, se tiene:

$MS_{Tratamientos} = \sigma^2 + n\ \sigma_\tau^2$

$MS_E = \sigma^2$

entonces $\hat{\sigma}^2 = MS_E$ y $\hat{\sigma}_\tau^2 = \dfrac{MS_{Tratamientos} - MS_E}{n}$, para el **caso desbalanceado**, debe tomarse $n$ como $n_0 = \dfrac{1}{a-1}\left[\sum_{i=1}^{a} n_i - \dfrac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}\right]$

Si alguna estimación de las componentes de varianza es negativa, se debe tomar en cuenta las siguientes opciones:

a) Aceptar la estimación y utilizarla como evidencia de que el valor verdadero de la componente de varianza es cero.
b) Volver a estimarla usando algún método que siempre produzca estimaciones positivas.
c) Considerarla como evidencia de que el Modelo Lineal propuesto es incorrecto y reexaminar el problema.

**Ejemplo 10**

Se llevó a cabo un experimento en el cual se seleccionaron cinco tipos de gasolinas de forma aleatoria (A,B,C,D y E) y se les midió el número de octanos que poseían, obteniéndose de cada tipo cuatro observaciones. A continuación se presentan los datos obtenidos:

| Gasolina | Observaciones | | | | $y_{i.}$ |
|---|---|---|---|---|---|
| A | 91.7 | 91.2 | 90.9 | 90.6 | 364.4 |
| B | 91.7 | 91.9 | 90.9 | 90.9 | 365.4 |
| C | 92.4 | 91.2 | 91.6 | 91.0 | 366.2 |
| D | 91.8 | 92.2 | 92.0 | 91.4 | 367.4 |
| E | 93.1 | 92.9 | 92.4 | 92.4 | 370.8 |

$y_{..} = 1834.2$

**Solución**

Por tratarse de un experimento de Modelo de Efectos Aleatorios, las hipótesis a probar son:

$H_0$ : $\sigma_\tau^2 = 0$ No existe variabilidad en los tratamientos.

$H_1$ : $\sigma_\tau^2 > 0$ Existe variabilidad en los tratamientos.

**Variable Respuesta: Efecto de las gasolinas.**

**El significado verbal es:**

**$H_0$ :** No existe diferencia en el efecto que produce cada tipo de gasolina en relación al número de octanos que contienen.

**$H_1$:** Existe diferencia en el efecto que produce cada tipo de gasolina en relación al número de octanos que contienen.

Los cálculos de sumas de cuadrados para formar la tabla de Análisis de Varianza son similares al Modelo de Efectos Fijos.

**Datos**

$a = 5$ , $n = 4$ , $N = 20$ , $i = 1,2,3,4,5$ , $j = 1,2,3,4$

**Cálculos Matemáticos**

**Medias de los Tratamientos**

$$\bar{y}_{i.} = \frac{y_{i.}}{n}$$

$$\bar{y}_{1.} = \frac{364.4}{4} = 91.10 \quad , \quad \bar{y}_{2.} = \frac{365.2}{4} = 91.35 \quad , \quad \bar{y}_{3.} = \frac{366.2}{4} = 91.55$$

$$\bar{y}_{4.} = \frac{367.4}{4} = 91.85 \quad , \quad \bar{y}_{5.} = \frac{370.8}{4} = 92.70 \quad , \quad \bar{y}_{..} = \frac{1834.2}{20} = 91.71$$

**Sumas de Cuadrados**

$SS_T = \sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}^2 - \frac{y_{..}^2}{N}$ = [ $(91.7)^2+(91.2)^2+(90.9)^2+\ldots..+(92.4)^2+(92.4)^2$] - $\frac{(1834.2)^2}{20}$

= 168223.96 - $\frac{(1834.2)^2}{20}$

= 168223.96 - 168214.48

$SS_T$ = 9.478

$SS_{Tratamientos} = \frac{\sum_{i=1}^{5} y_{i.}^2}{n} - \frac{y_{..}^2}{N} = \frac{(364.4)^2 + (365.4)^2 + (366.2)^2 + (367.4)^2 + (370.8)^2}{4} - \frac{(1834.2)^2}{20}$

= 168220.59 - $\frac{(1834.2)^2}{20}$

= 168220.59 - 168214.48

$SS_{Tratamientos}$ = 6.11

$SS_E = SS_T - SS_{Tratamientos}$

$SS_E$ = 9.478 - 6.11

$SS_E$ = 3.368

**Medias de Cuadrados**

$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} = \frac{6.11}{5-1} = \frac{6.11}{4}$ = 1.5275

$MS_E = \frac{SS_E}{N-a} = \frac{3.368}{20-5} = \frac{3.368}{15}$ = 0.224

**Estadística**

$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{1.5275}{0.224}$ = 6.82

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Gasolinas | 6.11 | 4 | 1.5275 | 6.82 |
| Error | 3.368 | 15 | 0.224 | |
| Total | 9.478 | 19 | | |

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 4 grados de libertad ($a$-1) en el numerador y 15 grados de libertad ($N$-$a$) en el denominador.

$F_{\alpha,a-1,N-a} = F_{0.05,4,15} = 3.06$

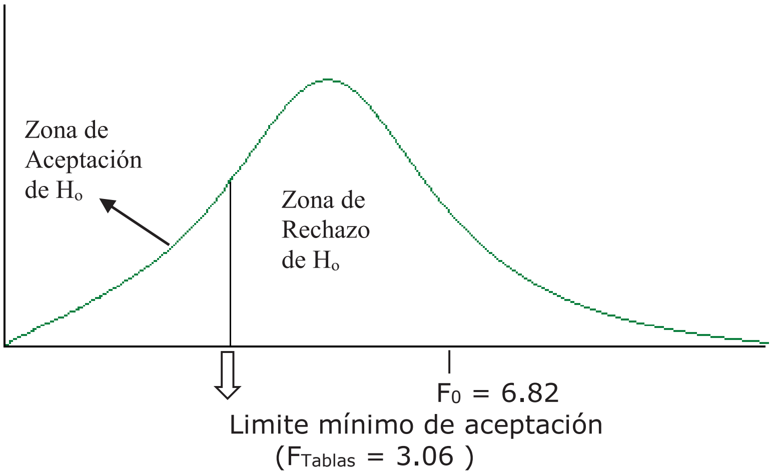
Comparando el $F_0$ calculado en el Análisis de Varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

6.82 > 3.06

Por tanto se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$).

También se puede observar gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

Existe diferencia en el efecto que produce cada tipo de gasolina en relación al número de octanos que contienen.

La estimación para los componentes de varianza y la estimación de la variación de cualquier observación de la muestra son:

$\hat{\sigma}^2 = MS_E$

$= 0.224$

$$\hat{\sigma}_\tau^2 = \frac{MS_{Tratamientos} - MS_E}{n} = \frac{1.5275 - 0.224}{4}$$

$= 0.326$

$\hat{V}(y_{ij}) = \hat{\sigma}_\tau^2 + \hat{\sigma}^2 = 0.326 + 0.224$

$= 0.549$

Además, se puede encontrar un intervalo de confianza para $\sigma^2$. Si las observaciones son independientes y están normalmente distribuidas, $\frac{(N-a)MS_E}{\sigma^2}$ tiene una distribución $\chi^2_{N-a}$.

Por lo tanto: P( $\chi^2_{1-(\frac{\alpha}{2}),N-a} \leq \frac{(N-a)MS_E}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},N-a}$ ) = 1 - $\alpha$

Luego un intervalo de confianza para $\sigma^2$ a un nivel del 100(1 - $\alpha$ )% es:

$$\frac{(N-a)MS_E}{\chi^2_{\frac{\alpha}{2},N-a}} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi^2_{1-(\frac{\alpha}{2}),N-a}}$$ **(Ver apéndice (4))**

Por otra parte, no es posible construir un intervalo de confianza para $\sigma^2_\tau$, por no poderse obtener una expresión cerrada para la distribución de la combinación lineal de variables aleatorias que se forman.

Pero es muy fácil encontrar una expresión exacta para un intervalo de confianza para la expresión $\frac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2}$. Esta expresión indica la proporción de la varianza de una observación; ya que V($y_{ij}$) = $\sigma^2_\tau$ + $\sigma^2$, que resulta de la diferencia entre los tratamientos.

En el diseño balanceado para desarrollar este intervalo, se debe notar que $MS_{Tratamientos}$ y $MS_E$ son variables aleatorias independientes. Además se puede demostrar que :

$$\frac{\dfrac{MS_{Tratamientos}}{n\sigma^2_\tau + \sigma^2}}{\dfrac{MS_E}{\sigma^2}} \approx F_{\alpha,a-1,N-a}$$

Luego

$$P\left( F_{1-\frac{\alpha}{2},a-1,N-a} \leq \frac{MS_{Tratamiento}}{MS_E} \frac{\sigma^2}{n\sigma^2_\tau + \sigma^2} \leq F_{\frac{\alpha}{2},a-1,N-a} \right) = 1 - \alpha$$

Al reordenar esta expresión se obtiene: **(Ver apéndice (5))**.

P( L $\leq \frac{\sigma^2_\tau}{\sigma^2} \leq$ U ) = 1 - $\alpha$ , donde

$$L = \frac{1}{n}\left( \frac{MS_{Tratamientos}}{MS_E} \frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} - 1 \right)$$

$U = \frac{1}{n}\left(\frac{MS_{Tratamientos}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}} - 1\right)$; es decir, L y U son limites inferior y superior de confianza al 100(1-α)% respectivamente, para $\frac{\sigma_\tau^2}{\sigma^2}$.

El valor de $F_{1-\left(\frac{\alpha}{2}\right),a-1,N-a}$ no se encuentra directamente en la tabla F; por lo tanto, se debe calcular de la siguiente manera: $F_{1-\left(\frac{\alpha}{2}\right),a-1,N-a} = \frac{1}{F_{\frac{\alpha}{2},N-a,a-1}}$

En consecuencia, el intervalo de confianza del 100(1-α)% para $\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$ está dado por:

$$\frac{L}{1+L} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leq \frac{U}{1+U}$$ **(Ver apéndice (6))**

**Ejemplo 11**

Para llevar a cabo una aplicación de las fórmulas antes planteadas, consideremos el ejemplo 10, el de tipos de gasolina.

**Datos**

$a$ = 5 , $n$ = 4 , $N$ = 20 , $\chi^2_{\frac{0.05}{2},20-5} = \chi^2_{0.025,15}$ = 27.49

$MS_E$ = 0.224 , α = 0.05 , $\chi^2_{1-(\frac{0.05}{2}),20-5} = \chi^2_{0.975,15}$ = 6.27

**Intervalo de confianza para σ².**

$$\frac{(N-a)MS_E}{\chi^2_{\frac{\alpha}{2},N-a}} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi^2_{1-(\frac{\alpha}{2}),N-a}}$$

**Sustituyendo**

$$\frac{(20-5)\,(0.224)}{\chi^2_{\frac{0.05}{2},20-5}} \leq \sigma^2 \leq \frac{(20-5)(0.224)}{\chi^2_{1-(\frac{0.05}{2}),20-5}}$$

$$\frac{3.36}{\chi^2_{0.025,15}} \leq \sigma^2 \leq \frac{3.36}{\chi^2_{0..975,15}}$$

$$\frac{3.36}{27.49} \leq \sigma^2 \leq \frac{3.36}{6.27}$$

$$0.122 \leq \sigma^2 \leq 0.536$$

Significa que la variabilidad del efecto total de los tipos de gasolina se encuentra entre el 12% y el 53.6%.

**Intervalo de confianza para** $\dfrac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$ **es:**

$$\frac{L}{1+L} \le \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \le \frac{U}{1+U}$$

**Datos**

$a = 5$ , $n = 4$ , $N = 20$ , $MS_E = 0.224$ , $\alpha = 0.05$ , $MS_{Tratamientos} = 1.5275$

$$F_{\frac{\varepsilon}{2},a-1,N-a} = F_{\frac{0.05}{2},5-1,20-5} = F_{0.025,4,15} = 3.80$$

$$F_{1-(\frac{\varepsilon}{2}),a-1,N-a} = F_{1-\frac{0.05}{2},5-1,20-5} = F_{0.975,4,15} = \frac{1}{F_{0.025,15,4}} = \frac{1}{8.66} = 0.115473441$$

**Calculando los límites inferior y superior de confianza**

$$L = \frac{1}{n}\left(\frac{MS_{tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}}-1\right) = \frac{1}{4}\left(\frac{1.5275}{0.224}\frac{1}{3.80}-1\right) = \frac{1}{4}(1.7945-1) = \frac{1}{4}(0.7945)$$

L = 0.1986

$$U = \frac{1}{n}\left(\frac{MS_{tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}-1\right) = \frac{1}{4}\left(\frac{1.5275}{0.224}\frac{1}{0.115473}-1\right) = \frac{1}{4}(59.05445-1)$$

$$= \frac{1}{4}(58.05445) = 14.51$$

U = 14.51

**Obteniendo el intervalo**

$$\frac{L}{1+L} \le \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \le \frac{U}{1+U}$$

**Sustituyendo se tiene:**

$$\frac{0.1986}{1+0.1986} \le \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \le \frac{14.51}{1+14.51}$$

$$\frac{0.1986}{1.1986} \le \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \le \frac{14.51}{15.51}$$

$$0.1657 \le \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \le 0.9355$$

Significa que la variabilidad de la proporción de una observación entre los tipos de gasolina se justifica entre 16% y 93% de la varianza observada en el efecto que produce.

**Cuadro Comparativo del Modelo de Efectos Fijos y el Modelo de Efectos Aleatorios**

| Aspecto | Modelo de Efectos Fijos | Modelo de Efectos Aleatorios |
|---|---|---|
| Modelo | $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ <br> $\sum_{i=1}^{a} \tau_i = 0$ | $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ <br> $\tau_i \sim N(0,\sigma_\tau)$ |
| Efecto de $\tau$ | Contrastes desconocidos | Variables aleatorias |
| Forma de selección de los tratamientos a considerar en el Experimento. | El experimentador selecciona específicamente los "*a*" tratamientos que intervienen en el experimento. | La selección de los "*a*" tratamientos se hace muestreando al azar dentro de cierta población de tratamientos. |
| Planteamiento de Hipótesis | $H_o : \mu_1 = \mu_2 = \ldots\ldots = \mu_a$ <br> $H_1 : \mu_i \neq \mu_j$, para al menos un par (i,j) <br> ó <br> $H_o: \tau_1 = \tau_2 = \ldots = \tau_a = 0$ <br> $H_1: \tau_i \neq 0$ para al menos un i $i=1,2,\ldots,a$ | $H_o: \sigma^2_\tau = 0$ <br> $H_1: \sigma^2_\tau > 0$ |
| Se pretende estimar | $\tau$ | $\sigma^2_\tau$ |
| Análisis de Varianza | Igual | Igual |
| Conclusiones | Se hacen solamente para los tratamientos involucrados en el experimento; es decir, las conclusiones son válidas respecto a los niveles del factor considerados de manera específica en el análisis. | Se hacen para todos los tratamientos que conforman la población de donde fueron tomados los tratamientos; es decir, que las conclusiones pueden extenderse a la población de los niveles del factor. |

## 9. SELECCIÓN DEL TAMAÑO MUESTRAL

La selección del tamaño de la muestra es una decisión muy importante en todo problema de diseño experimental; es decir, determinar el número de réplicas que deben hacerse en el experimento para cada tratamiento.

Si al experimentador le interesan los efectos pequeños en lugar de grandes, generalmente se requiere más réplicas por cada tratamiento. Esta metodología se puede aplicar en general a situaciones más complicadas, en que aquí se plantee para diseños de un sólo factor.

A continuación se discutirán algunos métodos para determinar el tamaño adecuado de la muestra.

## 9.1 MODELO DE EFECTOS FIJOS

### 9.1.1 Curvas Características de Operaciones

Las curvas características son herramientas útiles para seleccionar el número de réplicas, de tal manera que su diseño sea sensible a diferencias potenciales entre los tratamientos.

Concretamente, las curvas características de operación es una gráfica de la probabilidad del error tipo II de una prueba estadística, para un tamaño de muestra particular, contra el parámetro que refleja la extensión en la cual la hipótesis nula es falsa.

La probabilidad de error tipo II en el caso de muestras del mismo tamaño en cada tratamiento (caso balanceado); está reflejado en el poder de la prueba, que se define de la siguiente manera:

$$1 - \beta = P\{\text{Rechazar } H_o \,/\, H_o \text{ es falsa}\}$$
$$= P\{F_o > F_{\alpha,a-1,N-a} \,/\, H_o \text{ es falsa}\}$$

Para evaluar la probabilidad establecida, se necesita conocer la distribución de la estadística $F_o$, en el caso que la hipótesis nula sea falsa. Se puede demostrar que, si $H_o$ es falsa, la estadística $F_o = \dfrac{MS_{Tratamientos}}{MS_E}$ se distribuye como una variable aleatoria F no centrada con $a$-1 y $N$-$a$ grados de libertad y un parámetro de centralización $\delta$. Si $\delta$=0, la distribución F no centrada se transforma en la usual distribución F (centrada).

Las gráficas de las curvas características se utilizan para calcular la probabilidad del poder de la prueba. En las cuales se encuentra la probabilidad de error tipo II ($\beta$) contra el parámetro $\varphi$, el cual está definido de la siguiente manera:

$\varphi^2 = \dfrac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$, en donde el valor $\varphi^2$ está relacionado con el parámetro de descentralización $\delta$.

Las Curvas Características tienen valores disponibles de $\alpha$ = 0.01 y $\alpha$ = 0.05, y para diversos valores de los grados de libertad del numerador ($V_1$) y del denominador ($V_2$).

Para utilizar las Curvas Características de Operación, el experimentador debe especificar el valor de $\varphi$; para el Modelo de Efectos Fijos. Esto a menudo resulta difícil, pero una forma de determinar $\varphi$ es elegir los valores de las medias de tratamiento para los cuales se desea rechazar $H_o$ con una probabilidad alta.

Es decir, que $\mu_1, \mu_2, \ldots, \mu_a$ son las medias de los tratamientos propuestos, el valor de $\varphi$ se encontrará utilizando la ecuación anterior, en donde $\iota_i = \mu_i - \bar{\mu}$ y $\bar{\mu} = \frac{1}{a}\sum_{i=1}^{a}\mu_i$ es el promedio de las medias individuales de tratamiento. También, se necesita una estimación para $\sigma^2$, la cual puede obtenerse en base a experiencias pasadas, una estimación propuesta o en experimentos previos se deberá usar $MS_E$.

Si no se tiene seguridad del valor de $\sigma^2$, se toman valores posibles de $\sigma^2$ en un intervalo y se determina el tamaño de la muestra estudiando el efecto que tiene el parámetro $\sigma^2$ sobre el tamaño de la muestra antes de tomar una decisión final.

**Procedimiento**

- Definir el $\alpha$.
- Encontrar el promedio de las medias individuales de tratamiento.
- Encontrar los efectos de los tratamientos.
- Encontrar $\varphi^2$.
- Encontrar la aproximación de n utilizando las curvas características.
- Hacer conclusión.

**Ejemplo 12**

Suponga que $\mu_1 = 50$ , $\mu_2 = 60$ , $\mu_3 = 50$ y $\mu_4 = 60$ son las medias de cuatro poblaciones normales. ¿Cuántas observaciones deben tomarse de cada población para que la probabilidad de rechazar la hipótesis nula de igualdad de medias sea al menos de 0.90?. Suponga que $\alpha = 0.05$ y que una estimación razonable de la varianza del error es $\sigma^2 = 25$.

**Solución**

**Datos**

$\mu_1 = 50$ , $\mu_2 = 60$ , $\mu_3 = 50$ , $\mu_4 = 60$ , $\alpha = 0.05$ , $\sigma^2 = 25$ , $1\text{-}\beta = 0.90$

- **Encontrando el promedio de los tratamientos.**

$$\sum_{i=1}^{a} \mu_i = \sum_{i=1}^{4} \mu_i = 50 + 60 + 50 + 60 = 220$$

$$\bar{\mu} = \frac{1}{4}\sum_{i=1}^{4} \mu_i = \frac{1}{4}(220) = 55$$

- **Calculando los efectos de los tratamientos.**

$\iota_i = \mu_i - \bar{\mu}$

$\iota_1 = \mu_1 - \bar{\mu} = 50 - 55 = -5$ $\qquad \iota_2 = \mu_2 - \bar{\mu} = 60 - 55 = 5$

$\iota_3 = \mu_3 - \bar{\mu} = 50 - 55 = -5$ $\qquad \iota_4 = \mu_4 - \bar{\mu} = 60 - 55 = 5$

Por lo tanto $\sum_{i=1}^{4} \tau_i^2 = (-5)^2 + (5)^2 + (-5)^2 + (5)^2 = 100$

$$\varphi^2 = \frac{n\sum_{i=1}^{4}\tau_i^2}{4(25)} = \frac{n(100)}{100} = n$$

- **Se utilizará la curva característica de operación con los siguientes datos:**

  $a - 1 = 4 - 1 = 3$ (grados de libertad del numerador $V_1$)

  $N - 4 = 4\ (n\text{-}1)$ (grados de libertad del denominador $V_2$)

  $\varphi^2 = n$ , $\alpha = 0.05$ , $a = 4$

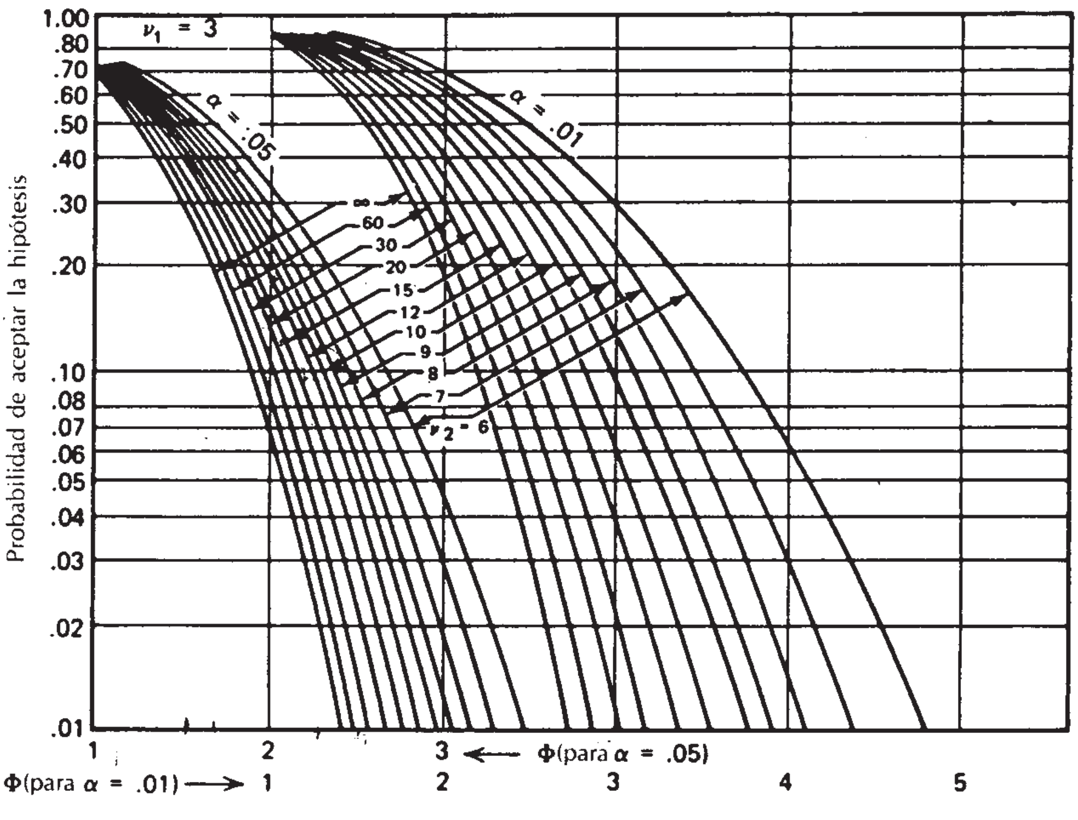
  Tomando una primera aproximación para el número de réplicas de n=3, entonces se obtiene:

  $V_1 = 3$ , $V_2 = 4(3\text{-}1) = 4(2) = 8$ , $\varphi^2 = 3$ , $\varphi = 1.73$

  Los valores que se van ha utilizar en la gráfica de esta primera aproximación para encontrar la probabilidad de error tipo II ($\beta$) son:

  $\alpha = 0.05$ , $V_1 = 3$ , $V_2 = 8$ , $\varphi = 1.73$

  La gráfica para estos valores es la siguiente:



De la gráfica anterior se obtiene que $\beta \approx 0.38$, lo cual se concluye que $n$=3 réplicas no son suficientes, porque la potencia de la prueba es aproximadamente $1 - \beta \approx 1-0.38 = 0.62$, lo cual resulta menor a la requerida que es de 0.90.

Utilizando el procedimiento anterior se obtiene la siguiente tabla.

| $n$ | $\varphi^2$ | $\varphi$ | $V_2 = a(n-1)$ | $\beta$ | Poder $(1-\beta)$ |
|---|---|---|---|---|---|
| 3 | 3 | 1.71 | 8 | 0.38 | 0.62 |
| 4 | 4 | 2 | 12 | 0.18 | 0.82 |
| 5 | 5 | 2.24 | 16 | 0.09 | 0.91 |

- **Conclusión**

Para $n$= 5 se obtiene que el poder de la prueba es 0.91 y se estableció que 0.90 era aceptable; por lo tanto, se requiere al menos 5 réplicas para tener una prueba con la potencia deseada.

**Ejemplo 13**

Suponga que cinco medias han sido comparadas en un Análisis de Varianza con $\alpha = 0.01$.
Al experimentador le gustaría conocer cuántas réplicas debe correr, si le parece importante rechazar $H_0$ con una probabilidad de al menos 0.90, si $\sum_{I=1}^{5} \frac{\tau_i^2}{\sigma^2} = 5.0$.

**Solución**

**Datos**

$\alpha = 0.01$ , $a = 5$ , $\sum_{I=1}^{5} \frac{\tau_i^2}{\sigma^2} = 5.0$

- **Calculando $\varphi^2$**

$$\varphi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2} = \frac{5n}{5} = n$$

- **Se utilizará las curvas características de operación del Modelo de Efectos Fijos con los siguientes datos:**

  a – 1 = 5 – 1 = 4 (grados de libertad del numerador $V_1$)

  $N$ – 5 = 5 ($n$-1) (grados de libertad del denominador $V_2$)

  $\varphi^2 = n$ , $\alpha = 0.01$ , a = 5

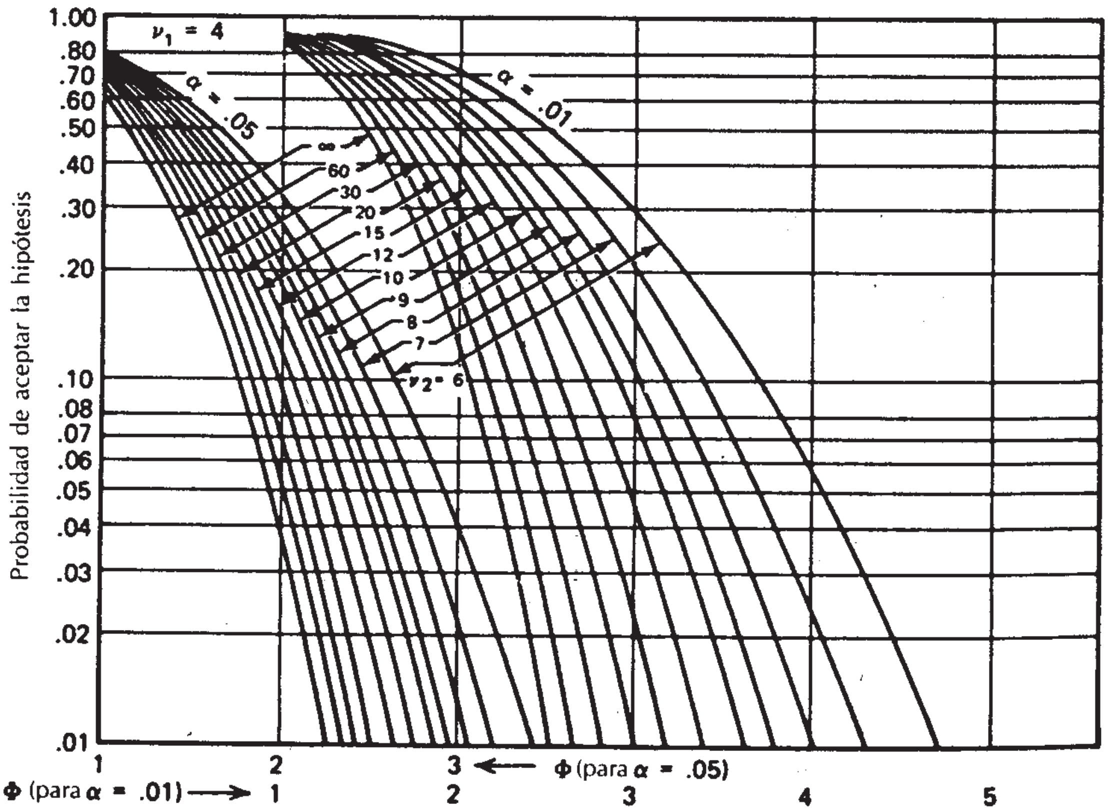
  Tomando una primera aproximación para el número de réplicas de $n$=4, entonces se obtiene:

$V_1 = 4$ , $V_2 = 5(4\text{-}1) = 5(3) = 15$ , $\varphi^2 = 4$ , $\varphi = 2$

Los valores que se van ha utilizar en la gráfica de esta primera aproximación para encontrar la probabilidad de error tipo II ($\beta$) son:

$\alpha = 0.01$ , $V_1 = 4$ , $V_2 = 15$ , $\varphi = 2$

La gráfica para estos valores es la siguiente:




De la gráfica anterior se obtiene que $\beta \approx 0.38$, lo cual se concluye que $n$=4 réplicas no son suficientes, porque la potencia de la prueba es aproximadamente $1 - \beta \approx 1\text{-}0.38 = 0.62$, lo cual resulta menor a la requerida que es de 0.90.

Utilizando el procedimiento anterior se obtiene la siguiente tabla.

| $n$ | $\varphi^2$ | $\varphi$ | $V_2 = a(n\text{-}1)$ | $\beta$ | Poder ($1\text{-}\beta$) |
|---|---|---|---|---|---|
| 4 | 4 | 2 | 15 | 0.38 | 0.62 |
| 5 | 5 | 2.24 | 20 | 0.18 | 0.82 |
| 6 | 6 | 2.45 | 25 | 0.06 | 0.94 |

- **Conclusión**

Para $n$ = 6 se obtiene que el poder de la prueba es 0.94 y se estableció que 0.90 era aceptable; por lo tanto, se requiere al menos 6 réplicas para tener una prueba con la potencia deseada.

El inconveniente de este enfoque es que resulta difícil obtener el conjunto de medias de tratamiento sobre el cual se basará la decisión sobre el tamaño de la muestra.

Para solucionar este inconveniente se debe seleccionar el tamaño de la muestra, de manera que la hipótesis nula se rechace; si la diferencia entre cualquier par de medias de tratamiento excede a un valor específico y cuando mucho sea D, se puede demostrar que el valor mínimo de $\varphi^2$ es: $\varphi^2 = \dfrac{nD^2}{2a\sigma^2}$ .

Ya que éste es el valor mínimo de $\varphi^2$, el tamaño de las muestras que se obtienen con las curvas características toma un valor conservador; es decir, que proporciona una potencia igual, al menos a la especificada por el experimentador.

**Ejemplo 14**

Del ejemplo anterior, el experimentador desea rechazar la hipótesis nula con una probabilidad de 0.90 como mínimo, si la diferencia entre cualquier par de medias de tratamiento es a lo sumo igual a 10. Suponiendo que $\sigma^2$ =25.

El valor mínimo de $\varphi^2$ es: $\varphi^2 = \dfrac{n(10)^2}{2(4)(25)} = 0.5n$ y al realizar el análisis como en el ejemplo anterior se concluye que $n$=6 réplicas son necesarias para obtener el nivel de sensibilidad deseado cuando $\alpha$ = 0.01.

### 9.1.2 Especificación de un incremento en la Desviación Estándar.

La desviación estándar de una observación escogida al azar es $\sigma$; si no existen diferencias entre las medias de tratamiento.

Pero si las medias de tratamiento son diferentes, la desviación estándar de una observación seleccionada al azar, viene dada por: $\sqrt{\sigma^2 + \sum_{i=1}^{a} \dfrac{\tau_i^2}{a}}$

Si se toma *p* como el porcentaje de que no debe sobrepasar la desviación estándar de una observación; es decir, que equivale a tomar

$$\frac{\sqrt{\sigma^2 + \sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\sigma} = 1 + 0.01p \quad \text{, donde } p = \text{porcentaje}.$$

después de hacer cálculos algebraicos en la ecuación anterior se tiene:

$$\frac{\sqrt{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\sigma} = \sqrt{(1+0.01p)^2 - 1}$$ **(Ver Apéndice (7))**

Ahora, multiplicando por $\sqrt{n}$

$$\sqrt{n}\left(\frac{\sqrt{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\sigma}\right) = \left(\sqrt{(1+0.01p)^2 - 1}\right)\sqrt{n}$$

Entonces $\varphi = \dfrac{\sqrt{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\frac{\sigma}{\sqrt{n}}} = \sqrt{(1+0.01p)^2 - 1}\,\sqrt{n}$

Se calcula $\varphi$ para un valor específico de $p$; y se utilizan las Curvas Características de Operación (Modelo de Efectos Fijos), para determinar el valor de $\beta$ y luego encontrar el poder de la prueba para establecer el tamaño de muestra requerido que cumpla con la probabilidad establecida.

**Procedimiento.**

- Encontrar el valor de $\varphi$
- Encontrar la aproximación de n utilizando las curvas características.
- Hacer conclusión.

**Ejemplo 15**

Supongamos que en el ejemplo de los programas de estudio se desea detectar un incremento del 25% en la desviación estándar con probabilidad mínima de 80% y que $\alpha$ = 0.05. ¿Cuál es el tamaño de muestra requerido?.

**Solución.**

$p = 25$ , $a = 3$, $N = an$ , $N - a = an - a = a(n-1)$

- **Encontrar el valor de $\varphi$**

$$\varphi = \sqrt{(1+0.01p)^2 - 1}\,\sqrt{n} = \sqrt{(1+0.01(25))^2 - 1}\,\sqrt{n} = \sqrt{0.5625}\,\sqrt{n} = 0.75\sqrt{n}$$

- **Se utilizará la curva característica de operación con los siguientes datos:**

  $a$ – 1 = 3 – 1 = 2 (grados de libertad del numerador $V_1$)

  $N$ – 3 = 3 ($n$-1) (grados de libertad del denominador $V_2$)

  $\varphi = 0.75\sqrt{n}$ , $\alpha = 0.05$ , $a = 3$

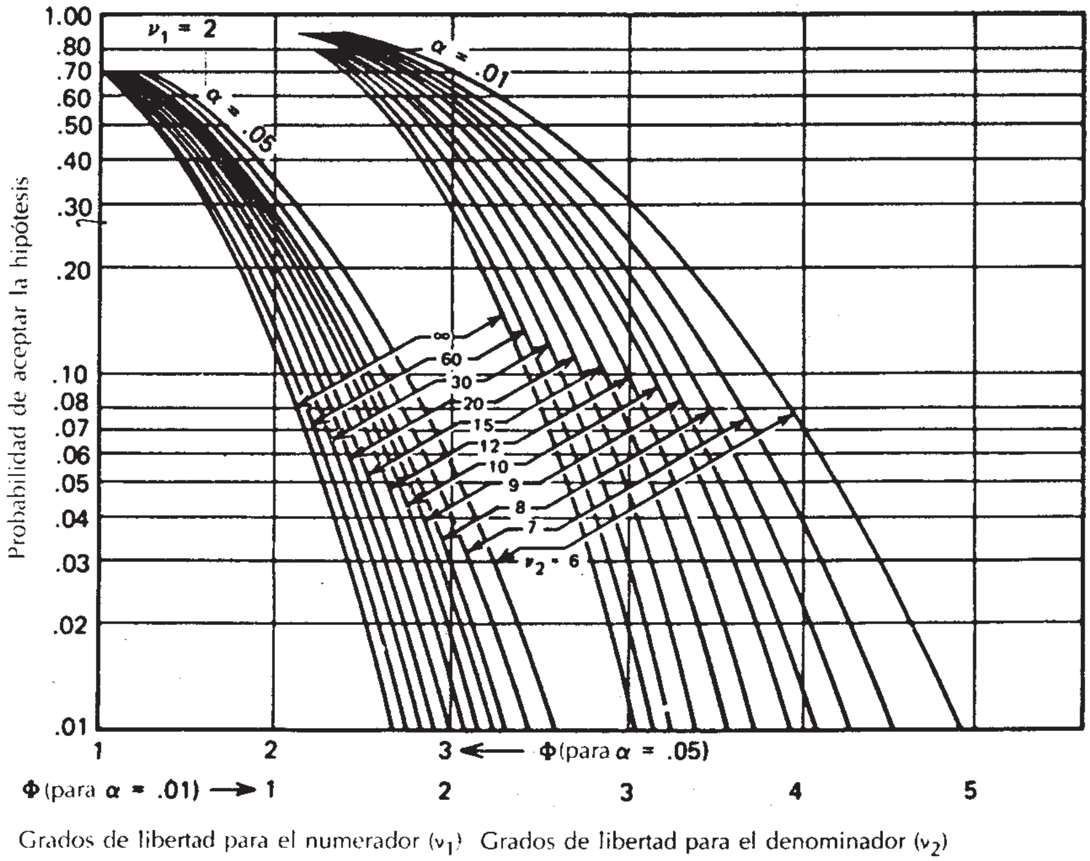
Tomando una primera aproximación para el número de réplicas de n=5, entonces se obtiene:

$V_1 = 2$ , $V_2 = 3(5-1) = 3(4) = 12$ , $\varphi = 0.75\sqrt{5} = 1.68$

Los valores que se van ha utilizar en la gráfica de ésta primera aproximación para encontrar la probabilidad de error tipo II ($\beta$) son:

$\alpha = 0.05$ , $V_1 = 2$ , $V_2 = 12$ , $\varphi = 1.68$

La gráfica para estos valores es la siguiente:



De la gráfica anterior se obtiene que $\beta \approx 0.40$, lo cual se concluye que $n$= 5 réplicas no son suficientes, porque la potencia de la prueba es aproximadamente $1 - \beta \approx 1-0.40 = 0.60$, lo cual resulta menor a la requerida que es de 0.80.

Utilizando el procedimiento anterior se obtiene la siguiente tabla.

| $n$ | $\varphi$ | $V_2 = a(n-1)$ | $\beta$ | Poder $(1-\beta)$ |
|---|---|---|---|---|
| 5 | 1.68 | 12 | 0.40 | 0.60 |
| 6 | 1.84 | 15 | 0.25 | 0.75 |
| 7 | 1.98 | 18 | 0.18 | 0.82 |

- **Conclusión**

  Para $n$=7 se obtiene que el poder de la prueba es 0.82 y se estableció que 0.80 era aceptable; por lo tanto, se requiere al menos 7 réplicas para tener una prueba con la potencia deseada.

## 9.2 MODELOS DE EFECTOS ALEATORIOS

### 9.2.1 Curvas Características de Operación.

La probabilidad de error tipo II en el caso de muestras del mismo tamaño en cada tratamiento, está reflejado en el poder de la prueba, que se define de la siguiente manera:

$\beta = 1 - P\{\text{Rechazar } H_o / H_o \text{ es falsa}\}$

$= 1 - P\{F_o > F_{\alpha,a-1,N-a} / \sigma_\tau^2 > 0\}$

Para llevar acabo la evaluación de esta probabilidad, es necesario saber la distribución de la estadística $F_0 = \dfrac{MS_{Tratamientos}}{MS_E}$ cuando la hipótesis alternativa es verdadera. Además, se puede demostrar que si $H_1$ es verdadera ($\sigma_\tau^2 > 0$), el $F_0$ tiene una distribución F, con $a$-1 y $N$-$a$ grados de libertad.

La sensibilidad de la prueba puede determinarse utilizando las Curvas Características de Operación para el Análisis de Varianza del Modelo de Efectos Aleatorios. Estas curvas están elaboradas para diversos valores de grados de libertad del numerador y denominador, y para valores de $\alpha = 0.01$ y $\alpha = 0.05$.

En estas curvas se encuentra graficada la probabilidad de error tipo II, contra el parámetro $\lambda$, el cual está definido por:

$$\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}}$$

Este parámetro contiene los dos parámetros desconocidos $\sigma_{\tau}^{2}$ y $\sigma^2$, pero pueden ser estimados. Al tener una idea de cuanta variabilidad es importante detectar en la población de tratamientos se puede estimar $\sigma_{\tau}^{2}$ y $\sigma^2$ a través de un juicio o experiencias anteriores. Algunas veces es conveniente definir $\sigma_{\tau}^{2}$ en función de la razón $\frac{\sigma_{\tau}^{2}}{\sigma^{2}}$; porque $\sigma^2$ suele conocerse.

**Procedimiento.**

- Definir el $\alpha$
- Encontrar $\lambda$
- Encontrar la aproximación de n utilizando las curvas características.
- Hacer conclusión.

**Ejemplo 16**

Retomando el ejemplo planteado de los tipos de gasolinas. Si suponemos que la diferencia entre los tipos de gasolina es lo suficientemente grande, como para producir un incremento en la desviación estándar del doble de la varianza del error en cualquier observación. ¿Cuál debe ser el tamaño de la muestra para detectar este aumento con probabilidad mínima de 0.85?.

**Solución**

**Datos**

$a$= 5 , $\sigma_{\tau}^{2}$ = 2 $\sigma^2$ , 1 - $\beta$ = 0.85 , $N= an$

- **Definir el** $\alpha$ **=** 0.01
- **Encontrar** $\lambda$

$$\lambda = \sqrt{1+\frac{n\sigma_{\tau}^{2}}{\sigma^{2}}} = \sqrt{1+\frac{n(2\sigma^{2})}{\sigma^{2}}} = \sqrt{1+2n}$$

- **Se utilizará la curva característica de operación con los siguientes datos:**

  a – 1 = 5 – 1 = 4 (grados de libertad del numerador $V_1$)

  N – 5 = 5 (n-1) (grados de libertad del denominador $V_2$)

$$\lambda = \sqrt{1+2n}$$

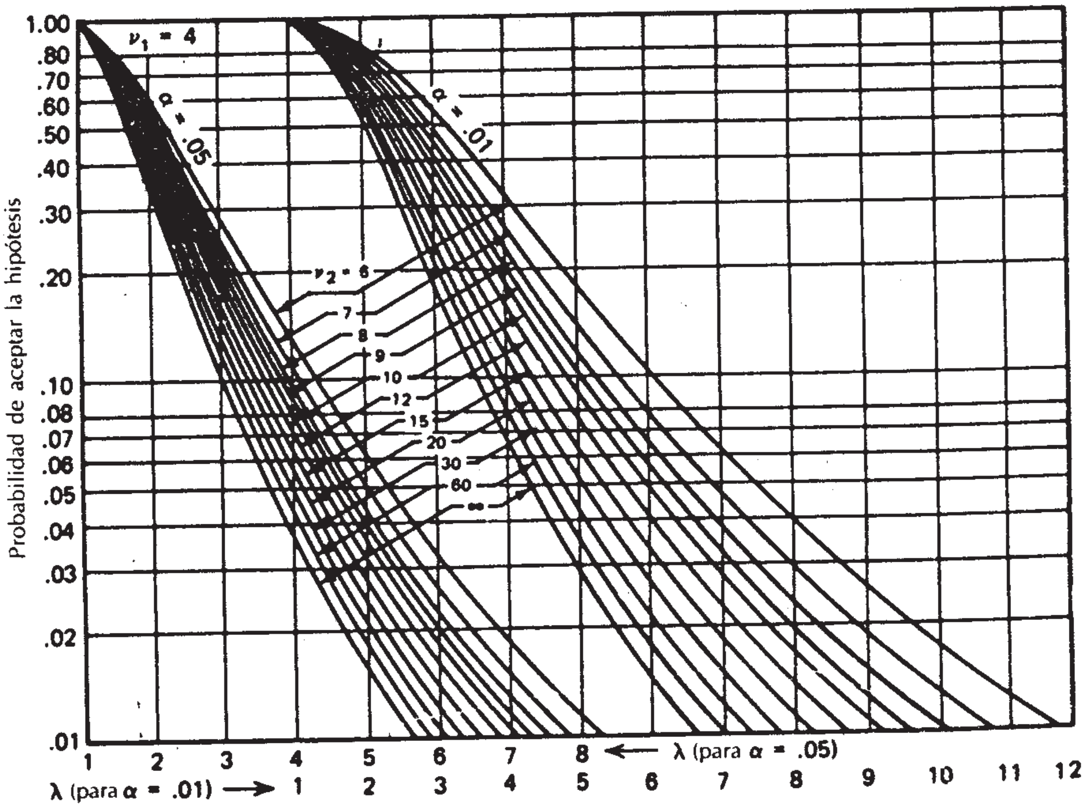
Tomando una primera aproximación para el número de réplicas de n=2, entonces se obtiene:

$V_1$ = 4 , $V_2$ = 5(2-1) = 5(1) = 5 , $\lambda = \sqrt{1+2n} = \sqrt{1+2(2)} = \sqrt{5}$

Los valores que se van ha utilizar en la gráfica de esta primera aproximación para encontrar la probabilidad de error tipo II ($\beta$) son:

$\alpha$ = 0.01 , $V_1$ = 4 , $V_2$ = 5 , $\lambda$ = 2.24

La gráfica para estos valores es la siguiente:




De la gráfica anterior se obtiene que $\beta \approx 0.80$, lo cual se concluye que $n$ = 2 réplicas no son suficientes, porque la potencia de la prueba es aproximadamente $1 - \beta \approx 1\text{-}0.80 = 0.20$, lo cual resulta menor a la requerida que es de 0.85.

Utilizando el procedimiento anterior se obtiene la siguiente tabla.

| $n$ | $\lambda$ | $V_2 = a(n\text{-}1)$ | $\beta$ | Poder ($1\text{-}\beta$) |
|---|---|---|---|---|
| 2 | 2.24 | 5 | 0.8 | 0.2 |
| 3 | 2.64 | 10 | 0.5 | 0.5 |
| 4 | 3.00 | 15 | 0.3 | 0.7 |
| 5 | 3.32 | 20 | 0.2 | 0.8 |
| 6 | 3.60 | 25 | 0.12 | 0.88 |

- **Conclusión**

Para $n$= 6 se obtiene que el poder de la prueba es 0.88 y se estableció que 0.85 era aceptable; por lo tanto, se requiere al menos 6 réplicas para tener una prueba con la potencia deseada.

### 9.2.2 Especificación de un incremento en la Desviación Estándar.

De manera similar el de efectos fijos, la desviación estándar de una observación escogida al azar es $\sigma$; si no existen diferencias entre las medias de tratamiento.

Pero si las medias de tratamiento son diferentes, la desviación estándar de una observación seleccionada al azar, viene dada por: $\sqrt{\sigma^2 + \sigma_\tau^2}$ . Si se decidiera rechazar la hipótesis nula para un incremento superior a un porcentaje fijado p, en la desviación estándar de una observación, entonces: $\frac{\sqrt{\sigma^2 + \sigma_\tau^2}}{\sigma}$ = 1 + 0.01p por lo tanto, $\frac{\sigma_\tau^2}{\sigma^2}$ = (1 + 0.01p)$^2$ – 1 **(Ver Apéndice(8))**, luego sustituyendo en la ecuación del parámetro y utilizando el método de las curvas características $\left(\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}}\right)$ se tiene $\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}} = \sqrt{1 + n[(1+0.01p)^2 - 1]}$

Se calcula $\lambda$ para un valor específico de $p$; luego se utilizan las Curvas Características de Operación del Análisis de Varianza del Modelo de Efectos Aleatorios para determinar el valor de $\beta$, y luego encontrar el poder de la prueba para establecer el tamaño de muestra requerido que cumpla con el poder de la prueba establecida.

**Procedimiento**

- Encontrar el valor de $\lambda$
- Encontrar la aproximación de n utilizando las curvas características.
- Hacer conclusión.

**Ejemplo 17**

Si suponemos que en el ejemplo de los tipos de gasolina se desea observar un incremento del 25% en la desviación estándar con una probabilidad mínima de 80% y que $\alpha$ = 0.05. ¿Cuál es el tamaño de la muestra requerido?.

**Solución.**

**Datos**

$p$ = 25 , $a$ = 5, $\alpha$ = 0.05 , 1 - $\beta$ = 0.80 , $N$= $an$

- **Encontrando el valor de $\lambda$**

$$\lambda = \sqrt{1+n[(1+0.01p)^2 - 1]} = \sqrt{1+n[(1+0.01(25))^2 - 1]} = \sqrt{1+n(0.5625)}$$

- **Se utilizará la curva característica de operación con los siguientes datos:**

$a$ – 1 = 5 – 1 = 4 (grados de libertad del numerador $V_1$)

$N$– 5 = 5 ($n$-1) (grados de libertad del denominador $V_2$)

$\lambda = \sqrt{1+n(0.5625)}$ , $\alpha$ = 0.05, $a$ = 5

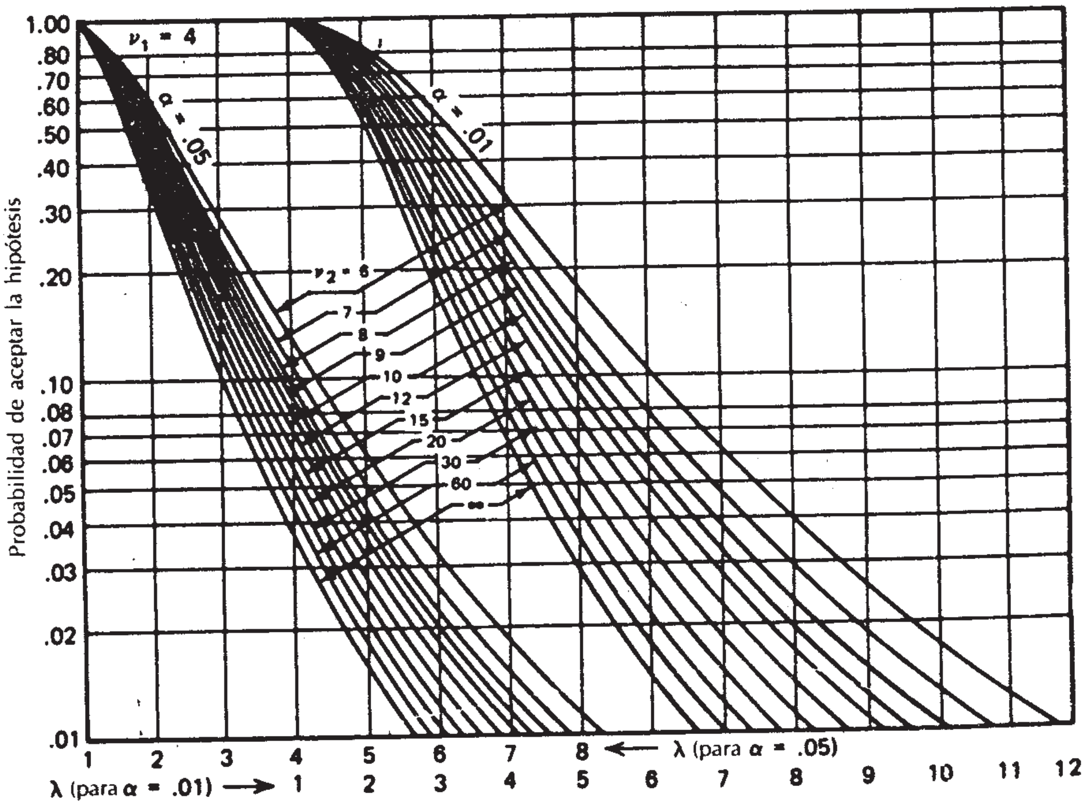
Tomando una primera aproximación para el número de réplicas de $n$=4, entonces se obtiene:

$V_1 = 4$ , $V_2 = 5(4-1) = 5(3) = 15$ , $\lambda = \sqrt{1+n(0.5625)} = \sqrt{1+4(0.5625)} = 1.80$

Los valores que se van ha utilizar en la gráfica de esta primera aproximación para encontrar la probabilidad de error tipo II ($\beta$) son:

$\alpha = 0.05$ , $V_1 = 4$ , $V_2 = 15$ , $\lambda = 1.80$

La gráfica para estos valores es la siguiente:




De la gráfica anterior se obtiene que $\beta \approx 0.50$, lo cual se concluye que $n$ = 4 réplicas no son suficientes, porque la potencia de la prueba es aproximadamente $1 - \beta \approx 1-0.50 = 0.50$, lo cual resulta menor a la requerida que es de 0.80.

Utilizando el procedimiento anterior se obtiene la siguiente tabla.

| $n$ | $\lambda$ | $V_2 = a(n\text{-}1)$ | $\beta$ | Poder $(1\text{-}\beta)$ |
|---|---|---|---|---|
| 4 | 1.80 | 15 | 0.50 | 0.50 |
| 5 | 1.95 | 20 | 0.35 | 0.65 |
| 6 | 2.09 | 25 | 0.30 | 0.70 |
| 7 | 2.22 | 30 | 0.25 | 0.75 |
| 8 | 2.34 | 35 | 0.18 | 0.82 |

- **Conclusión**

  Para $n$= 8 se obtiene que el poder de la prueba es 0.82 y se estableció que 0.80 era aceptable; por lo tanto, se requiere al menos 8 réplicas para tener una prueba con la potencia deseada.

## 9.3 MÉTODO DE ESTIMACIÓN POR INTERVALO DE CONFIANZA.

El método supone que los resultados finales que el experimentador desea reportar aparecen en términos de intervalos de confianza y que está dispuesto a especificar de antemano la anchura de estos intervalos.

Para ello, se debe utilizar la ecuación que determina el error estándar de cada promedio y la distribución t, encontrándose la precisión del intervalo de la siguiente manera.

$$\pm\ t_{\frac{\alpha}{2},N-a}\sqrt{\frac{2MS_E}{n}}$$

**Ejemplo 18**

Retomando el ejemplo 4, del porcentaje de carbón.

Encontrar el número de réplicas necesarias si se desea que el intervalo de confianza del 95% para la diferencia de la resistencia media a la tensión entre dos tratamientos sea de $\pm$ 5, y que una estimación a priori de $\sigma$ es 3.

**Solución**

**Datos**

$\alpha$ = 0.05 , $N$ = 16 , $a$ = 4 , $MS_E \approx \sigma^2$ =9 , $\sigma$ = 3 , $N - a = a\,(n - 1)$

$$\pm\ t_{\frac{\alpha}{2},N-a}\sqrt{\frac{2MS_E}{n}}$$

$$\pm\ t_{\frac{0.05}{2},N-a}\sqrt{\frac{2(9)}{n}}$$

$\pm\ t_{0.025,N-a}\sqrt{\dfrac{18}{n}}$

Puesto que $N$ - $a$ = $a$ ($n$ – 1) y suponiendo que se propone $n$ = 6 réplicas , se tiene la precisión del intervalo:

$N$ - $a$ = 4(6-1) = 4(5)=20

$\pm\ t_{0.025,20}\sqrt{\dfrac{18}{6}}$

$\pm$ (2.086) $\sqrt{3}$

$\pm$ 3.61 este valor es más preciso que el propuesto (±5)

Suponiendo ahora $n$=5 réplicas, la precisión del intervalo es:

$N$ - $a$ = 4(5-1) = 4(4) = 16

$\pm\ t_{0.025,16}\sqrt{\dfrac{18}{5}}$

$\pm$ (2.120) $\sqrt{3.6}$

$\pm$ 4.02 este valor es más preciso que el propuesto (±5)

Suponiendo ahora $n$= 4 réplicas, la precisión del intervalo es:

$N$ - $a$ = 4(4-1) = 4(3) = 12

$\pm\ t_{0.025,12}\sqrt{\dfrac{18}{4}}$

$\pm$ (2.179) $\sqrt{4.5}$

$\pm$ 4.62 este valor es más preciso que el propuesto (±5)

Suponiendo ahora $n$= 3 réplicas, la precisión del intervalo es:

$N$ - $a$ = 4(3-1) = 4(2) = 8

$\pm\ t_{0.025,8}\sqrt{\dfrac{18}{3}}$

$\pm$ (2.306) $\sqrt{6}$

$\pm$ 5.65

**Conclusión**

Con $n$= 3 réplicas el intervalo calculado sobre pasa el valor de precisión establecido ($\pm$5), entonces el mínimo de tamaño de muestra que debe tomarse es de $n$= 3 ; ya que conduce a la precisión deseada.

## 10. CODIFICACIÓN DE LAS OBSERVACIONES

Existen problemas de Diseño de Experimento en donde los resultados que se obtienen (valores de las observaciones) suelen ser demasiado grandes o son valores con punto decimal, etc; y los cálculos Matemáticos para llevar acabo el Análisis de Varianza suelen ser tediosos o complicados en el caso que realicen manualmente, en este tipo de problemas se deben codificar los datos para simplificar los cálculos Matemáticos que se llevan acabo en el Análisis de Varianza. Esta codificación consiste en restar, sumar, dividir o multiplicar un valor constante a cada una de los valores de las observaciones o realizar cualquier operación aritmética que el experimentador o analista estime conveniente; en esta codificación puede combinarse una o más codificaciones a la vez, siempre y cuando se cumpla el objetivo de simplificar los cálculos Matemáticos. Es fácil comprobar que cualquier codificación que se realice en las observaciones no hace variar los resultados que se obtienen en la Tabla de Análisis de Varianza.

**Ejemplo 19**

En el ejemplo 10 se presentó la tabla siguiente:

| Gasolina | Observaciones | | | |
|---|---|---|---|---|
| A | 91.7 | 91.2 | 90.9 | 90.6 |
| B | 91.7 | 91.9 | 90.9 | 90.9 |
| C | 92.4 | 91.2 | 91.6 | 91.0 |
| D | 91.8 | 92.2 | 92.0 | 91.4 |
| E | 93.1 | 92.9 | 92.4 | 92.4 |

Si a esta tabla de datos se codifica restando 90 a cada una de las observaciones, nos queda la siguiente tabla:

| Gasolina | Observaciones | | | | $y_{i.}$ |
|---|---|---|---|---|---|
| A | 1.7 | 1.2 | 0.9 | 0.6 | 4.4 |
| B | 1.7 | 1.9 | 0.9 | 09 | 5.4 |
| C | 2.4 | 1.2 | 1.6 | 1.0 | 6.2 |
| D | 1.8 | 2.2 | 2.0 | 1.4 | 7.4 |
| E | 3.1 | 2.9 | 2.4 | 2.4 | 10.8 |

$y_{..} = 34.2$

Ahora se realizan los cálculos Matemáticos para llevar a cabo el Análisis de Varianza, de la misma forma como se hizo anteriormente:

**Datos**

$a$ = 5 , $n$ = 4 , $N$= 20 , $i$ = 1,2,3,4,5 , $j$ = 1,2,3,4

**Cálculos Matemáticos**

**Medias de los Tratamientos**

$$\bar{y}_{i.} = \frac{y_{i.}}{n}$$

$\bar{y}_{1.} = \frac{4.4}{4}$ = 1.1 , $\bar{y}_{2.} = \frac{5.4}{4}$ = 1.35 , $\bar{y}_{3.} = \frac{6.2}{4}$ = 1.55

$\bar{y}_{4.} = \frac{7.4}{4}$ = 1.85 , $\bar{y}_{5.} = \frac{10.8}{4}$ = 2.7 , $\bar{y}_{..} = \frac{34.2}{20}$ = 1.71

**Sumas de Cuadrados**

$SS_T = \sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}^2 - \frac{y_{..}^2}{N}$

$= 67.96 - \frac{(34.2)^2}{20}$

$= 67.96 - 58.482$

$SS_T = 9.478$

$SS_{Tratamientos} = \sum_{i=1}^{5} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$

$= \frac{(4.4)^2 + (5.4)^2 + (6.2)^2 + (7.4)^2 + (10.8)^2}{4} - \frac{(34.2)^2}{20}$

$= 64.59 - 58.482$

$SS_{Tratamientos} = 6.108$

$SS_E = SS_T - SS_{Tratamientos}$

$= 9.478 - 6.108$

$SS_E = 3.37$

**Medias de Cuadrados**

$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} = \frac{6.108}{5-1} = \frac{6.108}{4}$ = 1.527

$MS_E = \frac{SS_E}{N-a} = \frac{3.37}{20-5} = \frac{3.37}{15}$ = 0.224

**Estadística**

$$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{1.5275}{0.224} = 6.82$$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Gasolinas | 6.108 | 4 | 1.5275 | 6.82 |
| Error | 3.37 | 15 | 0.224 | |
| Total | 9.478 | 19 | | |

Como puede observarse los valores obtenidos en la Tabla de Análisis de Varianza anterior coinciden a los obtenidos en el ejemplo desarrollado sin codificar los datos; por lo tanto, la codificación de los datos no afecta los resultados de la Tabla de Análisis de Varianza; solamente facilita los cálculos Matemáticos que se utilizan para llevar a cabo la elaboración de la Tabla de Análisis de Varianza.

## 11. PROBLEMAS RESUELTOS

**PROBLEMA 1**

En un proceso químico de fabricación de sinter (un aglomerado), se cuenta con tres diferentes tipos de máquinas: A: manual , B: semiautomática y C: automática.
Se mide el contenido de impureza (%) de los productos de una semana y se obtienen los siguientes resultados.

| Tratamiento (nivel) | Observaciones |
|---|---|
| Máquina A | 3.09 , 3.18 , 0.92 , 2.10 , 1.95 , 3.27 |
| Máquina B | 2.19 , 1.92 , 4.65 , 2.16 , 3.15 , 2.88 , 4.23 , 4.85 , 3.90 , 3.13 ,4.02 |
| Máquina C | 3.16 , 4.92 , 3.01 , 4.38 , 5.04 , 3.59 , 4.62 , 2.21 , 3.69 , 2.83 ,4.17 3.62 , 3.81 , 4.65 , 3.97 , 3.61 |

Pruebe con un nivel de significancia del 5%, si existe diferencia en el contenido de impureza del sinter fabricado en cada máquina.

**Solución**

**Variable de estudio**: Porcentaje de Impureza en el sinter

$H_o$ : $\mu_A = \mu_B = \mu_C$ (No existe diferencia entre las tres máquinas)

$H_1$ : $\mu_A \neq \mu_B \neq \mu_C$ (Existe diferencia entre las tres máquinas)

**El significado verbal de las hipótesis es:**

$H_o$ : No existe diferencia significativa en el contenido de impureza del sinter fabricado en cada una de las máquinas.

$H_1$ : Existe diferencia significativa en el contenido de impureza del sinter fabricado en cada una de las máquinas.

Este experimento es un ejemplo de un Diseño Unifactorial Desbalanceado; porque cada tratamiento tiene un número diferente de observaciones.

**Datos**

Tratamientos: $a = 3$

Número de observaciones: $n_1 = 6$ , $n_2 = 11$ $n_3 = 16$

Número total de observaciones: $N = n_1+n_2+n_3 = 6+11+16 = 33$

$i = 1,2,3$ $j = 1,2,....,n_i$

## Cálculos Matemáticos

**Totales**

$y_{i.} = \sum_{j=1}^{n_i} y_{ij}$

$y_{1.} = \sum_{j=1}^{6} y_{ij}$ = 3.09 + 3.18 + ....+3.27 = 14.51

$y_{2.} = \sum_{j=1}^{11} y_{ij}$ = 2.19 + 1.92 + ...+4.02 = 37.08

$y_{3.} = \sum_{j=1}^{16} y_{ij}$ = 3.16 + 4.92 +...+3.61 = 61.28

$y_{..} = \sum_{i=1}^{3} y_{i.}$ = 14.51 + 37.08 + 61.28 = 112.87

**Medias de los Tratamientos**

$\bar{y}_{i.} = \frac{y_{i.}}{n_i}$

$\bar{y}_{1.} = \frac{y_{1.}}{6} = \frac{14.51}{6} = 2.42$ $\qquad$ $\bar{y}_{2.} = \frac{y_{2.}}{11} = \frac{37.08}{11} = 3.37$

$\bar{y}_{3.} = \frac{y_{3.}}{16} = \frac{61.28}{16} = 3.83$ $\qquad$ $\bar{y}_{..} = \frac{y_{..}}{33} = \frac{112.87}{33} = 3.42$

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} = [(3.09)^2+(3.18)^2+(0.92)^2+.....+(3.97)^2+(3.61)^2] - \frac{(112.87)^2}{33}$$

= 418.99 – 386.05

$SS_T$ = 32.937

$$SS_{Tratamientos} = \sum_{i=1}^{3} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} = \left[\frac{(14.51)^2}{6} + \frac{(37.08)^2}{11} + \frac{(61.28)^2}{16}\right] - \frac{(112.87)^2}{33}$$

= 294.78 - 386.05

$SS_{Tratamientos}$ = 8.73

$SS_E = SS_T - SS_{Tratamientos}$

= 32.937 - 8.73

$SS_E$ = 24.207

**Medias de Cuadrados**

$$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} = \frac{8.73}{2} = 4.365$$

$$MS_E = \frac{SS_E}{N-a} = \frac{24.207}{30} = 0.8069$$

**Estadística**

$$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{4.365}{0.756} = 5.77$$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Máquinas | 8.73 | 2 | 4.365 | 5.41 |
| Error (Dentro de Tratamientos) | 24.207 | 30 | 0.8069 | |
| Total | 32.937 | 32 | | |

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 2 grados de libertad ($a$-1) en el numerador y 30 grados de libertad ($N$-$a$) en denominador.

$F_{\alpha,a-1,N-a} = F_{0.05,2,30} = 3.32$

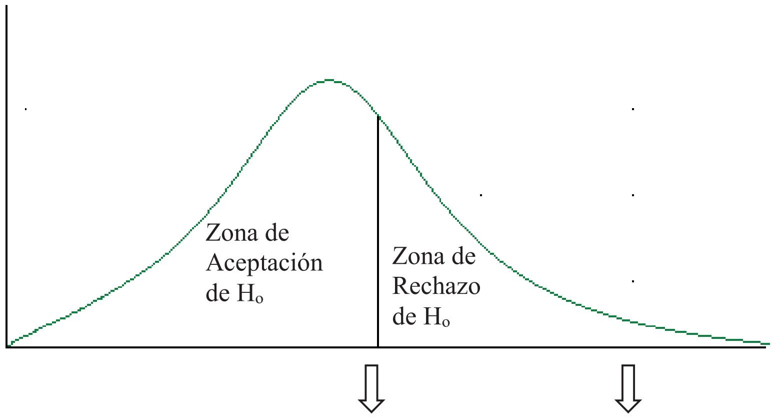
Comparando el $F_0$ calculado en el análisis de varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

5.77 > 3.32

Por tanto, se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$).

También, se puede observar gráficamente, de la siguiente manera:




Limite mínimo de aceptación $F_0 = 5.41$
($F_{Tablas} = 3.32$)

Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

Existe diferencia significativa en el contenido de impureza del sinter fabricado en cada una de las máquinas.

**PROBLEMA 2**

Como en el problema 1, la hipótesis nula fue rechazada; lo cual significa que existe diferencia significativas entre las tres máquinas.

Entonces se podría ser interés saber cuales son las parejas de medias que difieren. Para ello se utilizará el Método de comparación de Parejas de Medias de Tratamientos llamado Método de la Mínima Diferencia Significativa (LSD).

**Solución**

Por ser un Diseño Desbalanceado se aplicará la siguiente fórmula para encontrar el LSD

$$LSD = t_{\frac{\alpha}{2},N-a}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

**Datos.**

$\alpha = 0.05$ , $MS_E = 0.756$ , $N = 33$ , $n_1 = 6$ , $n_2 = 11$ , $n_3 = 16$ , $a = 3$

$\bar{y}_{1.} = 2.42$ , $\bar{y}_{2.} = 3.37$ , $\bar{y}_{3.} = 3.83$ , $t_{\frac{\alpha}{2},N-a} = t_{\frac{0.05}{2},33-3} = t_{0.025,30} = 2.042$

- **Encontrando el valor del LSD para cada una de las parejas de medias, con la fórmula establecida.**

Para $\bar{y}_{1.}$ y $\bar{y}_{2.}$ el LSD es: $(2.042)\sqrt{0.756\left(\frac{1}{6}+\frac{1}{11}\right)}=$ (2.042)(0.441278936) = 0.901

Para $\bar{y}_{1.}$ y $\bar{y}_{3.}$ el LSD es: $(2.042)\sqrt{0.756\left(\frac{1}{6}+\frac{1}{16}\right)}=$ (2.042)(0.4162329) = 0.849

Para $\bar{y}_{2.}$ y $\bar{y}_{3.}$ el LSD es: $(2.042)\sqrt{0.756\left(\frac{1}{11}+\frac{1}{16}\right)}=$ (2.042)(0.3405542)= 0.695

- **Calculando la diferencia de los promedios.**

1 vrs 2 : $|\bar{y}_{1.} - \bar{y}_{2.}|$ = |2.42 – 3.37| = |-0.95| > 0.901

1 vrs 3 : $|\bar{y}_{1.} - \bar{y}_{3.}|$ = |2.42 – 3.83| = |-1.41| > 0.849

2 vrs 3 : $|\bar{y}_{2.} - \bar{y}_{3.}|$ = |3.37 – 3.83| = |-0.46| < 0.695

Se dice que una pareja de medias difieren significativamente si el valor absoluto de las diferencias de los promedios de los tratamientos correspondientes es mayor que LSD encontrado para cada pareja de medias.

- **Conclusiones**

c) Se observa que la pareja de medias que no difiere significativamente son la media dos y la media tres; ya que |–0.46 | < 0.695; por lo tanto, no existe diferencia significativa entre la máquina semiautomática y la máquina automática.

d) La pareja de medias que difieren significativamente son la uno y la dos; como también la uno y la tres; ya que el valor absoluto de las diferencia de los promedios a resultado ser mayor que el valor encontrado del LSD correspondiente ; por lo tanto, existe diferencia significativa entre la máquina manual y la máquina semiautomática y también entre la máquina manual y la máquina automática.

**PROBLEMA 3**

Retomando los resultados del problema 1, se desea encontrar un intervalo de confianza del 95% para el contenido medio de impureza del sinter fabricado en la máquina automática.

**Solución**

Por ser un Diseño Desbalanceado y tratarse de un intervalo de confianza para una sola media se utilizará la siguiente fórmula:

$\bar{y}_{i.} \pm t_{\frac{\alpha}{2},N-a}\sqrt{\frac{MS_E}{n}}$, tomando el valor de $n = \dfrac{a}{\sum_{i=1}^{a}\frac{1}{n_i}}$

**Datos**

$\alpha$ = 0.05 , $MS_E$ = 0.756 , $N$ = 33 , $n_1$ = 6 , $n_2$ = 11 , $n_3$ = 16 , $a$ = 3

$\bar{y}_{1.}$ = 2.42 , $\bar{y}_{2.}$ = 3.37 , $\bar{y}_{3.}$ = 3.83 , $t_{\frac{\alpha}{2},N-a} = t_{\frac{0.05}{2},33-3} = t_{0.025,30} = 2.042$

$$n = \frac{3}{\sum_{i=1}^{3}\frac{1}{n_i}} = \frac{3}{\frac{1}{6}+\frac{1}{11}+\frac{1}{16}} = \frac{3}{0.320075757} = 9.37 \approx 9$$

**Sustituyendo**

$$\bar{y}_{3.} \pm t_{\frac{0.05}{2},33-3}\sqrt{\frac{MS_E}{9}}$$

$$3.83 \pm t_{0.025,30}\sqrt{\frac{0.756}{9}}$$

3.83 ± (2.042) (0.289827534)

3.83 ± 0.592

Por tanto, el intervalo buscado es: $3.238 \leq \mu_C \leq 4.422$; es decir, que se tiene el 95% de confianza que en 100 muestras del mismo tamaño tomado anteriormente ($n$=9) ; 95 de esas muestras la media del contenido de impureza del sinter fabricado en la máquina automática caerá dentro del intervalo de [3.238 , 4.422].

**PROBLEMA 4**

Si en el Diseño Experimental del Problema 1; se contara con nueve máquinas, tres de cada tipo y se seleccionaran aleatoriamente una de las tres de cada uno de los tipos para medir el contenido de impureza del sinter fabricado en cada máquina.

a) Escriba el nombre del Diseño Experimental a que se refiere el problema 1; así planteado.
b) Realizar el análisis del experimento detallando las diferencia de cómo fue analizado en el Problema 1.

**Solución**

a) El problema 1, así planteado se refiere a un Modelo de Efectos Aleatorios.
b) Como es un Diseño de Experimentos de Modelo de Efectos Aleatorios su análisis tendrá algunas diferencias como fue analizado en el problema 1.

**Las diferencias son las siguientes:**

i) El planteamiento de las hipótesis.

Las hipótesis en este caso se deben plantear en relación a la variabilidad entre las máquinas. Es decir.

$H_0$ : $\sigma^2_\tau = 0$ ( No existe variabilidad entre las máquinas),

$H_1$ : $\sigma^2_\tau > 0$ (Existe variabilidad entre las máquinas)

ii) En la conclusión ya que son aplicadas a todas máquinas de esos tipos.

Como el análisis de varianza nos indica que se rechazar Ho; entonces en este sentido se concluye que. Existe diferencia significativa en el contenido de impureza del sinter fabricado en cada uno de los tipos de máquinas.

**PROBLEMA 5**

Considerando el problema 1, como un Modelo de Efectos Aleatorios. Encontrar:

a) La estimación de los componentes de varianza y la estimación de la variación de cualquier observación de la muestra.
b) Un intervalo de confianza para $\sigma^2$ del 99%.

**Datos.**

$a = 3$ , $N = 33$ , $\chi^2_{\frac{0.01}{2},33-3} = \chi^2_{0.005,30} = 53.67$ , $MS_E = 0.756$ , $MS_{Tratamientos} = 4.365$

$\alpha = 0.01$ , $\chi^2_{1-(\frac{0.01}{2}),33-3} = \chi^2_{0.995,30} = 13.79$

**Solución a)**

**Estimación de los Componentes de Varianza.**

$\hat{\sigma}^2$ = $MS_E$

= 0.756

$$\hat{\sigma}_\tau^2 = \frac{MS_{Tratamientos} - MS_E}{n}$$

Como se trata de un Diseño Desbalanceado debe tomarse $n$ como

$$n_0 = \frac{1}{a-1}\left[\sum_{i=1}^{a} n_i - \frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}\right]$$

**Sustituyendo**

$$n_0 = \frac{1}{3-1}\left[33 - \frac{413}{33}\right] = \frac{1}{2}\left[33 - 12.515151\right] = \frac{1}{2}\ (20.484849) = 10.2424 \approx 10$$

$$\hat{\sigma}_\tau^2 = \frac{4.365 - 0.756}{10} = \frac{3.609}{10} = 0.361$$

$$\hat{V}(y_{ij}) = \hat{\sigma}_\tau^2 + \hat{\sigma}^2 = 0.361 + 0.756$$

$= 1.117$

**Solución b)**

**El Intervalo de confianza para $\sigma^2$ se encuentra de la siguiente manera.**

$$\frac{(N-a)MS_E}{\chi^2_{\frac{\alpha}{2},N-a}} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi^2_{1-(\frac{\alpha}{2}),N-a}}$$

**Sustituyendo**

$$\frac{(33-3)\ (0.756)}{\chi^2_{\frac{0.01}{2},33-3}} \leq \sigma^2 \leq \frac{(33-3)(0.756)}{\chi^2_{1-(\frac{0.01}{2}),33-3}}$$

$$\frac{22.68}{\chi^2_{0.005,30}} \leq \sigma^2 \leq \frac{22.68}{\chi^2_{0..995,30}}$$

$$\frac{22.68}{53.67} \leq \sigma^2 \leq \frac{22.68}{13.79}$$

$$0.422 \leq \sigma^2 \leq 1.645$$

Significa que la variabilidad del efecto total del contenido de impureza del sinter fabricado en los tipos de máquinas se encuentra entre el 42.2% y el 164.5%.

# UNIDAD PROGRAMÁTICA III:

# “ DISEÑOS POR BLOQUES ”

# DISEÑOS POR BLOQUES

## 1. DESCRIPCIÓN

En todo Diseño de Experimento se desea que el Error Experimental sea lo más pequeño posible, ya que refleja tanto el Error aleatorio como la variabilidad entre las unidades experimentales o grupos de Unidades Experimentales; es decir, se trata de sustraer del Error Experimental la variabilidad producida por las unidades experimentales o grupos de Unidades Experimentales.

Para lograr lo anterior, las unidades experimentales se deben distribuir aleatoriamente en grupos o bloques, de tal manera que dentro de un bloque sean relativamente homogéneas y que el número de ellas dentro de un bloque sea igual al número de tratamientos por investigar. A esta forma de organizar y distribuir las unidades Experimentales se le denomina **Diseño por Bloques.**

El principio de bloqueo se basa en que las unidades experimentales dentro de cada bloque o grupo deben ser parecidas entre si (que exista homogeneidad dentro del bloque) y que los bloques debieran ser diferentes entre si (que exista heterogeneidad entre los bloques); y además que las unidades experimentales se deben distribuir aleatoriamente en los grupos o bloques de tal manera que dentro de un bloque sean relativamente homogéneas; lo que significa que el bloqueo o agrupamiento del material experimental debe hacerse de tal forma que las unidades experimentales dentro de un bloque sean tan homogéneas como sea posible y que las diferencias entre las unidades experimentales sean explicadas, en mayor proporción, por las diferencias entre los bloques. Esta técnica es una generalización de la prueba de comparación de medias de muestras apareadas, donde cada pareja es un bloque. Para la validez del análisis, se supone que no existe interacción entre los bloques y los tratamientos.

En experimentos de campo, cuando no se tenga información segura sobre la uniformidad del terreno, siempre es preferible introducir bloques, que servirán para separar la variabilidad debido a dicha heterogeneidad.

Los Diseños que cumplen estas características y que serán objeto de estudio en esta unidad, se clasifican en:

1) Diseños Aleatorizados por Bloques Completos.

2) Diseños Aleatorizados por Bloques Incompletos Balanceados.

## 2. DISEÑOS ALEATORIZADOS POR BLOQUES COMPLETOS
## DESCRIPCIÓN

Este Diseño se da cuando cada bloque tiene tantas unidades Experimentales como Tratamientos, y todos los tratamientos son asignados al azar dentro de cada bloque.

En este tipo de Diseño la palabra **"Completo"** significa que todos los tratamientos son probados en cada bloque, y los bloques forman una unidad experimental más homogénea con la cual se comparan los tratamientos, y **"Aleatorizado"** porque dentro de cada bloque los tratamientos son asignados de forma aleatoria a las unidades Experimentales; es decir, que todas las Unidades Experimentales de un mismo bloque tienen la misma probabilidad de recibir cualquiera de los tratamientos. Esta estrategia de Diseño mejora efectivamente la precisión en las comparaciones al eliminar la variabilidad en los tratamientos.

El procedimiento general para este Diseño consiste en seleccionar "$b$" bloques y ejecutar una réplica completa del experimento en cada bloque, por lo tanto, habrá "$a$" observaciones (una por nivel del factor) en cada bloque. Se realiza una observación por tratamiento en cada bloque y el orden en que los tratamientos son asignados a las unidades experimentales dentro de cada bloque se hace aleatoriamente.

Este Diseño en la práctica se aplica en situaciones muy numerosas y su utilidad se puede detectar con mucha facilidad; es por ello, que es uno de los Diseños más ampliamente utilizados.

Entre algunas situaciones que pueden ser controladas sistemáticamente mediante el análisis por bloques, son las unidades de equipo de prueba o maquinaria; ya que tienen diferencias en sus características de operación y constituyen un factor típico que es necesario controlar, y además lotes de materia prima , personas o tiempo, también constituyen fuentes de variabilidad en un experimento.

**Ejemplo 1**

En la Escuela Nacional de Agricultura (ENA) se llevó a cabo un experimento para indicar los efectos relativos de tres fertilizantes diferentes en el rendimiento de una cierta variedad de fríjol de soya. Para lo anterior se dispuso de 18 parcelas diferentes y debido a la localización de éstas se consideró ventajoso agruparlas en bloques de tres parcelas cada uno. Después los tratamientos se asignaron al azar a las parcelas dentro de cada bloque. Al final de la época de la cosecha se obtuvieron los rendimientos promedios en quintales (qq.) por parcela dentro de cada bloque.

**Interpretación del Ejemplo 1**

El ejemplo planteado anteriormente es considerado como un Diseño Aleatorizado por Bloques Completos, ya que se dispone de 18 parcelas diferentes y por su localización se realizaron 6 grupos de tres parcelas cada uno, los cuales constituyeron los bloques, luego se distribuyen los tres tipos de fertilizante (tratamientos) de una forma aleatoria en las tres parcelas de cada uno de los bloques y de esta forma fue posible hacer un estudio del efecto de los tres fertilizantes en el rendimiento de fríjol de soya, disminuyendo así, el error experimental y la variabilidad de las parcelas en el experimento.

## 2.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

En el supuesto que se tienen "*a*" tratamientos y se han seleccionado "*b*" bloques y llevado a cabo una réplica completa del experimento en cada uno de los bloques. La representación típica de los datos para este tipo de Diseño es la siguiente:

| Tratamientos | Bloques | | | | | $y_{i.}$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | ...... | $b$ | |
| 1 | $y_{11}$ | $y_{12}$ | $y_{13}$ | ...... | $y_{1b}$ | $y_{1.}$ |
| 2 | $y_{21}$ | $y_{22}$ | $y_{23}$ | ...... | $y_{2b}$ | $y_{2.}$ |
| . | . | . | . | ...... | . | . |
| . | . | . | . | ...... | . | . |
| . | . | . | . | ...... | . | . |
| $a$ | $y_{a1}$ | $y_{a2}$ | $y_{a3}$ | ...... | $y_{ab}$ | $y_{a.}$ |
| $y_{.j}$ | $y_{.1}$ | $y_{.2}$ | $y_{.3}$ | ...... | $y_{.b}$ | $y_{..}$ |

Un valor de la tabla $y_{ij}$ representa la observación del tratamiento *i* en el j-ésimo bloque. En general habrán "$b$" observaciones en el tratamiento *i*.

donde:

$y_{i.}$ :Totales por renglones (tratamientos) o total de observaciones del tratamiento $i$.

$y_{.j}$ :Totales por columna (Bloques) o total de observaciones del bloque $j$.

$y_{..}$ :Total General o total de todas las observaciones.

$\bar{y}_{i.}$ : Promedio del Renglón(tratamiento) o promedio total de las observaciones del tratamiento $i$.

$\overline{y}_{.j}$ : Promedio de columnas (Bloques) o promedio de las observaciones del bloque $j$.

$\overline{y}_{..}$ : Promedio general (Promedio de todas las observaciones)

**Matemáticamente se expresan de la siguiente manera:**

$y_{i.} = \sum_{j=1}^{b} y_{ij}$ , $y_{.j} = \sum_{i=1}^{a} y_{ij}$ , $y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}$ ó $y_{..} = \sum_{i=1}^{a} y_{i.} = \sum_{j=1}^{b} y_{.j}$ , $\overline{y}_{i.} = \dfrac{y_{i.}}{b}$ , $\overline{y}_{.j} = \dfrac{y_{.j}}{a}$

$\overline{y}_{..} = \dfrac{y_{..}}{N}$ , $i$ = 1,2,.....,$a$ , $j$ = 1,2,.....,$b$

donde $N$ = $ab$, número total de observaciones.

## 2.2 MODELO ESTADÍSTICO

Se pueden escribir las observaciones de la tabla anterior por medio del siguiente Modelo Estadístico Lineal:

$$y_{ij} = \mu + \iota_i + \beta_j + \varepsilon_{ij} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

donde:

$y_{ij}$ : Es la observación de la variable respuesta medida en el i-ésimo tratamiento y j-ésimo bloque.

$\mu$ : Es la media general.

$\iota_i$ : Es el efecto del i-ésimo tratamiento.

$\beta_j$ : Es el efecto del j-ésimo bloque

$\varepsilon_{ij}$ : Es el término usual NID(0,$\sigma^2$) del error aleatorio.

$i$ : Variando de 1 hasta el número de Tratamientos ($a$).

$j$ : Variando de 1 hasta el número de Bloques ($b$).

Inicialmente se considera que los tratamientos y los bloques son factores fijos. Además, los efectos de tratamiento y de bloque se consideran como desviaciones de la media general, por lo tanto: $\sum_{i=1}^{a} \hat{\tau}_i = 0$ y $\sum_{j=1}^{b} \hat{\beta}_j = 0$.

## 2.3 SUMAS Y MEDIAS DE CUADRADOS

El Análisis de Varianza consiste en el estudio de la descomposición de la variabilidad total de los datos en sus partes que está compuesta.

Sea:

$SS_T$ : Suma total de cuadrados corregida.

$SS_{Tratamientos}$ : Suma de cuadrados debida a los tratamientos.

$SS_{Blques}$ : Suma de cuadrados debida a los bloques.

$SS_E$ : Suma de cuadrados debida al error

La suma total de cuadrados corregida se puede expresar de la siguiente forma:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2$$

Al descomponer esta sumatoria y después de hacer algunos pasos algebraicos y considerar que los tres términos que contienen productos cruzados son iguales a cero (Ver Douglas C. Montgomery, año 1991, Páginas 121 y 122) se obtiene lo siguiente:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2$$

Esta expresión representa una descomposición de la suma total de cuadrados y está formada por la suma de cuadrados de las diferencias entre los promedios de los tratamientos y el promedio general, que constituye una medida de la diferencia entre las medias de tratamiento. El siguiente término es la suma de cuadrados de la diferencia entre los promedios de los bloques y el promedio general, lo cual constituye una medida de la diferencia entre las medias de bloques, y el último término que es la doble sumatoria de las observaciones y la media general restando los promedios de los tratamientos y los de los bloques; constituye el error aleatorio.

Por lo tanto:

$$SS_T = SS_{Tratamientos} + SS_{Bloques} + SS_E$$

donde

$SS_T$ : Tiene $N$-1 grados de libertad, porque existen $N$ observaciones en total y un sólo parámetro a estimar que es $\mu$ .

$SS_{Tratamientos}$ : Tiene $a$-1 grados de libertad, porque existen " $a$ " niveles del factor y un sólo parámetro a estimar que es $\iota_i$ .

$SS_{Bloques}$ : Tiene b-1 grados de libertad, porque existen “ $a$ ” bloques y un sólo parámetro a estimar que es $\beta_j$ .

$SS_E$ : Tiene $(a$-1)($b$-1) grados de libertad, porque existen $ab$ celdas que proporcionan $ab$-1 grados de libertar y la suma de cuadrados del error no es mas que la suma de cuadrados entre las celdas, menos la suma de cuadrados de tratamiento y la suma de cuadrados de bloques; entonces los grados de libertad del error serán: $ab$-1-($a$-1)-($b$-1) = ($a$-1)($b$-1).

Los grados de libertad de la suma total debe ser igual a la suma de los grados de libertad de $SS_{Tratamientos}$ , $SS_{Bloques}$ y $SS_E$; es decir, $N$-1 = ($a$ -1)+( $b$-1)+( $a$-1)( $b$-1) = $ab$-1.

**Matemáticamente estas sumas se obtienen de la siguiente manera:**

$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N}$ ó $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2$ **(Ver Apéndice 1)**

$SS_{Tratamientos} = \sum_{i=1}^{a}\frac{y_{i.}^2}{b} - \frac{y_{..}^2}{N}$ ó $SS_{Tratamientos} = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$ **(Ver Apéndice 2)**

$SS_{Bloques} = \sum_{j=1}^{b}\frac{y_{.j}^2}{a} - \frac{y_{..}^2}{N}$ ó $SS_{Bloques} = a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2$ **(Ver Apéndice 3)**

La suma de cuadrados del error se encuentra por diferencia:

$SS_E = SS_T - SS_{Tratamientos} - SS_{Bloques}$ ó $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2$

Las medias de cuadrados o cuadrados medios, que se definen en función de las sumas de cuadrados y los grados de libertad; es decir, cada suma de cuadrados dividida entre sus grados de libertad es igual a una media de cuadrados.

**Matemáticamente se expresan de la manera siguiente:**

$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1}$ , $MS_{Bloques} = \frac{SS_{Bloques}}{b-1}$ , $MS_E = \frac{SS_E}{(a-1)(b-1)}$

donde :

$MS_{Tratamientos}$ : Suma de Cuadrados Medios entre Tratamientos.

$MS_{Bloques}$ : Suma de Cuadrados Medios entre Bloques.

$MS_E$ : Suma de Cuadrados Medios del Error.

Si se usa la suposición de normalidad de los errores y el resultado del teorema de **COCHRAN** es posible demostrarse que $\frac{SS_{Tratamientos}}{\sigma^2}$, $\frac{SS_{Bloques}}{\sigma^2}$ y $\frac{SS_E}{\sigma^2}$ son variables aleatorias independientes con distribución ji-cuadrada.

Igualmente que en la Unidad anterior, puede mostrarse que los valores esperados de las medias de cuadrados pueden definirse así:

$$E(MS_{Tratamientos}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\iota_i^2}{a-1} \quad , \quad E(MS_{Bloques}) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1} \quad , \quad E(MS_E) = \sigma^2$$

## 2.4 ANÁLISIS ESTADÍSTICO

A continuación se plantearan las bases estadísticas para el estudio de este tipo de Diseño.

En un Experimento donde se utilice el Diseño Aleatorizado por Bloques, puede darse el caso que los tratamientos y los bloques sean factores fijos o que ambos sean factores aleatorios; las diferencias para su estudio son mínimas.

### Tratamientos y Bloques Fijos

El Análisis Estadístico comienza definiendo la hipótesis que se desean probar, en este modelo se desea comparar las medias de tratamiento, por lo tanto las hipótesis son:

$H_o$ : $\mu_1 = \mu_2 = \ldots = \mu_a$
$H_1$ : $\mu_i \# \mu_j$ , para $i \neq j$.

Una forma equivalente de expresar las hipótesis anteriores en función de los efectos de tratamiento, es la siguiente:

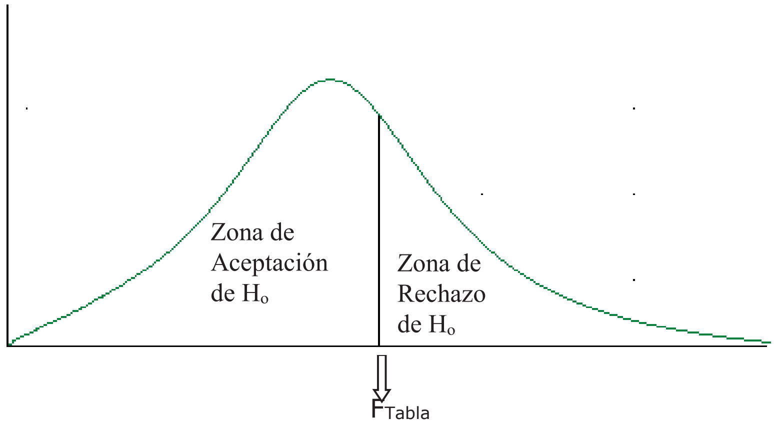
$H_o$ : $\iota_1 = \iota_2 = \ldots = \iota_a = 0$
$H_1$ : $\iota_i \neq 0$ , para al menos un $i$

Ya que la media del i-ésimo tratamiento es $\mu_i = \frac{1}{b}\sum_{j=1}^{b}(\mu + \tau_i + \beta_j) = \mu + \iota_i$

Luego para llevar a cabo la prueba de las hipótesis planteadas anteriormente, es necesario realizar el Análisis de Varianza y para la cual se debe usar la estadística siguiente:

$$F_o = \frac{\frac{SS_{Tratamientos}}{a-1}}{\frac{SS_E}{N-a}} = \frac{MS_{Tratamientos}}{MS_E}$$

, la cual tiene una distribución $F_{\alpha,(a-1),(a-1)(b-1)}$ si la hipótesis nula es verdadera. Por lo tanto, la región crítica es el extremo superior de la distribución F, como se observa a continuación.




De tal modo la hipótesis nula ($H_o$) se rechazará si: $F_o > F_{\alpha,(a-1),(a-1)(b-1)}$

Donde $F_o$ se obtiene a través del Análisis de Varianza y $F_{\alpha,(a-1),(a-1)(b-1)}$ por medio de la tabla F. La siguiente tabla Resume el Análisis de Varianza.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Tratamientos | $SS_{Tratamientos}$ | $a-1$ | $MS_{Tratamientos}$ | $F_o = \dfrac{MS_{Tratamientos}}{MS_E}$ |
| Bloques | $SS_{Bloques}$ | $b-1$ | $MS_{Bloques}$ | |
| Error | $SS_E$ | $(a-1)(b-1)$ | $MS_E$ | |
| Total | $SS_T$ | $N-1$ | | |

En este tipo de Diseño como se ha estudiado existen dos fuentes de variación controlables (Tratamientos y Bloques) y una incontrolable (Error).

Las hipótesis que se han planteado es para comparar las medias de tratamientos; pero puede resultar de interés la comparación de las medias de los bloques; ya que si no hay gran diferencia entre ellas, el análisis por bloques quizá no es necesario en el experimento. Si se analizan los valores esperados de las medias de cuadrados, puede parecer que la hipótesis $H_0 : \beta_j = 0$, puede ser probada utilizando $F_o = \dfrac{MS_{Bloques}}{MS_E}$ y $F_{\alpha,(b-1),(a-1)(b-1)}$. Pero la aleatorización fue realizada sólo a los tratamientos dentro de los bloques, no así los bloques a los tratamientos; es por eso que los bloques representan una restricción para la aleatorización, en consecuencia ¿Qué efecto tiene esto sobre la estadística $F_o = \frac{MS_{Bloques}}{MS_E}$?.

**Hay diversos argumentos a esta pregunta**

- **Box, Hunter y Hunter (1978)**

  Concluye que la prueba para comparar bloques no puede ser incluida bajo este argumento a consecuencia de la aleatorización; pero que si los errores son NID($0,\sigma^2$), la estadística $F_o = \frac{MS_{Bloques}}{MS_E}$ puede usarse para comparar las medias de los bloques.

- **Anderson y Mcean (1974)**

  Argumentan que la restricción de aleatorización impide que esta estadística puede ser útil para comparar las medias de los bloques, y que la estadística F, en realidad es una prueba para la igualdad de las medias de los bloques más la restricción de aleatorización.

  Cómo la restricción de aleatorización queda en tela de juicio, ¿Qué hacer en la práctica?.

  En general no es conveniente tomar $F_o = \frac{MS_{Bloques}}{MS_E}$ como una prueba F exacta. Es por

eso que no aparece en la tabla de Análisis de Varianza. Pero el examen de la razón entre $MS_{Bloques}$ y $MS_E$ se puede utilizar para investigar de forma aproximada el efecto de la variable bloque. Un valor grande de esta razón, indica que el factor bloque tiene un efecto grande y que la reducción de ruido obtenida al analizar por bloques posiblemente fue útil al mejorar la precisión de las comparaciones entre las medias de tratamiento.

**Ejemplo 2**

Se probaran 5 raciones respecto a sus diferencias en el engorde de novillos. Se dispone de 20 novillos para el experimento, que se distribuyen en 4 bloques (5 novillos por bloque) con base a sus pesos, al iniciar la prueba de engorde. Los 5 tratamientos (raciones) se asignaron al azar dentro de cada bloque. Los novillos más pesados se agruparon en un bloque, en otro se agruparon los 5 siguientes más pesados y así sucesivamente. Se obtuvieron los siguientes datos:

| Tratamientos (Raciones) | Bloques | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 0.9 | 1.4 | 1.4 | 2.3 |
| 2 | 3.6 | 3.2 | 4.5 | 4.1 |
| 3 | 0.5 | 0.9 | 0.5 | 0.9 |
| 4 | 3.6 | 3.6 | 3.2 | 3.6 |
| 5 | 1.8 | 1.8 | 0.9 | 1.4 |

**Solución**

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

$H_o$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (No existe diferencia entre las Raciones)

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$ (Existe diferencia entre las Raciones)

**Variable Respuesta:** Engorde de los Novillos.

**El significado verbal es:**

$H_o$ : La cantidad de ración no influye en el engorde de los novillos.

$H_1$: La cantidad de ración influye en el engorde de los novillos.

**Datos**

Tratamientos: $a$ = 5 , Bloques : $b$ = 4

Número total de observaciones: $N$ = 5x4 = 20 , $i$ = 1,2,3,4,5 , $j$ = 1,2,3,4

Para que los cálculos matemáticos resulten más fáciles, la siguiente tabla muestra los datos de la tabla anterior codificados (restándoles 1 y multiplicando el resultados por 10).

| Tratamientos (Raciones) | Bloques | | | | $y_{i.}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | -1 | 4 | 4 | 13 | 20 |
| 2 | 26 | 22 | 35 | 31 | 114 |
| 3 | -5 | -1 | -5 | -1 | -12 |
| 4 | 26 | 26 | 22 | 26 | 100 |
| 5 | 8 | 8 | -1 | 4 | 19 |
| $y_{.j}$ | 54 | 59 | 55 | 73 | $y_{..}$= 241 |

## Cálculos Matemáticos

**Totales Tratamientos**

$y_{i.} = \sum_{j=1}^{4} y_{ij}$

$y_{1.} = \sum_{j=1}^{4} y_{1j}$ = -1 + 4 + 4 + 13 = 20

$y_{2.} = \sum_{j=1}^{4} y_{2j}$ = 26 + 22 + 35 + 31 = 114

$y_{3.} = \sum_{j=1}^{4} y_{3j}$ = -5 + (-1) + (-5) + (-1) = -12

$y_{4.} = \sum_{j=1}^{4} y_{4j}$ = 26 + 26 + 22 + 26 = 100

$y_{5.} = \sum_{j=1}^{4} y_{5j}$ = 8 + 8 + (-1) + 4 = 19

**Totales de Bloques**

$y_{.j} = \sum_{i=1}^{5} y_{ij}$

$y_{.1} = \sum_{i=1}^{5} y_{i1}$ = -1 + 26 + (-5) + 26 + 8 = 54

$y_{.2} = \sum_{i=1}^{5} y_{i2}$ = 4 + 22 + (-1) + 26 + 8 = 59

$y_{.3} = \sum_{i=1}^{5} y_{i3}$ = 4 + 35 + (-5) + 22 + (-1) = 55

$y_{.4} = \sum_{i=1}^{5} y_{i4}$ = 13 + 31 + (-1) + 26 + 4 = 73

$y_{..} = \sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}$ =(-1) + 4 + 4 + 13 +….+ (-1) + 4 = 241

**Medias de Tratamientos**

$\bar{y}_{i.} = \frac{y_{i.}}{b}$

$\bar{y}_{1.} = \frac{y_{1.}}{4} = \frac{20}{4}$ = 5

$\bar{y}_{2.} = \frac{y_{2.}}{4} = \frac{114}{4}$ = 25.5

$\bar{y}_{3.} = \frac{y_{3.}}{4} = \frac{(-12)}{4}$ = -3

$\bar{y}_{4.} = \frac{y_{1.}}{4} = \frac{100}{4} =$ 25

$\bar{y}_{5.} = \frac{y_{1.}}{4} = \frac{19}{4}$ = 4.75

**Medias de Bloques**

$\bar{y}_{.j} = \frac{y_{.j}}{a}$

$\bar{y}_{.1} = \frac{y_{.1}}{5} = \frac{54}{5}$ = 10.8

$\bar{y}_{.2} = \frac{y_{.2}}{5} = \frac{59}{5}$ = 11.8

$\bar{y}_{.3} = \frac{y_{3.}}{5} = \frac{55}{5}$ = 11

$\bar{y}_{.4} = \frac{y_{.4}}{5} = \frac{73}{5}$ = 14.6

$\bar{y}_{..} = \frac{y_{..}}{20} = \frac{241}{20}$ = 12.05

**Sumas de Cuadrados**

$SS_T = \sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}^2 - \frac{y_{..}^2}{N}$ = [(-1)$^2$ + (4)$^2$ + (4)$^2$ +…..+ (-1)$^2$ + (4)$^2$] - $\frac{(241)^2}{20}$

= 6257 - 2904.05

$SS_T$ = 3352.95

$SS_{Tratamientos} = \frac{\sum_{i=1}^{5} y_{i.}^2}{b} - \frac{y_{..}^2}{N} = \frac{(20)^2 + (114)^2 + (100)^2 + (19)^2}{4} - \frac{(241)^2}{20}$

= 5975.25 - 2904.05

$SS_{Tratamientos}$ = 3071.2

$$SS_{Bloques} = \frac{\sum_{j=1}^{4} y_{.j}^2}{a} - \frac{y_{..}^2}{N} = \frac{(54)^2+(59)^2+(55)^2+(73)^2}{5} - \frac{(241)^2}{20}$$

= 2950.20 - 2904.05

$SS_{Bloques}$ = 46.15

$SS_E = SS_T - SS_{Tratamientos} - SS_{Bloques}$

$SS_E$ = 3352.92 - 3071.20 - 46.15

$SS_E$ = 235.6

**Medias de Cuadrados**

$$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} = \frac{3071.20}{5-1} = \frac{3071.20}{4} = 767.8$$

$$MS_{Bloques} = \frac{SS_{Bloques}}{b-1} = \frac{46.15}{4-1} = \frac{46.15}{3} = 15.38$$

$$MS_E = \frac{SS_E}{(a-1)(b-1)} = \frac{235.6}{(4)(3)} = \frac{235.6}{12} = 19.63$$

**Estadística**

$$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{767.8}{19.63} = 39.1136 \approx 39.11$$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Raciones | 3071.20 | 4 | 767.8 | 39.11 |
| Novillos | 46.15 | 3 | 15.38 | |
| Error | 235.60 | 12 | 19.63 | |
| Total | 3352.95 | 19 | | |

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05) para encontrar el $F_{Tablas}$ (Tablas Fisher) con 4 grados de libertad ($a$-1) en el numerador y 12 grados de libertad ($a$-1)($b$-1) en denominador.

$F_{\alpha,a-1,(a-1)(b-1)} = F_{0.05,4,12}$ = 3.26

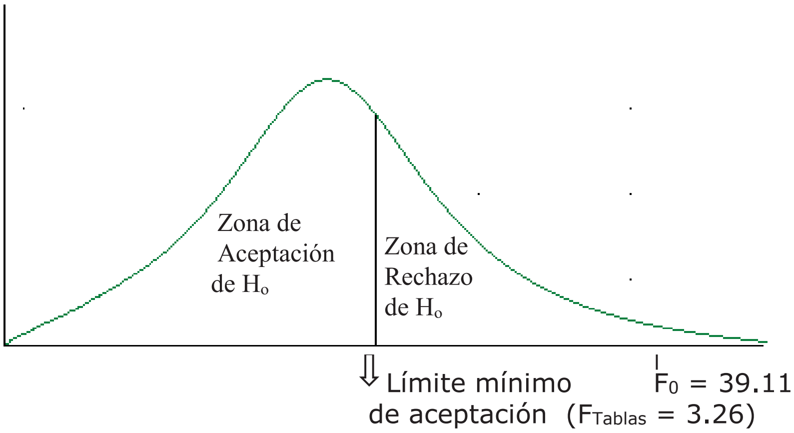
Comparando el $F_0$ calculado del Análisis de Varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

$39.11 > 3.26$

Por tanto, se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$).

También puede observarse gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

Las cinco raciones no son igualmente efectivas en el engorde de novillos o la cantidad de ración influye significativamente en el engorde de los novillos.

## 2.5 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Al utilizar el método de mínimos cuadrados se pueden obtener los estimadores de $\mu$, $\iota_i$ y $\beta_j$ del modelo:

$$y_{ij} = \mu + \iota_i + \beta_j + \varepsilon_{ij}.$$

Se deben tomar en cuenta las restricciones $\sum_{i=1}^{a} \hat{\tau}_i = 0$ y $\sum_{j=1}^{b} \hat{\beta}_j = 0$, por lo tanto se obtienen las estimaciones siguientes: (Ver Douglas C. Montgomery, año 1991, Página 135).

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \quad , \quad i = 1,2,\ldots,a$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..} \quad , \quad j = 1,2,\ldots,b$$

Lo que significa que la media general puede ser estimada usando el promedio total de las observaciones, que cualquiera de los efectos de los tratamientos son sólo la diferencia entre el promedio del tratamiento y el promedio total, y cualquiera de los efectos de los bloques son sólo la diferencia entre el promedio del bloque y el promedio total.

Estas estimaciones encontradas se pueden utilizar para encontrar los valores estimados o ajustados de $y_{ij}$ sustituyendo la ecuación del modelo de la siguiente manera:

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j$$

$$\hat{y}_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})$$

$$= \bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} + \bar{y}_{.j} - \bar{y}_{..}$$

es decir

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

**Ejemplo 3**

Al usar el ejemplo 2; en el cual se prueban cinco raciones respecto a sus diferencias en el engorde de novillos, para encontrar las estimaciones de la media general, los efectos de las raciones (tratamientos) y la distribución de los novillos (bloques) se tiene:

**Datos**

$\bar{y}_{1.}$ = 5.0 , $\bar{y}_{2.}$ = 28.5 , $\bar{y}_{3.}$ = -3.0 , $\bar{y}_{4.}$ = 25 , $\bar{y}_{5.}$ = 4.75 (Promedios de Tratamientos)

$\bar{y}_{.1}$ = 10.8 , $\bar{y}_{.2}$ = 11.8 , $\bar{y}_{.3}$ = 11 , $\bar{y}_{.4}$ = 14.6 (Promedios de Bloques)

$\bar{y}_{..}$ = 12.05 , $i$ = 1,2,3,4,5 , $j$ = 1,2,3,4

**Solución**

$\hat{\mu} = \bar{y}_{..}$ = 12.05 (**Media general**)

**Estimación de los Efectos de los Tratamientos**

$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

$\hat{\tau}_1 = \bar{y}_{1.} - \bar{y}_{..}$ = 5 - 12.05 = -7.05 (**Efecto de la ración 1**)

Significa que el engorde de los novillos disminuirá en 7.05 con la aplicación de la ración número 1.

$\hat{\tau}_2 = \bar{y}_{2.} - \bar{y}_{..}$ = 28.5 - 12.05 = 16.45 (**Efecto de la ración 2**)

Significa que el engorde de los novillos aumentará en 16.45 con la aplicación de la ración número 2.

$\hat{\tau}_3 = \bar{y}_{3.} - \bar{y}_{..}$ = -3 - 12.05 = -15.05 (**Efecto de la ración 3**)
Significa que el engorde de los novillos disminuirá en 15.05 con la aplicación de la ración número 3.

$\hat{\tau}_4 = \bar{y}_{4.} - \bar{y}_{..}$ = 25 -12.05 = 12.95 (**Efecto de la ración 4**)
Significa que el engorde de los novillos aumentará en 12.95 con la aplicación de la ración número 2.

$\hat{\tau}_5 = \bar{y}_{5.} - \bar{y}_{..}$ = 4.75 - 12.05 = -7.3 (**Efecto de la ración 5**)
Significa que el engorde de los novillos disminuirá en 7.3 con la aplicación de la ración número 5.

**Estimación de los Efectos de los Bloques**

$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$

$\hat{\beta}_1 = \bar{y}_{.1} - \bar{y}_{..}$ = 10.8 - 12.05 = -1.25 (**Efecto del bloque 1**)
Significa que el engorde de los novillos disminuirá en 1.25 al haber agrupado los novillos en el bloque 1. (Novillos más pesados).

$\hat{\beta}_2 = \bar{y}_{.2} - \bar{y}_{..}$ = 11.8 - 12.05 = -0.25 (**Efecto del bloque 2**)

Significa que el engorde de los novillos disminuirá en 0.25 al haber agrupado los novillos en el bloque 2. (Novillos menos pesados que los del bloque 1).

$\hat{\beta}_3 = \bar{y}_{.3} - \bar{y}_{..}$ = 11 - 12.05 = -1.05 (**Efecto del bloque 3**)
Significa que el engorde de los novillos disminuirá en 1.05 al haber agrupado los novillos en el bloque 3. (Novillos menos pesados que los del bloque 2).

$\hat{\beta}_4 = \bar{y}_{.4} - \bar{y}_{..}$ = 14.6 - 12.05 = 2.25 (**Efecto del bloque 4**)

Significa que el engorde de los novillos aumentará en 1.25 al haber agrupado los novillos en el bloque 4. (Novillos menos pesados que los del bloque 3).

## 2.6 COMPARACIÓN ENTRE TRATAMIENTOS

Si en un Diseño Aleatorizado por Bloques los tratamientos son fijos y el Análisis de Varianza indica que existe una diferencia significativa entre las medias de tratamiento, el experimentador estará interesado en realizar comparaciones adicionales en grupos de medias de tratamiento, para determinar cuales son las medias que difieren. Cualquier método estudiado en el Diseño Unifactorial (Unidad II) puede ser utilizado para este fin; con algunas variantes.

Para llevar a cabo las comparaciones entre grupos de tratamiento para un Diseño Aleatorizado por Bloques se debe sustituir el número de réplicas o repeticiones ($n$) por el número de bloques ($b$) en las fórmulas utilizadas en cada uno de los métodos estudiados en la Unidad anterior y además se debe utilizar los grados de libertad del error que están definidos por ($a$-1)($b$ -1) para un Diseño Aleatorizado por Bloques.

A continuación se presentarán los métodos descritos en la Unidad II, expresando solamente las variantes que se deben incorporar para llevar a cabo la comparación de medias de tratamiento para un Diseño Aleatorizado por Bloques. Las hipótesis a probar, el procedimiento y conclusiones se harán de igual manera que en la Unidad II.

**1) Comparación de Medias de Tratamientos Individuales**

**a) Contrastes Ortogonales**

La suma de cuadrados de un contraste viene dada por:

$$SS_c = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{b\sum_{i=1}^{a} c_i^2}$$ , con un sólo grado de libertad.

La estadística de prueba ($F_{Tablas}$) que se debe utilizar para rechazar la hipótesis tiene una distribución $F$ con 1 y ($a$-1)($b$-1) grados de libertad, es decir $F_{\alpha,1,(a-1)(b-1)}$.

**b) Método de Scheffé para comparar todos los contrastes.**

El error estándar del contraste será:

$$S_{c_k} = \sqrt{\frac{MS_E}{b}\sum_{i=1}^{a} c_{ik}^2}$$

El valor crítico con el que $C_k$ será comparado está dado por:

$$S_{\alpha,k} = S_{c_k}\sqrt{(a-1)F_{\alpha,a-1,(a-1)(b-1)}}$$

**2) Comparación de Parejas de Medias de Tratamientos.**

**a) Método de la Mínima Diferencia Significativa (LSD)**

El LSD estará dado de la siguiente manera: LSD $= t_{\frac{\alpha}{2},(a-1)(b-1)}\sqrt{\frac{2MS_E}{b}}$

**b) Prueba de Intervalos Múltiples de Duncan**

El error estándar de cada promedio se calcula de la siguiente forma:

$$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{b}}$$

Para encontrar los intervalos significativos $r_\alpha(p,f)$, para $p$=2,3,…,$a$, $\alpha$ sigue siendo el nivel de significancia y $f$ el número de grados de libertad del error que son ($a$-1)($b$-1). De igual manera para encontrar los mínimos intervalos significativos $R_p = r_\alpha(p, f)\ S_{\bar{y}_{i.}}$ con p= 2,3,…, $a$, se tomará f como el número de grados de libertad del error ($a$-1)($b$-1).

**c) Prueba de Tukey**

El valor crítico de todas las comparaciones vendrá dado por: $T_\alpha = q_\alpha(a, f)\,S_{\bar{y}_{i.}}$ donde $S_{\bar{y}_{i.}}$ es el error estándar de cada promedio y está dado por $S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{b}}$ y $f$ = ($a$-1)($b$-1), los grados de libertad del error, $\alpha$ el nivel de significancia y " $a$ " el número de tratamientos.

**3) Comparación de Tratamientos con un control.**

La hipótesis nula se rechaza si $|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a\text{-}1,f)\sqrt{\frac{2MS_E}{b}}$, donde $d_\alpha(a\text{-}1,f)$ se encuentra en la tabla de Dunnett con $\alpha$ que es el nivel de significancia y $f$ los grados de libertad del error, que están dados por ($a$-1)($b$-1).

**Ejemplo 4**

Como en el ejemplo de las raciones, se rechazó la hipótesis nula ($H_o$); es necesario determinar que medias son significativamente diferentes; para observar esas diferencias entre las medias se utilizará la Prueba de Intervalos Múltiples de Duncan.

**Datos**

$\alpha$ = 0.05 , $MS_E$ = 19.63 , $a$ = 5 , $b$ = 4 , N = 20 ,

Grados de libertad del error: ($a$-1)($b$-1) = (5-1)(4-1)= 4x3 = 12

$\bar{y}_{1.} = 5.0$ , $\bar{y}_{2.} = 25.5$, $\bar{y}_{3.} = -3.0$ , $\bar{y}_{4.} = 25.0$ , $\bar{y}_{5.} = 4.75$

**Solución**

Medias de tratamientos ordenadas ascendentemente:

$\bar{y}_{3.}$ = -3.0 , $\bar{y}_{5.}$ = 4.75 , $\bar{y}_{1.}$ = 5.0 , $\bar{y}_{4.}$ = 25.0 , $\bar{y}_{2.}$ = 25.5

- **Obtención del error estándar de cada media.**

$$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{19.63}{4}} = 2.22$$

- **Tomando un $\alpha$ = 0.05 y 12 grados de libertad ($f$), de las tablas de Intervalos Significativos de Duncan, se obtienen los siguientes valores de los intervalos significativos, para $p$=2,3,4,5.**

$r_{0.05}(2,12) = 3.08$ , $r_{0.05}(3,12) = 3.23$ , $r_{0.05}(4,12) = 3.33$ , $r_{0.05}(5,12) = 3.36$

- **Calculando los mínimos intervalos significativos para p = 2,3,4,5.**

$R_2 = r_{0.05}\,(2,12)\ S_{\bar{y}_{i.}} = (3.08)(2.22) = 6.84$

$R_3 = r_{0.05}\,(3,12)\ S_{\bar{y}_{i.}} = (3.23)(2.22) = 7.17$

$R_4 = r_{0.05}\,(4,12)\ S_{\bar{y}_{i.}} = (3.33)(2.22) = 7.39$

$R_5 = r_{0.05}\,(5,12)\ S_{\bar{y}_{i.}} = (3.36)(2.22) = 7.46$

- **Realizando las comparaciones y las diferencias de las medias.**

2 vrs 3 : $\bar{y}_{2.} - \bar{y}_{3.} = 25.5 - (-3.00) = 28.50 > 7.46\ (R_5)$*

2 vrs 5 : $\bar{y}_{2.} - \bar{y}_{5.} = 25.5 - 4.75 = 20.75 > 7.39\ (R_4)$ *

2 vrs 1 : $\bar{y}_{2.} - \bar{y}_{1.} = 25.5 - 5.00 = 20.50 > 7.17(R_3)$*

2 vrs 4 : $\bar{y}_{2.} - \bar{y}_{4.} = 25.5 - 25.00 = 0.50 < 6.84\ (R_2)$

4 vrs 3 : $\bar{y}_{4.} - \bar{y}_{3.} = 25.0 - (-3.00) = 28.00 > 7.39\ (R_4)$ *

4 vrs 5 : $\bar{y}_{4.} - \bar{y}_{5.} = 25.0 - 4.75 = 20.25 > 7.17\ (R_3)$ *

4 vrs 1 : $\bar{y}_{4.} - \bar{y}_{1.} = 25.00 - 5.00 = 20.00 > 6.84\ (R_2)$*

1 vrs 3 : $\bar{y}_{1.} - \bar{y}_{3.} = 5.00 - (-3.00) = 8.00 > 7.17\ (R_3)$*

1 vrs 5 : $\bar{y}_{1.} - \bar{y}_{5.} = 5.00 - 4.75 = 0.25 < 6.84\ (R_2)$

5 vrs 3 : $\bar{y}_{5.} - \bar{y}_{3.} = 4.75 - (-3.00) = 7.75 > 6.84\ (R_2)$*

Dos medias son significativamente diferentes si la diferencias observada es mayor que el intervalo mínimo significativo correspondiente.

**Conclusiones**

a) Como el valor de las diferencias de las medias dos y cuatro, una y cinco resultó ser menor que el mínimo intervalo significativo correspondiente; entonces se dice que no existe diferencias significativas entre las raciones dos y cuatro, una y cinco.

b) Por el contrario las demás diferencias de medias (*) resultaron ser mayores que el mínimo intervalo significativo correspondiente; entonces se dice que existe diferencia significativa entre ellas.

## 2.7 TRATAMIENTOS Y BLOQUES ALEATORIOS

Cuando se lleva a cabo un Experimento de un Diseño Aleatorizado por Bloques, al momento de seleccionar los tratamientos y los bloques se pueden dar los siguientes casos:

1) **Tratamientos y Bloques sean Fijos.**
2) **Tratamientos Fijos y Bloques aleatorios.**
3) **Tratamientos Aleatorios y Bloques Fijos.**
4) **Tratamientos y Bloques Aleatorios.**

El primer caso, ya fue estudiado; sin embargo si se dan los demás casos el procedimiento de Análisis de Varianza se lleva a cabo como el primer caso, sólo existen unas pequeñas variantes; que es en la interpretación de los resultados obtenidos del Análisis; ya que las conclusiones pueden ser generalizadas a toda la población de tratamientos y bloques de donde fueron seleccionados aleatoriamente y el planteamiento de las hipótesis, dependiendo si los tratamientos o bloques son aleatorios o fijos.

Además, si los bloques son aleatorios se espera que las comparaciones entre los tratamientos sean las mismas en toda la población de bloques de la que se eligieron aleatoriamente los que se usaron en el experimento, y la esperanza de la media de cuadrados estará definida por $E(MS_{Bloques}) = \sigma^2 + a\sigma_\beta^2$, en donde $\sigma_\beta^2$ es la componente de varianza del efecto de los bloques; y en cualquier caso $E(MS_{Tratamientos})$ siempre es independiente del efecto de los bloques y por lo tanto, el estadístico de prueba para estudiar la variabilidad entre los tratamientos siempre será: $F_0 = \dfrac{MS_{Tratamientos}}{MS_E}$.

Puede suceder que el efecto del tratamiento i afecte adversamente al bloque j, produciendo un rendimiento extraordinariamente bajo, pero no afecta a los otros tratamientos, entonces se dice que ha ocurrido una iteración entre los tratamientos y los bloques. La iteración también se puede dar cuando la respuesta se mide en una escala equivocada; pero esto no afecta la prueba para las medias de los tratamientos, ya que si los bloques son aleatorios, tanto el valor esperado de la media de cuadrados de tratamiento y la media de cuadrados del error contienen el efecto de la iteración; y por esta razón que la prueba para la diferencia en el nivel medio de los tratamientos se realiza comparando la media de cuadrados de tratamientos y la media de cuadrados del error.

## 2.8 SELECCIÓN DEL TAMAÑO MUESTRAL

En un Diseño Aleatorizado por bloques la selección del tamaño de la muestra o del número de bloques debe ser una decisión muy importante, ya que al aumentar el número de bloques, también aumenta el número de réplicas y los grados de libertad de la media de cuadrados del error, haciendo el Diseño más sensible.

En este tipo de Diseño no tiene sentido hablar o encontrar el número de réplicas, sólo el número de bloques, ya que debe cumplir que el tratamiento se debe aplicar una sola vez en cada uno de los bloques, y por consecuencia al encontrar el número de bloques se está determinando el número de réplicas en el Experimento.

Cualquier método estudiado en el Diseño Unifactorial (Unidad II) es válido, solamente que se deben tomar en cuenta algunas variantes.

Para llevar a cabo la selección de la muestra o número de bloques en cada una de las fórmulas utilizadas en los métodos estudiados en la Unidad anterior, se debe cambiar el número de réplicas ($n$) por el número de bloques ($b$) y se deben utilizar los grados de libertad del error ($a$-1) ($b$-1).

A continuación se presentan los métodos estudiados en la Unidad II, expresando solamente las variantes que se deben realizar. El procedimiento y conclusiones se harán de manera similar.

### 1) FACTOR FIJO

#### a) Curvas Características de Operaciones

El valor de $\varphi$ se calcula de la siguiente manera: $\varphi^2 = \dfrac{b\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$, en donde $a$-1 son los grados de libertad del numerador y ($a$-1)($b$-1) los grados de libertad del denominador.

En caso que resulte difícil obtener el conjunto de medias de tratamientos, el valor de $\varphi^2$ vendrá dado por: $\varphi^2 = \frac{bD^2}{2a\sigma^2}$, donde $a$-1 son los grados de libertad del numerador y ($a$ -1)($b$-1) los grados de libertad del denominador.

**b) Especificación de un Incremento en la Desviación Estándar.**

El valor de $\varphi$ viene dado por: $\varphi = \sqrt{(1+0.01p)^2 - 1}\ \sqrt{b}$, donde $a$-1 son los grados de libertad del numerador y ($a$ -1)($b$-1) los grados de libertad del denominador.

**2) FACTOR ALEATORIO**

**a) Curvas Características de Operaciones**

El valor de $\lambda$ se calcula de la siguiente manera: $\lambda = \sqrt{1 + \frac{b\sigma_\tau^2}{\sigma^2}}$, donde $a$-1 son los grados de libertad del numerador y ($a$ -1)($b$-1) los grados de libertad del denominador.

**b) Especificación de un Incremento en la Desviación Estándar.**

El valor de $\lambda$ viene dato por $\lambda = \sqrt{1 + b(1+0.01p)^2 - 1}$, donde a-1 son los grados de libertad del numerador y ($a$ -1)($b$-1) los grados de libertad del denominador.

**3) Método de Estimación por Intervalo de Confianza**

La precisión del intervalo viene dada por: $\pm\ t_{\frac{\alpha}{2},(a-1)(b-1)} \sqrt{\frac{2MS_E}{b}}$

**Ejemplo 5**

Retomando el ejemplo de las raciones para Novillos.

Encontrar el número necesario de bloques si se desea que el intervalo de confianza del 95%, para la diferencia media de las raciones sea $\pm$ 3 y que una estimación a priori de $\sigma$ es 2.

**Solución**

Para solucionar este ejemplo se pondrá en práctica el método de Estimación por Intervalo de Confianza.

**Datos**

$\alpha$ = 0.05 , $a$ = 5 , $MS_E \approx \sigma^2$ , $\sigma$ = 2 , $\sigma^2$ = 4

**Precisión del intervalo**

$$\pm\ t_{\frac{\alpha}{2},(a-1)(b-1)}\sqrt{\frac{2MS_E}{b}}$$

$$\pm\ t_{\frac{0.05}{2},(5-1)(b-1)}\sqrt{\frac{2(4)}{b}}$$

$$\pm\ t_{0.025,(4)(b-1)}\sqrt{\frac{8}{b}}$$

Suponiendo que se propone $b$= 6 bloques , se tiene la precisión del intervalo:

$$\pm\ t_{0.025,(4)(6-1)}\sqrt{\frac{8}{6}}$$

$$\pm\ t_{0.025,(4)(5)}\sqrt{1.333333}$$

$$\pm\ t_{0.025,20}\sqrt{1.333333}$$

± (2.086)(1.154700538)

± 2.41 este valor es más preciso que el propuesto (±3)

Suponiendo ahora $b$= 5 bloques, la precisión del intervalo es:

$$\pm\ t_{0.025,(4)(5-1)}\sqrt{\frac{8}{5}}$$

$$\pm\ t_{0.025,(4)(4)}\sqrt{1.6}$$

$$\pm\ t_{0.025,16}\sqrt{1.6}$$

± (2.120)(1.264911064)

± 2.68 este valor es más preciso que el propuesto (±3)

Suponiendo ahora $b$= 4 bloques, la precisión del intervalo es:

$$\pm\ t_{0.025,(4)(4-1)}\sqrt{\frac{8}{4}}$$

$$\pm\ t_{0.025,(4)(3)}\sqrt{2}$$

$$\pm\ t_{0.025,12}\sqrt{2}$$

± (2.179)(1.414213562)

± 3.08 este valor se aproxima mas a la precisión establecida.

**Conclusión**

Con el valor de $b$= 4 bloques el intervalo calculado sobrepasa el valor de precisión establecido (± 3), entonces el mínimo número de bloques que deben tomarse es $b$= 4; ya que conduce a la precisión deseada.

## 2.9 EFICIENCIA RELATIVA DE UN DISEÑO ALEATORIZADO POR BLOQUES

El análisis o estudio de un Experimento puede ser llevado a cabo a través de un Diseño Unifactorial o por un Diseño Aleatorizado por Bloques. Sin embargo, se puede dar el caso que no se obtenga la misma sensibilidad. En general al utilizar un Diseño Unifactorial la $MS_E$ podría ser mayor que al utilizar un Diseño Aleatorizado por Bloques; ya que el Diseño Aleatorizado por Bloques reduce suficientemente la cantidad de ruido para lograr detectar diferencias significativas entre los tratamientos.

En un Diseño Aleatorizado por Bloques resulta útil estimar la eficiencia relativa, para compararlo con el Diseño Unifactorial. El valor que resulta de esta estimación se puede interpretar como el incremento del número de réplicas necesarias que hay que llevar a cabo en un Diseño Unifactorial para que pueda ser usado en lugar de un Diseño Aleatorizado por Bloques, y así mantener la misma sensibilidad en ambos Diseños.

La forma de definir la eficiencia relativa es mediante la siguiente fórmula:

$$\mathrm{R} = \frac{(df_b+1)\,(df_r+3)}{(df_b+3)\,(df_r+1)}\frac{\sigma_r^2}{\sigma_b^2}$$

donde:

$\sigma_r^2$: Es la varianza del error experimental del Diseño Unifactorial .

$\sigma_b^2$: Es la varianza del error experimental del Diseño Aleatorizado por Bloques.

$df_r$: Los grados de libertad del error del Diseño Unifactorial

$df_b$: Los grados de libertad del error del Diseño Aleatorizado por Bloques.

es decir, $df_r = N - a$ y $df_b = (a-1)(b-1)$

Como puede observarse para calcular la eficiencia relativa, se deben llevar a cabo estimaciones para $\sigma_r^2$ y $\sigma_b^2$, las cuales es posible estimarlas de la siguiente forma:

$\hat{\sigma}_b^2 \approx MS_E$ del Diseño Aleatorizado por Bloques y $\hat{\sigma}_r^2 = \dfrac{(b-1)MS_{Bloques} + b(a-1)MS_E}{ab-1}$ es un estimador insesgado de la varianza del error de un Diseño Unifactorial.

**Ejemplo 6**

Para poner en práctica este procedimiento, se retomará el ejemplo de las raciones para encontrar la eficiencia relativa de este Diseño Experimental y determinar el número de réplicas en caso de hacer su análisis como un Diseño Unifactorial y mantener la misma sensibilidad en ambos Diseños.

**Datos**

$MS_E = 19.63$ , $a = 5$ , $b = 4$ , $MS_{Bloques} = 15.38$ , $N = 20$

$df_r = N\text{-}a = 20 - 5 = 15$

$df_b = (a\text{-}1)(b\text{-}1) = (5\text{-}1)(4\text{-}1) = (4)(3) = 12$

**Solución**

**Estimación de las varianzas**

$$\hat{\sigma}_b^2 \approx MS_E = 19.63$$

$$\hat{\sigma}_r^2 = \frac{(b-1)MS_{Bloques} + b(a-1)MS_E}{ab-1} = \frac{(4-1)(15.38) + 4(5-1)(19.63)}{5x4-1} = \frac{3(15.38) + 4x4(19.63)}{20-1}$$

$$= \frac{46.14 + 314.08}{19}$$

$$\hat{\sigma}_r^2 = \frac{360.22}{19} = 18.96$$

**Calculando la Eficiencia Relativa del Diseño Aleatorizado por Bloques.**

$$R = \frac{(df_b + 1)\,(df_r + 3)}{(df_b + 3)\,(df_r + 1)}\,\frac{\hat{\sigma}_r^2}{\hat{\sigma}_b^2} = \frac{(12+1)\,(15+3)}{(12+3)\,(15+1)}\,\frac{18.96}{19.63} = \frac{(13)(18)}{(15)(16)}(0.965868568)$$

$$= \frac{234}{240}(0.965868568) = 0.941721854 \approx 1$$

**Conclusión**

Significa que se debe usar una réplica más si se utiliza un Diseño Unifactorial para lograr la misma sensibilidad que la obtenida al analizar el experimento de las raciones por medio de un Diseño Aleatorizado por Bloques.

En este ejemplo, es independiente utilizar un Diseño u otro, porque no se necesitan muchas réplicas más, al utilizar el Diseño Unifactorial. Pero existen experimentos en los cuales se debe tomar en cuenta la eficiencia Relativa, ya que se puede tomar una decisión incorrecta al hacer un análisis por bloques. La diferencia está en la perdida de grados de libertad del error.

En un Diseño Aleatorizado por Bloques el error tiene ($a$-1)($b$-1) grados de libertad y un Diseño Unifactorial con $b$ réplicas, el error tendrá $a$(b-1) grados de libertad. Por consiguiente tomar una decisión incorrecta al analizar por bloques tiene un costo de $a(b\text{-}1)\text{-}(a\text{-}1)(b\text{-}1) = b\text{-}1$ grados de libertad para el error; es decir se pierden $b$-1 grados de libertad para el error.

## 2.10 ESTIMACIÓN DE VALORES FALTANTES

Algunas de las observaciones en uno de los bloques puede hacer falta, cuando se utiliza un Diseño Aleatorizado por Bloques Completos; esto puede suceder debido a algún descuido o error, o por razones fuera de control del experimentador, como la pérdida de alguna unidad experimental. Una observación faltante genera un problema en su análisis, ya que hace que el Diseño este desbalanceado, y se dice que los tratamientos y los bloques no son ortogonales, porque todos los tratamientos no ocurren en todos los bloques.

Existen varias formas de solucionar este problema, una de ellas es realizar un análisis aproximado en el que se estima la observación faltante y luego se lleva a cabo el Análisis de Varianza tomando la observación estimada como si fuera un dato real. Este análisis aproximado consiste en hacer estimaciones de los valores faltantes, de manera que se minimice la media de cuadrados del error.

Supóngase que falta la observación $y_{ij}$ que corresponde al tratamiento *i* y al bloque *j*; y se representa por $x$.

El procedimiento que se lleva a cabo para estimar $y_{ij}$ , es el siguiente:

1) Se calcula el gran total con la observación faltante y se representa por $y'_{..}$ .
2) Se obtienen los totales del tratamiento y del bloque con el dato faltante que se representa por $y'_{i.}$ y $y'_{.j}$ respectivamente.
3) Se calcula el estimador de la observación faltante de la siguiente forma:

$$\hat{y}_{ij} = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}$$

Para más detalles (Ver Douglas C. Montgomery, año 1991, Páginas 133 y 134)

Puede suceder que falte más de una observación en el experimento. Existen dos formas para encontrar estas observaciones:

**Primera**

Utilizar el procedimiento descrito anteriormente iterativamente. Por ejemplo, supóngase que hacen falta dos observaciones, la forma de llevar a cabo la estimación de las dos observaciones, es estimando arbitrariamente el primer valor faltante y se usa este valor como un dato real para estimar el segundo, luego se hace una segunda estimación para el primer dato faltante utilizando la estimación del segundo; con la estimación encontrada para el primero se vuelve a estimar el segundo.

Este procedimiento continúa hasta obtener la convergencia en los valores estimados; es decir, hasta que resulten valores parecidos en cada iteración.

**Segunda.**

Escribir la suma de cuadrados del error en función de los datos faltes, derivar con respecto a cada uno, igualar a cero y resolver las ecuaciones que resultan.

En general, para cualquier problema que falten datos, el número de los grados de libertad del error se debe reducir en uno por cada dato que es estimado.

**Ejemplo 7**

Un ingeniero industrial realiza un experimento para estudiar el tiempo que tarda el ojo en enfocar. Esta interesado en la relación que existe entre la distancia del objeto al ojo y el tiempo que el ojo tarda en enfocar. Cuatro diferentes distancias resultan de interés (4,6,8 y 10). Existen cinco sujetos disponibles para el experimento. Como puede haber diferencia entre los sujetos, él decide aplicar un Diseño Aleatorizado por Bloque. Supongamos que por algún motivo el ingeniero industrial no pudo recopilar los datos correspondientes a la distancia 6, sujeto 3 y a la distancia 8, sujeto 5. Los datos recopilados se muestran a continuación.

| Distancias (pies) | Sujetos | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | $y'_{i.}$ |
| 4 | 10 | 6 | 6 | 6 | 6 | 34 |
| 6 | 7 | 6 | $x$ | 1 | 6 | 20 |
| 8 | 5 | 3 | 3 | 2 | **x** | 13 |
| 10 | 6 | 4 | 4 | 2 | 3 | 19 |
| $y'_{.j}$ | 28 | 19 | 13 | 11 | 15 | $y'_{..}$ =86 |

**Solución**

Como puede observarse en este ejemplo se supone que faltan dos observaciones de la tabla, que son $y_{23}$ y $y_{35}$. Se utilizará el procedimiento de análisis aproximado iterativamente para estimar estos valores faltantes.

**Datos**

$a = 4$ , $b = 5$

**Primera iteración.**

Supongamos arbitrariamente que $y_{23}$ = 8, sustituimos este valor en la tabla y entonces $y'_{..} = 94$; ya que $y'_{..}$ = 86 con los dos valores faltantes (86+8 = 94).

Luego se estima $y_{35}$, con la fórmula.

$$\hat{y}_{ij} = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}$$, con i=3 y j=5, $y'_{3.}$ = 13 $y'_{.5}$ = 15

$$\hat{y}_{35} = \frac{4y'_{3.} + 5y'_{.5} - y'_{..}}{(4-1)(5-1)} = \frac{(4)(13)+(5)(15)-94}{(3)(4)} = \frac{33}{12} = 2.75$$

**Segunda iteración.**

Tomando $y_{35}$ = 2.75, sustituyendo este valor en la tabla, entonces $y'_{..}$ = 88.75 , $y'_{2.}$ = 20 y $y'_{.3} = 13$. Luego se estima $y_{23}$, con la fórmula.

$$\hat{y}_{23} = \frac{4y'_{2.} + 5y'_{.3} - y'_{..}}{(4-1)(5-1)} = \frac{(4)(20)+(5)(13)-88.75}{(3)(4)} = \frac{56.25}{12} = 4.69$$

Tomando $y_{23}$ = 4.69, sustituyendo este valor en la tabla, entonces $y'_{..}$ = 90.69 , $y'_{3.}$ = 13 y $y'_{..5} = 15$. Luego se estima $y_{35}$, con la fórmula.

$$\hat{y}_{35} = \frac{4y'_{3.} + 5y'_{.5} - y'_{..}}{(4-1)(5-1)} = \frac{(4)(13)+(5)(15)-90.69}{(3)(4)} = \frac{36.31}{12} = 3.02$$

Tomando $y_{35}$ = 3.02, sustituyendo este valor en la tabla, entonces $y'_{..}$ = 89.02 , $y'_{2.}$ = 20 y $y'_{..3}$ = 13. Luego se estima $y_{23}$, con la fórmula.

$$\hat{y}_{23} = \frac{4y'_{2.} + 5y'_{.3} - y'_{..}}{(4-1)(5-1)} = \frac{(4)(20)+(5)(13)-89.02}{(3)(4)} = \frac{55.98}{12} = 4.665$$

Tomando $y_{23}$ = 4.665, sustituyendo este valor en la tabla, entonces $y'_{..}$ = 90.66 , $y'_{3..}$ = 13 y $y'_{..5} = 15$. Luego se estima $y_{35}$, con la fórmula.

$$\hat{y}_{35} = \frac{4y'_{3.} + 5y'_{.5} - y'_{..}}{(4-1)(5-1)} = \frac{(4)(13)+(5)(15)-90.66}{(3)(4)} = \frac{36.34}{12} = 3.028$$

Si se vuelve a tomar $\hat{y}_{35}$ = 3.02 se llegará a obtener aproximadamente el mismo valor encontrado de $y_{23}$ = 4.66.

Por lo tanto, $\hat{y}_{35}$ = 3.02 y $\hat{y}_{23}$ = 4.66 son las estimaciones de los valores faltantes y la tabla con las estimaciones encontradas se muestra a continuación.

| Distancias (pies) | Sujetos | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | $y'_{i.}$ |
| 4 | 10 | 6 | 6 | 6 | 6 | 34 |
| 6 | 7 | 6 | **4.66** | 1 | 6 | 20 |
| 8 | 5 | 3 | 3 | 2 | **3.02** | 16.02 |
| 10 | 6 | 4 | 4 | 2 | 3 | 19 |
| $y'_{.j}$ | 28 | 19 | 17.66 | 11 | 15 | $y'_{..}$ =93.68 |

## 3. DISEÑOS QUE SE BASAN EN EL PRINCIPIO DE ANÁLISIS POR BLOQUES COMPLETOS

### 3.1 DISEÑO DE CUADRADO LATINO.

Este Diseño es una extensión del Diseño por Bloques Completos, puesto que en él se imponen las mismas restricciones que se han visto para los Diseños por Bloques Completos; agregándose éstas en las columnas; es decir, que cada tratamiento seleccionado al azar es aplicado una vez en cada bloque y cada columna. Además el número de bloques debe ser igual al número de columnas y tratamientos en estudio. A este tipo de Diseño se le conoce con el nombre de "**Diseño de Cuadrado Latino"**, porque los tratamientos están representados en el Diseño por las Letras Latinas. El cual permite agrupar las unidades experimentales en más de un factor o fuente de variación independiente de los tratamientos y se usa para eliminar dos fuentes de variabilidad problemáticas; es decir, que permite analizar sistemáticamente por bloques en dos direcciones en donde intervienen dos factores de agrupamiento comúnmente llamados factores renglón y columnas.

**Algunas consideraciones que se deben tomar en cuenta en un Diseño de Cuadrado Latino son:**

- En número de repeticiones es impuesto, ya que tiene que ser igual al número de tratamientos, renglón y columnas.
- No se debe usar cuando hay más de 8 tratamientos, ya que se requiere un costo elevado de la experimentación por la restricción que el número de tratamientos = número de renglones = Número de columnas = número de repeticiones.
- No es conveniente usarlo cuando el número de tratamientos es menor de 4, ya que el Diseño tendrá relativamente un número de grados de libertad muy pequeño y no se puede estimar el error experimental, pero se puede repetir un cierto número de veces para hallar la estimación del error experimental.
- Presenta complicación en los cálculos del Análisis de Varianza cuando faltan datos.

- En general este Diseño se aplica cuando:
  Existen tres factores en estudio.
  El número de niveles para cada factor es el mismo.
  No son esperables interacciones entre los factores.

**Ejemplo 8**

Se desean probar 4 máquinas con el objetivo de observar si existe diferencia en la capacidad de producción de una cierta pieza manufacturada. Se conoce que diferentes trabajadores y diferentes períodos de tiempos en un día de trabajo tendrá un efecto sobre la producción. Se eligen a 4 operadores (columnas) y 4 períodos de tiempo (renglones) y se asignan al azar las máquinas correspondientes a cada una de las celdas del cuadrado respetando la restricción que cada máquina se usa por un sólo operador en cada período de tiempo.

**Interpretación del Ejemplo 8**

Para determinar si existe diferencia en la capacidad de producción de las 4 máquinas, se observa que existen dos fuentes de variabilidad extrañas que afectan la producción de las máquina que son los operadores y los períodos de tiempo en que se elabora la pieza, los cuales son independientes de las máquinas. Por lo tanto, debemos hacer el análisis en dos direcciones tanto para los operadores como para los períodos de tiempo; y así lograr eliminar estas dos fuentes de variabilidad que intervienen en el Diseño. Para llevar a cabo el experimento se toman 4 operadores y 4 períodos de tiempo y las máquinas son asignadas al azar al cuadrado que forman los operadores y los períodos, cumpliendo que una máquina es usada por un operador en cada período de tiempo.

### 3.1.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

En general un Cuadrado Latino contiene p tratamientos, p columnas y p renglones, por lo tanto tendrá p unidades experimentales, de tal manera que haya sólo una representación de cada tratamiento en cada columna y en cada fila.
La representación de los datos para este tipo de Diseño es la siguiente:

| Renglones | Columnas | | | | | $y_{i..}$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | ……… | $p$ | |
| 1 | A = $y_{111}$ | B = $y_{122}$ | C = $y_{133}$ | ……… | p = $y_{1pp}$ | $y_{1..}$ |
| 2 | B = $y_{221}$ | C = $y_{232}$ | D = $y_{243}$ | ……… | A = $y_{21p}$ | $y_{2..}$ |
| 3 | C = $y_{331}$ | D = $y_{342}$ | E = $y_{353}$ | ……… | B = $y_{32p}$ | $y_{3..}$ |
| .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | ………<br>………<br>……… | .<br>.<br>. | .<br>.<br>. |
| $p$ | P = $y_{pp1}$ | A = $y_{p12}$ | B = $y_{p23}$ | ……… | $(p$-1) = $y_{p(p-1)p}$ | $y_{p..}$ |
| $y_{..k}$ | $y_{..1}$ | $y_{..2}$ | $y_{..3}$ | ……… | $y_{..p}$ | $y_{...}$ |

Como puede observarse de la tabla general de datos, un Cuadrado Latino es un cuadrado "$pxp$"; es decir, que contiene "$p$" renglones y " $p$" columnas.

Cada una de las $p^2$ celdas contiene una de las $p$ letras del alfabeto latino que corresponden a los tratamientos, y cada letra aparece una sola vez, en cada fila y en cada columna.

Un valor de la tabla $y_{ijk}$ representa la observación correspondiente al i-ésimo renglón, la $k$-ésima columna y el j-ésimo tratamiento.

Puesto que hay solamente una observación en cada celda, solamente dos subíndices $(i,k)$ o $(i,j)$ de las tres restricciones $(i,j,k)$ son necesarias para denotar una observación en particular.

En general habrán "$p$" observaciones en cada renglón y cada columna.

Sea.

$y_{i..}$ : Totales por renglones.

$y_{.j.}$ : Totales de tratamientos (Letras Latinas).

$y_{..k}$ : Totales por columnas.

$y_{...}$ :Total general.

$\bar{y}_{i..}$: Promedio de Renglón.

$\bar{y}_{.j.}$: Promedio de Tratamiento.

$\bar{y}_{..k}$: Promedio de Columna.

$\bar{y}_{...}$: Promedio General.

**Matemáticamente se expresan de la siguiente manera:**

La notación de paréntesis en los subíndice de las fórmulas significa que no importa dicho subíndice que se encuentra entre el paréntesis para obtener los totales.

$$y_{i..} = \sum_{j=1}^{p} y_{ij(k)} = \sum_{k=1}^{p} y_{i(j)k} \quad , \quad y_{.j.} = \sum_{i=1}^{p} y_{ij(k)} = \sum_{k=1}^{p} y_{(i)jk} \quad , \quad y_{..k} = \sum_{i=1}^{p} y_{i(j)k} = \sum_{j=1}^{p} y_{(i)jk}$$

$$y_{...} = \sum_{i=1}^{p}\sum_{j=1}^{p} y_{ij(k)} \quad , \quad \bar{y}_{i..} = \frac{y_{i..}}{p} \quad , \quad \bar{y}_{.j.} = \frac{y_{.j.}}{p} \quad , \quad \bar{y}_{..k} = \frac{y_{..k}}{p} \quad , \quad \bar{y}_{...} = \frac{y_{...}}{N}$$

$$y_{...} = \sum_{i=1}^{p} y_{i..} = \sum_{j=1}^{p} y_{.j.} = \sum_{k=1}^{p} y_{..k}$$ donde $N = pxp = p^2$

## 3.1.2 CUADRADO LATINO ESTÁNDAR

En general un Cuadrado Latino consta de las Letras Latinas del alfabeto que representan los tratamientos, si éstas se escriben en orden alfabético; el Diseño se conoce como Cuadrado Latino Estándar. Siempre es posible obtener un cuadrado latino estándar si el primer renglón se escribe en orden alfabético, y cada renglón siguiente se escribe como el anterior, desplazando un lugar hacia la izquierda. Existe la tabla general de Cuadros Latinos y a la vez se describe cuantos Cuadrados Latinos estándar existen por cada cuadrado $pxp$. (Ver Douglas C. Montgomery, año 1991, Página 143).

Ya que en este Diseño se aplica el principio de la aleatorización, se debe seleccionar en forma aleatoria el cuadrado en particular a utilizar. Pero existe un número muy grande de cuadrados latinos de un tamaño particular, esto significa que es muy tedioso estar numerando todos los cuadrados y seleccionar uno al azar. Para llevar a cabo la aleatorización se debe seleccionar un cuadrado latino de una tabla de estos Diseños, que proporciona Fisher y Yates, y luego acomodar el orden de los renglones, columnas y letras aleatoriamente. A continuación se presenta un procedimiento para seleccionar aleatoriamente un Cuadrado Latino.

**Procedimiento**

1) Tomar al azar un Cuadrado Latino cualquiera de los presentados en la tabla.
2) Mediante números aleatorios permutar al azar el orden de las filas y columnas.
3) Asignar al azar los tratamientos a las letras.

**Ejemplo 9**

1) Si se tiene un Cuadrado Latino de $p$ = 4 y se elige al azar el segundo Cuadrado Latino de orden cuatro de la tabla. Es decir:

   A B C D
   D C B A
   B A D C
   C D A B

2) Utilizando los números aleatorios se obtiene el orden para las filas de (2,3,1 y 4) y para las columnas de (3,1,4 y 2). El cuadrado resultante de reordenar las filas será:

   D C B A
   B A D C
   A B C D
   C D A B

   Y luego permutando las columnas se tiene el cuadrado.

   B D A C
   D B C A
   C A D B
   A C B D

3) Por último, asignando al azar los tratamientos a cada una de las letras, se tendrá el Diseño de Cuadrado Latino buscado.

### 3.1.3 MODELO ESTADÍSTICO

El modelo estadístico Lineal que resulta de un Diseño de Cuadrado Latino es el siguiente:

$$y_{ijk} = \mu + \alpha_i + \iota_j + \beta_k + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,p \\ k = 1,2,\ldots,p \end{cases}$$

donde:

$y_{ijk}$ : Es la observación ijk-ésima.

$\mu$ : Es la media general.

$\alpha_i$ : Es el i-ésimo efecto del renglón.

$\iota_j$ : Es el j-ésimo efecto del tratamiento.

$\beta_k$ : Es el k-ésimo efecto de columna.

$\varepsilon_{ijk}$ : Es el término usual NID(0,$\sigma^2$) del error aleatorio.

Este modelo es completamente Aditivo, es decir, no existe iteración entre los renglones, las columnas y los tratamientos.

Además los efectos de los renglones, columnas y los tratamientos se consideran como desviaciones de la media general; por lo tanto deben cumplir: $\sum_{i=1}^{p}\hat{\alpha}_i = 0$ , $\sum_{j=1}^{p}\hat{\tau}_j = 0$ y $\sum_{k=1}^{p}\hat{\beta}_k = 0$ .

### 3.1.4 SUMAS Y MEDIAS DE CUADRADOS

Como en todo Diseño el Análisis de Varianza consiste en descomponer la suma total de cuadrados de las N= $p^2$ observaciones en sus componentes que las forman:

Sea:

$SS_T$ : Suma total de cuadrados corregida

$SS_{Renglones}$ : Suma de cuadrados debida a los renglones.

$SS_{Columnas}$ : Suma de cuadrados debida a las columnas.

$SS_{Tratamientos}$ : Suma de cuadrados debida a los tratamientos

$SS_E$ : Suma de cuadrados debida al error

La suma total de cuadrados corregida, que es la medida de variabilidad total de los datos, puede ser escrita como:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \bar{y}_{...})^2$$

Al descomponer esta sumatoria como se hizo en el Diseño Aleatorizado por Bloques Completos, nos queda que:

$$SS_T = p^2\sum_{i=1}^{p}(\bar{y}_{i..} - \bar{y}_{...})^2 + p^2\sum_{j=1}^{p}(\bar{y}_{.j.} - \bar{y}_{...})^2 + p^2\sum_{k=1}^{p}(\bar{y}_{..k} - \bar{y}_{...})^2 +$$
$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$$

Por lo tanto

$$SS_T = SS_{Renglones} + SS_{Tratamientos} + SS_{Columnas} + SS_E$$

Donde estas sumas se expresan matemáticamente de la siguiente manera:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 - \frac{y_{...}^2}{N} \quad , \quad SS_{Tratamientos} = \sum_{j=1}^{p}\frac{y_{.j.}^2}{p} - \frac{y_{...}^2}{N}$$

$$SS_{Renglones} = \sum_{i=1}^{p}\frac{y_{i..}^2}{p} - \frac{y_{...}^2}{N} \quad , \quad SS_{Columnas} = \sum_{k=1}^{p}\frac{y_{..k}^2}{p} - \frac{y_{...}^2}{N}$$

y el $SS_E$ se obtiene por diferencia, de la siguiente forma:

$SS_E = SS_T - SS_{Renglones} - SS_{Tratamientos} - SS_{Columnas}$

donde

$SS_T$ : Tiene $p^2$ - 1 grados de libertad, porque existen un total de $N = p^2$ observaciones en total y un sólo parámetro a estimar que es $\mu$ .

$SS_{Renglones}$ : Tiene $p-1$ grados de libertad, porque existen p renglones y un sólo parámetro a estimar que es $\alpha_i$ .

$SS_{Columnas}$ : Tiene $p-1$ grados de libertad, porque existen p columnas y un sólo parámetro a estimar que es $\beta_k$ .

$SS_{Tratamientos}$ : Tiene $p$-1 grados de libertad, porque existen p tratamientos y un sólo parámetro a estimar que es $\iota_j$ .

$SS_E$ : Tiene ($p$-2)( $p$-1) grados de libertad, porque existen $p^2$ celdas que proporcionan $p^2$-1 grados de libertad, y como la suma de cuadrados del error es igual a la suma de cuadrados entre las celdas menos la suma de cuadrados de renglones, la suma de cuadrados de columnas y la suma de cuadrados de tratamientos; entonces los grados de libertad del error será: $p^2-1-(p-1)-(p-1)-(p-1) = p^2- 3p+2=(p-1)(p -2)$.

Las Medias de Cuadrados igual que en cualquier Diseño Experimental están determinadas por las sumas de cuadrados divididas por sus respectivos grados de libertad;es decir, cada suma de cuadrados dividida entre sus grados de libertad es igual a una media de cuadrados.

**Matemáticamente se expresan de la manera siguiente:**

$$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{p-1} \quad , \quad MS_{Renglones} = \frac{SS_{Renglones}}{p-1}$$

$$MS_{Columnas} = \frac{SS_{Columnas}}{p-1} \quad , \quad MS_E = \frac{SS_E}{(p-2)(p-1)}$$

donde :

$MS_{Tratamientos}$ : Suma de Cuadrados Medios entre Tratamientos.

$MS_{Renglones}$ : Suma de Cuadrados Medios entre Renglones.

$MS_{Columnas}$ : Suma de Cuadrados Medios entre las Columnas.

$MS_E$ : Suma de Cuadrados Medios del Error.

## 3.1.5 ANÁLISIS ESTADÍSTICO

Las hipótesis que se prueban siempre tienen que hacerse en relación a los tratamientos (Letras Latinas); o sea en la igualdad de las medias de tratamientos, con la diferencia que en este Diseño intervienen en el estudio otras dos fuentes de variación extrañas que son los de renglones y columnas.

Las hipótesis a probar serán:

$H_o$ : $\mu_1 = \mu_2 = \ldots = \mu_p$

$H_1$ : $\mu_i \neq \mu_j$ para al menos un par (i,j)

ó de forma equivalente

$H_o$ : $\iota_1 = \iota_2 = \ldots = \iota_p$

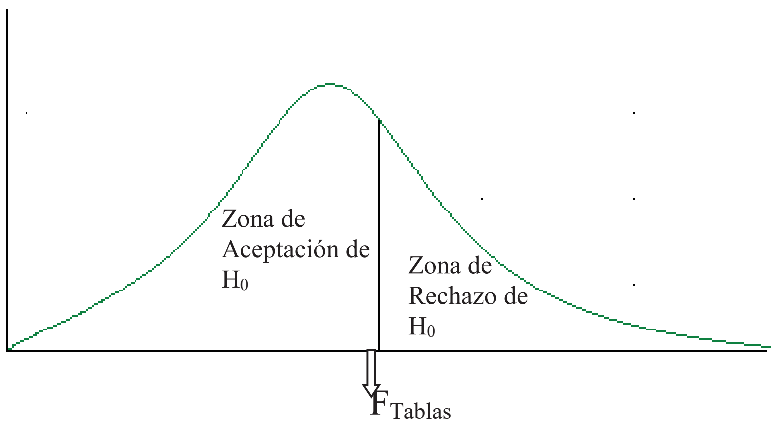
$H_1$ : $\iota_i \neq \iota_j$ para al menos un par (i,j)

Aunque se podría plantear las hipótesis relacionadas con los renglones y columnas; pero sin embargo, posiblemente estas hipótesis no sean apropiadas porque los renglones y las columnas no han sido asignadas aleatoriamente; es decir, tienen restricciones de aleatorización.

Se debe utilizar el Análisis de Varianza para probar estas hipótesis y tomando en cuenta las suposiciones usuales de que $\varepsilon_{ijk}$ es NID$(0,\sigma^2)$ y que cada una de las sumas de cuadrados al dividirlas entre $\sigma^2$, son variables aleatorias independientes con distribución ji-cuadrada.

La hipótesis $H_0$ es frecuentemente la de interés central; por lo tanto, para probarla se utiliza el siguiente estadística:

$F_0 = \dfrac{MS_{Tratamientos}}{MS_E}$, la cual tiene una distribución $F_{\alpha,(p-1),(p-2)(p-1)}$ si la hipótesis nula es verdadera. Por lo tanto, la región crítica es el extremo superior de la distribución F, como se observa a continuación.



De tal modo la hipótesis nula ($H_o$) se rechazará si:

$F_o > F_{\alpha,(p-1),(p-2)(p-1)}$

Donde $F_o$ se obtiene a través del Análisis de Varianza y $F_{\alpha,(p-1),(p-2)(p-1)}$ se obtiene a través de la tabla F.

La tabla siguiente resume el Análisis de Varianza para el Diseño de Cuadrado Latino.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Tratamientos (Letras Latinas) | $SS_{Tratamientos}$ | $p-1$ | $MS_{Tratamientos}$ | $F_o = \frac{MS_{Tratamientos}}{MS_E}$ |
| Renglones | $SS_{Renglones}$ | $p-1$ | $MS_{Renglones}$ | |
| Columnas | $SS_{Columnas}$ | $p-1$ | $MS_{Columnas}$ | |
| Error | $SS_E$ | $(p-2)(p-1)$ | $MS_E$ | |
| Total | $SS_T$ | $p^2-1$ | | |

**Ejemplo 10**

Retomando el ejemplo 8.

Se desean probar 4 máquinas con el objetivo de observar si existe diferencia en la capacidad de producción de una cierta pieza manufacturada. Se conoce que diferentes trabajadores y diferentes períodos de tiempos en un día de trabajo tendrá un efecto sobre la producción. Se eligen a 4 operadores (columnas) y 4 períodos de tiempo (renglones) y se asignan al azar las máquinas correspondientes a cada una de las celdas del cuadrado respetando la restricción que cada máquina se usa por un solo operador en cada período de tiempo. Los datos que se obtuvieron son los siguientes:

| Período de Tiempo | Operadores | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | $y_{i..}$ |
| 1 | C = 31 | D = 43 | A = 67 | B =36 | 177 |
| 2 | D = 39 | A = 96 | B = 40 | C = 48 | 223 |
| 3 | B = 57 | C = 33 | D = 40 | A = 84 | 214 |
| 4 | A = 85 | B = 46 | C = 48 | D = 50 | 229 |
| $y_{..k}$ | 212 | 218 | 195 | 218 | $y_{...}$= 843 |

**Solución**

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

$H_o$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (No hay diferencia entre las máquinas)

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ (Existe diferencia entre las máquinas)

**Variable Respuesta:** Producción de piezas

**El significado verbal es:**

$H_o$: En la producción de una cierta pieza manufacturada no existe diferencia significativa en la capacidad de las cuatro máquinas.

$H_1$: En la producción de una cierta pieza manufacturada existe diferencia significativa en la capacidad de las cuatro máquinas.

**Datos**

$p = 4$, Número total de observaciones: $N = p^2 = (4)^2 = 16$

**Cálculos Matemáticos**

**Cálculo de los totales de tratamientos (Letras Latinas)**

| Tratamientos | $y_{.j.}$ |
|---|---|
| A | 67 + 96 + 84 + 85 = 332 |
| B | 36 + 40 + 57 + 46 = 179 |
| C | 31 + 48 + 33 + 48 = 160 |
| D | 43 + 39 + 40 + 50 = 172 |

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} y_{ijk}^2 - \frac{y_{...}^2}{N} = (31)^2 + (43)^2 + \ldots + (50)^2 - \frac{(843)^2}{4x4} = 5959.44$$

$$SS_{Renglones} = \sum_{i=1}^{4}\frac{y_{i..}^2}{p} - \frac{y_{...}^2}{N} = \frac{(177)^2+(223)^2+(214)^2+(229)^2}{4} - \frac{(843)^2}{16} = 408.19$$

$$SS_{Columnas} = \sum_{k=1}^{4}\frac{y_{..k}^2}{p} - \frac{y_{...}^2}{N} = \frac{(212)^2+(218)^2+(195)^2+(218)^2}{4} - \frac{(843)^2}{16} = 88.69$$

$$SS_{Tratamientos} = \sum_{j=1}^{4}\frac{y_{.j.}^2}{p} - \frac{y_{...}^2}{N} = \frac{(332)^2+(179)^2+(160)^2+(172)^2}{4} - \frac{(843)^2}{16} = 4946.69$$

$SS_E = SS_T - SS_{Renglones} - SS_{Columnas} - SS_{Tratamientos}$

$SS_E = 5959.44 - 4946.69 - 408.19 - 88.69$

$SS_E = 515.87$

**Tabla de Análisis de Varianza.**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Máquinas | 4946.69 | 3 | 1648.90 | 19.18 |
| Períodos de Tiempo | 408.19 | 3 | 136.06 | |
| Operadores | 88.69 | 3 | 29.56 | |
| Error | 515.87 | 6 | 85.98 | |
| Total | 5959.44 | 15 | | |

Utilizando un nivel de significancía del 5% ($\alpha$=0.05), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 3 grados de libertad en el numerador ($p$-1) y 6 grados de libertad en denominador ($p$-2)($p$-1).

$F_{\alpha,p-1,(p-2)(p-1)} = F_{0.05,3,6} = 4.76$

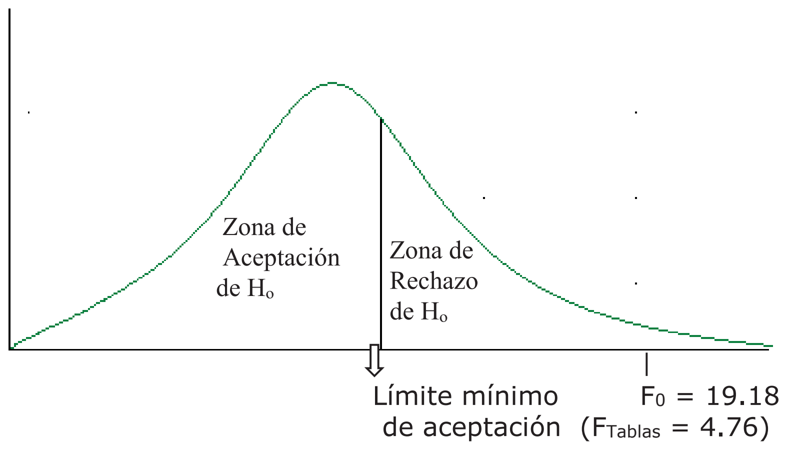
Comparando el $F_0$ calculado en el Análisis de Varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

$19.18 > 4.76$

Por lo tanto, se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$);

También puede observarse gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

En la producción de una cierta pieza manufacturada existe diferencia significativa en la capacidad de las cuatro máquinas.

### 3.1.6 VALOR FALTANTE

En alguna ocasión puede suceder que ocurra que falte una observación en un Diseño de Cuadrado Latino $p$x$p$; lo cual no es conveniente porque se vuelve un Diseño Desbalanceado y ya no podría ser estudiado como un Diseño Completo; por lo tanto, para estimar esta observación se debe utilizar la siguiente fórmula:

$\hat{y}_{ijk} = \dfrac{p(y'_{i..} + y'_{.j.} + y'_{..k}) - 2y'_{...}}{(p-2)(p-1)}$, donde $y'_{i..}$, $y'_{.j.}$ y $y'_{..k}$ son los totales del renglón, tratamiento y columna respectivamente con la observación faltante y $y'_{...}$ es el total general con el valor faltante.

El procedimiento es el mismo que se utilizó en el Diseño por Bloques Aleatorios Completo para encontrar un valor faltante.

De igual forma que se hizo para el caso anterior se pueden encontrar dos observaciones faltantes utilizando esta fórmula reiteradamente.

### 3.1.7 DESVENTAJAS DE LOS CUADRADOS LATINOS PEQUEÑOS

Los cuadrados latinos con un número pequeño de tratamientos, columnas y renglones, por ejemplo menores o iguales que cuatro (4x4) proporcionan un número relativamente reducido de grados de libertad del error que es la desventaja que ellos poseen.

Pero esta desventaja puede corregirse repitiendo el cuadrado Latino n veces, con el objetivo de aumentar los grados de libertad del error.

Las formas en que se pueden llevar a cabo las n repeticiones es en relación a los factores extraños del experimento, de la siguiente manera:

1) Utilizando los mismos Renglones y Columnas en cada réplica.
2) Usando los mismos Renglones pero diferentes Columnas en cada réplica o de manera equivalente, los mismos Renglones pero diferentes Columnas.
3) Utilizando diferentes Renglones y diferentes Columnas.

El Análisis de Varianza para cada caso es diferente; ya que se debe tomar en cuenta la forma en que fue realizada la repetición.

**Para el caso 1.**

La observación $y_{ijkl}$ corresponde a el Renglón i, Tratamiento $j$, la Columna $k$ y la réplica $l$, y existe un número total de $N=np^2$ observaciones; donde n es el número de réplica.

La Tabla de Análisis de Varianza a utilizar en este caso es la siguiente:

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Tratamientos | $\sum_{j=1}^{p}\frac{y_{.j..}^2}{np} - \frac{y_{....}^2}{N}$ | $p$-1 | $\frac{SS_{Tratamientos}}{p-1}$ | $F_o = \frac{MS_{Tratamientos}}{MS_E}$ |
| Renglones | $\sum_{i=1}^{p}\frac{y_{i...}^2}{np} - \frac{y_{....}^2}{N}$ | $p$-1 | $\frac{SS_{Renglones}}{p-1}$ | |
| Columnas | $\sum_{k=1}^{p}\frac{y_{..k.}^2}{np} - \frac{y_{....}^2}{N}$ | $p$-1 | $\frac{SS_{Columnas}}{p-1}$ | |
| Replicaciones | $\sum_{l=1}^{p}\frac{y_{...l}^2}{p^2} - \frac{y_{....}^2}{N}$ | $n$-1 | $\frac{SS_{Replicaciones}}{p-1}$ | |
| Error | Resta | ($p$-1)[$n$($p$+1)-3] | $\frac{SS_E}{(p-1)[n(p+1)-3]}$ | |
| Total | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijkl}^2 - \frac{y_{....}^2}{N}$ | $Np^2$-1 | | |

donde

$$\text{Resta} = \left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijkl}^2 - \frac{y_{....}^2}{N}\right) - \left(\sum_{j=1}^{p}\frac{y_{.j..}^2}{np} - \frac{y_{....}^2}{N} + \sum_{i=1}^{p}\frac{y_{i...}^2}{np} - \frac{y_{....}^2}{N} + \sum_{k=1}^{p}\frac{y_{..k.}^2}{np} - \frac{y_{....}^2}{N} + \sum_{l=1}^{p}\frac{y_{...l}^2}{p^2} - \frac{y_{....}^2}{N}\right)$$

**Para el caso 2**

Existe p nuevos renglones en cada réplica.

La fuente de variación para los renglones miden en realidad, la variabilidad entre los renglones dentro de las n réplicas.

La Tabla de Análisis de Varianza a utilizar en este caso es la siguiente:

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Tratamientos | $\sum_{j=1}^{p} \frac{y_{.j..}^2}{np} - \frac{y_{....}^2}{N}$ | $p$-1 | $\frac{SS_{Tratamientos}}{p-1}$ | $F_o = \frac{MS_{Tratamientos}}{MS_E}$ |
| Renglones | $\sum_{l=1}^{n}\sum_{i=1}^{p} \frac{y_{i..l}^2}{p} - \sum_{l=1}^{n} \frac{y_{...l}^2}{p^2}$ | $n(p$-1) | $\frac{SS_{Renglones}}{n(p-1)}$ | |
| Columnas | $\sum_{k=1}^{p} \frac{y_{..k.}^2}{np} - \frac{y_{....}^2}{N}$ | $p$-1 | $\frac{SS_{Columnas}}{p-1}$ | |
| Replicaciones | $\sum_{l=1}^{p} \frac{y_{...l}^2}{p^2} - \frac{y_{....}^2}{N}$ | $n$-1 | $\frac{SS_{Replicaciones}}{n-1}$ | |
| Error | Resta | ($p$-1)($np$-1) | $\frac{SS_E}{(p-1)(np-1)}$ | |
| Total | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{N}$ | $Np^2$-1 | | |

donde

$$\text{Resta} = \left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{N}\right) - \left(\sum_{j=1}^{p} \frac{\mathrm{y}_{.j..}^2}{\mathrm{np}} - \frac{\mathrm{y}_{....}^2}{\mathrm{N}} + \sum_{l=1}^{n}\sum_{i=1}^{p} \frac{y_{i..l}^2}{p} - \sum_{l=1}^{n} \frac{y_{...l}^2}{p^2} + \sum_{k=1}^{p} \frac{\mathrm{y}_{..k.}^2}{\mathrm{np}} - \frac{\mathrm{y}_{....}^2}{\mathrm{N}} + \sum_{l=1}^{p} \frac{\mathrm{y}_{...1}^2}{\mathrm{p}^2} - \frac{\mathrm{y}_{....}^2}{\mathrm{N}}\right)$$

**Para el caso 3.**

La variabilidad que producen tanto los Renglones como las Columnas, mide la variación de estos factores dentro de las réplicas.

La Tabla de Análisis de Varianza a utilizar en este caso es la siguiente:

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Tratamientos | $\sum_{j=1}^{p} \frac{y_{.j..}^2}{np} - \frac{y_{....}^2}{N}$ | $p$-1 | $\frac{SS_{Tratamientos}}{p-1}$ | $F_o = \frac{MS_{Tratamientos}}{MS_E}$ |
| Renglones | $\sum_{l=1}^{n}\sum_{i=1}^{p} \frac{y_{i..l}^2}{p} - \sum_{l=1}^{n} \frac{y_{...l}^2}{p^2}$ | $n(p$-1) | $\frac{SS_{Renglones}}{n(p-1)}$ | |

Continuación de la tabla de Análisis de Varianza.

| | | | | |
|---|---|---|---|---|
| Columnas | $\sum_{l=1}^{n}\sum_{k=1}^{p}\frac{y_{..kl}^2}{p}-\sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}$ | $n(p\text{-}1)$ | $\frac{SS_{Columnas}}{n(p-1)}$ | |
| Replicaciones | $\sum_{l=1}^{n}\sum_{k=1}^{p}\frac{y_{..kl}^2}{p}-\sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}$ | $n$-1 | $\frac{SS_{\text{Re}\,plicaciones}}{n-1}$ | |
| Error | Resta | $(p\text{-}1)[n(p\text{-}1)\text{-}1]$ | $\frac{SS_E}{(p-1)[n(p-1)-1]}$ | |
| Total | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijkl}^2$ - $\frac{y_{....}^2}{N}$ | $Np^2$-1 | | |

donde

Resta= $\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijkl}^2-\frac{y_{....}^2}{N}\right)$-

$$\left(\sum_{j=1}^{p}\frac{y_{.j..}^2}{np}-\frac{y_{....}^2}{N}+\sum_{l=1}^{n}\sum_{i=1}^{p}\frac{y_{i..l}^2}{p}-\sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}+\sum_{l=1}^{n}\sum_{k=1}^{p}\frac{y_{..kl}^2}{p}-\sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}+\sum_{l=1}^{p}\frac{y_{...l}^2}{p^2}-\frac{y_{....}^2}{N}\right)$$

## 3.1.8 ESTIMACIÓN DE LOS PARÁMETROS

Al igual que los Diseños estudiados anteriormente se puede utilizar el método de mínimos cuadrados para obtener las estimaciones de $\mu$, $\alpha_i$ , $\iota_j$ y $\beta_k$ del modelo:

$$y_{ijk} = \mu + \alpha_i + \iota_j + \beta_k + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,p \\ k = 1,2,\ldots,p \end{cases}$$

Al utilizar el método de mínimos cuadrados se debe tomar en cuenta las restricciones: $\sum_{i=1}^{p}\hat{\alpha}_i = 0$ , $\sum_{j=1}^{p}\hat{\tau}_j = 0$ y $\sum_{k=1}^{p}\hat{\beta}_k = 0$ , por lo tanto, las estimaciones de los parámetros son:

$$\hat{\mu} = \bar{y}_{...} \ , \quad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} \ , \quad \hat{\tau}_j = \bar{y}_{.j.} - \bar{y}_{...} \ , \quad \hat{\beta}_k = \bar{y}_{..k} - \bar{y}_{...}$$

Lo que significa que la media general puede ser estimada por el promedio total de las observaciones, que los efectos de los tratamientos son sólo la diferencia entre el promedio del tratamiento y el promedio total, y cualquiera de los efectos de los renglones son sólo la diferencia entre el promedio de los renglones y el promedio total; mientras que el efecto de las columnas son sólo la diferencia entre el promedio de renglones y el promedio total.

Estas estimaciones encontradas se pueden utilizar para encontrar los valores estimados o ajustados de $y_{ijk}$ sustituyendo en la ecuación del modelo de la siguiente manera:

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\tau}_j + \hat{\beta}_k$$

$$\hat{y}_{ijk} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...})$$

$$= \bar{y}_{...} + \bar{y}_{i..} - \bar{y}_{...} + \bar{y}_{.j.} - \bar{y}_{...} + \bar{y}_{..k} - \bar{y}_{...}$$

es decir

$$\hat{y}_{ijk} = \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - 2\bar{y}_{...}$$

Si se desea obtener las estimaciones mencionadas, utilizar las fórmulas planteadas anteriormente.

## 3.2 DISEÑO DE CUADRADOS GRECO-LATINOS

Si se tiene un cuadrado Latino $pxp$, y se le sobrepone otro Cuadrado Latino en el cual los tratamientos se representan con las letras griegas como $\alpha$, $\beta$, $\gamma$, $\delta$, $\varepsilon$, etc, este Diseño se denomina Diseño de Cuadrado Greco-Latino. Estos dos cuadrados deben ser ortogonales; es decir, que cada letra griega aparezca una sola vez con cada Letra Latina.

Este Diseño se puede utilizar para controlar sistemáticamente tres fuentes extrañas de variabilidad que están representadas en los renglones, columnas y letras griegas; es decir, que el análisis se hace por bloques en tres direcciones, y nos permite analizar el factor renglón, columna, letras griegas y Letras Latinas cada uno con p niveles, usando solamente $p^2$ ensayos.

**Ejemplo 11**

Un ingeniero lleva a cabo un experimento para comparar cinco procedimientos de fabricación de ladrillo de concreto, en el cual usa material de cinco mezclas preparadas cada una de ellas en cada uno de cinco días consecutivos, y preparadas en cinco máquinas diferentes.

**Interpretación del Ejemplo 11**

Como puede observarse este ejemplo corresponde a un Diseño de Cuadrado Greco-Latino; ya que el ingeniero con este experimento pretende investigar que factores intervienen en la fabricación de ladrillo de concreto; para ello utiliza cinco tipos de mezclas (renglones), los cinco días consecutivos en que se preparan dichas mezclas (columnas). Si el ingeniero toma primeramente las cinco máquinas denotadas por las Letras Latinas (A,B,C,D y E) y las relaciona con los tipos de mezclas y los cinco días consecutivos de preparación se forma un Cuadrado Latino 5x5.

Pero si después toma los cinco procedimientos denotados por las letras griegas ($\alpha$,$\beta$, $\gamma$, $\sigma$, $\varepsilon$), y los relaciona con los cinco tipos de mezclas y los cinco días consecutivos de preparación también se forma un Cuadrado Latino 5x5.

En cualquiera de los casos existen dos factores extraños en estudio (mezclas y días), en cada renglón y columna aparece un sola vez una Letra Griega o Latina.

Pero para hacer el estudio completo el ingeniero sobrepone o traslapa estos dos Cuadrados Latinos formando así un Diseño de Cuadrado Greco-Latino; el cual cumple que a cada Letra Latina le corresponde una Letra Griega, y así ya puede estudiar los cuatro factores que son fila, columna, Letras Griegas y Letras Latinas o poner tipos de mezclas, días de preparación, procedimiento y efecto de máquinas.

### 3.2.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS.

Un cuadrado Greco-Latino tendrá la misma representación de los datos de un Cuadrado Latino, con la diferencia que cada Letra Latina ira acompañada de una Letra Griega. Ejemplo de un cuadrado Greco-Latino 4x4.

| Renglones | Columnas | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Total ($y_{i\ldots}$) |
| 1 | A$\alpha$ | B$\beta$ | C$\gamma$ | D$\delta$ | $y_{1\ldots}$ |
| 2 | B$\delta$ | A$\gamma$ | D$\beta$ | C$\alpha$ | $y_{2\ldots}$ |
| 3 | C$\beta$ | D$\alpha$ | A$\delta$ | B$\gamma$ | $y_{3\ldots}$ |
| 4 | D$\gamma$ | C$\delta$ | B$\alpha$ | A$\beta$ | $y_{4\ldots}$ |
| Total ($y_{\ldots l}$) | $Y_{\ldots 1}$ | $y_{\ldots 2}$ | $Y_{\ldots 3}$ | $y_{\ldots 4}$ | $y_{\ldots.}$ |

Como puede observarse en la tabla de datos un cuadrado Greco-latino es un cuadrado $pxp$; con p renglones y $p$ columnas cada una de las celdas $p^2$ contiene una de las letras del Alfabeto y una Letra Griega.

En general un valor de la tabla $y_{ijkl}$ es la observación que corresponde al renglón $i$, la columna $l$, la Letra Latina $j$ y la Letra Griega $k$; y por tanto, habrán $p$ observaciones en cada renglón y cada columna.

Sea:

$y_{i\ldots}$ : Total por renglón

$y_{\ldots l}$ : Total por Columna

$y_{.j..}$ : Total de Letras Latinas

$y_{..k.}$: Total de Letras Griegas

$y_{....}$ : Total General

$\bar{y}_{i...}$ : Promedio por renglón.

$\bar{y}_{...l}$ : Promedio por columna.

$\bar{y}_{.j..}$: Promedio de Letras Latinas.

$\bar{y}_{..k.}$: Promedio de Letras Griegas.

$\bar{y}_{....}$: Promedio General.

donde $y_{.j..}$ y $y_{..k.}$ se obtiene de la siguiente manera:

| Letra Latina | Total ($y_{.j..}$) | Letra Griega | Total ($y_{..k.}$) |
|---|---|---|---|
| A | $y_{.1..}$ | $\alpha$ | $y_{..1.}$ |
| B | $y_{.2..}$ | $\delta$ | $y_{..2.}$ |
| C | $y_{.3..}$ | $\beta$ | $y_{..3.}$ |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| $p$ | $y_{.p..}$ | $p$ | $y_{..P.}$ |

Sólo dos de los cuatro subíndices son necesarios para identificar completamente cualquier observación; matemáticamente se expresan de la siguiente manera:

**Total de renglones**

$$y_{i...} = \sum_{l=1}^{p} y_{i(jk)l} \quad \text{ó} \quad y_{i...} = \sum_{j=1}^{p} y_{ij(kl)} \quad \text{ó} \quad y_{i...} = \sum_{k=1}^{p} y_{i(j)k(l)}$$

**Total de columnas**

$$y_{...l} = \sum_{i=1}^{p} y_{i(jk)l} \quad \text{ó} \quad y_{...l} = \sum_{j=1}^{p} y_{(i)j(k)l} \quad \text{ó} \quad y_{...l} = \sum_{k=1}^{p} y_{(ij)kl}$$

**Total de Letras Latinas**

$$y_{.j..} = \sum_{i=1}^{p} y_{ij(kl)} \quad \text{ó} \quad y_{.j..} = \sum_{l=1}^{p} y_{(i)j(k)l} \quad \text{ó} \quad y_{.j..} = \sum_{k=1}^{p} y_{(i)jk(l)}$$

**Total de Letras Griegas**

$$y_{..k.} = \sum_{i=1}^{p} y_{i(j)k(l)} \quad \text{ó} \quad y_{..k.} = \sum_{i=1}^{p} y_{ij(kl)} \quad \text{ó} \quad y_{..k.} = \sum_{j=1}^{p} y_{(i)jk(l)}$$

**Total General**

$$y_{....} = \sum_{i=1}^{p}\sum_{l=1}^{p} y_{i(jk)l} \quad \text{ó} \quad y_{....} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk(l)} \quad \text{ó} \quad y_{....} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} y_{ijkl} \quad \text{ó} \quad y_{....} = \sum_{i=1}^{p} y_{i...}$$

ó $y_{....}= \sum_{j=1}^{p} y_{.j..}$ ó $y_{....}= \sum_{k=1}^{p} y_{..k.}$ ó $y_{....}= \sum_{l=1}^{p} y_{...l}$

**Nota:** Los subíndices que se encuentra entre paréntesis en las fórmulas anteriores significan que no se toman en cuenta para encontrar cada uno de los totales.

$$\bar{y}_{i...} = \frac{y_{i...}}{p} \quad , \quad \bar{y}_{...l} = \frac{y_{...l}}{p} \quad , \quad \bar{y}_{.j..} = \frac{y_{.j..}}{p} \quad , \quad \bar{y}_{..k.} = \frac{y_{..k.}}{p} \quad \text{y} \quad \bar{y}_{....} = \frac{y_{....}}{p^2}$$

### 3.2.2 MODELO ESTADÍSTICO

El Modelo Estadístico que resulta de un Diseño de Cuadrado Greco-latino es el siguiente:

$$y_{ijkl} = \mu + \alpha_i + \iota_j + \beta_k + \omega_l + \varepsilon_{ijkl} \qquad \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,p \\ k = 1,2,\ldots,p \\ l = 1,2,\ldots,p \end{cases}$$

donde:

$y_{ijkl}$ : Es la observación que corresponde al renglón $i$, la columna $l$, la Letra Latina $j$ y la Letra Griega $k$.

$\mu$ : Es la Media General.

$\alpha_i$ : Es el efecto del i-ésimo Renglón.

$\iota_j$ : Es el efecto del tratamiento $j$ de las Letras Latinas.

$\beta_k$ : Es el efecto del tratamiento $k$ de las Letras Griegas.

$\omega_l$ : Es el efecto de la Columna $l$.

$\varepsilon_{ijkl}$ : Es el componente del Error Aleatorio con Distribución NID$(0,\sigma^2)$.

### 3.2.3 SUMAS Y MEDIAS DE CUADRADOS

Igualmente que cualquiera de los Diseños estudiados el Análisis de Varianza trata de la descomposición de la suma de cuadrados total en sus componentes que la forman:

Sea.

$SS_T$ : Suma total de cuadrados corregida.

$SS_{Renglones}$: Suma de cuadrados debida a los Renglones.

$SS_L$ : Suma de cuadrados debida a los tratamientos de las Letras Latinas.

$SS_G$ : Suma de cuadrados debida a los tratamientos de las Letras Griegas.

$SS_{Columnas}$ : Suma de cuadrados debida a las Columnas.

$SS_E$ : Suma de cuadrados debida al error.

La $SS_T$ es la medida de la variabilidad total de los datos y puede escribirse como:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}(y_{ijkl} - \bar{y}_{....})^2$$

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}(y_{ijkl} - \bar{y}_{....})^2 = p^2\sum_{i=1}^{p}(\bar{y}_{i...} - \bar{y}_{....})^2 + p^2\sum_{j=1}^{p}(\bar{y}_{.j..} - \bar{y}_{....})^2 + p^2\sum_{k=1}^{p}(\bar{y}_{..k.} - \bar{y}_{....})^2 +$$
$$p^2\sum_{l=1}^{p}(\bar{y}_{...l} - \bar{y}_{....})^2 + \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}(y_{ijkl} - \bar{y}_{i...} - \bar{y}_{.j..} - \bar{y}_{..k.} - \bar{y}_{...l} + \bar{y}_{....})^2$$

Al descomponer esta sumatoria, nos queda que:

$SS_T = SS_{Renglones} + SS_L + SS_G + SS_{Columnas} + SS_E$

**Matemáticamente se expresan como:**

$SS_{Renglones} = \sum_{i=1}^{p}\frac{y_{i...}^2}{p} - \frac{y_{....}^2}{N}$ , $SS_L = \sum_{j=1}^{p}\frac{y_{.j..}^2}{p} - \frac{y_{....}^2}{N}$ , $SS_G = \sum_{k=1}^{p}\frac{y_{..k.}^2}{p} - \frac{y_{....}^2}{N}$

$SS_{Columnas} = \sum_{l=1}^{p}\frac{y_{...l}^2}{p} - \frac{y_{....}^2}{N}$ , $SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}y_{ijkl}^2 - \frac{y_{....}^2}{N}$

$SS_E = SS_T - SS_L - SS_G - SS_{Renglones} - SS_{Columnas}$ , con $N = p^2$

El factor representado por las Letras Griegas es ortogonal a los Renglones, Columnas y los tratamientos de la Letra Latina; porque cada letra griega ocurre una sola vez en cada Renglón, en cada Columna y para cada Letra Latina. Por lo tanto, la suma de cuadrados debida al factor Letra Griega puede calcularse usando los totales de la Letra Griega.

$SS_T$ : Esta suma tiene $p^2$ -1 grados de libertad; porque existen $N = p^2$ observaciones en total y un sólo parámetro a estimar $\mu$.

$SS_L$ : Tienen $p$-1 grados de libertad; porque existen $p$ Letras Latinas (tratamientos) y sólo hay un parámetro a estimar $\iota_j$.

$SS_G$ : Tienen $p$-1 grados de libertad; porque existen $p$ Letras Greco-latinas (tratamientos) y sólo hay un parámetro a estimar $\beta_k$.

$SS_{Renglones}$ : Tienen $p$-1 grados de libertad; porque existen $p$ renglones y sólo hay un parámetro a estimar $\alpha_i$ .

$SS_{Columnas}$ : Tienen $p$-1 grados de libertad; porque existen $p$ columnas y sólo hay un parámetro a estimar $\omega_l$ .

$SS_E$ : Esta suma tiene $(p\text{-}3)(p\text{-}1)$ grados de libertad; porque existen $p^2$ celdas que proporcionan $p^2 - 1$ grados de libertad, y como la suma de cuadrados del error es igual a la suma de cuadrados entre las celdas, menos la suma de cuadrados de las letras latinas, la suma de cuadrados de letras griegas, la suma de cuadrados de los renglones y la suma de cuadrados de columnas; entonces los grados de libertad de la suma de cuadrados del error será:

$$p^2 - 1 - (p-1) - (p-1) - (p-1) - (p-1) = p^2 - 1 - p + 1 - p + 1 - p + 1 - p + 1$$

$$= p^2 - 4p + 3 = (p-3)(p-1).$$

Las Medias de Cuadrados igual que cualquier Diseño Experimental están determinadas por las sumas de cuadrados divididas por sus respectivos grados de libertad;es decir, cada suma de cuadrados dividida entre sus grados de libertad es igual a una Media de Cuadrados.

**Matemáticamente se expresan de la manera siguiente:**

$$MS_{Renglones} = \frac{SS_{Renglones}}{p-1}, \quad MS_L = \frac{SS_L}{p-1}, \quad MS_G = \frac{SS_G}{p-1}$$

$$MS_{Columnas} = \frac{SS_{Columnas}}{p-1}, \quad MS_E = \frac{SS_E}{(p-3)(p-1)}$$

Donde

$MS_{Renglones}$ : Suma de Cuadrados Medios de Renglones.
$MS_L$ : Suma de Cuadrados Medios de las Letras Latinas.
$MS_G$ : Suma de Cuadrados Medios de las Letras Griegas.
$MS_{Columnas}$ : Suma de Cuadrados Medios de las Columnas.
$MS_E$ : Suma de Cuadrados Medios del Error.

### 3.2.4 ANÁLISIS ESTADÍSTICO

En realidad en este tipo de Diseño pueden probarse hipótesis de igualdad de renglones, columnas, Letras Latinas y Letras Griegas. Pero las hipótesis de los Renglones y Columnas posiblemente no sea apropiado probar; porque los renglones y las columnas no han sido asignadas aleatoriamente (representan restricciones de aleatorización). Por lo tanto, las hipótesis más apropiadas de probar son:

**Igualdad de Tratamientos asignados a las Letras Latinas**

$H_o : \iota_1 = \iota_2 = . . .= \iota_p$

$H_1 : \iota_i \neq \iota_j$ , para al menos un par (i,j)

**Igualdad de Tratamientos asignados a las Letras Griegas**

$H_o : \beta_1 = \beta_2 =\ldots= \beta_p$

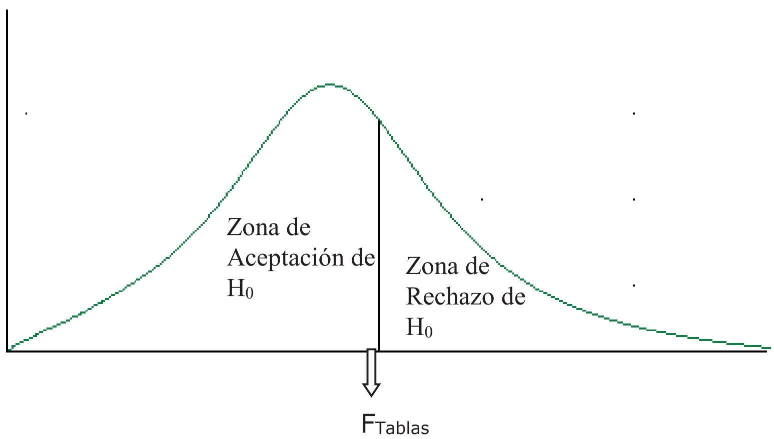
$H_1 : \beta_i \neq \beta_j$ , para al menos un par (i,j)

Se debe utilizar el Análisis de Varianza para probar estas hipótesis, y tomando en cuenta las suposiciones usuales de que $\varepsilon_{ijkl}$ es NID(0,$\sigma^2$), que cada una de las sumas de cuadrados al dividirlas entre $\sigma^2$, son variables aleatorias independientes con distribución ji-cuadrada.

La hipótesis $H_0$ es la de interés central; por lo tanto, para probar las hipótesis antes descritas se deben utilizar los siguientes Estadísticos:

$F_0 = \dfrac{MS_L}{MS_E}$ (Letra Latina) y $F_0 = \dfrac{MS_G}{MS_E}$ (Letra Griega)

El $F_0$ tiene una distribución $F_{\alpha,(p-1),(p-3)(p-1)}$ si la hipótesis nula es verdadera. Por lo tanto la región crítica es el extremo superior de la distribución F, como se observa a continuación.




De tal modo la hipótesis nula ($H_o$) se rechazará si:

$F_o > F_{\alpha,(p-1),(p-3)(p-1)}$

Donde $F_o$ se obtiene a través del Análisis de Varianza y $F_{\alpha,(p-1),(p-3)(p-1)}$ se obtiene a través de la tabla F.

La tabla siguiente resume el Análisis de Varianza para el Diseño de Cuadrado Greco-latino.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Tratamientos de Letra Latina | $SS_L$ | $p-1$ | $MS_L = \dfrac{SS_L}{p-1}$ | $F_o = \dfrac{MS_L}{MS_E}$ |
| Tratamientos de Letra Griega | $SS_G$ | $p-1$ | $MS_G = \dfrac{SS_G}{p-1}$ | $F_o = \dfrac{MS_G}{MS_E}$ |
| Renglones | $SS_{Renglones}$ | $p-1$ | $MS_{Renglones} = \dfrac{SS_{Renglones}}{p-1}$ | |
| Columnas | $SS_{Columnas}$ | $p-1$ | $MS_{Columnas} = \dfrac{SS_{Columnas}}{p-1}$ | |
| Error | $SS_E$ | $(p-3)(p-1)$ | $MS_E = \dfrac{SS_E}{(p-3)(p-1)}$ | |
| Total | $SS_T$ | $p^2-1$ | | |

### 3.2.5 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Al igual que los Diseños estudiados anteriormente se puede utilizar el método de mínimos cuadrados para obtener las estimaciones de $\mu$, $\alpha_i$ , $\iota_j$ , $\beta_k$ y $\omega_l$ del Modelo:

$$y_{ijkl} = \mu + \alpha_i + \iota_j + \beta_k + \omega_l + \varepsilon_{ijkl} \qquad \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,p \\ k = 1,2,\ldots,p \\ l = 1,2,\ldots,p \end{cases}$$

Al utilizar el método de mínimos cuadrados se debe tomar en cuenta las restricciones: $\sum_{i=1}^{p} \hat{\alpha}_i = 0$ , $\sum_{j=1}^{p} \hat{\tau}_j = 0$ , $\sum_{k=1}^{p} \hat{\beta}_k = 0$ y $\sum_{l=1}^{p} \hat{\omega}_l = 0$; por lo tanto, las estimaciones de los parámetros son:

$$\hat{\mu} = \bar{y}_{....} \qquad \hat{\alpha}_i = \bar{y}_{i...} - \bar{y}_{....} \qquad \hat{\tau}_j = \bar{y}_{.j..} - \bar{y}_{....} \qquad \hat{\beta}_k = \bar{y}_{..k.} - \bar{y}_{....} \qquad \hat{\omega}_l = \bar{y}_{...l} - \bar{y}_{....}$$

Lo que significa que la media general puede ser estimada por el promedio total de las observaciones, cualquiera de los efectos de los renglones son sólo la diferencia entre el promedio de los renglones y el promedio total, que los efectos de los tratamientos (Letras Latinas) son sólo la diferencia entre el promedio del tratamiento (Letras Latinas) y el promedio total, que los efectos de los tratamientos (Letras Griegas) son sólo la diferencia entre el promedio del tratamiento (Letras Griegas) y el promedio total, mientras que el efecto de las columnas son sólo la diferencia entre el promedio de columnas y el promedio total.

Estas estimaciones encontradas se pueden utilizar para encontrar los valores estimados o ajustados de $y_{ijkl}$ sustituyendo en el ecuación del modelo de la siguiente manera:

$$\hat{y}_{ijkl} = \hat{\mu} + \hat{\alpha}_i + \hat{\tau}_j + \hat{\beta}_k + \hat{\omega}_l$$

$$\hat{y}_{ijkl} = \bar{y}_{....} + (\bar{y}_{i...} - \bar{y}_{....}) + (\bar{y}_{.j..} - \bar{y}_{....}) + (\bar{y}_{..k.} - \bar{y}_{....}) + (\bar{y}_{...l} - \bar{y}_{....})$$

$$= \bar{y}_{....} + \bar{y}_{i...} - \bar{y}_{....} + \bar{y}_{.j..} - \bar{y}_{....} + \bar{y}_{..k.} - \bar{y}_{....} + \bar{y}_{...l} - \bar{y}_{....}$$

es decir

$$\hat{y}_{ijkl} = \bar{y}_{i...} + \bar{y}_{.j..} + \bar{y}_{..k.} + \bar{y}_{...l} - 3\bar{y}_{....}$$

**Ejemplo 12**

Un experimento podría usarse para comparar 5 procedimientos ($\alpha$,$\beta$, $\gamma$, $\sigma$, $\varepsilon$) para fabricar ladrillos de concreto, usando material de 5 mezclas preparadas cada una de ellas en cada uno de 5 días consecutivos, y preparada en 5 máquinas diferentes (A,B,C,D,E). Los dados se presentan en la siguiente tabla:

| Renglones (Mezclas) | Columnas (Días) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | $y_{i...}$ |
| 1 | A$\alpha$ = 1 | B$\gamma$ = 0 | C$\varepsilon$ = 4 | D$\beta$ = 0 | E$\delta$ = 1 | 6 |
| 2 | B$\beta$ = 3 | C$\delta$ = 4 | D$\alpha$ = 1 | E$\gamma$ = 5 | A$\varepsilon$ = 3 | 16 |
| 3 | C$\gamma$ = 2 | D$\varepsilon$ = 5 | E$\beta$ = 0 | A$\delta$ = 0 | B$\alpha$ = -1 | 6 |
| 4 | D$\delta$ = -1 | E$\alpha$ = 2 | A$\gamma$ = -1 | B$\varepsilon$ = 1 | C$\beta$ = 4 | 5 |
| 5 | E$\varepsilon$ = 0 | A$\beta$ = 1 | B$\delta$ = -2 | C$\alpha$ =-3 | D$\gamma$ = 1 | -3 |
| $y_{...l}$ | 5 | 12 | 2 | 3 | 8 | $Y_{....}$ = 30 |

**Solución**

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

**Para las letras Latinas**

$H_o$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (No hay diferencia entre las Máquinas)

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 = \mu_5$ (Existe diferencia entre las Máquinas)

**Para las letras Griegas**

$H_o$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (No hay diferencia entre los procedimientos)

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 = \mu_5$ (Existe diferencia entre los procedimientos)

**Variable Respuesta:** Fabricación de Ladrillos de concreto.

**El significado verbal de las hipótesis de Letras Latinas es:**

$H_o$: En la fabricación de ladrillos no influye significativamente el tipo de Máquina que los produce.

$H_1$: En la fabricación de ladrillos influye significativamente el tipo de Máquina que los produce.

**El significado verbal de las hipótesis de Letras Griegas es:**

$H_o$: En la fabricación de ladrillos no influyen significativamente los procedimientos que utilizan para fabricarlos.

$H_1$: En la fabricación de ladrillos influyen significativamente los procedimientos que utilizan para fabricarlos.

**Datos.**

$p = 5$ , Número total de observaciones: $N = p^2 = (5)^2 = 25$

**Cálculos Matemáticos**

**Cálculo de los totales de tratamientos (Letras Latinas y Letras Griegas)**

| Letras Latinas | Total $y_{.j..} = \sum_{i=1}^{p} y_{ij(kl)}$ | Letras Griegas | Total $y_{..k.} = \sum_{i=1}^{p} y_{ij(kl)}$ |
|---|---|---|---|
| A | 1+3+0-1+1 = 4 | $\alpha$ | 1+1-1+2-3 = 0 |
| B | 0+3-1+1-2 = 1 | $\beta$ | 0+3+0+4+1 = 8 |
| C | 4+4+2+4-3 = 11 | $\gamma$ | 0+5+2-1+1 = 7 |
| D | 0+1+5-1+1 = 6 | $\delta$ | 1+4+0-1-2 = 2 |
| E | 1+5+0+2+0 = 8 | $\varepsilon$ | 4+3+5+1+0 = 13 |

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{k=1}^{5}\sum_{l=1}^{5} y_{ijkl}^2 - \frac{y_{....}^2}{N} = (1)^2 + (0)^2 + \ldots + (1)^2 - \frac{(30)^2}{5x5} = 110.0$$

$$SS_L = \sum_{j=1}^{5} \frac{y_{.j..}^2}{p} - \frac{y_{....}^2}{N} = \frac{(4)^2 + (1)^2 + (11)^2 + (6)^2 + (8)^2}{5} - \frac{(30)^2}{25} = 11.6$$

$SS_G = \sum_{k=1}^{5} \frac{y_{..k.}^2}{p} - \frac{y_{....}^2}{N} = \frac{(0)^2 + (8)^2 + (7)^2 + (2)^2 + (13)^2}{5} - \frac{(30)^2}{25} = 21.2$

$SS_{Renglones} = \sum_{i=1}^{5} \frac{y_{i...}^2}{p} - \frac{y_{....}^2}{N} = \frac{(6)^2 + (16)^2 + (6)^2 + (5)^2 + (-3)^2}{5} - \frac{(30)^2}{25} = 36.4$

$SS_{Columnas} = \sum_{l=1}^{5} \frac{y_{...l}^2}{p} - \frac{y_{....}^2}{N} = \frac{(5)^2 + (12)^2 + (2)^2 + (3)^2 + (8)^2}{5} - \frac{(30)^2}{25} = 13.2$

$SS_E = SS_T - SS_L - SS_G - SS_{Renglones} - SS_{Columnas}$

$SS_E = 110 - 11.6 - 21.2 - 36.4 - 13.2$

$SS_E = 27.6$

**Medias de Cuadrados**

$MS_L = \frac{SS_L}{p-1} = \frac{11.60}{4} = 2.9$ , $MS_G = \frac{SS_G}{p-1} = \frac{21.20}{4} = 5.3$

$MS_{Renglones} = \frac{SS_{Renglones}}{p-1} = \frac{36.40}{4} = 9.1$ , $MS_{Columnas} = \frac{SS_{Columnas}}{p-1} = \frac{13.20}{4} = 3.3$

$MS_E = \frac{SS_E}{(p-3)(p-1)} = \frac{27.60}{8} = 3.45$

**Estadísticas de prueba**

$F_0 = \frac{MS_L}{MS_E} = \frac{2.90}{3.45} = 0.84$ Letras Latinas , $F_0 = \frac{MS_G}{MS_E} = \frac{5.30}{3.45} = 1.54$ Letras Griegas

**Tabla de Análisis de Varianza.**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Máquinas (Letras Latinas) | 11.60 | 4 | 2.90 | 0.84 |
| Procedimientos (Letra Griega) | 21.20 | 4 | 5.30 | 1.54 |
| Mezclas (Renglones) | 36.40 | 4 | 9.10 | |
| Días (Columnas) | 13.20 | 4 | 3.30 | |
| Error | 27.60 | 8 | 3.45 | |
| Total | 110.0 | 24 | | |

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 4 grados de libertad (p-1) en el numerador y 8 grados de libertad (p-3)(p-1) en denominador.

$F_{\alpha,p-1,(p-3)(p-1)} = F_{0.05,4,8} = 3.84$

Comparando el $F_0$ calculado de el Análisis de Varianza y el $F_{Tablas}$ , se puede observar que:

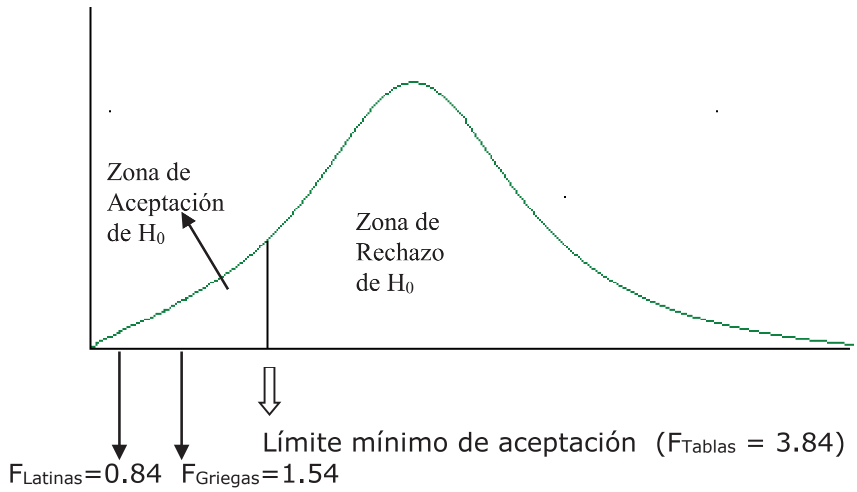
$F_0 < F_{Tablas}$

0.84 < 3.84 Letras Latinas (Máquinas)

1.54 < 3.84 Letras Griegas (Procedimientos)

Por tanto, en ambos casos se acepta la hipótesis nula ($H_0$) y se rechaza la hipótesis alternativa ($H_1$).

También puede observarse gráficamente, de la siguiente manera:




Se observa que ambos valores de $F_0$ caen en la zona de Aceptación de $H_0$.

**Conclusiones**

1) En la fabricación de ladrillos no influye significativamente el tipo de Máquina que los produce.
2) En la fabricación de ladrillos no influyen significativamente los procedimientos que se utilizan para fabricarlos.

### 3.2.6 SELECCIÓN ALEATORIA DE UN CUADRADO GRECO-LATINO

Para llevar a cabo la selección aleatoria de un Cuadrado Greco-latino se debe realizar el siguiente procedimiento:

1) Escoger al azar dos de los Cuadrados Latinos de la tabla que son ortogonales y del mismo orden.
2) Al primer Cuadrado Latino cambiarle las letras A,B,C,D,…, por las letras $\alpha,\beta,\gamma,\delta,\ldots.$,
3) Luego superponerlos, obteniendo como resultado un sólo cuadrado, que es el Cuadrado Greco-latino.
4) Permutar las filas y las columnas al azar como en el Cuadrado Latino.
5) Asignar también al azar las letras Griegas y Latinas a los tratamientos.

**Ejemplo 13**

1) Si se escogen los Cuadrado Latino de $p$=3, y se elige al azar el Primero y segundo cuadrado Latino de orden tres de la tabla. Es decir:

Primero Segundo

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | A | B | C |
| B | C | A | C | A | B |
| C | A | B | B | C | A |

2) Al primer Cuadrado Latino se cambia a las letras griegas $\alpha,\beta,\gamma$; obteniéndose el siguiente cuadrado Latino.

Primero Segundo

| | | | | | |
|---|---|---|---|---|---|
| $\alpha$ | $\beta$ | $\gamma$ | A | B | C |
| $\beta$ | $\gamma$ | $\alpha$ | C | A | B |
| $\gamma$ | $\alpha$ | $\beta$ | B | C | A |

3) Superponiendo los dos Cuadrados Latinos, se obtiene el siguiente Cuadrado Greco-latino:

| | | |
|---|---|---|
| $\alpha$A | $\beta$B | $\gamma$C |
| $\beta$C | $\gamma$A | $\alpha$B |
| $\gamma$B | $\alpha$C | $\beta$A |

4) Utilizando los números aleatorios se obtiene el orden para las filas de (1,3 y 2) y para las columnas de (2,1 y 3). El cuadrado resultante de reordenar las filas será:

| | | |
|---|---|---|
| $\alpha$A | $\beta$B | $\gamma$C |
| $\gamma$B | $\alpha$C | $\beta$A |
| $\beta$C | $\gamma$A | $\alpha$B |

Y luego permutando las columnas se tiene el cuadrado.

$\beta$B $\alpha$A $\gamma$C
$\gamma$A $\beta$C $\alpha$B
$\alpha$C $\gamma$B $\beta$A

5) Por último asignando al azar las Letras Latina y Letras Griegas a los tratamientos, y así se tendrá el Diseño de Cuadrado Grecolatino buscado.

## 4. DISEÑOS POR BLOQUES INCOMPLETOS BALANCEADOS

### DESCRIPCIÓN

Puede darse el caso que en algunos experimentos que usan Diseños Aleatorizados por Bloques no se puedan llevar a cabo los ensayos de todas las combinaciones de tratamientos dentro de cada bloque; ya sea por la escasez de los recursos del experimento, por la situación económica, o por el tamaño físico de los bloques. Para analizar estos tipos de experimentos se usa el Diseño Aleatorizado por Bloques en los que cada tratamiento no está presente en cada bloque; y a este tipo de Diseño se conocen como Diseño Aleatorizado por Bloques Incompletos.

Un Diseño particular de ellos son los Diseños por Bloques Incompletos Balanceados; el cual consiste en un Diseño por Bloques Incompleto en el que cualquier par de tratamientos ocurren juntos el mismo número de veces.

Si se tienen "$a$" tratamientos y se pueden probar $k$ ($k$<$a$) tratamientos en cada bloque entonces un Diseño Balanceado por Bloques Incompletos puede ser construido tomando $\binom{a}{k}$ bloques y asignándose una combinación de tratamientos diferentes a cada bloque. Sin embargo frecuentemente es posible obtener un Diseño Balanceado con menos de $\binom{a}{k}$ bloques.

**Ejemplo 14**

Siete diferentes concentraciones de madera están siendo estudiadas para determinar su efecto sobre la resistencia del papel producido. Sin embargo, la planta piloto sólo puede producir tres ensayos diarios. Como puede existir variación a causa de los días, el analista observa el experimento siete días consecutivos.

**Interpretación.**

Este ejemplo es un Diseño Aleatorizado por Bloques, en donde los días se toman como bloques (7 días) y las concentraciones de fibra como los tratamientos (7 concentraciones).

Pero existe un problema, es que en la planta piloto sólo pueden llevarse a cabo tres ensayos diarios; es decir, que sólo se pueden probar tres concentraciones por día (bloque) y como son siete concentraciones quedaran cuatro sin probar cada día; de las concentraciones se seleccionan tres al azar de las siete para probarse. Es por este motivo que el Diseño Aleatorizado por Bloque se convierte en un Diseño Aleatorizado por Bloques Incompletos Balanceado.

### 4.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

Como se han seleccionado "$b$" bloques y se tienen "$a$" tratamientos de los cuales se prueban $k$ en cada bloque. La representación de los datos para este tipo de Diseño es la siguiente.

| Tratamientos | Bloques | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | ... | b | $y_{i.}$ |
| 1 | $y_{11}$ | — | — | ... | — | $y_{1.}$ |
| 2 | — | $y_{22}$ | $y_{23}$ | ... | $y_{2b}$ | $y_{2.}$ |
| 3 | $y_{31}$ | — | $y_{33}$ | .. | — | $y_{3.}$ |
| .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. | .<br>.<br>. |
| A | — | $y_{a2}$ | — | ... | $y_{ab}$ | $y_{a.}$ |
| $y_{.j}$ | $y_{.1}$ | $y_{.2}$ | $y_{.3}$ | ... | $y_{.b}$ | $y_{..}$ |

Como se observa en la tabla se prueban $k$ tratamientos en cada bloque, que cada tratamiento ocurre r veces en el Diseño (o se repite r veces) y que hay un total de N = $ar= bk$ observaciones. Entonces el número de veces que cada par de tratamientos ocurre a la vez en los bloques viene dado por:

$$\lambda = \frac{r(k-1)}{a-1}, \text{ donde } \lambda < r < b \text{ y debe ser un número entero.}$$

La deducción de $\lambda$, se da cuando se toma cualquier tratamiento; este ocurre en $r$ bloques, y hay otros $k$-1 tratamientos en cada uno de estos bloques, existen $r(k\text{-}1)$ observaciones en un bloque que contienen a este tratamiento. Estas $r(k\text{-}1)$ observaciones deben representar al resto de los $a$ -1 tratamientos $\lambda$ veces. Por lo tanto, $\lambda(a\text{-}1) = r(k\text{-}1)$.

Las medias de tratamiento observadas no proporcionan estimadores sesgados de $\mu_i = \mu + \iota_i$, debido a que los efectos de tratamientos y bloques no son ortogonales en el Diseño de Bloques Incompletos; porque no aparecen todos los tratamientos en cada bloque.

Por lo tanto, se utilizarán los estimadores de los efectos de tratamiento como estimadores de las medias de los tratamientos.

También no sería correcto para el Diseño Incompleto calcular la partición de suma de cuadrados para los tratamientos igual que para los Diseños de Bloques Completos.

Sea:

$y_{i.}$ : Total de observaciones del tratamiento i.

$y_{.j}$ :Total de observaciones del bloque j.

$y_{..}$ : Total de todas las observaciones.

$\bar{y}_{..}$ : Promedio de todas las observaciones.

**Matemáticamente se expresan de la siguiente manera:**

$$y_{i.} = \sum_{j=1}^{b} y_{ij} \ , \ y_{.j} = \sum_{i=1}^{a} y_{ij} \qquad y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} \ \text{ ó } \ y_{..} = \sum_{i=1}^{a} y_{i.} = \sum_{j=1}^{b} y_{.j} \ , \quad \bar{y}_{..} = \frac{y_{..}}{N}$$

$i = 1,2,\ldots, a$

$j = 1,2,\ldots, b$

donde $N = ar = bk$, número total de observaciones.

**Nota**: El valor faltante del renglón o columna numéricamente se toma como cero en las sumatorias para encontrar los totales.

## 4.2 MODELO ESTADÍSTICO

Las observaciones de la tabla anterior se pueden escribir por medio del siguiente Modelo Estadístico Lineal:

$$y_{ij} = \mu + \iota_i + \beta_j + \varepsilon_{ij} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

donde:

$y_{ij}$ : Es la i-ésima observación del j-ésimo bloque.

$\mu$ : Es la media general.

$\iota_i$ : Efecto del i-ésimo tratamiento.

$\beta_j$ : Efecto del j-ésimo bloque.

$\varepsilon_{ij}$ : Componente el error aleatorio NID(0,$\sigma^2$).

Los tratamientos y los bloques se consideran que son factores fijos. Además, los efectos de los tratamientos y de bloques se consideran como desviaciones de la media general, por lo tanto: $\sum_{i=1}^{a} \hat{\tau}_i = 0$ y $\sum_{j=1}^{b} \hat{\beta}_j = 0$

## 4.3 SUMAS Y MEDIAS DE CUADRADOS

En este tipo de experimento se debe corregir (o ajustar) la suma total de cuadrados, mediante la suma de cuadrados de tratamiento, con el objetivo de separar los efectos de tratamiento y de bloque; ya que cada tratamiento ocurre en un conjunto diferente de r bloques. Y por esta razón las diferencias entre los totales de tratamiento no corregidos $y_{1.}$ , $y_{2.}$ , ...., $y_{a.}$ también son afectadas por las diferencias entre los bloques.

El Análisis de Varianza trata sobre la descomposición de la variabilidad total de los datos, la cual estará expresada mediante la suma total de cuadrados corregida ( o ajustada) de la siguiente manera:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{..})^2$$

Al igual que el Diseño por Bloques Completo esta variabilidad total puede descomponerse como:

$SS_T = SS_{\text{Tratamientos(ajustada)}} + SS_{\text{Bloques}} + SS_E$

La suma de cuadrados de tratamiento corregida(o ajustada) se lleva a cabo de la siguiente forma:

Se debe calcular el total corregido del i-ésimo tratamiento ($Q_i$) utilizando la siguiente fórmula:

$$Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij}\, y_{.j}\ , \text{ con } n_{ij} = \begin{cases} 1, \text{si el tratamiento i ocurre en el bloque j} \\ 0, \text{en caso contrario} \end{cases} \qquad i = 1,2,\ldots,a$$

Entonces $\frac{1}{k}\sum_{j=1}^{b} n_{ij}\, y_{.j}$ es el promedio de los totales de los bloques en los que se aplica el tratamiento i. Y la suma de los totales de tratamiento corregidos siempre sarán igual a cero.

Matemáticamente las sumas de cuadrados en que está descompuesta la variabilidad total se obtienen de la siguiente manera:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N}\ , \quad SS_{\text{Tratamientos(ajustada)}} = \frac{k\sum_{i=1}^{a} Q_i^2}{\lambda a}\ , \quad SS_{\text{Bloques}} = \sum_{j=1}^{b}\frac{y_{.j}^2}{k} - \frac{y_{..}^2}{N}$$

$SS_E$ se obtiene por diferencia de igual forma que los anteriores Diseños Experimentales, de la siguiente forma: $SS_E = SS_T - SS_{\text{Tratamientos(ajustada)}} - SS_{\text{Bloques}}$

donde

$SS_T$ : Esta suma de cuadrados corregida tiene $N$-1 grados de libertad, porque existen $N$ observaciones en total y un sólo parámetro a estimar $\mu$ .

$SS_{\text{Tratamientos(ajustada)}}$ : Tiene $a$-1 grados de libertad porque existen "$a$" tratamientos y un sólo parámetro a estimar $\iota_i$ .

$SS_{\text{Bloques}}$ : Tiene $b$-1 grados de libertad por haber "$b$" bloques y un sólo parámetro a estimar $\beta_j$.

$SS_E$ : Tiene $N$-$a$-$b$+1 grados de libertad; porque existen $N$ celdas que Proporcionan $N$-1 grados de libertar, y como la suma de cuadrados del error es igual a la suma de cuadrados entre las celdas menos la suma de cuadrados de tratamientos ajustada y la suma de cuadrados de los bloques; entonces los grados de libertar de la suma de cuadrados del error será: $N$-1-($a$-1)-($b$-1) = N - $a$- $b$ +1.

Si en algunas ocasiones se desea evaluar los efectos de los bloque, la variabilidad total de los datos puede ser descompuesta como:

$$SS_T = SS_{\text{Tratamientos}} + SS_{\text{Bloques(ajustada)}} + SS_E$$

En este caso se observa que $SS_{\text{Tratamientos}}$ no aparece corregida si no que la de bloques. La suma de bloques (corregida) se puede encontrar de la siguiente forma:

Primeramente se debe obtener el total corregido del j-ésimo bloque ($Q_j$), utilizando la siguiente fórmula

$$Q_j = y_{.j} - \frac{1}{r}\sum_{i=1}^{a} n_{ij}\, y_{i.}\ , \quad \text{con } n_{ij} = \begin{cases} 1, \text{si el bloque j ocurre en el tratamiento i} \\ 0, \text{en caso contrario} \end{cases} \qquad j = 1,2,\ldots,b$$

Entonces $\frac{1}{r}\sum_{i=1}^{a} n_{ij}\, y_{i.}$ es el promedio de los totales de los tratamientos en los que se aplica el bloque j. Por ser $a$=$b$ el Diseño Balanceado por Bloques Incompletos se dice que es Simétrico.

Matemáticamente las sumas de cuadrados en que está descompuesta la variabilidad total se obtienen de la siguiente manera:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N} \quad , SS_{Tratamientos} = \sum_{i=1}^{a} \frac{y_{i.}^2}{r} - \frac{y_{..}^2}{N} \quad , SS_{Bloques(ajustada)} = \frac{r\sum_{j=1}^{b} Q_j^2}{\lambda b}$$

$SS_E$ se obtiene por diferencia de igual forma que los anteriores Diseños Experimentales, de la siguiente forma: $SS_E = SS_T - SS_{Tratamientos} - SS_{Bloques(ajustada)}$

Porque los tratamientos y los bloques no son ortogonales $SS_T \neq SS_{Tratamientos(ajustada)} + SS_{Bloques(ajustada)} + SS_E$ .

Las medias de cuadrados se definen en función de las sumas de cuadrados y sus respectivos grados de libertad de la siguiente forma:

$$MS_{Tratamientos(ajustada)} = \frac{SS_{Tratamientos(ajustada)}}{a-1}$$

$$MS_{Bloques} = \frac{SS_{Bloques}}{b-1} \quad , \quad MS_E = \frac{SS_E}{N-a-b+1}$$

donde :

$MS_{Tratamiento(ajustada)}$ : Suma de cuadrados medios ajustada entre tratamientos.
$MS_{Bloques}$ : Suma de cuadrados medios entre bloques.
$MS_E$ : Suma de cuadrados medios del error.

## 4.4 ANÁLISIS ESTADÍSTICO

En este tipo de Diseño Experimental el análisis estadístico también inicia planteándose las hipótesis que se desean probar; éstas se deben hacer en relación a las medias de los tratamientos.

$H_o : \mu_1 = \mu_2 = \ldots\ldots = \mu_a$
$H_1 : \mu_i \neq \mu_j$ para cuando menos un par

Luego de haber planteado las hipótesis se debe llevar a cabo el Análisis de Varianza para aceptar o rechazar dichas hipótesis.

Y la estadística apropiada para probar la igualdad de los tratamientos, viene dada por:

$F_o = \frac{MS_{Tratamientos(ajustada)}}{MS_E}$, donde $F_o$ tiene una distribución $F_{\alpha,a-1,N-a-b+1}$ si la hipótesis nula es verdadera.

De modo que la hipótesis nula se debe rechazar si $F_o > F_{\alpha,a-1,N-a-b+1}$

$F_o$ se obtiene por medio del Análisis de Varianza y $F_{\alpha,a-1,N-a-b+1}$ por medio de la tabla F.

La tabla siguiente resume el Análisis de Varianza.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Tratamientos (corregidos) | $SS_{Tratamientos(ajustada)}$ | $a-1$ | $MS_{Tratamientos(ajustada)}$ | $F_o = \dfrac{MS_{Tratamientos(ajustada)}}{MS_E}$ |
| Bloques | $SS_{Bloques}$ | $b-1$ | $MS_{Bloques}$ | |
| Error | $SS_E$ | $N-a-b+1$ | $MS_E$ | |
| Total | $SS_T$ | $N-1$ | | |

**Ejemplo 15**

Un horticultor experimentó con la germinación de semillas de tomate a cuatro temperaturas diferentes (25°c, 30°c, 35°c y 40°c) en un Diseño de Bloques Incompleto Balanceado, porque sólo disponía de dos cámaras de cultivo para el estudio, cada corrida del experimento fue un bloque que consistía en dos cámaras de cultivo como unidades experimentales y se asignaron al azar dos temperaturas a las cámaras para cada corrida. Los siguientes datos son las tasas de germinación de las semillas de tomate.

| Tratamientos | Corridas (Bloques) | | | | | | |
|---|---|---|---|---|---|---|---|
| Temperaturas | 1 | 2 | 3 | 4 | 5 | 6 | $y_{i.}$ |
| 25°c | 24.65 | — | 29.17 | — | 28.90 | — | 82.72 |
| 30°c | — | 24.38 | 21.25 | — | — | 25.53 | 71.16 |
| 35°c | — | — | — | 5.90 | 18.27 | 8.42 | 32.59 |
| 40°c | 1.34 | 2.24 | — | 1.83 | — | — | 5.41 |
| $y_{.j}$ | 25.99 | 26.62 | 50.42 | 7.73 | 47.17 | 33.95 | $y_{..}$= 191.88 |

**Solución**

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

$H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4$ (No existe diferencia entre las temperaturas).

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ (Existe diferencia entre las temperaturas).

**Variable Respuesta:** Tasa de Germinación de la Semilla de Tomate.

**El significado verbal es:**

$H_o$ : El nivel de temperatura no influye significativamente en la tasa de germinación de la semilla de tomate.

$H_1$ : El nivel de temperatura influye significativamente en la tasa de germinación de la semilla de tomate.

**Datos**

$a = 4$, $b = 6$, $k = 2$ , $r = 3$ , $N = ar = bk = (4)(3) = 12$ , $\lambda = \frac{r(k-1)}{a-1} = \frac{3(2-1)}{4-1} = \frac{3}{3} = 1$

**Cálculos Matemáticos**

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N} = (24.65)^2 + (29.17)^2 + \ldots + (1.83)^2 - \frac{(191.88)^2}{12}$$

$$= 4441.1106 - 3068.1612$$

$$SS_T = 1372.95$$

$$SS_{Bloques} = \sum_{j=1}^{6} \frac{y_{.j}^2}{k} - \frac{y_{..}^2}{N}$$

$$= \frac{(25.99)^2 + (26.62)^2 + (50.42)^2 + (7.73)^2 + (47.17)^2 + (33.95)^2}{2} - \frac{(191.88)^2}{12}$$

$$= 3681.8226 - 3068.1612$$

$$SS_{Bloques} = 613.66$$

Para calcular la suma de cuadrados de tratamientos corregida. Primero hay que encontrar los totales de tratamientos corregidos de la siguiente manera:

$$Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{6} n_{ij}\, y_{.j} , \qquad i=1,2,3,4$$

$$Q_1 = y_{1.} - \frac{1}{2}(25.99+50.42+47.17) = 82.72 - 61.79 = 20.93$$

$$Q_2 = y_{2.} - \frac{1}{2}(26.62+50.42+33.95) = 71.16 - 55.495 = 15.66$$

$$Q_3 = y_{3.} - \frac{1}{2}(7.73+47.17+33.95) = 32.59 - 44.425 = -11.83$$

$$Q_4 = y_{4.} - \frac{1}{2}(25.99+26.62+7.73) = 5.41 - 30.17 = -24.76$$

$$SS_{\text{Tratamientos(ajustada)}} = \frac{k\sum_{i=1}^{4} Q_i^2}{\lambda a} = \frac{2\left[(20.93)^2 + (15.66)^2 + (-11.83)^2 + (-24.76)^2\right]}{(4)(1)}$$

$$= \frac{2(1436.307)}{4}$$

$SS_{\text{Tratamientos(ajustada)}}$ = 718.15

$SS_E = SS_T - SS_{\text{Tratamientos(ajustada)}} - SS_{\text{Bloques}}$

= 1372.95 - 718.15 - 613.66

$SS_E$ = 41.14

**Estadística**

$$F_0 = \frac{MS_{\text{Tratamientos(ajustada)}}}{MS_E} = \frac{239.38}{13.71} = 17.46$$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrado | $F_o$ |
|---|---|---|---|---|
| Temperaturas (corregidos) | 718.15 | 3 | 239.38 | 17.46 |
| Bloques (No ajustados) | 613.66 | 5 | 122.73 | |
| Error | 41.14 | 3 | 13.71 | |
| Total | 1372.95 | 11 | | |

Utilizando un nivel de significancía del 5% ($\alpha$ = 0.05), para encontrar el $F_{\text{Tablas}}$ (Tablas Fisher) con 3 grados de libertad ($a$ -1) en el numerador y 3 grados de libertad $N$- $a$- $b$ +1 en denominador.

$F_{\alpha,a-1,N-a-b+1} = F_{0.05,3,3}$ = 9.28

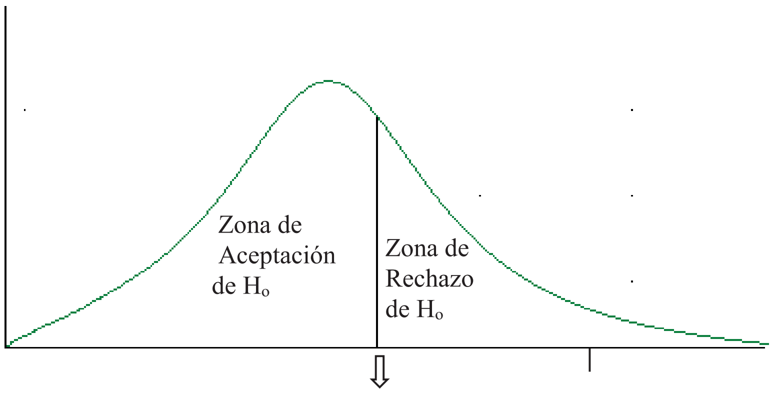
Comparando el $F_0$ calculado en el análisis de varianza y el $F_{\text{Tablas}}$ , se puede observar que:

$F_0 > F_{\text{Tablas}}$

17.46 > 9.28

Por tanto, se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$).

También se puede observar gráficamente, de la siguiente manera:




Límite mínimo de aceptación $F_0 = 17.46$ $(F_{Tablas} = 9.28)$

Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

Por lo tanto, el nivel de temperatura influye significativamente en la tasa de germinación de la semilla de tomate.

## 4.5 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Como el Modelo Estadístico Lineal para el Diseño por Bloques Incompletos Balanceado es: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$; también es posible obtener los estimadores de los parámetros del Modelo.

En este tipo de Diseños por ser los tratamientos y los bloque no ortogonales, para obtener los estimadores insesgados de mínima varianza de cada parámetro, es necesario hacer un análisis intrabloques e interbloques para obtener por medio del Método de Mínimos Cuadrados los estimadores de los parámetros de cada análisis y luego realizar la combinación de ellos, obteniendo así los verdaderos estimadores insesgados de mínima varianza de los parámetros del modelo.

El análisis que se ha efectuado anteriormente se conoce como **Intrabloque,** porque se eliminan las diferencias entre los bloques, y todos los contrastes de los efectos de tratamientos pueden expresarse en forma de comparaciones entre las observaciones en el mismo bloque.

Luego de utilizar las ecuaciones normales de mínimos cuadrados se obtienen las estimaciones de los parámetros del análisis intrabloques siguientes(Ver Douglas C. Montgomery, año 1991, Página 161).

$\hat{\mu} = \bar{y}_{..}$ (Media General) $\hat{\tau}_i = \dfrac{kQ_i}{\lambda a}$ (Efecto de los Tratamientos)

Los Diseños de Bloques Incompletos son no ortogonales debido a que no todos los tratamientos aparecen en todos los bloques, y las comparaciones entre bloques contienen cierta información sobre las comparaciones de tratamiento. Al método para obtener esta información adicional Yates le denomino **Análisis Interbloques**.

Los estimadores del análisis interbloques para $\mu$ y $\iota_i$ se determinan utilizando el método de Mínimos Cuadrados; los cuales son los siguientes (Ver Douglas C. Montgomery, año 1991, Página 162).

$\tilde{\mu} = \bar{y}_{..}$ Media General $\qquad \tilde{\tau}_i = \dfrac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..}}{r-\lambda}$, $i$ = 1,2,…,$a$ Efecto de los Tratamientos

Se puede demostrar que los estimadores del Análisis Interbloques { $\hat{\tau}_i$ } y los del Análisis Intrabloques{ $\tilde{\tau}_i$ } no están correlacionados.

Para obtener un sólo estimador insesgado de mínima varianza para cada $\iota_i$, se debe combinar los estimadores del Análisis Interbloque y los estimadores del Análisis Intrabloque. Se puede demostrar que $\hat{\tau}_i$ y $\tilde{\tau}_i$ son insesgados; y las varianzas de cada uno de ellos viene dada por:

V( $\hat{\tau}_i$ ) = $\dfrac{k(a-1)}{\lambda a^2}\sigma^2$ (Análisis Intrabloques)

V( $\tilde{\tau}_i$ ) = $\dfrac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2)$ (Análisis Interbloques)

Y después de efectuar la combinación lineal de ambos se obtiene el mejor estimador combinado para $\tau_i^*$ es: ( Ver Douglas C. Montgomery, año 1991, Página 163).

$$\tau_i^* = \begin{cases} \dfrac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + (\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..})\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + \lambda a(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)}, & \hat{\sigma}_\beta^2 > 0 \\[2ex] \dfrac{y_{i.} - (\frac{1}{a})y_{..}}{r}, & \hat{\sigma}_\beta^2 = 0 \end{cases}$$

donde $\sigma^2$ puede ser estimada por la media de cuadrados del error del Análisis de Varianza intrabloques o error intrabloques. Es decir: $\hat{\sigma}^2 = MS_E$

La estimación de $\sigma_\beta^2$ se encuentra usando la media de cuadrados de bloque corregida por los tratamientos. En general, para un Diseño Balanceado por Bloques Incompletos, la media de cuadrados de bloque corregida es:

$$MS_{Bloques(ajustada)} = \frac{\left( \frac{k\sum_{i=1}^{a} Q_i^2}{\lambda a} + \sum_{j=1}^{b} \frac{y_{.j}^2}{k} - \sum_{i=1}^{a} \frac{y_{i.}^2}{r} \right)}{b-1}$$

y su valor esperado (deducido por Graybill) es el siguiente:

$E(MS_{Bloques(ajustada)}) = \sigma^2 + \frac{a(r-1)}{b-1}\sigma_\beta^2$.

Para obtener la estimación de $\sigma_\beta^2$ se deben tomar en cuenta las siguientes consideraciones:

1) Si $MS_{Bloques(ajustada)} > MS_E$ entonces el estimador de $\sigma_\beta^2$ es:

$$\hat{\sigma}_\beta^2 = \frac{\left(MS_{Bloques(ajustada)} - MS_E\right)(b-1)}{a(r-1)}$$

2) Si $MS_{Bloques(ajustada)} \leq MS_E$ el valor de $\hat{\sigma}_\beta^2 = 0$

**Ejemplo 16**

Retomando los datos del ejemplo 15; se llevara a cabo la estimación de los parámetros del Modelo que representan estos datos:

**Datos**

$k = 2$ , $a = 4$ , $b = 6$ , $r = 3$ , $\lambda = 1$ , $N = 12$ , $y_{..} = 191.88$ , $\bar{y}_{..} = 15.99$

$Q_1 = 20.93$ , $Q_2 = 15.66$ , $Q_3 = -11.83$ , $Q_4 = -24.76$ , $MS_E = 13.71$

$y_{1.} = 82.72$ , $y_{2.} = 71.16$ , $y_{3.} = 32.59$ , $y_{4.} = 5.41$

$y_{.1} = 25.99$ , $y_{.2} = 26.62$ , $y_{.3} = 50.42$ , $y_{.4} = 7.73$ , $y_{.5} = 47.17$ , $y_{.6} = 33.95$

**Estimación de los Parámetros de Análisis Intrabloques**

$\hat{\mu} = \bar{y}_{..}$

$\hat{\mu} = 15.99$ (Estimación de la media general)

$\hat{\tau}_i = \frac{kQ_i}{\lambda a}$ (Estimación de los Efectos de los Tratamientos)

$\hat{\tau}_1 = \frac{kQ_1}{\lambda a} = \frac{2(20.93)}{(1)(4)} = \frac{41.86}{4} = 10.46$ (Estimación del efecto de la Temperatura a 25ºC)

$\hat{\tau}_2 = \frac{kQ_2}{\lambda a} = \frac{2(15.66)}{(1)(4)} = \frac{31.32}{4} = 7.83$ (Estimación del efecto de la Temperatura a 30ºC)

$\hat{\tau}_3 = \frac{kQ_3}{\lambda a} = \frac{2(-11.83)}{(1)(4)} = \frac{-23.66}{4} = -5.91$ (Estimación del efecto de la Temperatura a 35ºC)

$\hat{\tau}_4 = \frac{kQ_4}{\lambda a} = \frac{2(-24.76)}{(1)(4)} = \frac{-49.52}{4} =$ -12.38 (Estimación del efecto de la Temperatura a 40ºC)

**Estimación de los Parámetros de Análisis Interbloques**

$\tilde{\mu} = \bar{y}_{..}$

$\tilde{\mu} = 15.99$ (Estimación de la media general)

$\tilde{\tau}_i = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..}}{r-\lambda}$, i = 1,2,…,a (Efecto de los Tratamientos)

Estimación del efecto de la Temperatura a 25ºC

$$\tilde{\tau}_1 = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..}}{r-\lambda} = \frac{[25.99 + 50.42 + 47.17] - (2)(3)(15.99)}{3-1} = \frac{123.58 - 95.94}{2} = 13.82$$

Estimación del efecto de la Temperatura a 30ºC

$$\tilde{\tau}_2 = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..}}{r-\lambda} = \frac{[26.62 + 50.42 + 33.95] - (2)(3)(15.99)}{3-1} = \frac{110.99 - 95.94}{2} = 7.52$$

Estimación del efecto de la Temperatura a 35ºC

$$\tilde{\tau}_3 = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..}}{r-\lambda} = \frac{[7.73 + 47.17 + 33.95] - (2)(3)(15.99)}{3-1} = \frac{88.85 - 95.94}{2} = -3.55$$

Estimación del efecto de la Temperatura a 40ºC

$$\tilde{\tau}_4 = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..}}{r-\lambda} = \frac{[25.99 + 26.62 + 7.73] - (2)(3)(15.9988)}{3-1} = \frac{60.34 - 95.94}{2} = -17.80$$

**Estimación de los Parámetros Reales del Modelo**

$$\tau_i^* = \begin{cases} \dfrac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + (\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..})\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + \lambda a(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)}, & \hat{\sigma}_\beta^2 > 0 \\ \dfrac{y_{i.} - (\frac{1}{a})y_{..}}{r}, & \hat{\sigma}_\beta^2 = 0 \end{cases}$$

**Calculando la Media de Cuadrados de Bloques Ajustada.**

$$MS_{Bloques(ajustada)} = \frac{\left( \frac{k\sum_{i=1}^{4} Q_i^2}{\lambda a} + \sum_{j=1}^{6} \frac{y_{.j}^2}{k} - \sum_{i=1}^{4} \frac{y_{i.}^2}{r} \right)}{b-1}$$

El primer término del numerador de esta expresión es la suma de cuadrados de Tratamientos corregida; que fue calculada en ejercicio 15, es decir:

$$\frac{k\sum_{i=1}^{4} Q_i^2}{\lambda a} = 718.15$$

Y el segundo término del numerador es una parte de la suma de cuadrados de bloques no corregida; es decir:

$$\sum_{j=1}^{6} \frac{y_{.j}^2}{k} = 3681.8226$$

Calculando la último término del numerador se tiene:

$$\sum_{i=1}^{4} \frac{y_{i.}^2}{r} = \frac{(82.72)^2 + (71.16)^2 + (32.59)^2 + (5.4)^2}{3} = \frac{12997.72}{3} = 4332.5733 \text{ entonces}$$

$$MS_{Bloques(ajustada)} = \frac{[718.15 + 3681.8226 - 4332.5733]}{5} = \frac{67.399}{5} = 13.48$$

Como $MS_{Bloques(ajustada)} \leq MS_E$ entonces el valor de $\hat{\sigma}_\beta^2 = 0$

Por lo tanto, $\tau_i^* = \frac{y_{i.} - (\frac{1}{a})y_{..}}{r}$

$$\tau_1^* = \frac{y_{1.} - (\frac{1}{a})y_{..}}{r} = \frac{82.72 - (\frac{1}{4})(191.88)}{3} = \frac{82.72 - 47.97}{3} = \frac{34.75}{3} = 11.58$$

Significa que la tasa de germinación de la semilla de tomate tendrá un aumento de 11.58 al aplicar la temperatura de 25ºC.

$$\tau_2^* = \frac{y_{2.} - (\frac{1}{a})y_{..}}{r} = \frac{71.16 - (\frac{1}{4})(191.88)}{3} = \frac{71.16 - 47.97}{3} = \frac{23.19}{3} = 7.73$$

Significa que la tasa de germinación de la semilla de tomate tendrá un aumento de 7.73 al aplicar la temperatura de 30ºC.

$$\tau_3^* = \frac{y_{3.} - (\frac{1}{a})y_{..}}{r} = \frac{32.59 - (\frac{1}{4})(191.88)}{3} = \frac{32.59 - 47.97}{3} = \frac{-15.38}{3} = -5.13$$

Significa que la tasa de germinación de la semilla de tomate tendrá una disminución de 5.13 al aplicar la temperatura de 35ºC.

$$\tau_4^* = \frac{y_{4.} - (\frac{1}{a})y_{..}}{r} = \frac{5.41 - (\frac{1}{4})(191.88)}{3} = \frac{5.41 - 47.97}{3} = \frac{-42.56}{3} = -14.19$$

Significa que la tasa de germinación de la semilla de tomate tendrá una disminución de 14.19 al aplicar la temperatura de 40ºC.

Comparando los valores de los estimadores de los tratamientos intrabloques ( $\hat{\tau}_i$ ) con los estimadores de los tratamientos combinados ( $\tau_i^*$ ), se observa que son aproximadamente iguales; porque la varianza de las estimaciones del análisis interbloques es relativamente grande.

## 4.6 COMPARACIÓN ENTRE TRATAMIENTOS

Si en un Diseño Aleatorizado por Bloques Incompletos los tratamientos bajo estudio son fijos y el Análisis de Varianza indica que existe diferencia significativa entre las medias de tratamiento (se rechaza Ho), el experimentador estará interesado en realizar comparaciones adicionales en grupos de medias de tratamientos, para determinar cuales son las medias que difieren; cualquier método estudiado en el Diseño Unifactorial puede ser utilizado para este fin, con algunas variantes.

Para llevar a cabo las comparaciones entre grupos de tratamiento para un Diseño Aleatorizado por Bloques Incompletos, se debe sustituir el número de réplicas o repeticiones ($n$) por el número de tratamientos multiplicado por el número de veces que cada par de tratamientos ocurre en el mismo bloque ($\lambda a$), en las fórmulas utilizadas en cada uno de los métodos estudiados en la Unidad II y además se debe utilizar los grados de libertad del error que están definidos por ($N$-$a$-$b$+1) para un Diseño Aleatorizado por Bloques Incompletos. En los métodos en los cuales intervienen o se realizan las comparaciones, con los promedios de tratamientos; dichos promedios deben ser sustituidos por las medias de tratamientos corregidos ($\hat{\tau}_i$), los cuales se obtienen de la siguiente forma: $\hat{\tau}_i = \frac{kQ_i}{\lambda a}$, i=1,2,3,..., $a$; es decir, que las conclusiones serán siempre con respecto a los promedios de tratamientos sólo que en los cálculos se tomarán los valores de los tratamientos corregidos; ya que los tratamientos y los bloques no son ortogonales.

A continuación se presentaran los métodos descritos en la Unidad II, expresando solamente las variantes que se deben incorporar para llevar a cabo la comparación de medias de tratamientos para un Diseño Aleatorizado por Bloques Incompletos. Las hipótesis a probar, el procedimiento y conclusiones se harán de igual manera que en la Unidad II.

**1) Comparación de Medias de Tratamientos Individuales.**

**a) Contrastes Ortogonales.**

El valor numérico de los contrastes se encuentra de la siguiente manera: $c = \sum_{i=1}^{a} c_i Q_i$ . Los contrastes deben basarse en los totales de tratamiento corregidos (los $Q_i$ en lugar de los $y_{i.}$).

La suma de cuadrados de los contrastes viene dada por: $SS_c = \dfrac{k(\sum_{i=1}^{a} c_i Q_i)^2}{\lambda a \sum_{i=1}^{a} c_i^2}$, donde $Q_i$ son los totales de tratamientos corregidos.

La estadística de prueba ($F_{Tablas}$) que se debe utilizar para rechazar la hipótesis tiene una distribución F con 1 y N- $a$-$b$+1 grados de libertad, es decir: $F_{\alpha,1,\ N-a-b+1}$.

**b) Método de Scheffé para comparar todos los contrastes.**

El error estándar del contraste es: $S_{c_k} = \sqrt{\dfrac{kMS_E}{\lambda a} \sum_{i=1}^{a} c_{ik}}$

El valor crítico con el que $c_k$ será comparado está dado por: $S_{\alpha,k} = S_{c_k} \sqrt{(a-1)F_{\alpha,a-1,N-a-b+1}}$

**2) Comparación de Parejas de Medias de Tratamientos.**

**a) Método de la Mínima Diferencia Significativa (LSD)**

El LSD estará dado de la siguiente manera:

LSD = $t_{\frac{\alpha}{2},N-a-b+1} \sqrt{\dfrac{2kMS_E}{\lambda a}}$

**b) Prueba de Intervalos Múltiples de Duncan**

El error estándar de cada promedio corregido se calcula de la siguiente forma:

$$S_{\bar{y}_{i.}} = \sqrt{\frac{kMS_E}{\lambda a}}$$

Para encontrar los intervalos significativos $r_\alpha(p,f)$, para $p$ = 2,3, …,$a$; $\alpha$ sigue siendo el nivel de significancia y $f$ el número de grados de libertad del error que son $N$- $a$-$b$+1.

De igual manera para encontrar los mínimos intervalos significativos $R_p = r_\alpha(p,f)\ S_{\bar{y}_{i.}}$, con $p$ = 2, 3, …, $a$ se tomará f como el número de grados de libertad del error $N$- $a$-$b$+1.

**c) Prueba de Tukey**

El valor crítico de todas las comparaciones vendrá dado por : $T_\alpha = q_\alpha\ (a,f)\ S_{\bar{y}_{i.}}$,

donde $S_{\bar{y}_{i.}}$ es el error estándar de cada promedio y está dado por $S_{\bar{y}_{i.}} = \sqrt{\dfrac{kMS_E}{\lambda a}}$, y $f$ los grados de libertad del error $N$- $a$-$b$+1, $\alpha$ el nivel de significancía y "*a*" el número de tratamientos.

**3) Comparación de Tratamientos con un Control.**

La hipótesis nula se rechaza si $|\hat{\tau}_i - \hat{\tau}_a| > d_{\alpha(a-1,f)} \sqrt{\dfrac{2kMS_E}{\lambda a}}$, donde $d_{\alpha(a-1,f)}$ se encuentra en la tabla de Dunnett con $\alpha$ que es el nivel de significancia y $f$ los grados de libertad del error que están dados por $N$- $a$-$b$+1.

**Ejemplo 17**

Como en el ejemplo 15, se rechazo $H_0$ entonces se aplicará la Prueba de Intervalos Multiples de Duncan para determinar que medias son diferentes.

**Datos**

$MS_E$ = 13.71 , $a$ =4 , $b$ = 6 , $k$ = 2 , $r$ = 3 , $N$ = 12 , $\alpha$ = 0.05 ,

Ordenando las medias de tratamientos corregidas calculadas anteriormente (interbloques), en orden de menor a mayor (ascendente).

$\hat{\tau}_4 = -12.38$ $\hat{\tau}_3 = -5.91$ $\hat{\tau}_2 = 7.83$ $\hat{\tau}_1 = 10.46$

**Solución**

**Calculando los intervalos significativos**

P = 2,3,4 , f = 12-4-6+1 = 3 grados de libertad del error

$r_{0.05}(2,3) = 4.50$ , $r_{0.05}(3,3) = 4.50$ , $r_{0.05}(4,3) = 4.50$

**Error estándar de un tratamiento corregido.**

$$S_{\bar{y}_{i.}} = \sqrt{\frac{kMS_E}{\lambda a}} = \sqrt{\frac{2(13.71)}{(1)(4)}} = \sqrt{\frac{27.42}{(4)}} \quad \sqrt{6.85} = 2.62$$

**Calculando los Mínimos Intervalos Significativos**

$R_p = r_\alpha(p,f)\ S_{\bar{y}_{i.}}$ , con p = 2, 3, 4

$R_2 = r_{0.05}(2,3)\ S_{\bar{y}_{i.}} = (4.5)\ (2.62) = 11.79$

$R_3 = r_{0.05}(3,3)\ S_{\bar{y}_{i.}} = (4.5)\ (2.62) = 11.79$

$R_4 = r_{0.05}(4,3)\ S_{\bar{y}_{i.}} = (4.5)\ (2.62) = 11.79$

**Realizando las comparaciones Múltiples**

1 vrs 4 : $\hat{\tau}_1 - \hat{\tau}_4 = 10.46 - (-12.38) = 22.84 > 11.79$ *

1 vrs 3 : $\hat{\tau}_1 - \hat{\tau}_3 = 10.46 - (-5.91) = 16.37 > 11.79$ *

1 vrs 2 : $\hat{\tau}_1 - \hat{\tau}_2 = 10.46 - 7.83 = 2.63 < 11.79$

2 vrs 4 : $\hat{\tau}_2 - \hat{\tau}_4 = 7.83 - (-12.38) = 20.21 > 11.79$ *

2 vrs 3 : $\hat{\tau}_2 - \hat{\tau}_3 = 7.83 - (-5.91) = 13.74 > 11.79$ *

3 vrs 4 : $\hat{\tau}_3 - \hat{\tau}_4 = -5.91 -(-12.38) = 6.47 < 11.79$

Las diferencias se comparan solamente con un solo valor; ya que los Mínimos Intervalos Significativos resultaron tener el mismo valor.

**Conclusión**

Las medias de las temperaturas uno y cuatro, uno y tres, dos y cuatro, dos y tres son significativamente diferentes (*); sin embargo entre las medias de las temperaturas uno y dos, como las medias de las temperaturas tres y cuatro no existe diferencia significativa.

## 5. CUADRADO DE YOUDEN

Existen experimentos en los cuales se necesita eliminar dos fuentes de variabilidad problemática y puede darse el caso que el número de unidades experimentales exceda las restricciones del material experimental o el número de tratamientos puede exceder el tamaño de los bloques disponibles. En este caso se debe utilizar un Diseño de **Cuadrados Latinos "Incompletos".**

Los Diseños de Cuadrados Latinos Incompletos son aquellos en los que el número de columnas no es igual al número de renglones y tratamientos; es decir, es un Diseño de Cuadrado Latino que le falta un renglón, una columna o una diagonal.

Youden (1937-1940) desarrolló arreglos de Cuadrados Latinos Incompletos, es por ello que se conocen como **Cuadrados de Youden**.

En general un Cuadrado de Youden es un Diseño Balanceado por Bloques Incompletos, simétrico, en el que los renglones corresponden a los bloques y cada tratamiento ocurre exactamente una vez en cada columna o "posición" del bloque. Así, es posible construir Cuadrados de Youden de todos los Diseños de Bloques Incompletos Balanceados Simétricos.

La supresión arbitraria de más de una columna, renglón o diagonal de un Cuadrado Latino puede ocasionar que se destruya su balance. Por lo tanto, un Cuadrado de Youden es siempre un Cuadrado Latino del cual al menos una columna, renglón o diagonal falta; pero no siempre es necesariamente cierto que un Cuadrado Latino con más de una columna, renglón o diagonal removida es un Cuadrado de Youden.

Los parámetros del Diseño son: $a = b$, $r = k$ y $\lambda = \dfrac{r(k-1)}{a-1}$

Los grados de libertad serán:
Para la suma de Tratamientos son: $a - 1$
Para la suma de Bloques son : $b - 1$
Para la suma de posiciones son : $p - 1$
Para el total será : $N - 1$
Para el error será : $N-1-(a-1)-(b-1)-(p-1) = N-a-b-p+2$

**Ejemplo 18**

Se realizó un experimento para asegurar las resistencias relativas al desgaste de cuatro tipos de pieles: Caballo, Vaca, Conejo y Cabra.

Se usó una máquina en la cual se probaron las muestras en una de cualquiera de tres posiciones. Puesto que se conoce que diferentes ejecuciones del experimento dan resultados variables, se decidió hacer cuatro ejecuciones del mismo. Se obtuvieron los siguientes resultados.

| Posición | Ejecución | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | A=118 | B=127 | C=174 | D=130 |
| 2 | B=136 | C=141 | D=173 | A=170 |
| 3 | C=168 | D=129 | A=126 | B=125 |

**Interpretación**

Este ejemplo es considerado como un Cuadrado de YOUDEN, ya que en él se analizan las resistencias relativas al desgaste de 4 tipos de pieles (caballo, vaca, conejo y cabra), que son considerados como los tratamientos.

Se conoce que las diferentes ejecuciones del experimento dan resultados variables, por lo tanto las 4 ejecuciones se toman como columnas y las 3 posiciones en que se prueban las muestras son tomadas como filas. Se pueden observar que en el experimento cada tratamiento ocurre exactamente una vez en cada columna y cada fila; pero no posee igual número de filas y columnas, es por eso que es un Cuadrado Latino que le falta una fila; por lo tanto es un Diseño de Bloques Incompletos, en donde se cumple que : $a = b$, $r = k$ y $\lambda = \frac{r(k-1)}{a-1}$; ya que : $a$ = 4 Tratamientos (Letras Latinas) , : $b$ = 4 Número de bloques, : $r$ = 3 Número de veces que dos tratamientos aparecen juntos, k = 3 Número de tratamientos que se aplican en cada bloque.

## 5.1 MODELO ESTADÍSTICO

El modelo Lineal que representa los datos de un Cuadrado de YOUDEN está dado por:

$$y_{ijh} = \mu + \alpha_i + \iota_j + \beta_h + \varepsilon_{ijh}$$

en donde**:**

$\mu$ : Es la media general.

$\alpha_i$ : Es el i-ésimo efecto de bloque.

$\iota_j$ : Es el j-ésimo efecto de tratamiento (Letra Latina).

$\beta_h$ : Es el h-ésimo efecto de posición.

$\varepsilon_{ijh}$ : Es el término usual de error aleatorio NID($0,\sigma^2$).

Para este Diseño no se va ha describir todas las partes del análisis en detalle como en los anteriores Diseños Experimentales; ya que el análisis es similar a lo realizado en el Diseño de Cuadrados Latinos y el Diseño de Bloques Incompletos. Por ser una combinación de ambos, sólo se estudiaran la forma de calcular el total corregido del i-ésimo tratamiento y los totales de bloque corregido ($Q_j$ y $Q_i^{'}$) para observar su diferencia; como se detalla a continuación.

La suma de Cuadrados de tratamientos corregida viene dada por:

$SS_{\text{Tratamientos(ajustada)}} = \frac{k\sum_{i=1}^{a} Q_j^2}{\lambda a}$, en donde $Q_j$ se obtiene mediante la siguiente fórmula:

$Q_j = y_{.j.} - \frac{1}{k}\sum_{i=1}^{a} n_{ij}\, y_{i..}$ , j=1,2,...,b

Los $Q_j$ representan los totales corregidos de los tratamientos de la Letras Latinas.

Por ser simétrico el Cuadrado de YOUDEN es posible obtener la suma de Cuadrados de Bloques corregida la cual es:

$SS_{\text{Bloques(ajustada)}} = \dfrac{r\sum_{j=1}^{b}(Q_i^{'})^2}{\lambda b}$, en donde $Q_i^{'}$ se obtiene mediante la siguiente fórmula:

$Q_i^{'} = y_{i..} - \dfrac{1}{r}\sum_{j=1}^{b} n_{ij}\, y_{.j.}$ , i=1,2,...,a

Los $Q_i^{'}$ representan los totales corregidos de los bloques.

Para los $Q_j$ y $Q_i^{'}$ se debe tomar $n_{ij} = \begin{cases} 1 \text{ , si el tratamiento i ocurre en el bloque j} \\ 0 \text{ , en el caso contrario} \end{cases}$

**Ejemplo 19**

Para ilustrar lo planteado anteriormente se retomará los datos del ejemplo 18:

| Posición | Ejecución | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | $y_{..h}$ |
| 1 | A=18 | B=127 | C=174 | D=130 | 549 |
| 2 | B=136 | C=141 | D=173 | A=170 | 620 |
| 3 | C=168 | D=129 | A=126 | B=125 | 548 |
| $y_{i..}$ | 422 | 397 | 473 | 425 | $y_{...}$ = 1717 |

| $y_{.j.}$ |
|---|
| 414 (A) |
| 388 (B) |
| 483 (C) |
| 432 (D) |

**Datos**

$a$ = 4 , $b$ = 4 , $r$ = 3 , $k$ = 3 , $\lambda = \dfrac{3(3-1)}{4-1} = \dfrac{6}{3} = 2$

**Totales de tratamientos corregidos**

$Q_j = y_{.j.} - \dfrac{1}{k}\sum_{i=1}^{a} n_{ij}\, y_{i..}$ , j=1,2,3,4

$Q_1 = y_{.1.} - \dfrac{1}{3}\sum_{i=1}^{4} n_{i1} y_{i..} = 414 - \dfrac{1}{3}(422+473+425) = 414 - \dfrac{1}{3}(1320) = 414 - 440 = -26$

$Q_2 = y_{.2.} - \dfrac{1}{3}\sum_{i=1}^{4} n_{i2} y_{i..} = 388 - \dfrac{1}{3}(422+397+425) = 388 - \dfrac{1}{3}(1244) = 388 - 414.66 = -26.67$

$Q_3 = y_{.3.} - \frac{1}{3}\sum_{i=1}^{4} n_{i3} y_{i..} = 483 - \frac{1}{3}(422+397+473) = 483 - \frac{1}{3}(1292) = 483 - 430.67 = 52.33$

$Q_4 = y_{.4.} - \frac{1}{3}\sum_{i=1}^{4} n_{i4} y_{i..} = 432 - \frac{1}{3}(397+473+425) = 432 - \frac{1}{3}(1295) = 432 - 431.67 = 0.33$

$$SS_{Tratamientos(ajustada)} = \frac{k\sum_{i=1}^{4} Q_j^2}{\lambda a} = \frac{3\left[(-26)^2 + (-26.67)^2 + (52.33)^2 + (0.33)^2\right]}{2(4)} = \frac{3(4125.8267)}{8}$$

$= 1547.18$

**Totales de Bloques Corregidos**

$Q_i' = y_{i..} - \frac{1}{r}\sum_{j=1}^{b} n_{ij} y_{.j.}$ , i=1,2,3,4

$Q_1' = y_{1..} - \frac{1}{3}\sum_{j=1}^{4} n_{1j} y_{.j.} = 422 - \frac{1}{3}(414+388+483) = 422 - \frac{1}{3}(1285) = - 6.33$

$Q_2' = y_{2..} - \frac{1}{3}\sum_{j=1}^{4} n_{2j} y_{.j.} = 397 - \frac{1}{3}(388+483+432) = 397 - \frac{1}{3}(1303) = - 37.33$

$Q_3' = y_{3..} - \frac{1}{3}\sum_{j=1}^{4} n_{3j} y_{.j.} = 473 - \frac{1}{3}(414+483+432) = 473 - \frac{1}{3}(1329) = 30$

$Q_4' = y_{4..} - \frac{1}{3}\sum_{j=1}^{4} n_{4j} y_{.j.} = 423 - \frac{1}{3}(414+388+432) = 423 - \frac{1}{3}(1234) = 11.67$

## 6. PROBLEMAS RESUELTOS

**PROBLEMA 1. (Diseño Aleatorizado por Bloques Completos)**

Un ingeniero de control de calidad para manufacturas de componentes electrónicos, intuitivamente siente que hay mucha variabilidad entre los 36 hornos usados por la compañía manufacturera para la cual trabaja, en las pruebas de duración de los diversos componentes. Para determinar si tiene razón, escoge un sólo tipo de componente y obtiene los siguientes datos para las 2 temperaturas (T) normalmente usadas en las pruebas de duración de estas unidades; $T_1$ = 550ºF y $T_2$ = 600ºF. El componente se pone a funcionar en un horno hasta que falla. En este experimento se utilizaron 3 hornos (H) seleccionados aleatoriamente y se midió los minutos de duración del componente electrónico. Con un 5% de significancia, pruebe si el ingeniero tiene razón en su creencia de la diferencia entre los 36 hornos. Los datos obtenidos son los siguiente:

| Hornos | Temperaturas | |
|---|---|---|
| | 550ºF | 600ºF |
| $H_1$ | 246 | 180 |
| $H_2$ | 191 | 144 |
| $H_3$ | 187 | 134 |

**Solución**

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar; ya que hay dieciséis hornos y solo se toman tres aleatoriamente es un Modelo de Efectos aleatorios.

$H_o$ : $\sigma_\tau^2 = 0$ (No existe variabilidad entre los Hornos)

$H_1$ : $\sigma_\tau^2 > 0$ (Existe variabilidad entre los Hornos)

**Variable Respuesta:** Duración del componente electrónico.

**El significado verbal es:**

$H_o$ : La variabilidad entre los Hornos no influyen significativamente en la duración del componente electrónico probado.

$H_1$: La variabilidad entre los Hornos influyen significativamente en la duración del componente electrónico probado.

**Datos.**

$a = 3$ , $b = 2$ , $N = 3\text{x}2 = 6$ , $i = 1,2,3$ , $j = 1,2$

## Cálculos Matemáticos

**Totales Tratamientos**

$y_{i.} = \sum_{j=1}^{2} y_{ij}$

$y_{1.} = \sum_{j=1}^{2} y_{1j}$ = 246 +180 = 426

$y_{2.} = \sum_{j=1}^{2} y_{2j}$ = 191 + 144 = 335

$y_{3.} = \sum_{j=1}^{4} y_{3j}$ = 187 + 134 = 321

**Totales de Bloques**

$y_{.j} = \sum_{i=1}^{3} y_{ij}$

$y_{.1} = \sum_{i=1}^{3} y_{i1}$ = 246 + 191 + 187 = 624

$y_{.2} = \sum_{i=1}^{3} y_{i2}$ = 180 + 144 + 134 = 458

$y_{..} = \sum_{i=1}^{3}\sum_{j=1}^{2} y_{ij}$ = 246 + 180 + 191 +….+ 134 = 1082

**Medias de Tratamientos**

$\bar{y}_{i.} = \frac{y_{i.}}{b}$

$\bar{y}_{1.} = \frac{y_{1.}}{2} = \frac{426}{2}$ = 213

$\bar{y}_{2.} = \frac{y_{2.}}{2} = \frac{335}{2}$ = 167.5

$\bar{y}_{3.} = \frac{y_{3.}}{2} = \frac{321}{2}$ = 160.5

**Medias de Bloques**

$\bar{y}_{.j} = \frac{y_{.j}}{a}$

$\bar{y}_{.1} = \frac{y_{.1}}{3} = \frac{624}{3}$ = 208

$\bar{y}_{.2} = \frac{y_{.2}}{3} = \frac{458}{3}$ = 152.66

$\bar{y}_{..} = \frac{y_{..}}{6} = \frac{1082}{6}$ = 180.33

**Sumas de Cuadrados**

$SS_T = \sum_{i=1}^{3}\sum_{j=1}^{2} y_{ij}^2 - \frac{y_{..}^2}{N}$ = [(246)$^2$ + (180)$^2$ + (191)$^2$ +(144)$^2$+ (187)$^2$ + (134)$^2$] - $\frac{(1082)^2}{6}$

= 203058 – 195120.66

$SS_T$ = 7937.34

$SS_{Tratamientos} = \frac{\sum_{i=1}^{3} y_{i.}^2}{2} - \frac{y_{..}^2}{N} = \frac{(426)^2 + (335)^2 + (321)^2}{2} - \frac{(1082)^2}{6}$

= 198371 – 195120.66

$SS_{Tratamientos}$ = 3250.34

$SS_{Bloques} = \frac{\sum_{j=1}^{2} y_{.j}^2}{3} - \frac{y_{..}^2}{N} = \frac{(624)^2 + (458)^2}{3} - \frac{(1082)^2}{6}$

= 199713.33 – 195120.66

$SS_{Bloques}$ = 4592.67

$SS_E = SS_T - SS_{Tratamientos} - SS_{Bloques}$

$SS_E$ = 7937.34 – 3250.34 – 4592.67

$SS_E$ = 94.33

**Medias de Cuadrados**

$MS_{Tratamientos} = \dfrac{SS_{Tratamientos}}{a-1} = \dfrac{3250.34}{3-1} = \dfrac{3250.34}{2} = 1625.17$

$MS_{Bloques} = \dfrac{SS_{Bloques}}{b-1} = \dfrac{4592.67}{2-1} = \dfrac{4592.67}{1} = 4592.67$

$MS_E = \dfrac{SS_E}{(a-1)(b-1)} = \dfrac{94.33}{(2)(1)} = \dfrac{94.33}{2} = 47.16$

**Estadística.**

$F_o = \dfrac{MS_{Tratamientos}}{MS_E} = \dfrac{1625.17}{47.16} = 34.46$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Hornos | 3250.34 | 2 | 1625.17 | 34.46 |
| Temperaturas | 4592.67 | 1 | 4592.67 | |
| Error | 94.33 | 2 | 47.16 | |
| Total | 7937.34 | 5 | | |

Utilizando un nivel de significancia del 5% ($\alpha$ = 0.05) para encontrar el $F_{Tablas}$ (Tablas Fisher) con 2 grados de libertad ($a$-1) en el numerador y 2 grados de libertad ($a$-1)($b$-1) en denominador.

$F_{\alpha,a-1,(a-1)(b-1)} = F_{0.05,2,2} = 19.00$

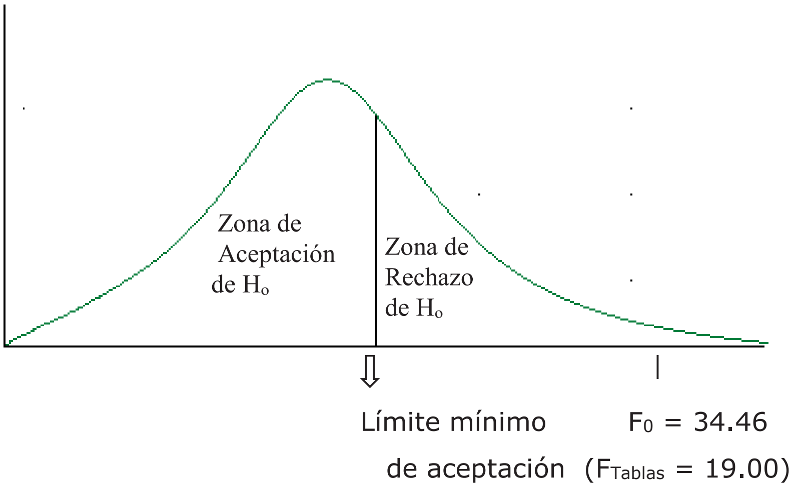
Comparando el $F_0$ calculado de el Análisis de Varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

$34.46 > 19.00$

Por tanto, se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$).

También puede observarse gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

**Conclusión**

Los tipos de hornos influyen significativamente en la duración del componente electrónico probado.

**PROBLEMA 2. ( Comparación de Medias de Tratamientos)**

Como en el Problema 1, fue rechazada la hipótesis nula. Supongamos que se desea saber cuales son las parejas de medias que son diferentes; para ello se utilizará el Método de la Mínima Diferencia Significativa.

**Datos.**

$\alpha$ = 0.05 , $MS_E$ = 47.16 , N = 6 , $a$ = 3 , $b$ = 2

$\bar{y}_{1.}$ = 213 , $\bar{y}_{2.}$ = 167.5 , $\bar{y}_{3.}$ = 160.5

- **Encontrando el valor del LSD, con la fórmula establecida.**

$$LSD = t_{\frac{\alpha}{2},(a-1)(b-1)}\sqrt{\frac{2MS_E}{b}} = t_{\frac{0.05}{2},(3-1)(2-1)}\sqrt{\frac{2(47.16)}{2}} = t_{0.025,2}\sqrt{47.16} = (4.303)(6.867)$$

LSD = 29.55

- **Calculando la diferencia de los promedios.**

1 vrs 2 : $|\bar{y}_{1.} - \bar{y}_{2.}|$ = |213 – 167.5| = |45.5| > LSD

1 vrs 3 : $|\bar{y}_{1.} - \bar{y}_{3.}|$ = |213 – 160.5| = |52.5| > LSD

2 vrs 3 : $|\bar{y}_{2.} - \bar{y}_{3.}|$ = |167.5 – 160.5| = |7.00| < LSD

Se dice que una pareja de medias difieren significativamente si el valor absoluto de las diferencias de los promedios de los tratamientos correspondientes es mayor que LSD = 29.55.

- **Conclusiones**
  a) Se observa que la pareja de medias que no difieren significativamente son la media dos y la media tres; ya que $|7.00| < 29.55$; por lo tanto, no existe diferencia significativa entre el Horno dos y tres.
  b) Las parejas de medias uno y dos, uno y tres el valor absoluto de las diferencias de los promedios a resultado ser mayor que el valor encontrado del LSD; por lo tanto, las dos parejas de medias difieren significtivamente; es decir que existe diferencia significativa entre el Horno uno con los Hornos dos y tres.

**PROBLEMA 3. (Diseño Aleatorizado por Bloques Completos)**

Retomando los datos del problema 1, se desea encontrar la eficiencia relativa de este Diseño Experimental y determinar el número de réplicas en caso de hacer su análisis como un Diseño Unifactorial y mantener la misma sensibilidad en ambos Diseños.

**Datos.**

$MS_E = 47.16$ , $a = 3$ , $b = 2$ , $MS_{Bloques} = 4592.67$ , $N = 6$

$df_r = N - a = 6 - 3 = 3$

$df_b = (a-1)(b-1) = (3-1)(2-1) = (2)(1) = 2$

**Solución**

**Estimación de las varianzas**

$$\hat{\sigma}_b^2 \approx MS_E = 47.16$$

$$\hat{\sigma}_r^2 = \frac{(b-1)MS_{Bloques} + b(a-1)MS_E}{ab-1} = \frac{(2-1)(4592.67) + 2(3-1)(47.16)}{3x2-1} = \frac{(4592.67) + 4(47.16)}{6-1}$$

$$= \frac{4592.67 + 188.64}{5}$$

$$\hat{\sigma}_r^2 = \frac{4782.31}{5} = 956.46$$

**Calculando la Eficiencia Relativa del Diseño Aleatorizado por Bloques.**

$$R = \frac{(df_b+1)\,(df_r+3)}{(df_b+3)\,(df_r+1)}\frac{\hat{\sigma}_r^2}{\hat{\sigma}_b^2} = \frac{(2+1)\,(3+3)}{(2+3)\,(3+1)}\frac{(956.46)}{(47.16)} = \frac{(3)(6)}{(5)(4)}\frac{(956.46)}{(47.16)} = \frac{18}{20}\frac{(956.46)}{(47.16)}$$

$$= \frac{17216.28}{943.20} = 18.25 \approx 18$$

- **Conclusión**

Significa que se debe usar dieciocho réplicas más si se utiliza un Diseño Unifactorial para lograr la misma sensibilidad que la obtenida al analizar el experimento por medio de un Diseño Aleatorizado por Bloques.

**PROBLEMA 4. (Cuadro de Youden)**

Un ingeniero industrial está estudiando el efecto de cinco niveles de iluminación, para analizar la ocurrencia de defectos en una operación de ensamble. Porque el tiempo puede ser un factor en el experimento, él ha decidido correr el experimento en cinco bloques, donde cada bloque es un día de la semana. Sin embargo, el departamento en el cual el experimento es conducido tiene cuatro estaciones de trabajo, estas estaciones representan una fuente potencial de variabilidad. El ingeniero decidió correr un cuadrado de YOUDEN con cinco filas (Días o bloques) cuatro columnas (Estaciones de trabajo) y cinco tratamientos (niveles de iluminación). Los datos codificados son mostrados en la siguiente tabla:

| Día Bloques | Estación de Trabajo | | | | $y_{i..}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | A = 3 | B = 1 | C = -2 | D = 0 | 2 |
| 2 | B = 0 | C = 0 | D = -1 | E = 7 | 6 |
| 3 | C = -1 | D= 0 | E = 5 | A = 3 | 7 |
| 4 | D = -1 | E = 6 | A = 4 | B = 0 | 9 |
| 5 | E = 5 | A = 2 | B = 1 | C = -1 | 7 |
| $y_{..k}$ | 6 | 9 | 7 | 9 | $y_{...}$= 31 |

Analice estos resultados con un nivel de significancia del 1%.

**Solución**

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

$H_o$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (No existe diferencia entre los niveles de iluminación).

$H_1$ : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$ (Existe diferencia entre los niveles de iluminación)

**Variable Respuesta:** Ocurrencia de defectos.

**El significado verbal es:**

$H_o$: Los niveles de iluminación no influyen significativamente en la ocurrencia de los efectos en una operación de ensamble.

$H_1$: Los niveles de iluminación influyen significativamente en la ocurrencia de los efectos en una operación de ensamble.

**Datos**

$a$ = 5 , $b$ = 5 , $r$ = 4 , k = 4 , $p$ = 4 y $\lambda = \frac{r(k-1)}{a-1} = \frac{4(4-1)}{5-1} = \frac{12}{4}$ = 3

| $y_{.j.}$ |
|---|
| 414 (A) |
| 388 (B) |
| 483 (C) |
| 432 (D) |

**Cálculos Matemáticos**

**Cálculo de los totales de tratamientos (Letras Latinas)**

| Tratamientos | $y_{.j.}$ |
|---|---|
| A | 3 + 2 + 4 + 3 = 12 |
| B | 0 + 1 + 1 + 0 = 2 |
| C | -1 + 0 – 2 -1 = -4 |
| D | -1 + 0 -1 + 0 = -2 |
| E | 5 + 6 + 5 + 7 = 23 |

**Cálculo de la suma de cuadrados corregidos de los tratamientos.**

$Q_j = y_{.j.} - \frac{1}{k}\sum_{i=1}^{a} n_{ij}\, y_{i..}$ , $j$=1,2,3,5

$Q_1 = y_{.1.} - \frac{1}{4}\sum_{i=1}^{5} n_{i1} y_{i..} = 12 - \frac{1}{4}(2 + 7 + 9 + 7) = 12 - \frac{1}{4}(25) = 12 - 6.25 = 5.75$

$Q_2 = y_{.2.} - \frac{1}{4}\sum_{i=1}^{5} n_{i2} y_{i..} = 2 - \frac{1}{4}(2 + 6 + 9 + 7) = 2 - \frac{1}{4}(24) = 2 - 6 = -4$

$Q_3 = y_{.3.} - \frac{1}{4}\sum_{i=1}^{5} n_{i3} y_{i..} = -4 - \frac{1}{4}(2 + 6 + 7 + 7) = -4 - \frac{1}{4}(22) = -4 - 5.5 = -9.5$

$Q_4 = y_{.4.} - \frac{1}{4}\sum_{i=1}^{5} n_{i4} y_{i..} = -2 - \frac{1}{4}(2 + 6 + 7 + 9) = -2 - \frac{1}{4}(24) = -2 - 6 = -8$

$Q_5 = y_{.5.} - \frac{1}{4}\sum_{i=1}^{5} n_{i4} y_{i..} = 23 - \frac{1}{4}(6 + 7 + 9 + 7) = 23 - \frac{1}{4}(29) = 23 - 7.25 = 15.75$

**Sumas de Cuadrados**

$SS_T = \sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{k=1}^{4} y_{ijk}^2 - \frac{y_{...}^2}{N} = (3)^2 + (1)^2 + \ldots + (-1)^2 - \frac{(31)^2}{20} = 183.00 - 48.05 = 134.95$

$$SS_{Tratamientos(ajustada)} = \frac{k\sum_{i=1}^{4} Q_j^2}{\lambda a} = \frac{4\left[(5.75)^2+(-4.00)^2+(-9.50)^2+(-8.00)^2+(15.75)^2\right]}{(3)(5)} = \frac{4(451.375)}{15}$$

$= 120.37$

$$SS_{Dias} = \sum_{i=1}^{5}\frac{y_{i..}^2}{k} - \frac{y_{...}^2}{N} = \frac{(2)^2+(6)^2+(7)^2+(9)^2+(7)^2}{4} - \frac{(31)^2}{20} = \frac{219}{4} - 48.05 = 54.75 - 48.05 = 6.70$$

$$SS_{Estaciones} = \sum_{h=1}^{4}\frac{y_{..h}^2}{b} - \frac{y_{...}^2}{N} = \frac{(6)^2+(9)^2+(7)^2+(9)^2}{5} - \frac{(31)^2}{20} = \frac{247}{5} - 48.05 = 49.40 - 48.05 = 1.35$$

$SS_E = SS_T - SS_{Tratamientos(ajustada)} - SS_{Dias} - SS_{Estaciones}$

$SS_E = 134.95 - 120.37 - 6.70 - 1.35$

$SS_E = 6.53$

**Medias de Cuadrados**

$$MS_{Tratamientos(ajustada)} = \frac{SS_{Tratamientos(ajustada)}}{a-1} = \frac{120.37}{5-1} = \frac{120.37}{4} = 30.09$$

$$MS_{Días} = \frac{SS_{Días}}{b-1} = \frac{6.70}{5-1} = \frac{6.70}{4} = 1.675$$

$$MS_{Estaciones} = \frac{SS_{Estaciones}}{p-1} = \frac{1.35}{4-1} = \frac{1.35}{3} = 0.45$$

$$MS_E = \frac{SS_E}{N-a-b-p+2} = \frac{6.53}{20-5-5-4+2} = \frac{6.53}{8} = 0.82$$

**Estadística.**

$$F_o = \frac{MS_{Tratamientos}}{MS_E} = \frac{30.09}{0.82} = 36.69$$

**Tabla de Análisis de Varianza.**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_o$ |
|---|---|---|---|---|
| Niveles de Iluminación (Ajustados) | 120.37 | 4 | 30.09 | 36.69 |
| Días (Sin ajustar) | 6.70 | 4 | 1.675 | |
| Estaciones de trabajo | 1.35 | 3 | 0.45 | |
| Error | 6.53 | 8 | 0.82 | |
| Total | 134.95 | 19 | | |

Utilizando un nivel de significancía del 1% ($\alpha$ = 0.01), para encontrar el $F_{Tablas}$ (Tablas Fisher) con 4 grados de libertad en el numerador ($a$-1) y 8 grados de libertad en denominador $N$-$a$-$b$-$p$+2.

$F_{\alpha,a-1,N-a-b-p+2} = F_{0.01,4,8} = 7.01$

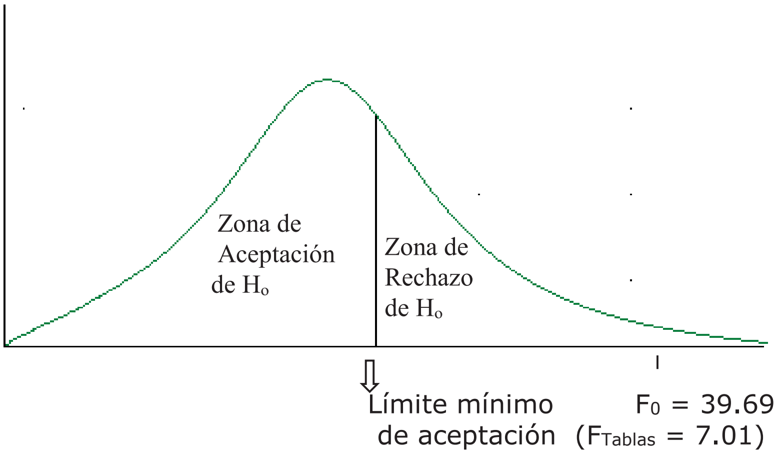
Comparando el $F_0$ calculado en el Análisis de Varianza y el $F_{Tablas}$ , se puede observar que:

$F_0 > F_{Tablas}$

36.69 > 7.01

Por lo tanto, se rechaza la hipótesis nula ($H_0$) y se acepta la hipótesis alternativa ($H_1$);

También puede observarse gráficamente, de la siguiente manera:




Se observa que el valor de $F_0$ cae en la zona de rechazo de $H_0$.

- **Conclusión**

Los niveles de iluminación influyen significativamente en la ocurrencia de los efectos en una operación de ensamble.

**PROBLEMA 5. (Diseño de Cuadrado Latino, valor faltante)**

Un instructor de Ingeniería Química está interesado en la medición de la tensión artificial entre Kerosene y Aceite. Decide correr la prueba con cinco estudiantes del laboratorio en cinco días consecutivos de tal forma que un mismo estudiante no realiza la prueba dos veces el mismo día. El instructor cree que existe variación de ésta tensión a distintos niveles de aceite y con tal propósito considera cinco de tales porcentajes. Los resultados son los siguientes:

| DIA | ESTUDIANTES | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | A = 21 | D = 38 | E = 50 | B = 28 | C = 37 |
| 2 | D = 40 | B = 30 | C = 38 | A = 24 | E = 43 |
| 3 | ******* | A = 20 | B = 30 | E = 49 | D = 41 |
| 4 | E = 45 | C = 33 | A = 28 | D = 37 | B = 35 |
| 5 | B = 27 | E = 40 | D = 43 | C = 34 | A = 33 |

Las letras latinas representan el porcentaje de aceite de la siguiente manera:
A .... 90% , B .... 80% , C .... 70% , D .... 20% y E .... 40%.

Supongamos que al llevar a cabo el experimento por algún motivo no fue posible obtener la observación que corresponde al día 3, estudiante 1, y porcentaje de aceite 70% (C). Con los datos de la tabla anterior obtener la estimación del valor faltante.

**Solución**

**Datos**

$P = 5$

**Cálculo de los totales con el valor faltante.**

| DIA | ESTUDIANTES | | | | | $y'_{i..}$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| 1 | A = 21 | D = 38 | E = 50 | B = 28 | C = 37 | 174 |
| 2 | D = 40 | B = 30 | C = 38 | A = 24 | E = 43 | 175 |
| 3 | | A = 20 | B = 30 | E = 49 | D = 41 | 140 |
| 4 | E = 45 | C = 33 | A = 28 | D = 37 | B = 35 | 178 |
| 5 | B = 27 | E = 40 | D = 43 | C = 34 | A = 33 | 177 |
| $y'_{.j.}$ | 133 | 161 | 189 | 172 | 189 | $y'_{...}$ =844 |

| Porcentajes | $y'_{.j.}$ |
|---|---|
| A | 21 + 20 + 28 + 24 + 33 = 126 |
| B | 27 + 30 + 30 + 28 + 35 = 150 |
| C | 33 + 33 + 38 + 34 + 37 = 142 |
| D | 40 + 38 + 43 + 37 + 41 = 199 |
| E | 45 + 40 + 50 + 49 + 43 = 227 |

Como se trata de un Diseño de Cuadrado Latino y el valor que se desea estimar es el $\hat{y}_{331}$ se debe utilizar la siguiente fórmula:

$\hat{y}_{ijk} = \dfrac{p(y'_{i..} + y'_{.j.} + y'_{..k}) - 2y'_{...}}{(p-2)(p-1)}$, donde $y'_{i..}$, $y'_{.j.}$ y $y'_{..k}$ son el total de renglón, tratamiento y columna respectivamente con la observación faltante y $y'_{...}$ es el total general con el valor faltante.

Sustituyendo se tiene:

$$\hat{y}_{331} = \frac{p(y'_{i..} + y'_{.j.} + y'_{..k}) - 2y'_{...}}{(p-2)(p-1)} = \frac{5(y'_{3..} + y'_{.3.} + y'_{..1}) - 2y'_{...}}{(5-2)(5-1)} = \frac{5(140+142+133) - 2(844)}{3x4} = \frac{5(415) - 2(844)}{12}$$

$$= \frac{2075-1688}{12} = \frac{382}{12} = 32.25$$

$\hat{y}_{331}$ = 32.25

Por lo tanto, $\hat{y}_{331}$ = 32.25 es la estimación del valores faltante de la observación de la tabla.

# UNIDAD PROGRAMÁTICA IV:

# “ DISEÑOS FACTORIALES ”

# INTRODUCCIÓN A LOS DISEÑOS FACTORIALES

## 1. DESCRIPCIÓN

En muchas situaciones experimentales resulta de interés estudiar los efectos producidos por dos o más factores simultáneamente; esto se logra con la ayuda de los Diseños Factoriales.

En general los Diseños Factoriales producen experimentos más eficientes, ya que cada observación proporciona información sobre todos los factores, y es posible ver las respuestas de un factor en diferentes niveles de otro factor en el mismo experimento. Por lo tanto, se entiende por **Diseño Factorial** a aquel diseño en el cual se pueden estudiar los efectos de dos o más factores de variación a la vez; es decir, que se puede investigar todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento.

Cada uno de los factores en estudio varían en su aplicación, a esta variación se le llama **Niveles del Factor**. Las combinaciones de los niveles de cada factor, forman los respectivos tratamientos.

En un diseño factorial, los factores en estudio se representan por letras mayúsculas (A,B,C,……) y los niveles de cada uno por sus respectivas letras minúsculas (a,b,c,……). Los cuales pueden tomar valores de 2,3,4, ……

Existen experimentos factoriales Balanceados y Desbalanceados; diremos que es balanceado cuando el número de réplicas es igual para cada uno de los tratamientos usados en el experimento; en caso contrario es Desbalanceado; también se puede dar el caso en que sólo exista una sola réplica para cada tratamiento. Los Diseños factoriales se pueden combinar con los Diseños Completamente al Azar (Unifactoriales), o con el Diseño de Bloques Aleatorios, etc., dependiendo de la naturaleza del experimento.

Entre las Ventajas de usar un diseño Factorial, se pueden mencionar las siguientes:

1) Ahorro y economía del recurso experimental; ya que cada unidad experimental provee información acerca de dos o más factores, lo que no sucede cuando se realiza con una serie de experimentos simples.
2) Da información respecto a las interacciones entre los diversos factores en estudio.
3) Permite realizar estimaciones de las interacciones de los factores, además de los efectos simples.
4) Permite estimar los efectos de un factor en diversos niveles de los otros factores, produciendo conclusiones que son válidas sobre toda la extensión de las condiciones experimentales.

La única desventaja es que si el número de niveles de algunos de los factores o el número de factores es demasiado grande, entonces el número de todas las combinaciones posibles de tratamientos de factores llega a ser un número grande, en consecuencia la variabilidad en el experimento podría ser grande. Estas dos situaciones, pueden hacer difícil detectar los efectos significativos en el experimento.

Se entiende por **efecto** de un factor al cambio en la respuesta media ocasionada por un cambio en el nivel de ese factor.

En los diseños factoriales existen tres efectos, los cuales son:

1) **Efecto Simple:** son comparaciones entre los niveles de un factor a un sólo nivel del otro factor.
2) **Efecto Principal:** son comparaciones entre los niveles de un factor promediados para todos los niveles del otro factor.
3) **Efecto de Interacción:** Miden las diferencias entre los efectos simples de un factor a diferentes niveles de otro factor; es decir, la diferencia en la respuesta entre los niveles de un factor no es la misma en todos los niveles de los otros factores.

En un experimento factorial se puede estimar y contrastar hipótesis acerca de las interacciones y los efectos principales.

Existe interacción entre los factores, si la diferencia en la respuesta entre los niveles de un factor no es la misma en todos los niveles de los otros factores.

Una interacción puede ser doble, triple, cuádruple, etc. según el número de factores que sean considerados en el experimento.

En general, el número de efectos principales y las interacciones se pueden determinar por el combinatorio: $\binom{k}{m} = \frac{k!}{m!(k-m)!}$ donde $k$: es el número de factores en el experimento, y para obtener los efectos principales $m$=1, se escribe $m$=2 para una interacción doble, $m$=3 para una interacción triple, etc. Considerando $m \leq k$.

También se puede utilizar el Triángulo de Pascal para encontrar el número de efectos principales y las interacciones:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
:::::::::::::::::::::::::::::::::::

Puede observarse que en el caso de un experimento factorial de cinco factores (línea señalada con la flecha), se tienen:

5 Efectos principales
10 Efectos Dobles
10 Efectos Triples
5 Efectos Cuádruples
1 Efecto Quíntuple

La interacción en la cual intervienen tres o más factores se considera una interacción de orden superior. Por ejemplo: ABC, ABD, BCD,ADCE…., etc.

La notación que se utiliza para indicar el número de niveles y el número de factores de un experimento factorial es sencilla; y podría simbolizarse en general de la siguiente manera: *mxnxoxpx…………*, en donde las letras indican el número de niveles de los factores y el número de veces que aparecen las letras indican el número de factores. En el siguiente cuadro se presentan algunos ejemplos de ésta notación:

| Notación del Experimento Factorial | Número de Factores | Niveles de los Factores | | | |
|---|---|---|---|---|---|
| | | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
| 3x4 | 2 | 3 | 4 | | |
| 2x4x3 | 3 | 2 | 4 | 3 | |
| 2x3x4 | 4 | 2 | 2 | 3 | 4 |
| $2^4$ = 2x2x2x2 | 4 | 2 | 2 | 2 | 2 |
| $3^4$ = 3x3x3x3 | 4 | 3 | 3 | 3 | 3 |

En los diseños factoriales se pueden estudiar los efectos de dos, tres y más factores a la vez, en esta unidad se estudiarán los efectos de dos factores en detalle, tres factores y se hará un planteamiento general para el estudio de los diseños de más de tres factores.

## 2. DISEÑO FACTORIAL DE DOS FACTORES

El diseño factorial más simple o sencillo es aquel que involucra en su estudio sólo dos factores o conjunto de tratamientos; es decir, que sólo se está interesado en los efectos que producen estos dos factores. A este tipo de diseño se le llama **Bifactorial.**

Si A y B son los factores que se van ha estudiar en un diseño factorial, el factor A tendrá "*a* niveles" y el factor B tendrá "*b* niveles", entonces cada repetición o réplica del experimento contiene todas las "*ab*" combinaciones de los tratamientos y en general hay "*n*" repeticiones, es necesario tener al menos dos réplicas ($n \geq 2$), para poder obtener la suma de cuadrados del error.

El orden en el cual se toman las "*abn*" observaciones es aleatorio; de modo que este es un diseño completamente aleatorizado.

Si se considera que los niveles del factor A son $a_1$, $a_2$ y los del factor B son $b_1$, $b_2$ entonces se tiene:

| Factor A | Factor B | |
|---|---|---|
| | $b_1$ | $b_2$ |
| $a_1$ | $a_1b_1$ | $a_1b_2$ |
| $a_2$ | $a_2b_1$ | $a_2b_2$ |

Los tratamientos estarán formados por las combinaciones de los niveles de ambos factores; es decir, que existen cuatro tratamientos los cuales son:

Tratamiento 1: $a_1b_1$ Tratamiento 2: $a_1b_2$

Tratamiento 3: $a_2b_1$ Tratamiento 4: $a_2b_2$

**Interpretación de los efectos principales y la interacción.**

- El efecto simple del Factor A cuando el Factor B toma el nivel $b_1$: es igual al cambio en la variable respuesta $y$; cuando se cambia del tratamiento $a_1b_1$ al tratamiento $a_2b_1$.
- El efecto simple del Factor A cuando el Factor B toma el nivel $b_2$: es igual al cambio en la variable respuesta $y$ , cuando se cambia del tratamiento $a_1b_2$ al tratamiento $a_2b_2$.
- El efecto principal del factor A: es el promedio de los efectos simples del factor A
- El efecto principal del factor B: es el promedio de los efectos simples del factor B
- La interacción del factor A y el factor B: es la diferencia del efecto simple del factor A cuando el factor B toma el nivel $b_2$ menos el efecto simple del factor A cuando el factor B toma el nivel $b_1$.

Es decir, si el efecto sobre la variable respuesta ($y$) debe cambiar del nivel $a_1$ al nivel $a_2$ del factor A es igual para los dos niveles del factor B, entonces la interacción es cero (no hay interacción o los factores operan independientemente).

Si el efecto simple del factor A depende del nivel del factor B, entonces existe interacción entre el factor A y el factor B.

Si el signo que resulta de la interacción es positivo se dice que existe "**efecto sinergizante" o "potencialización"**; sin embargo si es negativo se dice que existe "**efecto antagónico" o "interferencia"**.

**Forma Gráfica de observar la Interacción.**

Para observar la interacción entre el factor A y el factor B, se debe realizar una gráfica de la respuesta de los datos, contra los niveles del factor A, para ambos niveles del factor B, de la siguiente manera:

Figura 1

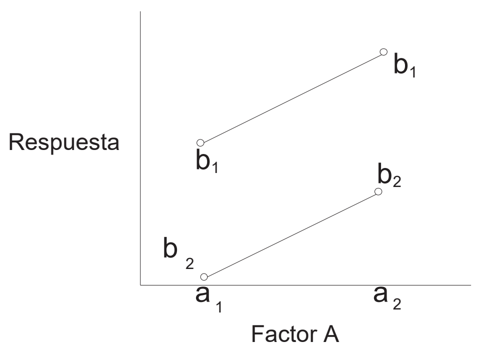
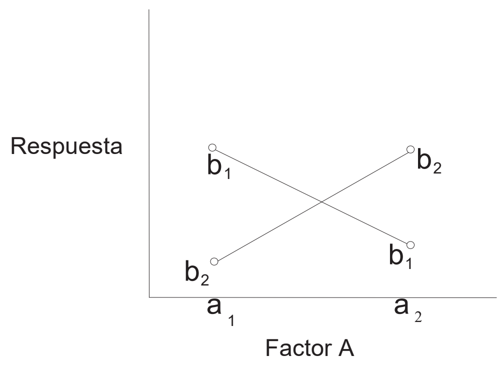



Figura 2




En la **Figura 1**, podemos observar que las rectas que se forman son aproximadamente paralelas lo cual indica que no existe interacción entre el Factor A y el Factor B; mientras que en la **Figura 2**, las dos rectas que se forman no son paralelas y se intersectan, lo que significa que existe interacción entre el Factor A y el Factor B.

Los "$a$" niveles del factor A y los "$b$" niveles del factor B, pueden ser elegidos aleatoriamente de poblaciones más grandes, entonces se dice que el factor A y el factor B son aleatorios; en este caso las inferencias pueden generalizarse a todos los niveles de las poblaciones bajo estudio, porque los niveles de los factores se eligieron al azar. También el experimentador puede elegir específicamente los " $a$" niveles del factor A y los " $b$" niveles del factor B usados en el experimento, entonces se dice que el factor A y el factor B son fijos; por lo tanto las inferencias hechas con base en el Análisis de Varianza pueden aplicarse solamente a los niveles específicos del factor A y el factor B que se probaron.

En consecuencia de lo anterior, se pueden dar los siguientes casos:

1) El factor A y factor B son fijos, entonces el Modelo que resulta es el Modelo de Efectos Fijos.
2) El factor A y factor B son aleatorios, entonces el Modelo que resulta es el Modelo de Efectos Aleatorios.
3) El factor A es aleatorio y el factor B es fijo, o viceversa, el Modelo que resulta es el Modelo Mixto.

**Ejemplo 1**

Un agrónomo desea planear un experimento usando el diseño de bloques al azar con 4 replicaciones. Él ha decidido usar 2 factores en el experimento: el factor Tipo de Suelo con 4 niveles y el factor Variedades de Maíz con 3 niveles. Él sabe que el diseño de bloques al azar tendrá 12 tratamientos, pero no ha tomado la decisión si debe usar un Modelo factorial fijo, aleatorio o mixto. Para tomar esta decisión, él enumera el siguiente listado de las alternativas posibles para este experimento.

***Situación 1***

El agrónomo selecciona 4 Tipos de Suelo y 3 Variedades de maíz según el objetivo del experimento.

Las conclusiones que se obtengan de este experimento solamente serán válidas para los 4 Tipos de Suelo y las 3 Variedades de maíz.

***Decisión:*** Usar un Modelo de Efectos Fijos

***Situación 2***

El agrónomo selecciona 4 Tipos de Suelo al azar a partir de un listado de todos los tipos de suelos que existen en la zona donde se va a hacer este experimento. Asimismo la selección de 3 Variedades de maíz se hace en forma aleatoria usando un listado que contiene todas las posibles variedades de maíz que existen en la zona.

Las conclusiones que se obtengan de este experimento solamente serán válidas para todos los Tipos de Suelo y todas las Variedades de maíz que existen en la zona experimental.

***Decisión:*** Usar un Modelo de Efectos Aleatorios

***Situación 3***

El agrónomo selecciona 4 Tipos de Suelos al azar a partir de todos los tipos de suelos que existen en la zona donde se va a hacer este experimento, pero usa 3 Variedades de maíz que fueron nombradas en el objetivo del experimento.

Las conclusiones que se obtengan de este experimento serán válidas para todos los Tipos de Suelo que existen en la zona de experimentación, pero solamente para 3 Variedades de maíz que fueron usadas en el experimento.
***Decisión:*** Usar un Modelo Mixto tomando Variedad de maíz como Factor Fijo y Tipos de Suelo como Factor Aleatorio.

***Nota:*** Obsérvese que existe también la posibilidad de usar otro Modelo Mixto en el cual se puede tomar el Factor Tipo de Suelo como Factor Fijo y el Factor Variedad de maíz como Factor Aleatorio.

A medida se vaya profundizando en el estudio de los diseños Bifactoriales, se irán enfatizando las diferencias que existen en su análisis para cada uno de los Modelos planteados anteriormente y siempre que se refiera al Modelo Mixto se considera al factor A como aleatorio y el factor B fijo.

**Ejemplo 2**
Se realizó un estudio para comparar las duraciones de escritura de cuatro marcas de plumas de primera calidad. Se pensó que la superficie de escritura podría afectar la duración, por lo que se seleccionaron tres superficies diferentes. Se utilizó una máquina de escribir para asegurar que las condiciones fueran homogéneas (por ejemplo, presión constante y un ángulo fijo). Para el estudio se obtuvieron dos duraciones (en minutos) para cada combinación de marcas de superficie.
**Interpretación.**
En el ejemplo, se puede observar que existen dos factores de interés el factor A que es la marca de la pluma; y como son cuatro marcas de plumas las que compara su duración entonces el factor A tiene 4 niveles. Mientras que el factor B son las superficies donde se prueba cada marca de las plumas y se utilizan tres superficies, por lo tanto el factor B tiene 3 niveles.
El objetivo del experimento es determinar si la superficie de la escritura afecta la duración de las plumas.

Este experimento así como esta planteado corresponde a un Modelo Mixto; ya que el Factor A es fijo, porque sólo hay cuatro marcas de interés para los experimentadores y todas están representadas en el experimento. El factor B, sin embargo es aleatorio, porque las tres superficies fueron seleccionadas de varias superficies y no son las tres superficies escogidas específicas en realidad las que son de interés.

## 2.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

Suponiendo que existen dos factores, el factor A con "$a$" niveles y el factor B con "$b$" niveles y cada nivel con "$n$" réplicas o repeticiones en el experimento; por lo tanto, la representación de los datos observados para un diseño Bifactorial será de la siguiente forma:

| Factor A | Factor B | | | |
|---|---|---|---|---|
| | 1 | 2 | ... | $b$ |
| 1 | $y_{111}, y_{112}, \ldots, y_{11n}$ | $Y_{121}, y_{122}, \ldots, y_{12n}$ | ... | $y_{1b1}, y_{1b2}, \ldots, y_{1bn}$ |
| 2 | $y_{211}, y_{212}, \ldots, y_{21n}$ | $Y_{221}, y_{222}, \ldots, y_{22n}$ | ... | $y_{2b1}, y_{2b2}, \ldots, y_{2bn}$ |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| $a$ | $y_{a11}, y_{a12}, \ldots, y_{a1n}$ | $y_{a21}, y_{a22}, \ldots, y_{a2n}$ | | $y_{ab1}, y_{ab2}, \ldots, y_{abn}$ |

Sea $y_{ijk}$ la respuesta observada cuando el factor A se encuentra en $i$-ésimo nivel ($i$=1,2,...,$a$), el factor B en el j-ésimo nivel (j=1,2,....,$b$) y la $k$-ésima observación ($k$=1,2,....,$n$). El número total de observaciones en el experimento será: $N = abn$, porque se realizan $n$ réplicas.

Sea:

$y_{i..}$ : El total de las observaciones bajo el $i$-ésimo nivel del factor A.

$y_{.j.}$ : El total de las observaciones bajo el j-ésimo nivel del factor B.

$y_{ij.}$ : El total de las observaciones de la $ij$-ésima celda.

$y_{...}$ : El total general de todas las observaciones.

$\bar{y}_{i..}$, $\bar{y}_{.j.}$, $\bar{y}_{ij.}$ y $\bar{y}_{...}$ : los promedios de renglón, columna, celda y general respectivamente.

Matemáticamente se expresan de la siguiente manera:

**Totales**

$$y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$$

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk}$$

**Promedios**

$$\bar{y}_{i..} = \frac{y_{i..}}{bn}, \quad i = 1,2,\ldots,a$$

$$\bar{y}_{.j.} = \frac{y_{.j.}}{an}, \quad j = 1,2,\ldots,b$$

$$\bar{y}_{ij.} = \frac{y_{ij.}}{n}, \quad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

$$y_{\dots} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \quad \text{ó} \quad \sum_{i=1}^{a} y_{i..} \quad \text{ó} \quad \sum_{j=1}^{b} y_{.j.} \qquad \bar{y}_{\dots} = \frac{y_{\dots}}{abn}$$

## 2.2 MODELO ESTADÍSTICO

Las observaciones descritas en el cuadro anterior, pueden ser representadas mediante el Modelo Lineal siguiente:

$$y_{ijk} = \mu + \iota_i + \beta_j + (\iota\beta)_{ij} + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \\ k = 1,2,\ldots,n \end{cases}$$

donde:

$y_{ijk}$ : Es la $ijk$-ésima observación de la variable respuesta.
$\mu$ : Es el efecto medio general.
$\iota_i$ : Es el efecto del $i$-ésimo nivel del Renglón (Factor A).
$\beta_j$ : Es el efecto del $j$-ésimo nivel de la Columna (Factor B).
$(\iota\beta)_{ij}$ : Es el efecto de la interacción entre $\iota_i$ y $\beta_j$ (interacción del Factor A y el Factor B).
$\varepsilon_{ijk}$ : Es el componente del error aleatorio

A continuación se presentan las suposiciones de los parámetros del Modelo Lineal, para cada uno de los Modelos en estudio:

**Modelo de Efectos Fijos**

Como ambos factores son fijos y se supone que los efectos de tratamiento se definen como desviaciones de la media general, por lo tanto $\sum_{i=1}^{a} \hat{\tau}_i = 0$ y $\sum_{j=1}^{b} \hat{\beta}_j = 0$; también los efectos de interacción son fijos y se definen de manera que: $\sum_{i=1}^{a} (\widehat{\tau\beta})_{ij} = 0$ , j = 1,2,…,$b$ y $\sum_{j=1}^{b} (\widehat{\tau\beta})_{ij} = 0$, i = 1,2,…,$a$. El término del error aleatorio $\varepsilon_{ijk} \sim NID(0,\sigma^2)$.

**Modelo de Efectos Aleatorios**

En este caso ambos factores son aleatorios entonces los parámetros del Modelo $\iota_i$ , $\beta_j$ , $(\iota\beta)_{ij}$ y $\varepsilon_{ijk}$ son variables aleatorias; es decir, $\iota_i$ es NID(0, $\sigma_\tau^2$ ) ,$\beta_j \sim$ NID(0, $\sigma_\beta^2$ ), $(\iota\beta)_{ij} \sim$ NID (0, $\sigma_{\tau\beta}^2$) y $\varepsilon_{ijk} \sim NID(0,\sigma^2)$. Y la varianza de cualquier observación es.:
$V(y_{ijk}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$.

**Modelo Mixto**

Se toma un factor como fijo y el otro aleatorio, suponiendo que el factor A es fijo y el factor B es aleatorio entonces $\tau_i$ es un efecto fijo, $\beta_j$ es un efecto aleatorio, se supone que la interacción $(\tau\beta)_{ij}$ es un efecto aleatorio. Por lo tanto $\sum_{i=1}^{a} \hat{\tau}_i = 0$ y $\beta_j \sim NID(0, \sigma_\beta^2)$. El efecto de la interacción $(\tau\beta)_{ij}$ es una variable aleatoria con distribución normal con media cero y varianza $\left[\frac{(a-1)}{a}\right]\sigma_{\tau\beta}^2$. Sin embargo, la suma del componente de la interacción sobre el factor fijo es igual a cero. Es decir, $\sum_{i=1}^{a} \widehat{(\tau\beta)}_{ij} = (\tau\beta)_{.j} = 0 \quad j = 1,2,\ldots,b$. Esto significa que a diferentes niveles del factor fijo ciertos elementos de la interacción no son independientes. Además puede mostrarse que: $Cov((\tau\beta)_{ij}, (\tau\beta)_{i'j}) = -\frac{1}{a}\sigma_{\tau\beta}^2, \quad i \neq i'$ y la covarianza entre $(\tau\beta)_{ij}$ y $(\tau\beta)'_{ij}$, es igual a cero, para $j \neq j'$. El término del error aleatorio $\varepsilon_{ijk} \sim NID(0,\sigma^2)$.

## 2.3 SUMAS Y MEDIAS DE CUADRADOS

Del estudio de la descomposición de la variabilidad total de los datos, en sus partes que esta compuesta; es de lo que se encarga el Análisis de Varianza.

Sea:

$SS_T$ : Suma total de cuadrados corregida

$SS_A$ : Suma de cuadrados debida a los renglones o al factor A

$SS_B$ : Suma de cuadrados debida a las columnas o al factor B.

$SS_{AB}$ : Suma de cuadrados debida a la interacción entre el factor A y el factor B.

$SS_E$ : Suma de cuadrados debida al error.

La suma total corregida puede expresarse como:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2$$

Al descomponer esta sumatoria y después de hacer algunos pasos algebraicos; y considerar que los seis productos cruzados son iguales a cero (Ver Douglas Montgomey, año 1991, Página 181), se obtiene:

$$SS_T = bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$
$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$

Se observa que la suma total de cuadrados se ha descompuesto en la suma de cuadrados debida a los renglones o al factor A, en la suma de cuadrados debida a las columnas o factor B, en una suma de cuadrados debida a la interacción entre el factor A y el factor B, y una suma debida al error.

Al analizar el último término del miembro derecho de la expresión se puede observar que es necesario tener al menos dos réplicas ($n \geq 2$) para poder obtener la suma de cuadrados del error.

Por lo tanto, $SS_T$ se puede expresar simbólicamente como:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

donde

$SS_T$ : Tiene $abn-1$ grados de libertad; porque existen N = $abn$ observaciones y un sólo parámetro a estimar que es $\mu$.

$SS_A$ :Tiene $a$-1 grados de libertad, porque el factor A tiene "$a$" niveles y sólo hay un parámetro a estimar que es $\iota_i$.

$SS_B$ : Tiene $b$-1 grados de libertad, porque el factor B tiene "$b$" niveles y sólo hay un parámetro a estimar que es $\beta_j$.

$SS_{AB}$ : Tiene $(a-1)(b-1)$ grados de libertar; ya que los grados de libertad de la interacción simplemente corresponde a los grados de libertad de cada celda (los cuales son $ab$-1) menos los grados de libertad de los dos efectos principales del Factor A y Factor B; es decir, $ab-1-(a-1)-(b-1) = (a-1)(b-1)$.

$SS_E$ : Tiene $ab(n-1)$ grados de libertad; porque dentro de cada una de las $ab$ celdas existen $n$-1 grados de libertad entre las $n$ réplicas.

**Matemáticamente estas sumas de cuadrados se obtienen de la siguiente manera:**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$ **(Apendice 1)**

**Sumas de cuadrados para los efectos principales.**

$$SS_A = \sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$$ **(Apendice 2)**

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$ **(Apendice 3)**

**Para obtener $SS_{AB}$, se hace en dos etapas.**

**Primero:** Se calcula la suma de cuadrados entre los totales de las "$ab$" celdas, llamada suma de cuadrados debido a los "subtotales", la cual contiene a la $SS_A$ y $SS_B$.

$$SS_{subtotales} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn}$$

**Segundo:** Se calcula $SS_{AB}$

$$SS_{AB} = SS_{subtotales} - SS_A - SS_B$$

La suma de cuadrados del error se encuentra por diferencia.

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B \quad \text{o} \quad SS_E = SS_T - SS_{subtotales}$$

Las medias de cuadrados se definen como la suma de cuadrados dividida entre sus correspondientes grados de libertad.

Sea:

$MS_A$ : Media de cuadrados del factor A.
$MS_B$ : Media de cuadrados del factor B.
$MS_{AB}$ : Media de cuadrados de la interacción del factor A y el factor B.
$MS_E$ : Media de cuadrados del error.

Las cuales vienen dadas por:

$$MS_A = \frac{SS_A}{a-1} \quad , \quad MS_B = \frac{SS_B}{b-1} \quad , \quad MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} \quad , \quad MS_E = \frac{SS_E}{ab(n-1)}$$

Los valores esperados de cada una de las medias de cuadrados para los Modelos anteriores son:

**Modelo de Efectos Fijos**

$$E(MS_A) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1} \qquad E(MS_B) = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_i^2}{b-1}$$

$$E(MS_{AB}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)} \qquad E(MS_E) = \sigma^2$$

**Modelo de Efectos Aleatorios**

$$E(MS_A) = \sigma^2 + n\,\sigma_{\tau\beta}^2 + bn\,\sigma_{\tau}^2 \qquad E(MS_B) = \sigma^2 + n\,\sigma_{\tau\beta}^2 + an\,\sigma_{\beta}^2$$

$$E(MS_{AB}) = \sigma^2 + n\,\sigma_{\tau\beta}^2 \qquad E(MS_E) = \sigma^2$$

**Modelo Mixto (Factor A fijo y Factor B aleatorio)**

$E(MS_A) = \sigma^2 + n\,\sigma^2_{\tau\beta} + \dfrac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$ $\qquad$ $E(MS_B) = \sigma^2 + an\,\sigma^2_{\beta}$

$E(MS_{AB}) = \sigma^2 + n\,\sigma^2_{\tau\beta}$ $\qquad$ $E(MS_E) = \sigma^2$

## 2.4 ANÁLISIS ESTADÍSTICO

En un diseño Bifactorial, tanto los factores (o tratamientos) de renglón como de columna tienen la misma importancia. A continuación se presentan las bases estadísticas para estos diseños.

Las hipótesis se prueban acerca de la igualdad de los efectos de tratamiento de renglón (Factor A), la igualdad de los efectos de tratamiento de columna (Factor B) y también es interesante determinar si los tratamientos de renglón y columna interaccionan (interacción AB).

El procedimiento para obtener la tabla de Análisis de Varianza de cada uno de los Modelos descritos es similar, la diferencia radica en el cálculo del estadístico ($F_0$). A continuación se presenta la tabla de Análisis de Varianza, detallando tal diferencia para cada uno de los Modelos y considerando en el de efecto mixto al factor A como fijo.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Media de Cuadrado | $F_0$ | | |
|---|---|---|---|---|---|---|
| | | | | Efecto Fijo | Efecto Aleatorio | Efecto Mixto |
| Factor A | $SS_A$ | $a-1$ | $MS_A$ | $F_0=\dfrac{MS_A}{MS_E}$ | $F_0=\dfrac{MS_A}{MS_{AB}}$ | $F_0=\dfrac{MS_A}{MS_{AB}}$ |
| Factor B | $SS_B$ | $b-1$ | $MS_B$ | $F_0=\dfrac{MS_B}{MS_E}$ | $F_0=\dfrac{MS_B}{MS_{AB}}$ | $F_0=\dfrac{MS_B}{MS_E}$ |
| Interacción AB | $SS_{AB}$ | $(a-1)(b-1)$ | $MS_{AB}$ | $F_0=\dfrac{MS_{AB}}{MS_E}$ | $F_0=\dfrac{MS_{AB}}{MS_E}$ | $F_0=\dfrac{MS_{AB}}{MS_E}$ |
| Error | $SS_E$ | $ab(n-1)$ | $MS_E$ | | | |
| Total | $SS_T$ | $abn-1$ | | | | |

Para determinar el estadístico ($F_0$) de los efectos, se deben observar los valores esperados de las medias de cuadrados definidas anteriormente; y evaluar el efecto que tiene la hipótesis nula.

El $F_0$ adecuado para probar la respectiva hipótesis será el numerador su respectiva media de cuadrados y el denominador una de las restantes que tenga igual valor esperado que la del numerador; cuando la hipótesis nula sea verdadera.

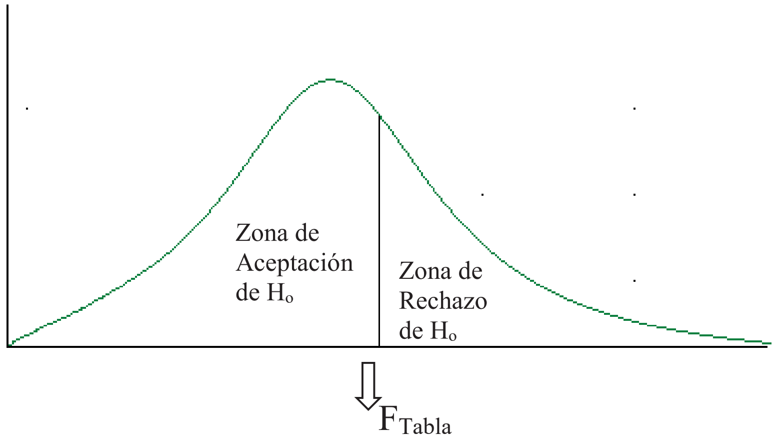
En la siguiente tabla se presentan los Modelos con sus hipótesis a probar y sus respectiva Región de rechazo de la Hipótesis nula, con el $F_{Tablas}$ que corresponden a la distribución F con sus grados de libertad del numerador y denominador para cada caso.

| Modelos | Hipótesis | Región de Rechazo |
|---|---|---|
| Efectos Fijos | $H_0$: $\iota_1 = \iota_2 = \ldots = \iota_a = 0$<br>$H_1$:cuando menos un $\iota_i \neq 0$ | $F_0 > F_{\alpha,a-1,ab(n-1)}$ |
| | $H_0$:$\beta_1 = \beta_2 = \ldots = \beta_b = 0$<br>$H_1$:cuando menos un $\beta_j \neq 0$ | $F_0 > F_{\alpha,b-1,ab(n-1)}$ |
| | $H_0$: $(\iota\beta)_{ij} = 0$, para todo ij<br>$H_1$: cuando menos un $(\iota\beta)_{ij} \neq 0$ | $F_0 > F_{\alpha,(a-1)(b-1),ab(n-1)}$ |
| Efectos Aleatorios | $H_0$: $\sigma^2_\tau = 0$<br>$H_1$: $\sigma^2_\tau > 0$ | $F_0 > F_{\alpha,a-1,(a-1)(b-1)}$ |
| | $H_0$: $\sigma^2_\beta = 0$<br>$H_1$: $\sigma^2_\beta > 0$ | $F_0 > F_{\alpha,b-1,(a-1)(b-1)}$ |
| | $H_0$: $\sigma^2_{\tau\beta} = 0$<br>$H_1$: $\sigma^2_{\tau\beta} > 0$ | $F_0 > F_{\alpha,(a-1)(b-1),ab(n-1)}$ |
| Efecto Mixto | $H_0$: $\iota_1 = \iota_2 = \ldots = \iota_a = 0$<br>$H_1$: cuando menos un $\iota_i \neq 0$ | $F_0 > F_{\alpha,a-1,(a-1)(b-1)}$ |
| | $H_0$: $\sigma^2_\beta = 0$<br>$H_1$: $\sigma^2_\beta > 0$ | $F_0 > F_{\alpha,b-1,ab(n-1)}$ |
| | $H_0$: $\sigma^2_{\tau\beta} = 0$<br>$H_1$: $\sigma^2_{\tau\beta} > 0$ | $F_0 > F_{\alpha,(a-1)(b-1),ab(n-1)}$ |

Lo cual la hipótesis $H_0$: $\iota_1 = \iota_2 = \ldots = \iota_a = 0$, significa que la diferencia de niveles del factor A no tienen efecto significativo sobre la variable respuesta y $H_0$: $\sigma^2_\tau = 0$, significa que la variabilidad en los niveles del factor A no tienen efecto significativo sobre la variable respuesta.

Para llevar a cabo las pruebas de hipótesis planteadas en la tabla, se deben tomar en cuenta las estadísticas correspondientes ($F_0$); las cuales tienen una distribución F, con sus respectivos grados de libertad como se muestra en la tabla ($F_{Tablas}$) y un nivel de significancia $\alpha$, si la hipótesis nula es verdadera.

Por lo tanto, la región crítica es el extremo superior de la distribución F, como se observa a continuación.




La hipótesis nula ($H_o$) se rechazará si, $F_o > F_{Tablas}$, que es la región de rechazo que se presenta en la tabla para cada una de las hipótesis a probar en cada uno de los casos.
Donde $F_o$ se obtiene a través del Análisis de Varianza y $F_{Tabla}$ se obtiene por medio de la tabla F.

**Ejemplo 3**

Se llevó a cabo un estudio del efecto de la temperatura sobre el porcentaje de encogimiento de telas teñidas, con dos réplicas para cada uno de cuatro tipos de tela en un diseño totalmente aleatorizado. Los datos son el porcentaje de encogimiento de dos réplicas de tela secadas a cuatro temperaturas; los cuales se muestran a continuación.

| Factor A (Tipos de tela) | Factor B (Temperatura) | | | |
|---|---|---|---|---|
| | 210°F | 215°F | 220°F | 225°F |
| 1 | 1.8<br>2.1 | 2.0<br>2.1 | 4.6<br>5.0 | 7.5<br>7.9 |
| 2 | 2.2<br>2.4 | 4.2<br>4.0 | 5.4<br>5.6 | 9.8<br>9.2 |
| 3 | 2.8<br>3.2 | 4.4<br>4.8 | 8.7<br>8.4 | 13.2<br>13.0 |
| 4 | 3.2<br>3.6 | 3.3<br>3.5 | 5.7<br>5.8 | 10.9<br>11.1 |

**Solución**

En este ejemplo el análisis se hará como un Modelo de Efectos Fijos; ya que el investigador define con su propio criterio las temperaturas y los tipos de telas que va ha utilizar para llevar a cabo este experimento.

**Variable Respuesta:** Porcentaje de encogimiento de la tela teñida.

**Planteamiento de las Hipótesis a probar:**

Con el objetivo de ejemplificar el Análisis de Varianza de este tipo de Modelo, se plantearan las tres hipótesis en forma Estadística que se desean probar.

a) $H_0 : \iota_1 = \iota_2 = \iota_3 = \iota_4 = 0$

$H_1$ : Cuando menos un $\iota_i \neq 0$

b) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_1$ : Cuando menos un $\beta_j \neq 0$

c) $H_0 : (\iota\beta)_{ij} = 0$, para todo ij

$H_1$ : Cuando menos un $(\iota\beta)_{ij} \neq 0$

**Forma verbal de las Hipótesis**

a) $H_0$ : El tipo de tela no influye en el porcentaje de encogimiento de la tela teñida.

$H_1$ : El tipo de tela influye en el porcentaje de encogimiento de la tela teñida .

b) $H_0$ : Los niveles de temperatura no influyen en el porcentaje de encogimiento de la tela teñida.

$H_1$ : Los niveles de temperatura influyen en el porcentaje de encogimiento de la tela teñida.

c) $H_0$ : La combinación del tipo de tela teñida y la temperatura no influye significativamente en el porcentaje de encogimiento de la tela.

$H_1$ : La combinación del tipo de tela teñida y la temperatura influye significativamente en el porcentaje de encogimiento de la tela.

**Datos**

$a = 4$ , $b = 4$ , $n = 2$ , $N = abn = 4x4x2 = 32$

**Cálculos Matemáticos**

**Totales de Celdas**

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk}$$

$y_{11.} = 3.9$ , $y_{12.} = 4.1$ , $y_{13.} = 9.6$ , $y_{14.} = 15.4$

$y_{21.} = 4.6$ , $y_{22.} = 8.2$ , $y_{23.} = 11.0$ , $y_{24.} = 19.0$

$y_{31.} = 6.0$ , $y_{32.} = 9.2$ , $y_{33.} = 17.1$ , $y_{34.} = 26.2$

$y_{41.} = 6.8$ , $y_{42.} = 6.8$ , $y_{43.} = 11.5$ , $y_{44.} = 22.0$

A continuación se presenta la tabla de datos con sus respectivos totales por celda que se calcularon anteriormente:

| Factor A | Factor B (Temperatura) | | | | |
|---|---|---|---|---|---|
| (Tipos de tela) | 210°F | 215°F | 220°F | 225°F | $y_{i..}$ |
| 1 | 3.9 | 4.1 | 9.6 | 15.4 | 33.0 |
| 2 | 4.6 | 8.2 | 11.0 | 19.0 | 42.8 |
| 3 | 6.0 | 9.2 | 17.1 | 26.2 | 58.5 |
| 4 | 6.8 | 6.8 | 11.5 | 22.0 | 47.1 |
| $y_{.j.}$ | 21.3 | 28.3 | 49.2 | 82.6 | $y_{...}$=181.4 |

**Sumas de Cuadrados.**

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{2} y_{ijk}^2 - \frac{y_{...}^2}{(4)(4)(2)}$$

$$= [(1.8)^2 + (2.1)^2 + (2.0)^2 + (2.1)^2 + \ldots + (10.9)^2 + (11.1)^2] - \frac{(181.4)^2}{32}$$

$= 1370.78 - 1028.31$

$SS_T = 342.47$

**Sumas de Cuadrados para los Efectos Principales**

$$SS_{\text{Tipos de tela}} = \sum_{i=1}^{4} \frac{y_{i..}^2}{(4)(2)} - \frac{y_{...}^2}{(4)(4)(2)}$$

$$= \frac{\left[(33.0)^2 + (42.8)^2 + (58.5)^2 + (47.1)^2\right]}{8} - \frac{(181.4)^2}{32}$$

$= 1070.19 - 1028.31$

$SS_{\text{Tipos de tela}} = 41.88$

$$SS_{\text{Temperatura}} = \sum_{j=1}^{4} \frac{y_{.j.}^2}{(4)(2)} - \frac{y_{...}^2}{(4)(4)(2)}$$

$$= \frac{\left[(21.3)^2 + (28.3)^2 + (49.2)^2 + (82.6)^2\right]}{8} - \frac{(181.4)^2}{32}$$

$= 1312.25 - 1028.31$

$SS_{\text{Temperatura}} = 283.94$

**Cálculo de Subtotales**

$$SS_{Subtotales} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{(4)(4)(2)}$$

$$= \frac{\left[(3.9)^2 + (4.1)^2 + (9.6)^2 + .......+ (22.0)^2\right]}{2} - \frac{(181.4)^2}{32}$$

$= 1369.98 - 1028.31$

$SS_{Subtotales} = 341.67$

**Cálculo de Interacción**

$SS_{Interacción} = SS_{Subtotales} - SS_{Tipos\ de\ tela} - SS_{Temperatura}$

$= 341.67 - 41.88 - 283.94$

$SS_{Interacción} = 15.85$

**Cálculo de la Suma de Cuadrados del Error.**

$SS_E = SS_T - SS_{Subtotales}$

$= 342.47 - 341.67$

$SS_E = 0.8$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Tipos de telas | 41.88 | 3 | 13.96 | 279.20 |
| Temperaturas | 283.94 | 3 | 94.65 | 1893.00 |
| Interacción | 15.85 | 9 | 1.76 | 35.20 |
| Error | 0.8 | 16 | 0.05 | |
| Total | 342.47 | 31 | | |

Tomando $\alpha = 0.05$, encontrando para cada hipótesis a probar sus respectivos $F_{Tablas}$, se tiene:

a) $F_{\alpha,a-1,ab(n-1)} = F_{0.05,4-1,4(4)(2-1)} = F_{0.05,3,16} = 3.24$

b) $F_{\alpha,b-1,ab(n-1)} = F_{0.05,4-1,4(4)(2-1)} = F_{0.05,3,16} = 3.24$

c) $F_{\alpha,(a-1)(b-1),ab(n-1)} = F_{0.05,(4-1)(4-1),4(4)(2-1)} = F_{0.05,9,16} = 2.54$

**Conclusiones**

Se tiene:

**Respecto a la hipótesis *a* (Factor A(Tipo de Tela))**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (279.20 > 3.24); por lo tanto, se rechaza $H_0$; es decir, el tipo de tela teñida influye significativamente en el porcentaje de encogimiento de ella.

**Respecto a la hipótesis *b* (Factor B(Temperatura))**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (1893 > 3.24); por lo tanto, se rechaza $H_0$; es decir, los niveles de temperatura influyen significativamente en el porcentaje de encogimiento de la tela teñida.

**Respecto a la hipótesis *c* (Interacción(Tipo de tela y Temperatura))**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (35.2 > 2.54); por lo tanto, se rechaza $H_0$; es decir, la combinación de el tipo de tela teñida y los niveles de temperatura influyen en el porcentaje de encogimiento de la tela teñida.

## 2.5 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

El Modelo Lineal de un Diseño Bifactorial está dado por $y_{ijk} = \mu + \iota_i + \beta_j + (\iota\beta)_{ij} + \varepsilon_{ijk}$, es posible encontrar las estimaciones de los parámetros de dicho Modelo, para los diferentes tipos de Modelos que se están estudiando.

**Modelo de Efectos Fijos**

Es posible utilizar el método de mínimos cuadrados para obtener las estimaciones de $\mu$, $\iota_i$, $\beta_j$ y $(\iota\beta)_{ij}$ que representan los parámetros del Modelo.

Al utilizar este método es importante tomar en cuenta las siguiente restricciones :

$\sum_{i=1}^{a} \hat{\tau}_i = 0$ , $\sum_{j=1}^{b} \hat{\beta}_j = 0$ , $\sum_{i=1}^{a} (\widehat{\tau\beta})_{ij} = 0$ ; $j = 1,2,\ldots,b$ , $\sum_{j=1}^{b} (\widehat{\tau\beta})_{ij} = 0$ , $i = 1,2,\ldots,a$ por ser desviaciones de la media general.

Al aplicar estas restricciones, las ecuaciones normales se simplifican considerablemente y se obtienen las estimaciones siguientes:

$$\hat{\mu} = \bar{y}_{...} \qquad \hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \quad i = 1,2,\ldots,a$$

$$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \quad j = 1,2,\ldots,b \qquad (\widehat{\tau\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \quad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

Por lo tanto, la media general puede ser estimada por el promedio total de las observaciones, los efectos de los tratamientos que se forman en los renglones (Factor A) se estiman mediante la diferencia entre el promedio de renglón y el promedio general, los efectos de tratamientos que se forman en las columnas (Factor B) se estiman por medio de la diferencia entre el promedio de la columna y el promedio general; y la $ij$-ésima interacción (Factor A y Factor B) se estima restando a la suma del promedio general el promedio de la $ij$-ésima celda el efecto del renglón $i$ y el de la columna $j$. Las estimaciones encontradas se pueden utilizar para encontrar los valores estimados o ajustados de $y_{ijk}$ sustituyéndolos en la ecuación del Modelo de la siguiente manera.

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \widehat{(\tau\beta)}_{ij} + \varepsilon_{ijk}.$$
$$= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})$$
$$= \bar{y}_{...} + \bar{y}_{i..} - \bar{y}_{...} + \bar{y}_{.j.} - \bar{y}_{...} + \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$
$$\hat{y}_{ijk} = \bar{y}_{ij.}$$

es decir, que el Modelo puede ser estimado por el promedio de la ij-ésima celda.

**Modelo de Efectos Aleatorios.**

Al igualar las medias de cuadrados observadas con los valores esperados correspondientes y resolviendo para los componentes de varianza; es posible obtener las estimaciones de los componentes de varianza (Ver Douglas Montgomery, año 1991, Página 199); los cuales son:

$$\hat{\sigma}_{\tau}^2 = \frac{MS_A - MS_{AB}}{bn} \quad , \quad \hat{\sigma}_{\beta}^2 = \frac{MS_B - MS_{AB}}{an} \quad , \quad \hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n} \quad , \quad \hat{\sigma}^2 = MS_E$$

También es posible realizar una estimación del intervalo de confianza para $\sigma^2$ utilizando la distribución ji-cuadrada, considerando que $\frac{ab(n-1)MS_E}{\sigma^2}$ tiene una distribución $\chi^2_{\frac{\alpha}{2},ab(n-1)}$

es decir: $P\left(\chi^2_{1-\frac{\alpha}{2},ab(n-1)} \leq \frac{ab(n-1)MS_E}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},ab(n-1)}\right) = 1 - \alpha$

entonces un intervalo de confianza para $\sigma^2$ a un nivel de confianza del $100(1 - \alpha)\%$ viene dado por: $\frac{ab(n-1)MS_E}{\chi^2_{\frac{\alpha}{2},ab(n-1)}} \leq \sigma^2 \leq \frac{ab(n-1)MS_E}{\chi^2_{1-\frac{\alpha}{2},ab(n-1)}}$

**Modelo de Efectos Mixtos**

En un Modelo mixto, se supone que el factor A es fijo y el factor B es aleatorio; por lo tanto, se pueden estimar los parámetros del factor fijo del Modelo mixto, los cuales serán similares a los del Modelo fijo, por ser el factor A fijo; es decir:

$\hat{\mu} = \bar{y}_{...}$

$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \quad i = 1,2,....,a$

También se deben estimar los componentes de varianza del factor aleatorio (Factor B) y la interacción de ambos factores, la interacción será aleatoria por ser el factor B aleatorio, entonces las estimaciones son similares a las del Modelo de Efectos Aleatorios; es decir:

$$\hat{\sigma}_{\beta}^2 = \frac{MS_B - MS_E}{an} \quad , \quad \hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n} \quad , \quad \hat{\sigma}^2 = MS_E$$

## 3. DISEÑO FACTORIAL DE TRES FACTORES

Estos tipos de Diseños experimentales son aquellos en los cuales se involucran en su estudio tres factores; es decir, que se esta interesado en los efectos que producen los tres factores en la variable respuesta en forma individual y conjunta (interacción). A los cuales se les llama **Diseños Trifactoriales**.

Sean, A,B y C los factores que se van a estudiar en un experimento; el factor A tiene "$a$" niveles, el factor B tiene "$b$" niveles y el factor C tiene "$c$" niveles; por lo tanto, cada repetición del experimento tiene todas la "$abc$" combinaciones de tratamiento y en general hay "$n$" repeticiones (n ≥ 2). El orden en que se toman las "*abcn*" observaciones en el experimento debe ser aleatorio, de modo que este es un Diseño completamente aleatorizado.

Existen tres efectos principales (A,B y C), tres efectos dobles (AB,AC y BC) y un efecto triple (ABC).

Los niveles de cada uno de los factores pueden ser elegidos de forma aleatoria, si es así los factores son aleatorios o ser elegidos específicamente por el experimentador; es decir, los que a él le interesan estudiar, entonces los factores son fijos.

Por lo tanto, de acuerdo a la forma en que son elegidos los niveles de los factores, así es el Modelo que resulta y el Análisis de Varianza que se lleva a cabo.

La tabla siguiente muestra los Modelos que se obtienen en cada uno de los casos.

**Tabla de Modelos del Diseño Trifactorial.**

| Combinación | Forma de elegir los niveles de los Factores | | Modelo |
|---|---|---|---|
| | Fijo | Aleatorio | |
| 1 | ABC | | Efecto Fijo |
| 2 | | ABC | Efecto Aleatorio |
| 3 | A | BC | Efecto Mixto |
| 4 | AB | C | Efecto Mixto |
| 5 | AC | B | Efecto Mixto |
| 6 | BC | A | Efecto Mixto |
| 7 | B | AC | Efecto Mixto |
| 8 | C | AB | Efecto Mixto |

**De la tabla anterior se observa que:**

1) Si al menos uno de los factores es aleatorio resulta un Modelo Mixto.
2) Cuando los tres factores son fijos el Modelo que resulta es de Efectos Fijos.
3) Cuando los tres factores son aleatorios resulta un Modelo de Efectos Aleatorios.
4) En tres factores en estudio, resultan 8 combinaciones.

Igual que en los Diseños Bifactoriales a medida se vaya profundizando en su estudio, se irán enfatizando las diferencias que existen en su análisis para cada uno de los Modelos que se presentaron en la tabla anterior.

## 3.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS.

Ya que existen tres factores, el factor A con "$a$" niveles, el factor B con "$b$"niveles y el factor C con "$c$" niveles y "$n$" réplicas en el experimento. Por lo tanto, la representación de los datos observados para un Diseño Trifactorial tiene la siguiente forma:

| Factor A | Factor B | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | | 2 | | | | ... | $b$ | | | |
| | Factor C | | | | Factor C | | | | ... | Factor C | | | |
| | 1 | 2 | ... | $c$ | 1 | 2 | ... | $c$ | ... | 1 | 2 | ... | $c$ |
| 1 | $X_{111}$ | $X_{112}$ | ... | $X_{11c}$ | $X_{121}$ | $X_{122}$ | ... | $X_{12c}$ | ... | $X_{1b1}$ | $X_{1b2}$ | ... | $X_{1bc}$ |
| 2 | $X_{211}$ | $X_{212}$ | ... | $X_{21c}$ | $X_{221}$ | $X_{222}$ | ... | $X_{22c}$ | ... | $X_{2b1}$ | $X_{2b2}$ | ... | $X_{2bc}$ |
| . | . | . | ... | . | . | . | ... | . | ... | . | . | ... | . |
| . | . | . | ... | . | . | . | ... | . | ... | . | . | ... | . |
| . | . | . | ... | . | . | . | ... | . | ... | . | . | ... | . |
| $a$ | $X_{a11}$ | $X_{a12}$ | | $X_{a1c}$ | $X_{a21}$ | $X_{a22}$ | | $X_{a2c}$ | ... | $X_{ab1}$ | $X_{ab2}$ | | $X_{abc}$ |

donde:

$X_{111} = y_{1111}, y_{1112}, \ldots, y_{111n}$ $\quad X_{121} = y_{1211}, y_{1212}, \ldots, y_{121n}$ $\quad X_{1b1} = y_{1b11}, y_{1b12}, \ldots, y_{1b1n}$

$X_{211} = y_{2111}, y_{2112}, \ldots, y_{211n}$ $\quad X_{221} = y_{2211}, y_{2212}, \ldots, y_{221n}$ $\quad X_{2b1} = y_{2b11}, y_{2b12}, \ldots, y_{2b1n}$

$X_{a11} = y_{a111}, y_{a112}, \ldots, y_{a11n}$ $\quad X_{a21} = y_{a211}, y_{a212}, \ldots, y_{a21n}$ $\quad X_{ab1} = y_{ab11}, y_{ab12}, \ldots, y_{ab1n}$

$X_{112} = y_{1121}, y_{1122}, \ldots, y_{112n}$ $\quad X_{122} = y_{1221}, y_{1222}, \ldots, y_{122n}$ $\quad X_{1b2} = y_{1b21}, y_{1b22}, \ldots, y_{b12n}$

$X_{212} = y_{2121}, y_{2122}, \ldots, y_{212n}$ $\quad X_{222} = y_{2221}, y_{2222}, \ldots, y_{222n}$ $\quad X_{2b2} = y_{2b21}, y_{2b22}, \ldots, y_{2b2n}$

$X_{a12} = y_{a121}, y_{a122}, \ldots, y_{a12n}$ $\quad X_{a22} = y_{a221}, y_{2a22}, \ldots, y_{a22n}$ $\quad X_{ab2} = y_{ab21}, y_{ab22}, \ldots, y_{ab2n}$

$X_{11c} = y_{11c1}, y_{11c2}, \ldots, y_{11cn}$ $\quad X_{12c} = y_{12c1}, y_{12c2}, \ldots, y_{12cn}$ $\quad X_{1bc} = y_{1bc1}, y_{1bc2}, \ldots, y_{1bcn}$

$X_{21c} = y_{21c1}, y_{21c2}, \ldots, y_{21cn}$ $\quad X_{22c} = y_{22c1}, y_{22c2}, \ldots, y_{22cn}$ $\quad X_{2bc} = y_{2bc1}, y_{2bc2}, \ldots, y_{2bcn}$

$X_{a1c} = y_{a1c1}, y_{a1c2}, \ldots, y_{a1cn}$ $\quad X_{a2c} = y_{a2c1}, y_{a2c2}, \ldots, y_{a2cn}$ $\quad X_{abc} = y_{abc1}, y_{abc2}, \ldots, y_{abcn}$

Una observación de la tabla $y_{ijkl}$ significa que el factor A se encuentra en el i-ésimo nivel (i=1,2,...,$a$), el factor B se encuentra en el $j$-ésimo nivel ($j$=1,2,..,$b$) y el factor C se encuentra en el $k$-ésimo nivel (k=1,2,...,$c$) y la $l$-ésima réplica ($l$=1,2,...,$n$) de los niveles de los factores A,B y C. **Por ejemplo**, la observación $y_{2321}$, es la observación de la primera réplica del nivel 2 del factor C, el tercer nivel del factor B y el segundo nivel del factor A.

El número total de observaciones en el experimento será $N = abcn$ ya que se realizan "$n$" réplicas.

Sea:

$y_{i...}$ : Total de las observaciones bajo el i-ésimo nivel del factor A.
$y_{.j..}$ : Total de las observaciones bajo el j-ésimo nivel del factor B.
$y_{..k.}$ : Total de las observaciones bajo el k-ésimo nivel del factor C.
$y_{ij..}$ : Total de las observaciones de la ij-ésima celda.
$y_{i.k.}$ : Total de las observaciones de la ik-ésima celda.
$y_{.jk.}$ : Total de las observaciones de la jk-ésima celda
$y_{ijk.}$ : Total de las observaciones de la ijk-ésima celda
$y_{....}$ : Total general de las observaciones.

$\bar{y}_{i...}$, $\bar{y}_{.j..}$, $\bar{y}_{..k.}$, $\bar{y}_{ij..}$, $\bar{y}_{i.k.}$, $\bar{y}_{.jk.}$, $\bar{y}_{ijk.}$, $\bar{y}_{....}$: los promedios de los totales, respectivos descritos anteriormente.

**Matemáticamente se expresan de la siguiente manera:**

$$y_{i...} = \sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl} \quad , \quad y_{.j..} = \sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl} \quad , \quad y_{..k.} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl} \quad , \quad y_{ij..} = \sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$$

$$y_{i.k.} = \sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl} \quad , \quad y_{.jk.} = \sum_{i=1}^{a}\sum_{l=1}^{n} y_{ijkl} \quad , \quad y_{ijk.} = \sum_{l=1}^{n} y_{ijkl}$$

$$y_{....} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl} \quad \text{ó} \quad \sum_{j=1}^{b}\sum_{k=1}^{c} y_{.jk.} \quad \text{ó} \quad \sum_{j=1}^{b} y_{.j..} \quad \text{ó} \quad \sum_{i=1}^{a} y_{i...} \quad \text{ó} \quad \sum_{k=1}^{c} y_{..k.}$$

$$\bar{y}_{i...} = \frac{y_{i...}}{bcn} \quad , \quad \bar{y}_{.j..} = \frac{y_{.j..}}{acn} \quad , \quad \bar{y}_{..k.} = \frac{y_{..k.}}{abn} \quad , \quad \bar{y}_{ij..} = \frac{y_{ij..}}{cn} \quad , \quad \bar{y}_{i.k.} = \frac{y_{i.k.}}{bn} \quad \bar{y}_{.jk.} = \frac{y_{.jk.}}{an}$$

$$\bar{y}_{ikl.} = \frac{y_{ijk.}}{n} \quad , \quad \bar{y}_{....} = \frac{y_{....}}{abcn}$$

## 3.2 MODELO ESTADÍSTICO

Las observaciones presentadas en el cuadro anterior pueden ser descritas por el Modelo Lineal siguiente:

$$y_{ijkl} = \mu + \iota_i + \beta_j + \gamma_k + (\iota\beta)_{ij} + (\iota\gamma)_{ik} + (\beta\gamma)_{jk} + (\iota\beta\gamma)_{ijk} + \varepsilon_{ijkl} \qquad \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,b \\ k = 1,2,\dots,c \\ l = 1,2,\dots,n \end{cases}$$

donde:

$\mu$ : Efecto medio general.
$\iota_i$ : Efecto del i-ésimo nivel del factor A.
$\beta_j$ : Efecto del j-ésimo nivel del factor B.
$\gamma_k$ : Efecto del k-ésimo nivel del factor C.

$(\iota\beta)_{ij}$ : Efecto de la interacción entre el i-ésimo nivel de A y el j-ésimo nivel de B.
$(\iota\gamma)_{ik}$ : Efecto de la interacción entre el i-ésimo nivel de A y el k-ésimo nivel de C.
$(\beta\gamma)_{jk}$ : Efecto de la interacción entre el j-ésimo nivel de B y el k-ésimo nivel de C.
$(\iota\beta\gamma)_{ijk}$ : Efecto de la interacción entre el i-ésimo nivel de A, j-ésimo nivel de B y el k-ésimo nivel de C.
$\varepsilon_{ijkl}$ : Componente del error aleatorio (Error Natural).

Para cada uno de los Modelos que resultan; el Modelo Lineal no cambia su forma, sólo existe diferencia en las suposiciones de sus parámetros.

**Modelo de Efectos Fijos**

Como se supone que los tres factores son fijos y que los efectos de tratamiento se definen como desviaciones de la media general, por lo tanto:

$\sum_{i=1}^{a}\hat{\tau}_i = \sum_{j=1}^{b}\hat{\beta}_j = \sum_{k=1}^{c}\hat{\gamma}_k = 0$; se supone que los efectos de interacción son fijos y que se definen de manera que: $\sum_{i=1}^{a}(\widehat{\tau\beta})_{ij} = \sum_{j=1}^{b}(\widehat{\tau\beta})_{ij} = 0$ , $\sum_{i=1}^{a}(\widehat{\iota\gamma})_{ik} = \sum_{k=1}^{c}(\widehat{\iota\gamma})_{ik} = 0$ , $\sum_{j=1}^{b}(\widehat{\beta\gamma})_{jk} = \sum_{k=1}^{c}(\widehat{\beta\gamma})_{jk} = 0$

$\sum_{i=1}^{b}(\iota\widehat{\beta}\gamma)_{ijk} = \sum_{j=1}^{b}(\iota\widehat{\beta}\gamma)_{ijk} = \sum_{k=1}^{c}(\iota\widehat{\beta}\gamma)_{ijk} = 0$ y el término del error aleatorio $\varepsilon_{ijk} \sim NID(0,\sigma^2)$.

**Modelo de Efectos Aleatorios**

En este caso como los tres factores son aleatorios entonces los parámetros del Modelo $\iota_i$ , $\beta_j$ , $\gamma_k$ , $(\iota\beta)_{ij}$ , $(\iota\gamma)_{ik}$ , $(\beta\gamma)_{jk}$ , $(\iota\beta\gamma)_{ijk}$ y $\varepsilon_{ijkl}$ son variables aleatorias; es decir: $\iota_i \sim NID(0,\sigma_\tau^2)$ , $\beta_j \sim NID(0, \sigma_\beta^2)$ , $\gamma_k \sim NID(0, \sigma_\gamma^2)$ , $(\iota\beta)_{ij} \sim NID(0, \sigma_{\tau\beta}^2)$, $(\iota\gamma)_{ik} \sim NID(0, \sigma_{\tau\gamma}^2)$, $(\beta\gamma)_{jk} \sim NID(0,\sigma_{\beta\gamma}^2)$, $(\iota\beta\gamma)_{ijk} \sim NID(0,\sigma_{\tau\beta\gamma}^2)$ y $\varepsilon_{ijkl} \sim NID(0,\sigma^2)$.

La varianza de cualquier observación vendrá dada por:

$$V(y_{ijkl}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_{\tau\beta}^2 + \sigma_{\tau\gamma}^2 + \sigma_{\beta\gamma}^2 + \sigma_{\tau\beta\gamma}^2 + \sigma^2$$

donde $\sigma_\tau^2$ , $\sigma_\beta^2$ , $\sigma_\gamma^2$ , $\sigma_{\tau\beta}^2$, $\sigma_{\tau\gamma}^2$ , $\sigma_{\beta\gamma}^2$ , $\sigma_{\tau\beta\gamma}^2$ , $\sigma^2$ se conocen como componentes de varianza.

**Modelo Mixto**

Según la tabla anterior de Modelos, cuando al menos un factor sea Aleatorio y los demás Fijos el Modelo que resulta es Mixto. Como para tres factores resultaron seis Modelos Mixtos, sólo se planteará uno de los casos; es decir, el caso en el que el factor A es fijo y los factores B y C son aleatorios.

Como el factor A es fijo, el parámetro $\iota_i$ es un efecto fijo; por lo tanto $\sum_{i=1}^{a} \hat{\tau}_i = 0$. El factor B y el factor C son aleatorios, entonces $\beta_j$ y $\gamma_k$ son de efectos aleatorios; es decir, $\beta_j \sim NID(0, \sigma_{\beta}^2)$ , $\gamma_k \sim NID(0, \sigma_{\gamma}^2)$. Los efectos de las interacciones dobles y triple será:

- El efecto de la interacción $(\iota\beta)_{ij}$ es una variable con distribución normal con media cero y varianza $\left[\frac{(a-1)}{a}\right]\sigma_{\tau\beta}^2$. Sin embargo, la suma del componente de la interacción sobre el factor fijo es igual a cero. Es decir, $\sum_{i=1}^{a}(\tau\beta)_{ij} = (\iota\beta)_{.j} = 0$ , $j = 1,2,\ldots,b$. Esto significa que a diferentes niveles del factor fijo ciertos elementos de la interacción no son independientes. Además puede mostrarse que: $Cov((\iota\beta)_{ij}, (\iota\beta)_{i'j}) = -\frac{1}{a}\sigma_{\tau\beta}^2$, $i \neq i'$ y la covarianza entre $(\iota\beta)_{ij}$ y $(\iota\beta)_{ij'}$ es igual a cero, para $j \neq j'$ .
- El efecto de la interacción $(\iota\gamma)_{ik}$ es una variable con distribución normal con media cero y varianza $\left[\frac{(a-1)}{a}\right]\sigma_{\tau\gamma}^2$. Sin embargo, la suma del componente de la interacción sobre el factor fijo es igual a cero. Es decir, $\sum_{i=1}^{a}(\tau\gamma)_{ik} = (\iota\gamma)_{.k} = 0$ , $k = 1,2,\ldots,c$. Esto significa que a diferentes niveles del factor fijo ciertos elementos de la interacción no son independientes. Además puede mostrarse que: $Cov((\iota\gamma)_{ik}, (\iota\gamma)_{i'k}) = -\frac{1}{a}\sigma_{\tau\gamma}^2$, $i \neq i'$ y la covarianza entre $(\iota\gamma)_{ik}$ y $(\iota\gamma)_{ik'}$ es igual a cero, para $k \neq k'$ .
- El efecto de la interacción $(\beta\gamma)_{jk} \sim NID(0, \sigma_{\beta\gamma}^2)$ , Por ser el factor B y factor C aleatorio.
- El efecto de la interacción $(\iota\beta\gamma)_{ijk}$ es una variable con distribución normal con media cero y varianza $\left[\frac{(a-1)}{a}\right]\sigma_{\tau\beta\gamma}^2$. Sin embargo, la suma del componente de la interacción sobre el factor fijo es igual a cero. Es decir, $\sum_{i=1}^{a}(\tau\beta\gamma)_{ijk} = (\iota\beta\gamma)_{.jk} = 0$ , $j = 1,2,\ldots,b$ y $k = 1,2,\ldots,c$. Esto significa que a diferentes niveles del factor fijo ciertos elementos de la interacción no son independientes. Además puede mostrarse que: $Cov((\iota\beta\gamma)_{ikj}, (\iota\beta\gamma)_{i'jk}) = -\frac{1}{a}\sigma_{\tau\beta\gamma}^2$, $i \neq i'$ y la covarianza entre $(\iota\beta\gamma)_{ijk}$ y $(\iota\beta\gamma)_{ij'k}$ , $(\iota\beta\gamma)_{ijk}$ y $(\iota\beta\gamma)_{ijk'}$ son iguales a cero, para $j \neq j'$ y $k \neq k'$ .

- El término del error aleatorio $\varepsilon_{ijk} \sim NID(0,\sigma^2)$.

## 3.3 SUMAS Y MEDIAS DE CUADRADOS

El Análisis de Varianza es igual que los Diseños estudiados anteriormente, trata de la descomposición de la variabilidad de los datos en sus partes que esta compuesta.

Sea:

$SS_T$ : Suma total de cuadrados corregida.

$SS_A$ : Suma de cuadrados debida al factor A.

$SS_B$ : Suma de cuadrados debida al factor B.

$SS_C$ : Suma de cuadrados debida al factor C.

$SS_{AB}$ : Suma de cuadrados debida a la interacción del factor A y el factor B.

$SS_{AC}$ : Suma de cuadrados debida a la interacción entre el factor A y el factor C.

$SS_{BC}$ : Suma de cuadrados debida a la interacción entre el factor B y el factor C.

$SS_{ABC}$ : Suma de cuadrados debida a la interacción entre el factor A, factor B y el factor C.

$SS_E$ : Suma de cuadrados debida al error.

La suma total de cuadrados corregida de un Diseño Trifactorial puede expresarse mediante la siguiente ecuación:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n}(y_{ijkl} - \bar{y}_{....})^2$$

Es posible descomponer esta sumatoria y observar las partes que la forma, realizando algunos pasos algebraicos un poco complicados, luego se obtiene:

$$SS_T = bcn\sum_{i=1}^{a}(\bar{y}_{i...} - \bar{y}_{....})^2 + acn\sum_{j=1}^{b}(\bar{y}_{.j..} - \bar{y}_{....})^2 + abn\sum_{k=1}^{c}(\bar{y}_{..k.} - \bar{y}_{....})^2 + cn\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij..} - \bar{y}_{....})^2$$

$$+ bn\sum_{i=1}^{a}\sum_{k=1}^{c}(\bar{y}_{i.k.} - \bar{y}_{....})^2 + an\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{.jk.} - \bar{y}_{....})^2 +$$

$$n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{ijk.} - \bar{y}_{i...} - \bar{y}_{.j..} - \bar{y}_{..k.} - \bar{y}_{i.k.} - \bar{y}_{.jk.} - \bar{y}_{ij..} + \bar{y}_{....})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n}(y_{ijkl}^2 - \bar{y}_{ijk.})^2$$

Se observa que la suma total de cuadrados se ha descompuesto en la suma de cuadrados debido al factor A, en la suma debida al factor B, en la suma debido al factor C, en la suma de cuadrados debida a la interacción del Factor A y Factor B, en la suma de cuadrados debido a la interacción de Factor A y Factor C, en la suma de cuadrados debido a la interacción de Factor B y Factor C, en la suma de cuadrados debida a la interacción Factor A, Factor B y Factor C, y en una suma debida al error.

Al observar el último término del miembro derecho de la expresión, es necesario para encontrar la suma de cuadrados del error al menos dos réplicas ($n \geq 2$).

En consecuencia $SS_T$ se puede expresar simbólicamente como:

$SS_T = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_E$

donde:

$SS_T$ : Tiene $abcn$-1 grados de libertad, porque existen N = $abcn$ observaciones y un sólo parámetro a estimar que es $\mu$.

$SS_A$ : Tiene $a$-1 grados de libertad; porque existen *a* niveles del factor A y un sólo parámetro a estimar que es $\iota_i$.

$SS_B$ : Tiene $b$-1 grados de libertad; porque existen *b* niveles del factor B y un sólo parámetro a estimar que es $\beta_j$.

$SS_C$ : Tiene $c$-1 grados de libertad; porque existen *c* niveles del factor C y un sólo parámetro a estimar que es $\gamma_k$ .

Los grados de libertad de cada suma de cuadrados de las interacciones son el producto de los grados de libertad del efecto principal de los factores incluidos; es decir:

| **Efecto** | **Suma de Cuadrados** | **Grados de Libertad** |
|---|---|---|
| AB | $SS_{AB}$ | $(a-1)(b-1)$ |
| AC | $SS_{AC}$ | $(a-1)(c-1)$ |
| BC | $SS_{BC}$ | $(b-1)(c-1)$ |
| ABC | $SS_{ABC}$ | $(a-1)(b-1)(c-1)$ |

$SS_E$ : Tiene $abc(n-1)$ grados de libertad; porque dentro de cada una de las $abc$ celdas existen $n$-1 grados de libertad entre las n réplicas.

Para obtener las fórmulas matemáticas de estas **Sumas de Cuadrados**, existen Reglas Generales (que serán estudiadas en el Diseño Factorial General) que pueden utilizarse para encontrarlas; ya que a medida aumenta el número de factores estas expresiones o fórmulas no se pueden determinar muy fácilmente.

**Matemáticamente estas Sumas de Cuadrados se expresan:**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{N} \quad \text{donde } N = abcn$$

Las suma de cuadrados de los efectos principales se obtiene usando los totales de cada uno de los factores.

$$SS_A = \sum_{i=1}^{a} \frac{y_{i...}^2}{bcn} - \frac{y_{....}^2}{N} \quad , \quad SS_B = \sum_{j=1}^{b} \frac{y_{.j..}^2}{acn} - \frac{y_{....}^2}{N} \quad , \quad SS_C = \sum_{k=1}^{c} \frac{y_{..k.}^2}{abn} - \frac{y_{....}^2}{N}$$

Para encontrar las Sumas de Cuadrados de las interacciones para dos factores, se deben obtener los totales de las celdas AxB, AxC y BxC, que son las sumas de cuadrados de los subtotales de los dos factores . Es decir, la tabla de datos originales se desglosa en tres tablas de dos sentidos como lo indican los productos de los factores, con el objetivo de calcular estas cantidades. Como se muestra a continuación.

$$S_{subtotales\ (AB)} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij..}^2}{cn} - \frac{y_{....}^2}{N} \quad , \quad SS_{subtotales\ (AC)} = \sum_{i=1}^{a}\sum_{k=1}^{c} \frac{y_{i.k.}^2}{bn} - \frac{y_{....}^2}{N}$$

$$SS_{subtotales\ (BC)} = \sum_{j=1}^{b}\sum_{k=1}^{c} \frac{y_{.jk.}^2}{an} - \frac{y_{....}^2}{N}$$

Por lo tanto, las Sumas de Cuadrados de interacciones de dos factores se encuentran utilizando las sumas de cuadrados de los subtotales calculados anteriormente de la siguiente manera:

$SS_{AB} = SS_{subtotales\ (AB)} - SS_A - SS_B$

$SS_{AC} = SS_{subtotales\ (AC)} - SS_A - SS_C$

$SS_{BC} = SS_{subtotales\ (BC)} - SS_B - SS_C$

Como puede observarse las Sumas de Cuadrados de las interacciones de dos factores se obtienen restándole a las Sumas de Cuadrados de los subtotales las Sumas de Cuadrados de los factores presentes en dicha suma.

Para calcular la suma de cuadrados de la interacción de los tres factores se usan los totales de las celdas en tres sentidos; que es la suma de cuadrados de los subtotales de los tres factores, de la siguiente forma:

$$SS_{subtotales\ (ABC)} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \frac{y_{ijk.}^2}{n} - \frac{y_{....}^2}{N}$$

Por lo tanto, la Suma de Cuadrados de la interacción de los tres factores viene dada por:

$SS_{ABC} = SS_{subtotales\ (ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SSA_{AC} - SS_{BC}$

Como puede observarse las Sumas de Cuadrados de la interacción de tres factores se obtienen; restándole a las Sumas de Cuadrados de los subtotales, las Sumas de Cuadrados de los factores principales y la Suma de los Cuadrados de las interacciones dobles entre ellos.

La suma de cuadrados del error se obtiene restando la suma de cuadrados de cada efecto principal e interacción a la suma total de cuadrados; es decir:

$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$ o Restando la Suma de Cuadrados de los subtotuales de los tres factores a la Suma Total de Cuadrados de la siguiente forma:

$SS_E = SS_T - SS_{subtotales(ABC)}$

El cálculo de las sumas de cuadrados no cambia para cada uno de los Modelos.

Las medias de Cuadrados son el cociente de las sumas de cuadrados entre sus grados de libertad.

Sea:

$MS_A$ : Media de cuadrados del factor A.

$MS_B$ : Media de cuadrados del factor B.

$MS_C$ : Media de cuadrados del factor C.

$MS_{AB}$ : Media de cuadrados de la interacción del factor A y el factor B.

$MS_{AC}$ : Media de cuadrados de la interacción del factor A y el factor C.

$MS_{BC}$ : Media de cuadrados de la interacción del factor B y el factor C.

$MS_{ABC}$ : Media de cuadrados de la interacción del factor A ,el factor B y el factor C.

$MS_E$ : Media de cuadrados del error (Variabilidad Natural).

Las cuales se definen como:

$$MS_A = \frac{SS_A}{a-1}, \quad MS_B = \frac{SS_B}{b-1}, \quad MS_C = \frac{SS_C}{c-1}, \quad MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$$

$$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}, \quad MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}, \quad MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)}, \quad MS_E = \frac{SS_E}{abc(n-1)}$$

Los valores esperados de las medias de cuadrados juegan un papel importante en el Análisis de Varianza.

Un examen de estos valores esperados permiten desarrollar los estadísticos apropiados para probar las hipótesis de cualquier parámetro del Modelo.

Estos estadísticos son razones de medias de cuadrados, elegidas de tal forma que el valor esperado de la media de cuadrados del numerador difiere del valor esperado de la media de cuadrados del denominador, solamente por la componente de varianza o el factor fijo de interés.

Los valores esperados de las medias de cuadrados de cualquier Modelo pueden obtenerse realizando la aplicación directa del valor operador esperado. Pero a medida aumenta el número de factores esto puede ser muy tedioso. Es posible aplicar algunas reglas que ayudan a obtener los valores esperados de las medias de cuadrados, las cuales serán estudiadas en el Diseño Factorial General.

### 3.4 ANÁLISIS ESTADISTICO.

En el análisis de un Diseño Trifactorial, los tres factores en estudio tienen igual importancia. A continuación se presentan las bases estadísticas para llevar a cabo el análisis para este tipo de Diseños.

Las hipótesis se deben plantear en relación a la igualdad de los efectos principales (A, B y C), además en relación a la igualdad de los efectos de las interacciones dobles (AB, AC, y BC) e interacción triple (entre los tres factores ABC).

En la siguiente tabla se presentan las hipótesis que se deben probar y su respectiva región de rechazo para el Modelo de Efectos Fijos.

**Tabla de Hipótesis de un Diseño Trifactorial**

| Factor | Hipótesis | Región de Rechazo |
|---|---|---|
| Factor A | $H_0$: $\iota_1 = \iota_2 = \ldots = \iota_a = 0$<br>$H_1$: cuando menos un $\iota_i \neq 0$ | $F_0 > F_{\alpha,a-1,abc(n-1)}$ |
| Factor B | $H_0$: $\beta_1 = \beta_2 = \ldots = \beta_b = 0$<br>$H_1$: cuando menos un $\beta_j \neq 0$ | $F_0 > F_{\alpha,b-1,abc(n-1)}$ |
| Interacción C | $H_0$: $\gamma_1 = \gamma_2 = \ldots = \gamma_b = 0$<br>$H_1$: cuando menos un $\gamma_k \neq 0$ | $F_0 > F_{\alpha,(c-1),abc(n-1)}$ |
| Interacción AB | $H_0$: $(\iota\beta)_{ij} = 0$, para todo ij<br>$H_1$: cuando menos un $(\iota\beta)_{ij} \neq 0$ | $F_0 > F_{\alpha,(a-1)(b-1),abc(n-1)}$ |
| Interacción AC | $H_0$: $(\iota\gamma)_{ik} = 0$, para todo ik<br>$H_1$: cuando menos un $(\iota\gamma)_{ik} \neq 0$ | $F_0 > F_{\alpha,(a-1)(c-1),abc(n-1)}$ |
| Interacción BC | $H_0$: $(\beta\gamma)_{jk} = 0$, para todo jk<br>$H_1$: cuando menos un $(\beta\gamma)_{jk} \neq 0$ | $F_0 > F_{\alpha,(b-1)(c-1),abc(n-1)}$ |
| Interacción ABC | $H_0$: $(\iota\beta\gamma)_{ijk} = 0$, para todo ijk<br>$H_1$: cuando menos un $(\iota\beta\gamma)_{ijk} \neq 0$ | $F_0 > F_{\alpha,(a-1)(b-1)(c-1),abc(n-1)}$ |

Por lo general, las hipótesis de efecto principal se prueban sólo si todas las hipótesis de las interacciones se juzgan no significativas; es decir, las hipótesis que se deben probar primero en un Diseño Factorial son las de interacciones, si estas resultan ser no significativas (aceptar $H_0$) se prueban las hipótesis de efectos principales.

Al igual que en un Diseño de dos factores, el procedimiento para obtener la Tabla de Análisis de Varianza en cada uno de los Modelos es similar.
A continuación se presenta la Tabla de Análisis de Varianza para un Modelo de Efectos Fijos.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Media de Cuadrado | $F_0$ |
|---|---|---|---|---|
| Factor A | $SS_A$ | $a-1$ | $MS_A$ | $F_0=\frac{MS_A}{MS_E}$ |
| Factor B | $SS_B$ | $b-1$ | $MS_B$ | $F_0=\frac{MS_B}{MS_E}$ |
| Factor C | $SS_C$ | $c-1$ | $MS_C$ | $F_0=\frac{MS_C}{MS_E}$ |
| Interacción AB | $SS_{AB}$ | $(a-1)(b-1)$ | $MS_{AB}$ | $F_0=\frac{MS_{AB}}{MS_E}$ |
| Interacción AC | $SS_{AC}$ | $(a-1)(c-1)$ | $MS_{AC}$ | $F_0=\frac{MS_{AC}}{MS_E}$ |
| Interacción BC | $SS_{BC}$ | $(b-1)(c-1)$ | $MS_{BC}$ | $F_0=\frac{MS_{BC}}{MS_E}$ |
| Interacción ABC | $SS_{ABC}$ | $(a-1)(b-1)(c-1)$ | $MS_{ABC}$ | $F_0=\frac{MS_{ABC}}{MS_E}$ |
| Error | $SS_E$ | $abc(n-1)$ | $MS_E$ | |
| Total | $SS_T$ | $abcn-1$ | | |

**Ejemplo 5**

En un experimento para investigar las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra, se utilizaron dos períodos (Edad A) diferentes de curado en combinación con dos Temperaturas(B) diferentes de curado y dos tierras(C) diferentes. Se hicieron dos réplicas para cada combinación de niveles de los tres factores, resultando los siguientes datos:

| Edad (A) | Temperatura (B) | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| | Tierra (C) | | Tierra (C) | |
| | 1 | 2 | 1 | 2 |
| 1 | 471<br>413 | 385<br>434 | 485<br>552 | 530<br>593 |
| 2 | 712<br>637 | 770<br>705 | 712<br>789 | 741<br>806 |

**Solución**

**Planteamiento de las hipótesis a probar.**

**Forma Estadística**

a) $H_0$: $\iota_1 = \iota_2 = 0$
$H_1$:cuando menos un $\iota_i \neq 0$

b) $H_0$:$\beta_1 = \beta_2 = 0$
$H_1$:cuando menos un $\beta_j \neq 0$

c) $H_0$: $\gamma_1 = \gamma_2 = 0$
$H_1$:cuando menos un $\gamma_k \neq 0$

d) $H_0$: $(\iota\beta)_{ij} = 0$, para todo ij
$H_1$: cuando menos un $(\iota\beta)_{ij} \neq 0$

e) $H_0$: $(\iota\gamma)_{ik} = 0$, para todo ik
$H_1$: cuando menos un $(\iota\gamma)_{ik} \neq 0$

f) $H_0$: $(\beta\gamma)_{jk} = 0$, para todo jk
$H_1$: cuando menos un $(\beta\gamma)_{jk} \neq 0$

g) $H_0$: $(\iota\beta\gamma)_{ijk} = 0$, para todo ijk
$H_1$:cuando menos un $(\iota\beta\gamma)_{ijk} \neq 0$

**Variable Respuesta:** Resistencia a la compresión de la mezcla de Cemento y Tierra.

**Forma Verbal**

a) $H_0$: La edad o períodos no influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.
$H_1$: La edad o períodos influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

b) $H_0$: La Temperatura no influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.
$H_1$: La Temperatura influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

c) $H_0$: Los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.
$H_1$: Los tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

d) $H_0$: La edad y la temperatura no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.
$H_1$: La edad y la temperatura influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

e) $H_0$: La edad y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.
$H_1$: La edad y los tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

f) $H_0$: La temperatura y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

$H_1$: La temperatura y los tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

g) $H_0$: La edad, Temperatura y tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

$H_1$: La edad, Temperatura y tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra

**Datos**

$a = b = c = 2$ , $n=2$ $N = abcn$ = 2x2x2x2= 16 , $i$=1,2 $j$=1,2 $k$=1,2

Como puede observarse en la tabla los datos obtenidos tienen valores grandes. A continuación se presentan la tabla de los datos codificados (divididos por 100); para llevar a cabo de una manera mas fácil los cálculos matemáticos.

| Edad (A) | Temperatura (B) | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| | Tierra (C) | | Tierra (C) | |
| | 1 | 2 | 1 | 2 |
| 1 | 4.71<br>4.13 | 3.85<br>4.34 | 4.85<br>5.52 | 5.30<br>5.93 |
| 2 | 7.12<br>6.37 | 7.70<br>7.05 | 7.12<br>7.89 | 7.41<br>8.06 |

**Cálculos Matemáticos**

**Totales por celdas** ($y_{ijk.} = \sum_{l=1}^{2} y_{ijkl}$)

$y_{111.}$ = 4.71 + 4.13 = 8.84

$y_{112.}$ = 3.85 + 4.34 = 8.19

$y_{121.}$ = 4.85 + 5.52 = 10.37

$y_{122.}$ = 5.30 + 5.93 = 11.23

$y_{211.}$ = 7.12 + 6.37 = 13.49

$y_{212.}$ = 7.70 + 7.05 = 14.75

$y_{221.}$ = 7.12 + 7.89 = 15.01

$y_{222.}$ = 7.41 + 8.06 = 15.47

**Totales del factor A** ($y_{i...} = \sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}$)

$y_{1...}$ = 8.84 + 8.19 + 10.37 + 11.23 = 38.63

$y_{2...}$ = 13.49 + 14.75 + 15.01 + 15.47 = 58.72

**Totales del factor B** ($y_{.j..} = \sum_{i=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}$)

$y_{.1..}$ = 22.33 + 22.94 = 45.27

$y_{.2..}$ = 25.38 + 26.70 = 52.08

**Totales del factor C** ($y_{..k.} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2} y_{ijkl}$)

$y_{..1.}$ = 22.33 + 25.38 = 47.71

$y_{..2.}$ = 22.94 + 26.70 = 49.64

**Totales de la interacción AxB** ($y_{ij..} = \sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}$)

$y_{11..}$ = 8.84 + 8.19 = 17.03 $y_{12..}$ = 10.37 + 11.23 = 21.60

$y_{21..}$ = 13.49 + 14.75 = 28.24 $y_{22..}$ = 15.01 + 15.47 = 30.48

**Totales de la interacción AxC** ($y_{i.k.} = \sum_{j=1}^{2}\sum_{l=1}^{2} y_{ijkl}$)

$y_{1.1.}$= 8.84 + 10.37 = 19.21 $y_{2.1.}$= 13.49 + 15.01 = 28.50

$y_{1.2.}$= 8.19 + 11.23 = 19.42 $y_{2.2.}$= 14.75 + 15.47 = 30.22

**Totales de la interacción BxC** ($y_{.jk.} = \sum_{i=1}^{2}\sum_{l=1}^{2} y_{ijkl}$)

$y_{.11.}$ = 8.84 + 13.49 = 22.33 $y_{.21.}$ = 10.37 + 15.01 = 25.38

$y_{.12.}$ = 8.19 + 14.75 = 22.94 $y_{.22.}$ = 11.23 + 15.47 = 26.70

**Total general**

$y_{....} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}$ = 4.71 + 4.13 + 3.85 + 4.34 + 4.85 +...+ 7.89 + 7.41 + 8.06 = 97.35

$y_{....} = \sum_{i=1}^{2} y_{i...}$ = 38.63 + 58.72 = 97.35 ó $y_{....} = \sum_{j=1}^{2} y_{.j..}$ = 45.27 + 52.08 = 97.35 ó

$y_{....} = \sum_{k=1}^{2} y_{..k.}$ = 47.71 + 49.64 = 97.35 ó

$y_{....} = \sum_{j=1}^{2}\sum_{k=1}^{2} y_{.jk.}$ = 22.33 + 22.94 + 25.38 + 26.70 = 97.35

A continuación se presenta la tabla de datos con los totales; los cuales fueron calculados anteriormente:

| Edad (A) | Temperatura (B) | | | | $y_{i...}$ |
|---|---|---|---|---|---|
| | 1 | | 2 | | |
| | Tierra (C) | | Tierra (C) | | |
| | 1 | 2 | 1 | 2 | |
| 1 | 8.84 | 8.19 | 10.37 | 11.23 | 38.63 |
| 2 | 13.49 | 14.75 | 15.01 | 15.47 | 58.72 |
| Totales BXC ($y_{.jk.}$) | 22.33 | 22.94 | 25.38 | 26.70 | $y_{....}$ = 97.35 |
| $y_{.j..}$ | 45.27 | | 52.08 | | |

**Totales AxB ( $y_{ij..}$)**

| Edad (A) | Temperatura (B) | |
|---|---|---|
| | 1 | 2 |
| 1 | 17.03 | 21.60 |
| 2 | 28.24 | 30.48 |

**Totales AxC ($y_{i.k.}$)**

| Edad (A) | Tierra (C) | |
|---|---|---|
| | 1 | 2 |
| 1 | 19.21 | 19.42 |
| 2 | 28.50 | 30.22 |

**Totales del factor C ($y_{..k.}$)**

| Tierra (C) | |
|---|---|
| 1 | 2 |
| 47.71 | 49.64 |

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}^2 - \frac{y_{....}^2}{abcn}$$

$$= (4.71)^2 + (4.73)^2 + (3.85)^2 + \ldots + (7.89)^2 + (7.41)^2 + (8.06)^2 - \frac{(97.35)^2}{16}$$

$$= 623.2289 - 592.3139063$$

$$= 30.91$$

Las sumas de cuadrados de los efectos principales se obtiene usando los totales de cada uno de los factores de la siguiente manera:

$$SS_{Edad(A)} = \sum_{i=1}^{2} \frac{y_{i...}^2}{bcn} - \frac{y_{....}^2}{N}$$

$$= \frac{(38.63)^2 + (58.72)^2}{8} - \frac{(97.35)^2}{16}$$

= 617.5394125 - 592.3139063

= 25.2255062 ≈ 25.23

$$SS_{Temperatura(B)} = \sum_{j=1}^{2} \frac{y_{.j..}^2}{acn} - \frac{y_{....}^2}{N}$$

$$= \frac{(45.27)^2 + (52.08)^2}{8} - \frac{(97.35)^2}{16}$$

= 595.2124125 - 592.3139063

= 2.8985062 ≈ 2.90

$$SS_{Tierra(C)} = \sum_{k=1}^{2} \frac{y_{..k.}^2}{abn} - \frac{y_{....}^2}{N}$$

$$= \frac{(47.71)^2 + (49.64)^2}{8} - \frac{(97.35)^2}{16}$$

= 592.5467125 - 592.3139063

= 0.2328062 ≈ 0.23

Las sumas de cuadrados de las interacciones de dos factores se calculan utilizando los totales respectivos de las celdas en dos sentidos; es decir, la suma de cuadrados de los subtotales.

**Cálculo de los Subtotales**

Para la interacción Edad-Temperatura, o interacción AxB, se utilizan los totales de las celdas AxB ($y_{ij..}$). asi:

$$SS_{subtotales(EdadxTemperatura)} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{y_{ij..}^2}{cn} - \frac{y_{....}^2}{N} = \frac{(17.03.)^2 + (21.60)^2 + (28.24)^2 + (30.48)^2}{4} - \frac{(97.35)^2}{16}$$

= 620.777225 - 592.3139063

= 28.4633187

Para la interacción Edad-Tierra, o interacción AxC, se utilizan los totales de las celdas AxC ($y_{i.k.}$). asi:

$$SS_{subtotales(EdadxTierra)} = \sum_{i=1}^{2}\sum_{k=1}^{2} \frac{y_{i.k.}^2}{bn} - \frac{y_{....}^2}{N} = \frac{(19.21)^2 + (19.42)^2 + (28.50)^2 + (30.22)^2}{4} - \frac{(97.35)^2}{16}$$

= 617.914725 - 592.3139063

= 25.6008187

Para la interacción Temperatura-Tierra, o interacción BxC, se utilizan los totales de las celdas BxC ($y_{.jk.}$). asi:

$$SS_{subtotales(TemperaturaxTierra)} = \sum_{j=1}^{2}\sum_{k=1}^{2}\frac{y_{.jk.}^2}{an} - \frac{y_{....}^2}{N} = \frac{(22.33)^2+(22.94)^2+(25.38)^2+(26.70)^2}{4} - \frac{(97.35)^2}{16}$$

$= 595.476725 - 592.3139063$

$= 3.1628187$

Por lo tanto, las sumas de Cuadrados de interacciones de dos factores se obtienen restándoles a los subtotales las sumas de cuadrados de los factores que están involucrados en dicha suma, de la siguiente manera:

$SS_{EdadxTemperatura} = SS_{subtotales(EdadxTemperatura)} - SS_A - SS_B$

$= 28.4633187 - 25.2255062 - 2.8985062$

$= 0.3392496 \approx 0.34$

$SS_{EdadxTierra} = SS_{subtotales(EdadxTierra)} - SS_A - SS_C$

$= 25.6008187 - 25.2255062 - 0.2328062$

$= 0.1425063 \approx 0.14$

$SS_{TemperaturaxTierra} = SS_{subtotales(TemperaturaxTierra)} - SS_B - SS_C$

$= 3.1628187 - 2.8985062 - 0.2328062$

$= 0.0315063 \approx 0.03$

La suma de los cuadrados de los tres factores Edad-Temperatura-Tierra, o interacción AxBxC se encuentra utilizando los totales de las celdas AxBxC ($y_{ijk.}$); es decir, la suma de cuadrados de los subtotales.

$$SS_{subtotales\ (ABC)} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\frac{y_{ijk.}^2}{n} - \frac{y_{....}^2}{N}$$

$$= \frac{(8.84)^2+(8.19)^2+(10.37)^2+(11.23)^2+(13.49)^2+(14.75)^2+(15.01)^2+(15.47)^2}{2} - \frac{(97.35)^2}{16}$$

$= 621.51755 - 592.3139063$

$= 29.2036437$

Por lo tanto, la Suma de Cuadrados de la interacción de tres factores se obtienen restándoles a los subtotales las Sumas de Cuadrados de los efectos principales y la Suma de Cuadrados de las interacciones dobles.

$SS_{ABC} = SS_{subtotales\ (ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$

$= 29.2036437-25.2255062-2.8985062-0.2328062-0.3392496-0.1425063-0.0315063$

$= 0.3335629 \approx 0.33$

La suma de cuadrados del error se obtiene, restándole a la Suma Total de Cuadrados la Suma de cuadrados de los subtotales de los tres factores.

$SS_E = SS_T - SS_{subtotales\ (ABC)}$

$= 30.914994 - 29.2036437$

$= 1.7113503$

o restando a la suma total de cuadrados las sumas de cuadrados de los efectos principales, las sumas de las interacciones dobles y la suma de cuadrados de la interacción triple.

$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$

$= 30.91499-25.22550-2.89850-0.23280-0.33924-0.1425063-0.03150-0.33356$

$= 1.7113503$

**Medias de Cuadrados**

Las Medias de Cuadrados se obtienen dividiendo las Sumas de Cuadrados por sus grados de libertad respectivos, como se muestra a continuación.

$MS_{Edad} = \frac{SS_{Edad}}{a-1} = \frac{25.23}{1} = 25.23$ $\quad$ $MS_{Temperatura} = \frac{SS_{Temperatura}}{b-1} = \frac{2.90}{1} = 2.90$

$MS_{Tierra} = \frac{SS_{Tierra}}{c-1} = \frac{0.23}{1} = 0.23$ $\quad$ $MS_{EdadxTemperatura} = \frac{SS_{EdadxTemperatura}}{(a-1)(b-1)} = \frac{0.34}{1} = 0.34$

$MS_{EdadxTierra} = \frac{SS_{EdadxTierra}}{(a-1)(c-1)} = \frac{0.14}{1} = 0.14$ $\quad$ $MS_{TemperaturaxTierra} = \frac{SS_{TemperaturaxTierra}}{(b-1)(c-1)} = \frac{0.03}{1} = 0.03$

$MS_{EdadxTemperaturaxTierra} = \frac{SS_{EdadxTemperaturaxTierra}}{(a-1)(b-1)(c-1)} = \frac{0.33}{1} = 0.33$ $\quad$ $MS_E = \frac{SS_E}{abc(n-1)} = \frac{1.71}{8} = 0.2139$

Los estadísticos para llevar a cabo la prueba de las hipótesis ($F_0$) se obtienen dividiendo sus respectivas Medias de Cuadrados por la Media de Cuadrados del error, de la siguiente manera:

**Para el factor Edad (A)**

$F_0 = \frac{MS_{Edad}}{MS_E} = \frac{25.23}{0.2139} = 117.95$

**Para el factor Temperatura (B)**

$F_0 = \frac{MS_{Temperatura}}{MS_E} = \frac{2.90}{0.2139} = 13.55$

**Para el factor Tierra (C)**

$F_0 = \frac{MS_{Tierra}}{MS_E} = \frac{0.23}{0.2139} = 1.07$

**Para la interacción Edad-Temperatura (AxB)**

$F_0 = \frac{MS_{EdadxTemperatura}}{MS_E} = \frac{0.34}{0.2139} = 1.59$

**Para la interacción Edad-Tierra (AxC)** **Para la interacción Temperatura-Tierra (BxC)**

$F_0 = \dfrac{MS_{EdadxTierra}}{MS_E} = \dfrac{0.14}{0.2139} = 0.65$ $F_0 = \dfrac{MS_{TemperaturaxTierra}}{MS_E} = \dfrac{0.03}{0.2139} = 0.14$

**Para la interacción Edad-Temperatura-Tierra (AxBxC)**

$$F_0 = \frac{MS_{EdadxTemperaturaxTierra}}{MS_E} = \frac{0.33}{0.2139} = 1.54$$

A continuación se presenta la Tabla de Análisis de Varianza obtenida mediante los cálculos realizados anteriormente.

**Tabla de Análisis de Varianza**

| Fuente de Variación | Sumas de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Edad (A) | 25.23 | 1 | 25.23 | 117.95 |
| Temperatura (B) | 2.90 | 1 | 2.90 | 13.55 |
| Tierra (C) | 0.23 | 1 | 0.23 | 1.07 |
| EdadxTemperatura (AxB) | 0.34 | 1 | 0.34 | 1.59 |
| EdadxTierra (AxC) | 0.14 | 1 | 0.14 | 0.65 |
| TemperturaxTierra(BxC) | 0.03 | 1 | 0.03 | 0.14 |
| EdadxTemperaturaxTierra (AxBxC) | 0.33 | 1 | 0.33 | 1.54 |
| Error | 1.71 | 8 | 0.2139 | |
| Total | 30.91 | 15 | | |

Tomando $\alpha$ = 0.05, encontrando para cada hipótesis a probar sus respectivos $F_{Tablas}$, se tiene:

Ya que $a = b = c = 2$, entonces $abc(n\text{-}1)$ = 2x2x2(2-1) = 8.

a) $F_{\alpha,a-1,abc(n-1)} = F_{0.05,2-1,8} = F_{0.05,1,8} = 5.32$

b) $F_{\alpha,b-1,abc(n-1)} = F_{0.05,2-1,8} = F_{0.05,1,8} = 5.32$

c) $F_{\alpha,c-1,abc(n-1)} = F_{0.05,2-1,8} = F_{0.05,1,8} = 5.32$

d) $F_{\alpha,(a-1)(b-1),abc(n-1)} = F_{0.05,(2-1)(2-1),8} = F_{0.05,1,8} = 5.32$

e) $F_{\alpha,(a-1)(c-1),abc(n-1)} = F_{0.05,(2-1)(2-1),8} = F_{0.05,1,8} = 5.32$

f) $F_{\alpha,(b-1)(c-1),abc(n-1)} = F_{0.05,(2-1)(2-1),8} = F_{0.05,1,8} = 5.32$

d) $F_{\alpha,(a-1)(b-1)(c-1),abc(n-1)} = F_{0.05,(2-1)(2-1)(2-1),8} = F_{0.05,1,8} = 5.32$

- **Conclusiones**

Se tiene:

**Respecto a la Hipótesis *a* (Factor A(Edad))**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (117.95 > 5.32); por lo tanto, se rechaza $H_0$; es decir, la edad o períodos influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

**Respecto a la Hipótesis *b* (Factor B (Temperatura))**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (13.55 > 5.32); por lo tanto, se rechaza $H_0$; es decir, la Temperatura influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

**Respecto a la Hipótesis *c* (Factor C (Tierra))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (1.07 < 5.32); por lo tanto, se acepta $H_0$; es decir, los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

**Respecto a la Hipótesis *d* (Interacción(Edad y Temperatura))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (1.58 < 2.54); por lo tanto, se acepta $H_0$; es decir, la edad y la temperatura no influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

**Respecto a la Hipótesis *e* (Interacción(Edad y Tierra))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (0.66 < 2.54); por lo tanto, se acepta $H_0$; es decir, la edad y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

**Respecto a la Hipótesis *f* (Interacción(Temperatura y Tierra))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (0.14 < 2.54); por lo tanto, se acepta $H_0$; es decir, la temperatura y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

**Respecto a la Hipótesis *g* (Interacción(Edad,Temperatura y Tierra))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (0.55 < 2.54); por lo tanto, se acepta $H_0$; es decir, la edad, Temperatura y tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

En los diseños factoriales que tienen tres o más factores; en el caso que todos los factores son fijos, fácilmente pueden formularse y probarse hipótesis a cerca de los efectos principales y de las interacciones utilizando los estadísticos ($F_0$) que se construyen de forma usual, dividiendo la media de cuadrados del efecto principal o de la interacción correspondiente entre la media de cuadrados del error (como se observa en la Tabla de Análisis de Varianza).

Pero para los Modelos Aleatorios y Mixtos, no existen Estadísticos exactos ($F_0$) para probar ciertos efectos.

Para llevar a cabo la prueba de las hipótesis de los Modelos de Efectos Aleatorios y Mixtos, que no se dispone de una Media de Cuadrados válida para el denominador de $F_0$, se debe utilizar las pruebas F aproximadas(que se estudiarán en el Diseño Factorial General); que son propuestas que se pueden tomar en cuenta para solucionar este problema.

## 4. DISEÑO FACTORIAL GENERAL

Todos los Análisis hechos para los Diseños de dos y tres Factores pueden extenderse al caso General en el que existen "$a$" niveles del Factor A, "$b$" niveles del Factor B, "$c$" niveles del Factor C,…., y así sucesivamente, ordenados en un Experimento Factorial. Por lo tanto, en general habrá un total de $abc \ldots n$ observaciones si existen $n$ réplicas en el Experimento Completo.

Una generalización del Modelo Lineal del Diseño Factorial General estaría dado por:

$$y_{ijk\ldots l} = \mu + \iota_i + \beta_j + \gamma_k + (\iota\beta)_{ij} + (\iota\gamma)_{ik} + (\beta\gamma)_{jk} + (\iota\beta\gamma)_{ijk} + \ldots\ldots + \varepsilon_{ijk\ldots l} \quad \begin{cases} i = 1,2,\ldots.,a \\ j = 1,2,\ldots.,b \\ k = 1,2,\ldots.,c \\ . \\ . \\ . \\ l = 1,2,\ldots.,n \end{cases}$$

## 4.1 REGLAS PARA LAS SUMAS DE CUADRADOS

Las reglas que se presentan a continuación son utilizadas para encontrar las fórmulas matemáticas de las Sumas de Cuadros, cuando se analizan diseños experimentales de tres o más factores; ya que a medida aumenta el número de factores estas expresiones o fórmulas no se pueden determinar fácilmente.

A continuación se plantean estas Reglas tomando como ejemplo el caso de tres factores, para encontrar la suma de cuadrados del factor C:

**Regla 1**

El término del error en el Modelo $\varepsilon_{ij....m}$ se representa por $\varepsilon_{(ij...)m}$, donde *m* corresponde al subíndice de la replicación. Para el Diseño de tres factores, esta regla significa que $\varepsilon_{ijkl}$ se transforma en $\varepsilon_{(ijk)l}$.

**Regla 2**

El Modelo Lineal del Diseño contiene la media general ($\mu$) y el término del error $\varepsilon_{(ij...)m}$, además todos los efectos principales y cualquier interacción que el experimentador suponga que existe. Si se dan todas la posibles interacciones entre los *k* factores, habrá $\binom{k}{2}$ interacciones de dos factores, $\binom{k}{3}$ interacciones de tres factores,…., 1 interacción de *k* factores.

En el caso que algún término de un factor aparece entre paréntesis, entonces significa que no habrá interacción entre ese factor y los otros factores de dicho término.

**Regla 3**

Los subíndices de cada término del Modelo deben dividirse en tres partes:

i) **Activos**: son aquellos que se hallan en el término y no están entre paréntesis.

ii) **Pasivos:** son los subíndices que están presentes en el Modelo pero que se encuentran entre paréntesis.

iii) **Ausentes:** son aquellos subíndices que están presentes en el Modelo pero que no ocurren en ese término en particular.

**Por ejemplo**: En el término $(\iota\gamma)_{ik}$, los subíndices $i$ y $k$ son activos, y los subíndices $j$ y $l$ son ausentes. En $\varepsilon_{(ijk)l}$ los subíndices $i,j$ y $k$ son subíndices pasivos mientras que $l$ es activo.

**Regla 4. Grados de Libertad**

Los grados de libertad de cualquier término del Modelo corresponden al producto del número de niveles asociados con cada subíndice pasivo, y el número de niveles menos uno, asociados con cada subíndice activo.

**Ejemplo:** Los grados de libertad de $(\iota\gamma)_{ik}$ son $(a-1)(c-1)$, ya que no hay subíndices pasivos y, $i,j$ y $k$ son subíndices activos y los grados de libertad de $\varepsilon_{(ijk)l}$ son $abc(n-1)$, $i,j$ y $k$ son pasivos y $l$ es activo.

**Regla 5. Suma de Cuadrados**

Para obtener las fórmulas con el fin de calcular las sumas de cuadrados de cualquier efecto, primero se deben desarrollar los grados de libertad de dicho efecto.

**Por ejemplo**: Los grados de libertad expandidos de $\gamma_k$ son simplemente $c-1$, para $(\iota\gamma)_{ik}$ es $(ac - a - c + 1)$; ya que los grados de libertad de $(\iota\gamma)_{ik}$ es $(a-1)(c-1)$.

Cada término de esta cantidad corresponde a una forma o procedimiento simbólico de una suma de cuadrados no corregida. Luego debe usarse el siguiente procedimiento:

a) Sea 1 el factor de corrección.

$$1 = \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\cdots\cdots\sum_{m=1}^{n} y_{ijk......m}\right)^2}{abc....n} = \frac{y^2_{........}}{abc....n}$$

, el número de puntos del numerador dependen del número de factores más uno.

para nuestro caso: $1 = \dfrac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} y_{ijkl}\right)^2}{abcn} = \dfrac{y^2_{....}}{abcn}$

b) Los signos de sumatoria deben acomodarse de manera que los relacionados con los subíndices de la forma simbólica de interés (en este caso C), aparezca primero. El resto de los elementos se deben encerrar entre paréntesis. Para C se obtiene.

$$\sum_{k=1}^{c}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl}\right)$$

c) La cantidad entre paréntesis debe transformarse en la notación estándar "punto en el subíndice", donde el punto que reemplaza al subíndice indica una sumatoria en este último: $\sum_{k=1}^{c}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl}\right) = \sum_{k=1}^{c}(y_{..k.})$

d) El término entre paréntesis debe elevarse al cuadrado y dividirse entre el producto del número de niveles de los subíndices "punto": $\sum_{k=1}^{c} \frac{y_{..k.}^2}{abn}$, esto equivale a la suma de cuadrados no corregida.

La suma de cuadrados corregida de un efecto se obtiene reemplazando cada forma simbólica por su correspondiente suma de cuadrados no corregida. Así, la suma de cuadrados para $\gamma_k$ se debe reemplazar la forma simbólica en $c$-1 por las sumas de cuadrados no corregidas, así:

$SS_C = \sum_{k=1}^{c} \frac{y_{..k.}^2}{abn} - \frac{y_{....}^2}{abcn}$, siendo está la suma del efecto principal de C en el Modelo de tres factores.

**Ejemplo 4**

Para observar la aplicación de la regla 5, se encontrará la deducción de la suma de cuadrados para $(\iota\gamma)_{ik}$ ,; es decir, $SS_{AC}$.

Los grados de libertad son $(a-1)(c-1) = ac - a - c + 1$

Por lo tanto, la suma de cuadrados, se obtiene de la siguiente forma.

| $ac$ | $-a$ | $-c$ | $+1$ |
|---|---|---|---|
| $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\frac{y_{....}^2}{abcn}$ (a) |
| $\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{i=1}^{a}\left(\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{k=1}^{c}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\frac{y_{....}^2}{abcn}$ (b) |
| $\sum_{i=1}^{a}\sum_{k=1}^{c}(y_{i.k.})$ | $\sum_{i=1}^{a}(y_{i...})$ | $\sum_{k=1}^{c}(y_{..k.})$ | $\frac{y_{....}^2}{abcn}$ (c) |
| $\sum_{i=1}^{a}\sum_{k=1}^{c}\frac{y_{i.k.}^2}{bn}$ | $\sum_{i=1}^{a}\frac{y_{i...}^2}{bcn}$ | $\sum_{k=1}^{c}\frac{y_{..k.}^2}{abn}$ | $\frac{y_{....}^2}{abcn}$ (d) |
| $SS_{AC} = \sum_{i=1}^{a}\sum_{k=1}^{c}\frac{y_{i.k.}^2}{bn} - \sum_{i=1}^{a}\frac{y_{i...}^2}{bcn} - \sum_{k=1}^{c}\frac{y_{..k.}^2}{abn} + \frac{y_{....}^2}{abcn}$ | | | |

**Nota:** $SS_{AC}$ se obtiene de la combinación de las sumas de cuadrados no corregida en el último renglón, de acuerdo con los signos en la parte superior de cada columna.

Entonces la Suma de Cuadrados es:

$SS_{AC} = SS_{subtotales(AC)} - SS_A - SS_C$

## 4.2 REGLAS PARA LAS MEDIAS DE CUADRADOS

Estas reglas son muy útiles para encontrar las Medias de Cuadrados para cualquiera de los Modelos que resultan de un Diseño Factorial de tres o más factores. A continuación se presentan estas reglas, que serán aplicadas a un Diseño con tres factores.

**Regla 1:** Cada uno de los efectos lleva asociado un componente de varianza (efecto aleatorio) o un factor fijo (efecto fijo). Si la interacción contiene al menos un efecto aleatorio, las interacción total se considera aleatoria. Los subíndices de un componente de varianza se representan mediante las letras griegas que identifican el efecto aleatorio particular. Por lo tanto, en un Modelo mixto en dos sentidos con el factor A es fijo y el factor B es aleatorio, el componente de varianza para el factor B es $\sigma_\beta^2$, y el componente de varianza de la interacción de AB es $\sigma_{\tau\beta}^2$. Un efecto fijo siempre se representa por la suma de los cuadrados de los componentes del Modelo asociadas con ese factor, dividida entre los grados de libertad correspondientes. En el caso que el factor A es fijo, el efecto del factor A es: $\frac{\sum_{i=1}^{a} \tau_i^2}{a-1}$.

**Regla 2:** **Medias de Cuadrados Esperados**. Para obtener los valores esperados de las medias de cuadrados debe prepararse la siguiente tabla. Existe un renglón por cada componente del Modelo (media de cuadrado), y una columna para cada subíndice. Sobre cada subíndice debe escribirse el número de niveles del factor asociado con él, y si el factor es fijo (F) o aleatorio (A). Las réplicas siempre se consideran aleatorias.

a) En cada renglón se escribe un “1” si uno de los subíndices pasivos del componente del renglón es igual al subíndice de la columna.
b) Si alguno de los subíndices del componente del renglón es igual al subíndice de la columna, se escribe en ese renglón un “0” si el encabezado de la columna es un factor fijo, y un “1” si es un factor aleatorio.

c) En las restantes columnas de los renglones se escribe el número de niveles correspondientes al encabezado de esa columna.

d) Para obtener el valor esperado de la media de cuadrados de cualquier componente del Modelo, primero se cubren todas las columnas encabezadas por los subíndices activos de ese componente. A continuación, en cada renglón que contenga al menos los mismos subíndices que el componente considerado, se calcula el producto de los números visibles y se multiplica por el factor fijo o aleatorio apropiado, obtenido mediante la Regla 1. La suma de estas cantidades corresponden al valor esperado de la media de cuadrados del componente del Modelo considerado.

Estas reglas pueden ser aplicadas a los Modelos que se van estudiando, para obtener los valores esperados de un **Modelo de Efectos Mixtos** de tres Factores, considerando que el factor A es fijo, el factor B y factor C es aleatorio. Se tiene:

| Factor | Factor A | Factor B | Factor C | |
|---|---|---|---|---|
| | F<br>$a$<br>$i$ | A<br>$b$<br>$j$ | A<br>$c$<br>k | R<br>$n$<br>$l$ |
| $\iota_I$ | 0 | $b$ | $c$ | $n$ |
| $\beta_j$ | $a$ | 1 | $c$ | $n$ |
| $\gamma_k$ | $a$ | $b$ | 1 | $n$ |
| $(\iota\beta)_{ij}$ | 0 | 1 | $c$ | $n$ |
| $(\iota\gamma)_{ik}$ | 0 | $b$ | 1 | $n$ |
| $(\beta\gamma)_{jk}$ | $a$ | 1 | 1 | $n$ |
| $(\iota\beta\gamma)_{ijk}$ | 0 | 1 | 1 | $n$ |
| $\varepsilon_{(ijk)l}$ | 1 | 1 | 1 | 1 |

Las aplicaciones de los literales de la regla dos, para el llenado de la tabla anterior se detalla a continuación:

a) Son los unos de las primeras tres columnas de la última fila, ya que los subíndices $i,j$ y $k$; en esa fila son pasivos.

b) Son los ceros y unos en todas las filas menos los unos descritos en el numeral anterior.

c) Los niveles de cada factor ($abcn$).

d) Para encontrar el valor esperados de la media de cuadrados del factor A, se debe cubrir la columna del renglón $i$, luego multiplicar los valores de las columnas por el componente de varianza o efecto, según la regla uno; esto se debe hacer para cada fila donde aparezca un subíndice "$i$" como activo y luego sumarse; agregándole además el $\sigma^2$ que es el componente de varianza de los errores.

De esta forma se obtienen los demás valores esperados, teniendo cuidado de cubrir la columna o número de columnas de los valores esperados que se desean encontrar.

Por lo tanto, de esta tabla se obtiene los siguientes resultados.

$$E(MS_A) = \sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma} + cn\sigma^2_{\tau\beta} + bcn\frac{\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_B) = \sigma^2 + an\sigma^2_{\beta\gamma} + acn\sigma^2_{\beta}$$

$$E(MS_C) = \sigma^2 + an\sigma^2_{\beta\gamma} + abn\sigma^2_{\gamma}$$

$$E(MS_{AB}) = \sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta}$$

$$E(MS_{AC}) = \sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma}$$

$$E(MS_{BC}) = \sigma^2 + an\sigma^2_{\beta\gamma}$$

$$E(MS_{ABC}) = \sigma^2 + n\sigma^2_{\tau\beta\gamma}$$

$$E(MS_E) = \sigma^2$$

Utilizando estas mismas reglas se obtienen las Medias de Cuadrados Esperados para un **Modelo de Efectos Fijos.**

$$E(MS_A) = \sigma^2 + \frac{bcn\sum_{i=1}^{a}\tau_i^2}{a-1} \qquad E(MS_B) = \sigma^2 + \frac{acn\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_C) = \sigma^2 + \frac{bcn\sum_{k=1}^{c}\gamma_k^2}{c-1} \qquad E(MS_{AB}) = \sigma^2 + \frac{cn\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MS_{AC}) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\sum_{k=1}^{c}(\tau\gamma)_{ik}^2}{(a-1)(c-1)} \qquad E(MS_{BC}) = \sigma^2 + \frac{an\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$$

$$E(MS_{ABC}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)} \qquad E(MS_E) = \sigma^2$$

## 4.3 PRUEBAS F APROXIMADAS

Estas pruebas, no son más que propuestas para solucionar el dilema de que no existen estadísticos exactos ($F_0$) para probar ciertos efectos de los Experimentos Factoriales con tres o más factores en el que interviene un Modelo Aleatorio o Mixto. A continuación se plantearan algunas de estas propuestas, especificando las inconvenientes que existen en cada una de ellas.

**PROPUESTA 1.**

Se puede suponer que ciertas interacciones son insignificantes (se acepta $H_0$), en un Diseño Factorial de tres Factores supongamos que se pueden despreciar todas las interacciones de dos factores; es decir, que es posible hacer $\sigma^2_{\tau\beta} = 0$, $\sigma^2_{\tau\gamma} = 0$ , $\sigma^2_{\beta\gamma} = 0$ ; y luego se deben llevar a cabo las pruebas para los efectos principales.

El inconveniente está en que para hacer esta suposición debe haber algo en la naturaleza del experimento, o un conocimiento a priori determinante, para considerar que una o más interacciones son insignificantes.

En general, estas suposiciones no se deben hacer, no hay que eliminar ciertas interacciones del Modelo sin tener una evidencia de que resulta apropiado hacerlo.

Se prueban primero las interacciones, y después se igualan a cero sino resultaran significativas (si $H_0$ se rechaza); luego se prueban los otros efectos del mismo experimento suponiendo las interacciones iguales a cero. Esta propuesta puede ser peligrosa; ya que en cualquier decisión en relación a las interacciones se pueden cometer los errores tipo I y tipo II.

**PROPUESTA 2.**

Se ponderan las medias de cuadrados del Análisis de Varianza para obtener una estimación del error con más grados de libertad. **Por ejemplo**, si el estadístico $F_0 = \dfrac{MS_{ABC}}{MS_E}$ , resulta que no fue significativo; es decir, que la hipótesis $H_0$, que $\sigma^2_{\tau\beta\gamma} = 0$, no se rechaza y tanto $MS_{ABC}$ como $MS_E$ proporcionan una estimación de la varianza del error $\sigma^2$ .

El experimentador puede decidir ponderar o combinar $MS_{ABC}$ y $MS_E$ de la siguiente manera:

$MS_E^{`} = \dfrac{abc(n-1)MS_E + (a-1)(b-1)(c-1)MS_{ABC}}{abc(n-1)+(a-1)(b-1)(c-1)}$, para que el valor esperado de $MS_E^{`}$ sea $\sigma^2$ ; es decir, $E(MS_E^{`}) = \sigma^2$ .

Se puede observar que $MS_E^{`}$ tiene $abc(n-1)+(a-1)(b-1)(c-1)$ grados de libertad y el $MS_E$ original $abc(n-1)$ grados de libertad.

El inconveniente de ponderar es que se puede cometer el error tipo II, y si se combina la media de cuadrados de un factor que en realidad es significativa con el error, se obtiene una nueva media de cuadrados residual ($MS_E$) muy grande. Esto hará difícil detectar otros efectos significativos. Por otro lado, si la media de cuadrados del error original tiene muy pocos grados de libertad (<6), el experimentador puede ganar mucho en ponderar; ya que en potencia se incrementa considerablemente la precisión de las pruebas posteriores.
Si la media de cuadrados del error original ($MS_E$) tiene menos de 6 grados de libertad debe ponderarse sólo si la estadística F de la media de cuadrados que se va a ponderar no es significativa a un valor grande de $\alpha$; es decir, $\alpha$=0.25.

**PROPUESTA 3 (Método de Satterthwaite)**

En el caso que no pueden omitirse ciertas interacciones y aún se necesitan hacer inferencias en relación a efectos cuyas pruebas exactas no existen, puede utilizarse el siguiente Método. Se utilizan combinaciones lineales de las medias de cuadrados, de la siguiente forma: $MS` = MS_r + .....+MS_s$ y $MS`` = MS_u + .....+MS_v$ , en donde las media de cuadrados para estas combinaciones lineales, se eligen de tal manera, que E(MS\`) – E(MS\`\`) sea igual al efecto (parámetro del modelo o componente de varianza) considerado en la hipótesis nula.

Y el estadístico para realizar la prueba es: $F_0 = \dfrac{MS`}{MS``}$, con una distribución aproximada de $F_{p,q}$, en donde $p = \dfrac{(MS_r + .... + MS_s)^2}{\dfrac{MS_r^2}{f_r} + .... + \dfrac{MS_s^2}{f_s}}$ y $q = \dfrac{(MS_u + .... + MS_v)^2}{\dfrac{MS_u^2}{f_u} + .... + \dfrac{MS_v^2}{f_v}}$ y $f_i$ corresponde a los grados de libertad asociados con la Media de Cuadrados $MS_i$. No existe la seguridad de que p y q sean enteros, por lo tanto será necesario interpolar en las tablas de la distribución F.
Satterthwaite hizo notar que pueden existir casos en que las medias de cuadrados en MS\` y MS\`\` pueden ser negativas (tener signo negativo).
Pero Gaylos y Hopper (1969) consideran que si $MS' = MS_1 - MS_2$ entonces la aproximación de Satterthwaite puede ser aplicada si:

$$\frac{MS_1}{MS_2} > F_{0.025, f_2, f_1} \times F_{0.50, f_2, f_2} \text{ con } f_1 \le 100 \text{ y } f_2 \ge \frac{f_1}{2}.$$

En el caso del **Modelo Aleatorio**, la construcción del estadístico $F_0$ para las pruebas de hipótesis sobre los componentes de varianza presenta algunas complicaciones, pues el cuadrado medio del error se puede usar para probar la hipótesis de que no hay interacción de tres factores, y el cuadrado medio de la interacción de tres factores para probar la hipótesis de la componente de interacción de dos factores, pero al observar con cuidado los Cuadrados

Medios Esperados de este Modelo calculados anteriormente, se puede ver que no hay un cuadrado medio legitimo para el denominador del estadístico $F_0$, al probar las hipótesis nulas sobre las componentes de la varianza que corresponden a los efectos principales.

Las dificultades encontradas en la construcción del Estadístico $F_0$ para algunas hipótesis se hacen evidentes de inmediato con la inspección de los Cuadrados Esperados e incluso en algunos casos, no se dispone de una media cuadrática válida para el denominador de $F_0$.

En la siguiente tabla se presentan las hipótesis que se deben probar y su respectiva región de rechazo para el **Modelo de Efectos Aleatorios**.

| Factor | Hipótesis | $F_0$ | Región de Rechazo |
|---|---|---|---|
| Factor A | $H_0$: $\sigma_{\tau}^2 = 0$ <br> $H_1$: $\sigma_{\tau}^2 > 0$ | Pendiente de la aplicación de pruebas F | Pendiente de la aplicación de pruebas F |
| Factor B | $H_0$: $\sigma_{\beta}^2 = 0$ <br> $H_1$: $\sigma_{\beta}^2 > 0$ | Pendiente de la aplicación de pruebas F | Pendiente de la aplicación de pruebas F |
| Factor C | $H_0$: $\sigma_{\gamma}^2 = 0$ <br> $H_1$: $\sigma_{\gamma}^2 > 0$ | Pendiente de la aplicación de pruebas F | Pendiente de la aplicación de pruebas F |
| Interacción AB | $H_0$: $\sigma_{\tau\beta}^2 = 0$ <br> $H_1$: $\sigma_{\tau\beta}^2 > 0$ | $F_0 = \dfrac{MS_{AB}}{MS_{ABC}}$ | $F_0 > F_{\alpha,(a-1)(b-1),(a-1)(b-1)(c-1)}$ |
| Interacción AC | $H_0$: $\sigma_{\tau\gamma}^2 = 0$ <br> $H_1$: $\sigma_{\tau\gamma}^2 > 0$ | $F_0 = \dfrac{MS_{AC}}{MS_{ABC}}$ | $F_0 > F_{\alpha,(a-1)(c-1),(a-1)(b-1)(c-1)}$ |
| Interacción BC | $H_0$: $\sigma_{\beta\gamma}^2 = 0$ <br> $H_1$: $\sigma_{\beta\gamma}^2 > 0$ | $F_0 = \dfrac{MS_{BC}}{MS_{ABC}}$ | $F_0 > F_{\alpha,(a-1)(c-1),(a-1)(b-1)(c-1)}$ |
| Interacción ABC | $H_0$: $\sigma_{\tau\beta\gamma}^2 = 0$ <br> $H_1$: $\sigma_{\tau\beta\gamma}^2 > 0$ | $F_0 = \dfrac{MS_{ABC}}{MS_E}$ | $F_0 > F_{\alpha,(a-1)(b-1)(c-1),abc(n-1)}$ |

En el **Modelo de Efectos Mixtos** la no existencia de una determinada media cuadrática válida para el denominador de $F_0$ para probar algunas hipótesis; depende de las combinaciones de los efectos de los factores de que esta formado el Modelo del experimento en estudio (fijo y aleatorio).

Como por ejemplo, al hacer un análisis de las medias de cuadrados esperados del Modelo Mixto, donde el Factor A es fijo, el factor B es aleatorio y el factor C es aleatorio; existen pruebas exactas para todos los efectos, menos para el efecto principal del factor A. A continuación se presenta la tabla de hipótesis, valores esperados de las medias de cuadrados y región de rechazo para el **Modelo de Efecto Mixto**.

| Factor | Hipótesis | Valor Esperado de las Medias de Cuadrados | $F_0$ | Región de Rechazo |
|---|---|---|---|---|
| Factor A | $H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$<br>$H_1$: cuando menos un $\tau_i \neq 0$ | $\sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma} + cn\sigma^2_{\tau\beta} + bcn\dfrac{\sum_{i=1}^{a}\tau_i^2}{a-1}$ | Pendiente de la aplicación de pruebas F | Pendiente de la aplicación de pruebas F |
| Factor B | $H_0: \sigma^2_{\beta} = 0$<br>$H_1: \sigma^2_{\beta} > 0$ | $\sigma^2 + an\sigma^2_{\beta\gamma} + acn\sigma^2_{\beta}$ | $F_0 = \dfrac{MS_B}{MS_{BC}}$ | $F_0 > F_{\alpha,(b-1),(b-1)(c-1)}$ |
| Factor C | $H_0: \sigma^2_{\gamma} = 0$<br>$H_1: \sigma^2_{\gamma} > 0$ | $\sigma^2 + an\sigma^2_{\beta\gamma} + abn\sigma^2_{\gamma}$ | $F_0 = \dfrac{MS_C}{MS_{BC}}$ | $F_0 > F_{\alpha,(c-1),(b-1)(c-1)}$ |
| Interacción AB | $H_0: \sigma^2_{\tau\beta} = 0$<br>$H_1: \sigma^2_{\tau\beta} > 0$ | $\sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta}$ | $F_0 = \dfrac{MS_{AB}}{MS_{ABC}}$ | $F_0 > F_{\alpha,(a-1)(b-1),(a-1)(b-1)(c-1)}$ |
| Interacción AC | $H_0: \sigma^2_{\tau\gamma} = 0$<br>$H_1: \sigma^2_{\tau\gamma} > 0$ | $\sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma}$ | $F_0 = \dfrac{MS_{AC}}{MS_{ABC}}$ | $F_0 > F_{\alpha,(a-1)(c-1),(a-1)(b-1)(c-1)}$ |
| Interacción BC | $H_0: \sigma^2_{\beta\gamma} = 0$<br>$H_1: \sigma^2_{\beta\gamma} > 0$ | $\sigma^2 + an\sigma^2_{\beta\gamma}$ | $F_0 = \dfrac{MS_{BC}}{MS_E}$ | $F_0 > F_{\alpha,(b-1)(c-1),abc(n-1)}$ |
| Interacción ABC | $H_0: \sigma^2_{\tau\beta\gamma} = 0$<br>$H_1: \sigma^2_{\tau\beta\gamma} > 0$ | $\sigma^2 + n\sigma^2_{\tau\beta\gamma}$ | $F_0 = \dfrac{MS_{ABC}}{MS_E}$ | $F_0 > F_{\alpha,(b-1)(c-1)(c-1),abc(n-1)}$ |
| Error | | $\sigma^2$ | | |

Para llevar a cabo la prueba de hipótesis para el factor A se debe utilizar las pruebas F aproximadas para obtener el $F_0$ válido y los grados de libertad del numerador y denominador de el $F_{Tablas}$ para contrastar la hipótesis correspondiente.

Por lo tanto, para realizar el Análisis de los diferentes Modelos Mixtos que resultan en un Diseño de Tres Factores; se debe tomar en cuenta el valor esperado de las medias de cuadrados de cada uno de los factores principales e interacciones, para observar en cuales existen estadísticos válidos y en cuales no; para llevar a cabo la prueba de la hipótesis.

**Ejemplo 6**

Se esta estudiando la caída de presión en una válvula de expansión de una turbina. El ingeniero de diseño considera que las variables importantes que influyen en esta caída de presión son la temperatura del gas de la entrada (Factor A), la velocidad de la turbina (Factor B), y la presión del gas a la entrada (Factor C). Estos tres factores se organizan en un Diseño Factorial con la temperatura del gas fija, la velocidad de la turbina y la presión aleatoria. Los datos codificados para dos réplicas se presentan a continuación:

| Presión (Factor C) | Temperatura del Gas (Factor A) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 60ºF | | | | 75ºF | | | | 90ºF | | | |
| | Velocidad (Factor B) | | | | Velocidad (Factor B) | | | | Velocidad (Factor B) | | | |
| | 150 | 200 | 225 | 300 | 150 | 200 | 225 | 300 | 150 | 200 | 225 | 300 |
| 50 psi | -2 | 0 | -1 | 4 | 14 | 6 | 1 | -7 | -8 | -2 | -1 | -2 |
| | -3 | -9 | -8 | 4 | 14 | 0 | 2 | 6 | -8 | 20 | -2 | 1 |
| 75 psi | -6 | -5 | -8 | -3 | 22 | 8 | 6 | -5 | -8 | 1 | -9 | -8 |
| | 4 | -1 | -2 | -7 | 24 | 6 | 2 | 2 | 3 | -7 | -8 | 3 |
| 85 psi | -1 | -4 | 0 | -2 | 20 | 2 | 3 | -5 | -2 | -1 | -4 | 1 |
| | -2 | -8 | -7 | 4 | 16 | 0 | 0 | -1 | -1 | -2 | -7 | 3 |

**Solución**

Como se mencionó en el enunciado del experimento el Factora A es fijo, el Factor B y Factor C es aleatorio; por lo tanto las hipótesis a probar son las siguientes:

**Forma Estadística.**

a) $H_0 : \iota_1 = \iota_2 = 0$

$H_1$ : cuando menos un $\iota_i \neq 0$

b) $H_0 : \sigma_{\beta}^2 = 0$

$H_1 : \sigma_{\beta}^2 > 0$

c) $H_0 : \sigma_{\gamma}^2 = 0$

$H_1 : \sigma_{\gamma}^2 > 0$

d) $H_0 : \sigma_{\tau\beta}^2 = 0$

$H_1 : \sigma_{\tau\beta}^2 > 0$

e) $H_0 : \sigma_{\tau\gamma}^2 = 0$

$H_1 : \sigma_{\tau\gamma}^2 > 0$

f) $H_0 : \sigma_{\beta\gamma}^2 = 0$

$H_1 : \sigma_{\beta\gamma}^2 > 0$

g) $H_0 : \sigma_{\tau\beta\gamma}^2 = 0$

$H_1 : \sigma_{\tau\beta\gamma}^2 > 0$

**Variable Respuesta:** Caída de presión en la una válvula.

**Forma Verbal**

a) $H_0$: La Temperatura del Gas no influye significativamente en la Caída de presión en una válvula de expansión de una turbina

$H_1$: La Temperatura del Gas influye significativamente en la Caída de presión en una válvula de expansión de una turbina

b) $H_0$: La Velocidad no influye significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

$H_1$: La Velocidad influye significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

c) $H_0$: La presión no influye significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

$H_1$: La presión influye significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

d) $H_0$: La Temperatura del Gas y la Velocidad no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

$H_1$: La Temperatura del Gas y la Velocidad influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

e) $H_0$: La Temperatura del Gas y la Presión no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

$H_1$: La Temperatura del Gas y la Presión influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

f) $H_0$: La Velocidad y la Presión no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

$H_1$: La Velocidad y la Presión influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

g) $H_0$: La Temperatura del Gas, la Velocidad y la Presión no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

$H_1$: La Temperatura del Gas, la Velocidad y la Presión influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

**Datos**

$a$=3, $b$=4 , $c$=3 , $n$=2 , $N = abcn$ = 3x4x3x2= 72 , $i$=1,2,3 $j$=1,2,3,4 $k$=1,2,3 , $n$=1,2

Este ejemplo corresponde a un Modelo de Efectos Mixtos por ser el Factor A fijo y los Factores B y C aleatorios; y se puede observar que su análisis será como el planteado anteriormente, en el que no existen pruebas exactas para todos los efectos en especial para el factor A; luego de haber observado los valores esperados de las Medias de Cuadrado que corresponden a este tipo de Diseño. Para probar la hipótesis del efecto del factor A ($H_0$: $\iota_1 = \iota_2 = 0$) se debe usar las pruebas F aproximadas de la siguiente manera:

Se calcula el estadístico F = $\dfrac{MS'}{MS``}$

donde:

MS` = $MS_A + MS_{ABC}$

MS`` = $MS_{AB} + MS_{AC}$

Ya que E(MS`) – E(MS``) = $\dfrac{bcn\sum_{i=1}^{a}\tau_i^2}{a-1}$

Por motivos de ahorrar los detalles de los cálculos, sólo se colocará a continuación el resultado de la Tabla de Análisis de Varianza:

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Media de Cuadrado | $F_0$ |
|---|---|---|---|---|
| Temperatura (Factor A) | 616.78 | 2 | 308.39 | 1.82 |
| Velocidad (Factor B) | 175.56 | 3 | 58.52 | 1.45 |
| Presión (Factor C) | 5.03 | 2 | 2.52 | 0.06 |
| Interacción AB | 809.44 | 6 | 134.91 | 7.00 |
| Interacción AC | 179.06 | 4 | 44.77 | 2.32 |
| Interacción BC | 242.19 | 6 | 40.37 | 1.16 |
| Interacción ABC | 231.07 | 12 | 19.26 | 0.56 |
| Error | 1248.00 | 36 | 34.67 | |
| Total | 3507.11 | 71 | | |

**Cálculo de MS´y MS´´**

MS` = $MS_A + MS_{ABC}$
= 308.39 + 19.26
= 327.65

MS`` = $MS_{AB} + MS_{AC}$
= 134.91 + 44.77
= 179.68

y $F = \dfrac{MS'}{MS``} = \dfrac{327.65}{179.68} = 1.82$

Los grados de libertad de esta estadística se calculan a continuación:

$$p = \frac{(MS_A + MS_{ABC})^2}{\frac{MS_A^2}{2} + \frac{MS_{ABC}^2}{12}} = \frac{(327.65)^2}{\frac{(308.39)^2}{2} + \frac{(19.26)^2}{12}} = 2.26$$

$$q = \frac{(MS_{AB} + MS_{AC})^2}{\frac{MS_{AB}^2}{6} + \frac{MS_{AC}^2}{4}} = \frac{(179.68)^2}{\frac{(134.91)^2}{6} + \frac{(44.77)^2}{4}} = 9.13$$

Tomando $\alpha$ = 0.05, encontrando para cada hipótesis a probar sus respectivos $F_{Tablas}$, se tiene:

Ya que $a$=3, $b$=4 , $c$=3 y $n$ = 2 entonces $abc(n$-1) =3x4x3(2-1) = 36.

a) $F_{\alpha,p,q} = F_{0.05,2.26,9.13} = F_{0.05,2,9}$ = 4.26

b) $F_{\alpha,b-1,(b-1)(c-1)} = F_{0.05,4-1,3x2} = F_{0.05,3,6}$ = 4.76

c) $F_{\alpha,c-1,(b-1)(c-1)} = F_{0.05,3-1,3x2} = F_{0.05,2,6}$ = 5.14

d) $F_{\alpha,(a-1)(b-1),(a-1)(b-1)(c-1)} = F_{0.05,(3-1)(4-1),2x3x2} = F_{0.05,6,12}$ = 3.00

e) $F_{\alpha,(a-1)(c-1),(a-1)(b-1)(c-1)} = F_{0.05,(3-1)(3-1),2x3x2} = F_{0.05,4,12}$ = 3.26

f) $F_{\alpha,(b-1)(c-1),abc(n-1)} = F_{0.05,(4-1)(3-1),3x4x3(2-1)} = F_{0.05,6,36}$ = 2.42

h) $F_{\alpha,(a-1)(b-1)(c-1),abc(n-1)} = F_{0.05,(3-1)(4-1)(3-1),3x4x3(2-1)} = F_{0.05,12,36}$ = 2.09

- **Conclusiones**

Se tiene:

**Respecto a la Hipótesis *a* (Factor A(Temperatura del Gas))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (1.82 < 4.26); por lo tanto, no se rechaza $H_0$; es decir, la Temperatura del Gas no influye significativamente en la Caída de presión en una válvula de expansión de una turbina.

**Respecto a la Hipótesis *b* (Factor B(Velocidad))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (1.45 < 4.76); por lo tanto, no se rechaza $H_0$; es decir, la Velocidad no influye significativamente en la Caída de presión en una válvula de expansión de una turbina.

**Respecto a la Hipótesis *c* (Factor C(Presión))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (0.06 < 5.14); por lo tanto, no se rechaza $H_0$; es decir, la Presión no influye significativamente en la Caída de presión en una válvula de expansión de una turbina.

**Respecto a la Hipótesis *d* (Interacción (Temperatura del Gas y Velocidad))**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (7.00 > 3.00); por lo tanto, se rechaza $H_0$; es decir, la Temperatura del Gas y la Velocidad influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

**Respecto a la Hipótesis *e* (Interacción (Temperatura del Gas y Presión))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (2.32 < 3.26); por lo tanto, no se rechaza $H_0$; es decir, la Temperatura del Gas y la Presión no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

**Respecto a la Hipótesis *f* (Interacción (Velocidad y Presión))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (1.16 < 2.42); por lo tanto, no se rechaza $H_0$; es decir, La Velocidad y la Presión no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

**Respecto a la Hipótesis g (Interacción (Temperatura del Gas, Velocidad y Presión))**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ ( 0.56< 2.09); por lo tanto, no se rechaza $H_0$; es decir, la Temperatura del Gas, la Velocidad y la Presión no influyen significativamente en la variabilidad de la Caída de presión en una válvula de expansión de una turbina.

## 5. DISEÑO FACTORIAL DESBALANCEADO O DESEQUILIBRADO

Este tipo de diseño se da cuando el número de réplicas por cada uno de los tratamientos no tiene igual número; es decir, en el cual no se recopilan el mismo número (n) de observaciones en cada una de las celdas; por lo tanto, el número de observaciones en las celdas es diferente.

Estos Diseños Factoriales pueden darse, por varias razones; entre ellas tenemos:

1) Cuando se dan problemas imprevistos a la hora de recopilar los datos y se pierde parte de la información.

2) Puede ser que el experimento se diseñe a propósito, en forma Desbalanceada; porque algunas combinaciones de tratamiento son más caras o difíciles de ensayar; por lo tanto, se realizan menos observaciones para esas celdas.
3) Se puede decidir obtener réplicas adicionales en algunas celdas; debido a que para el experimentador resulte de interés las combinaciones de tratamiento; ya sea porque representan condiciones nuevas o inexploradas.

En un Diseño Factorial desbalanceado la propiedad de Ortogonalidad de los efectos principales y de las interacciones no se cumple; por lo tanto, no es posible utilizar las técnicas usuales del Análisis de Varianza.

El análisis de los Diseños Factoriales Desbalanceados es mucho más difícil que el de los Diseños Balanceados.

Existen muchos métodos para analizar Diseños Factoriales Desbalanceados; que se basan según el caso en el número de réplicas. A continuación se plantean algunos de ellos, para el Modelo de efectos fijos:

Sea:

$n_{ij}$ : El número de observaciones en la $ij$-ésima celda.

$n_{i.}$ : El número de observaciones del $i$-ésimo renglón (el $i$-ésimo nivel del factor A), definido por: $\sum_{j=1}^{b} n_{ij}$

$n_{.j}$ : El número de observaciones de la $j$-ésima columna (el $j$-ésimo nivel del factor B), definido por: $\sum_{i=1}^{a} n_{ij}$

$n_{..}$ : El número total de observaciones, definido por: $\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}$

**Datos Proporcionales: Caso Simple.**

Este método se utiliza cuando se da el caso que los datos son proporcionales; es decir, cuando se cumple que el número de observaciones en la ij-ésima celda es: $n_{ij} = \dfrac{n_{i.}\, n_{.j}}{n_{..}}$.

Esta condición significa que el número de observaciones en cualquier par de renglones y columnas son proporcionales.

Este caso presenta poca dificultad para su análisis; ya que se puede emplear el Análisis de Varianza para diseños balanceados, solamente haciendo pequeñas modificaciones en las fórmulas para calcular las Sumas de Cuadrados, por ejemplo si se esta estudiando un Diseño Desbalanceado de dos factores; las Sumas de Cuadrados que se utilizan están definidas como se presentan a continuación:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} y_{ijk}^2 - \frac{y_{...}^2}{n_{..}} \qquad SS_A = \sum_{i=1}^{a}\frac{y_{i..}^2}{n_{i.}} - \frac{y_{...}^2}{n_{..}} \qquad SS_B = \sum_{j=1}^{b}\frac{y_{.j.}^2}{n_{.j}} - \frac{y_{...}^2}{n_{..}}$$

$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n_{ij}} - \frac{y_{...}^2}{n_{..}} - SS_A - SS_B$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n_{ij}}$$

### 5.1 MÉTODOS PARA TRANSFORMAR UN CASO DESBALANCEADO A BALANCEADO

**MÉTODOS APROXIMADOS.**

Cuando un diseño es Desbalanceado, es necesario encontrar una forma de volverlo Balanceado, para ello existen métodos aproximados para tal objetivo. Por supuesto estos métodos hacen que el análisis sea aproximado. En la práctica hay que decidir cuándo los datos no están muy alejados del caso desbalanceado, para asegurar que el grado de la aproximación introducido sea poco importante. Se debe suponer que cada celda tiene al menos una observación; es decir, $n_{ij} \geq 1$. A continuación, se presentan algunos de estos métodos.

**a) Estimación de las Observaciones Faltantes**

Si unas cuantas $n_{ij}$ son diferentes, resulta razonable utilizar un procedimiento para estimar los valores faltantes; es decir, que uno o pocos valores falten en cada celda; para que todas tengan el mismo número de réplicas.

Para un Modelo con interacción, la estimación del valor faltante de la ij-ésima celda que minimiza la suma de cuadrados del error es $\bar{y}_{ij}$; es decir, que el valor faltante se estima con el promedio de las observaciones de los datos disponibles en la celda de donde falta la observación y este valor estimado se toma como si fuera un dato observado.

El Análisis de Varianza se lleva a cabo de forma normal, con la modificación que se reducen los grados de libertad tantas veces como el número de observaciones faltantes que han sido estimadas.

**b) Eliminación de Datos**

Este método se utiliza cuando se da el caso que a la mayoría de celdas le faltan en relación a otras celdas una o varias observaciones; por lo tanto, resulta razonable eliminar un valor de la celda o celdas en lugar de estimarlo, que hacen que el diseño sea

Desbalanceado. Se deben elegir al azar la (las) observación(es) que se va o van a eliminar de todas las de la celda.

La observación eliminada puede no ser descartada e incorporarse al diseño nuevamente, quitar otra observación y repetir el análisis, se espera que estos dos análisis, reflejen interpretaciones similares, si no se da se puede sospechar que la observación eliminada corresponde a un residuo alejado, o a un dato con un serio error al recopilarse y debe ser tratado de acuerdo con estas circunstancias. En la realidad puede ser que no ocurra este problema, cuando se eliminan pocos datos y la variabilidad dentro de las celdas es pequeña.

**c) Método de los Promedios No Ponderados**

Método introducido por Yates (1934), en el cual los promedios de las celas se consideran como si fueran los datos y se les somete a un diseño estándar de un Diseño Balanceado, para obtener las sumas de cuadrados de los factores y sus interacciones. Este método se utiliza cuando los $n_{ij}$ no son muy diferentes. Sin embargo, la media de cuadrados del error se determina por :

$$MS_E = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij.})^2}{n_{..} - ab}$$

Este valor se utiliza para estimar $\sigma^2$, que es la varianza de las observaciones individuales de $y_{ijk}$

Ya que el análisis se ha realizado sobre los promedios de celda y la varianza del promedio de la ij-ésima celda es $\frac{\sigma^2}{n_{ij}}$, la media de cuadrados del error utilizada en el Análisis de Varianza debe estimar la varianza del promedio de $\bar{y}_{ij}$ ; es decir:

$$\bar{V}(\bar{y}_{ij.}) = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{\sigma^2}{n_{ij}}}{ab} = \frac{\sigma^2}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}$$

Al utilizar $MS_E$ como una estimación de $\sigma^2$ se tiene: $MS_E' = \frac{MS_E}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}$, que será la media de cuadrados del error (con n-ab grados de libertad) que debe utilizarse en el Análisis de Varianza.

**d) Método Exacto**

Si los métodos anteriores no son apropiados en un Diseño Factorial Desbalanceado, entonces existe el método exacto para resolver este problema, más que todo en el caso que hay celdas vacías (algunas nij = 0), cuando las nij son diferentes.

Este método no nos da resultados confiables ya que por la naturaleza del método las hipótesis que se prueban no siempre son análogas a las que corresponden al caso balanceado, y los resultados no siempre pueden interpretarse fácilmente.
(Para mayor información del método ver Douglas Mongomery, año 1991, Página 222).

En general, no es recomendable diseñar experimentos de Diseños Factoriales Desbalanceados; ya que el Análisis de Varianza resulta más complicado y no existen métodos adecuados confiables para volverlos balanceados y por lo tanto, el análisis y conclusiones no son mas que aproximaciones del Diseño Factorial Real.

## 6. UNA OBSERVACIÓN POR CELDA

En algunos estudios de investigación pueden haber ocasiones en las que se encuentran experimentos factoriales en la que sólo se dispone de una réplica o sea una observación por celda. Por ejemplo, si se estudian dos factores y una observación por celda, el Modelo Estadístico Lineal es el siguiente:

$$y_{ij} = \mu + \iota_i + \beta_j + (\iota\beta)_{ij} + \varepsilon_{ij} \quad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

Al observar los valores esperados de las medias de cuadrados se dice que la varianza del error experimental $\sigma^2$ no puede estimarse con sólo una réplica de las combinaciones de tratamiento; es decir, el efecto de interacción de los dos factores $(\iota\beta)_{ij} = 0$ y el error experimental no pueden separarse en forma clara, porque las particiones de las sumas de cuadrados para los efectos principales e interacciones de los factores son iguales a la suma de cuadrados total para las observaciones .

Lo que significa en el caso de dos factores tener: $SS_T = SS_A + SS_B + SS_{AB}$

Por lo tanto, $SS_E = 0$, ya que la suma de cuadrados de los subtotales son iguales a la suma de cuadrados totales.

En consecuencia, no existen prueba de hipótesis para los efectos principales, a no ser que el efecto de la interacción sea igual a cero. Si no hay interacción, entonces $(\iota\beta)_{ij} = 0$ para todo i y j; y por lo tanto, un posible Modelo en este caso será:

$$y_{ij} = \mu + \iota_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

Si no existe interacción o este Modelo es el apropiado entonces la media de cuadrados residuales esta determinada por:

$SS_{residual} = SS_T - SS_A - SS_B$, es un estimador insesgado de $\sigma^2$ y las hipótesis de los efectos principales pueden probarse mediante la comparación de $MS_A$ y $MS_B$ contra $MS_{residual}$ ; es decir,

$$F_0 = \frac{MS_A}{MS_{residual}} \quad \text{y} \quad F_0 = \frac{MS_B}{MS_{residual}}$$

Por lo tanto, la tabla de Análisis de Varianza para un Diseño Factorial de dos Factores, cuando se haga la suposición adicional de ignorar el efecto de interacción es:

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Renglón (A) | $SS_A = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$ | $a - 1$ | $MS_A = \frac{SS_A}{a-1}$ | $\frac{MS_A}{MS_{residual}}$ |
| Columna (B) | $SS_B = \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$ | $b - 1$ | $MS_B = \frac{SS_B}{b-1}$ | $\frac{MS_B}{MS_{residual}}$ |
| Residual | $SS_{residual} = SS_T - SS_A - SS_B$ | $(a-1)(b-1)$ | $MS_{residual} = \frac{SS_{residual}}{(a-1)(b-1)}$ | |
| Total | $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab}$ | $ab - 1$ | | |

Como la interacción de los factores principales o la ausencia de interacción no esta garantizada, es necesario utilizar algunas pruebas para evaluar la presencia o ausencia de interacción. A continuación se presenta una prueba elaborada por Tukey que ayudará a detectar la existencia de la interacción.

Primeramente se supone que la forma de la interacción es $(\iota\beta)_{ij} = \varphi\, \iota_i\, \beta_j$, en donde $\varphi$ es una constante desconocida. Si la interacción se define de esta forma, puede usarse el método de regresión para probar la significancia de este término. La prueba consiste en descomponer la suma de cuadrados residuales en un componente de un sólo grado de libertad debido a la no aditividad del Modelo (interacción) y en un componente para el error con $(a-1)-(b-1)-1$ grados de libertad.

Matemáticamente se tiene : $SS_N = \dfrac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}\, y_{i.}\, y_{.j} - y_{..}(SS_A + SS_B + \frac{y_{..}^2}{ab})\right]^2}{abSS_A SS_B}$, con un grado de libertad, y la suma de cuadrados del error es igual: $SS_E = SS_{residual} - SS_N$, con $(a-1)-(b-1)-1$ grados de libertad.

Se debe calcular el $F_0$ para probar la presencia de la interacción el cual se obtiene asi:

$$F_0 = \frac{SS_N}{\dfrac{SS_E}{(a-1)(b-1)-1}}$$

Por lo tanto, si $F_0 > F_{\alpha,1,(a-1)(b-1)-1}$, la hipótesis nula debe rechazarse; es decir, no existe interacción.

La tabla de Análisis de Varianza para un Diseño Factorial de dos factores cuando se haga la suposición adicional de ignorar el efecto de interacción es:

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Renglón (A) | $SS_A = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$ | $a - 1$ | $MS_A = \frac{SS_A}{a-1}$ | $\frac{MS_A}{MS_E}$ |
| Columna (B) | $SS_B = \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$ | $b - 1$ | $MS_B = \frac{SS_B}{b-1}$ | $\frac{MS_B}{MS_E}$ |
| No Aditividad | $SS_N$ | 1 | $MS_{No\ Aditividad} = \frac{SS_N}{1}$ | $\frac{MS_{No\ Aditividad}}{MS_E}$ |
| Error | $SS_E = SS_{residual} - SS_N$ | $(a-1)-(b-1)-1$ | $MS_{residual} = \frac{SS_E}{(a-1)-(b-1)-1}$ | |
| Total | $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab}$ | $ab - 1$ | | |

## 7. SELECCIÓN DEL TAMAÑO MUESTRAL

En los Diseños Factoriales también es importante determinar el tamaño apropiado de la muestra; es decir, el número de réplicas ($n$) que se van a llevar a cabo para obtener la sensibilidad deseada en el experimento. Para determinar ($n$) es posible usar las curvas características de Operación, vistas en detalle en el Diseño Unifactorial.

A continuación se presentan los parámetros $\varphi^2$; así como los valores apropiados de los grados de libertad de numerador y denominador que deben ser utilizados para los diversos Modelos de un Diseño Factorial de dos Factores.

| **Modelo Bifactorial de Efectos Fijos** | | | |
|---|---|---|---|
| Factor | $\varphi^2$ | Grados de libertad del numerador | Grados de libertad del denominador |
| A | $\dfrac{bn\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$ | $a$-1 | $ab(n$-1) |
| B | $\dfrac{an\sum_{j=1}^{b}\beta_j^2}{b\sigma^2}$ | $b$-1 | $ab(n$-1) |
| AB | $\dfrac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{\sigma^2\left[(a-1)(b-1)+1\right]}$ | ($a$-1)($b$-1) | $ab(n$-1) |
| **Modelo Bifactorial de Efectos Aleatorios** | | | |
| Factor | $\lambda$ | Grados de libertad del Numerador | Grados de libertad del Denominador |
| A | $\sqrt{1+\dfrac{bn\sigma_\tau^2}{\sigma^2+n\sigma_{\tau\beta}^2}}$ | $a$-1 | ($a$-1)($b$-1) |
| B | $\sqrt{1+\dfrac{an\sigma_\beta^2}{\sigma^2+n\sigma_{\tau\beta}^2}}$ | $b$-1 | ($a$-1)($b$-1) |
| AB | $\sqrt{1+\dfrac{n\sigma_{\tau\beta}^2}{\sigma^2}}$ | ($a$-1)($b$-1) | $ab(n$-1) |

| Modelo Bifactorial de Efectos Mixtos | | | |
|---|---|---|---|
| Factor | Parámetro | Grados de libertad del Numerador | Grados de libertad del Denominador |
| A(Fijo) | $\varphi^2 = \dfrac{bn\sum_{i=1}^{a}\tau_i^2}{a(\sigma^2 + n\sigma_{\tau\beta}^2)}$ | $a$-1 | $(a$-1$)(b$-1$)$ |
| B(Aleatorio) | $\lambda = \sqrt{1+\dfrac{an\sigma_\beta^2}{\sigma^2}}$ | $b$-1 | $ab(n$-1$)$ |
| AB | $\lambda = \sqrt{1+\dfrac{n\sigma_{\tau\beta}^2}{\sigma^2}}$ | $(a$-1$)(b$-1$)$ | $ab(n$-1$)$ |

Como puede observarse en las tablas los parámetros son $\varphi^2$ y $\lambda$, dependiendo si el factor es fijo o aleatorio.

En el **Modelo de Efectos Fijos** se debe utilizar los diagramas de las curvas características de operación del Modelo de efectos fijos. Para el **Modelo de Efectos Aleatorios** se debe utilizar los diagramas de las curvas características de operación del Modelo Aleatorio, mientras que para el Modelo Mixto se deben utilizar ambos diagramas dependiendo que parámetro se desea encontrar; a un valor de significancia dado ($\alpha$).

Como el valor de "n" debe ser único e igual para todas las celdas en el experimento, y al experimentador le interesan efectos pequeños en lugar de grandes en cada uno de los factores; por lo tanto, requiere un número de réplicas que le garantice tal efecto.

Para ello el experimentador debe seleccionar en relación a qué factor desea que sea sensible el diseño a diferencias potenciales importantes entre los tratamientos.

**Por ejemplo**, si se desea una sensibilidad para el factor A, tomará $\varphi^2$, utilizara las curvas características y el n que obtenga será el que proporcione el nivel deseado de sensibilidad del experimento en relación del Factor A.

Para el Modelo de Efectos Fijos las Curvas Características se pueden usar de una forma muy eficiente, al determinar el valor mínimo de $\varphi^2$ que corresponde a una diferencia especifica entre dos medias de tratamiento. Por lo tanto, si la diferencia entre dos medias del Factor A es $D$, el valor mínimo de $\varphi^2$ viene dado por : $\varphi^2 = \dfrac{nbD^2}{2a\sigma^2}$

Si la diferencia entre dos medias de columna (Factor B) es $D$, el valor mínimo de $\varphi^2$ será: $\varphi^2 = \dfrac{naD^2}{2b\sigma^2}$

Si el valor mínimo de $\varphi^2$ que corresponde a una diferencia $D$, entre cualquier para de efectos de interacción es:

$$\varphi^2 = \frac{nD^2}{2\sigma^2\left[(a-1)(b-1)+1\right]}$$

**Ejemplo 7**

Retomando el ejemplo 5 del porcentaje de encogimiento de las telas teñidas; y suponiendo que existen 3 tratamientos para los factores A y B.

Si se supone que antes de realizar el experimento se decide que habría que rechazar la hipótesis nula con probabilidad alta, si la diferencia máxima en el nivel medio de porcentaje de encogimiento de la tela, de cualquier par de efectos de interacción (Tipos de Tela y Temperatura) fuera igual a 10; y además se supone que la desviación estándar del porcentaje de encogimiento de las telas teñidas, de acuerdo a experimentos previos es aproximadamente igual a 2. Encontrar el número de réplicas.

**Solución**

**Datos**

$a = 3$ , $b = 3$ , $\sigma = 2$ , $\sigma^2 = 4$ , $D = 10$ , $\alpha = 0.05$

Como se decidió rechazar la hipótesis nula en relación a la interacción, el valor mínimo de $\varphi^2$, será:

$$\varphi^2 = \frac{n(10)^2}{2(4)\left[(4-1)(4-1)+1\right]} = \frac{n(100)}{2(4)(10)} = \frac{n(100)}{80}$$

$\varphi^2 = 1.25n$

Tomando una primera aproximación para el número de réplicas de n = 2, entonces se obtiene:

$(a\text{-}1)(b\text{-}1) = (3–1)(3\text{-}1) = 4$ (grados de libertad del numerador $V_1$)

$N - 5 = ab(n\text{-}1) = 3x3(2\text{-}1)= 9$ (grados de libertad del denominador $V_2$)

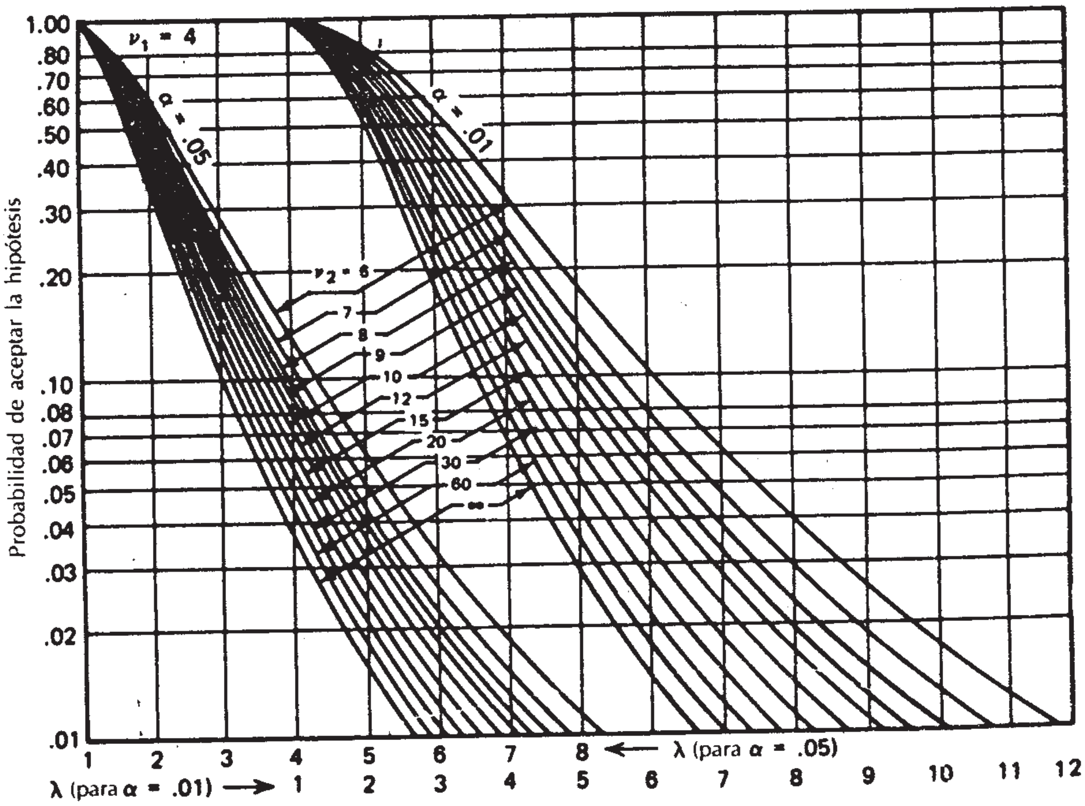
$\varphi^2 = 1.25n = 1.25(2) = 2.50$

$\varphi = 1.58$

Los valores que se van ha utilizar en la gráfica de esta primera aproximación para encontrar la probabilidad de error tipo II ($\beta$) son:

$\alpha = 0.05$ , $V_1 = 2$ , $V_2 = 9$ , $\varphi = 1.58$

La gráfica para éstos valores es la siguiente:




De la gráfica anterior se obtiene que $\beta \approx 0.50$, lo cual se concluye que $n$ = 2 réplicas no son suficientes, porque la potencia de la prueba es aproximadamente $1 - \beta \approx 1 - 0.50 = 0.50$, lo cual resulta pequeña.

Utilizando el procedimiento anterior se obtiene la siguiente tabla.

| $n$ | $\varphi^2$ | $\varphi$ | $V_1 = (a-1)(b-1)$ | $V_2 = ab(n-1)$ | $\beta$ | Poder $(1-\beta)$ |
|---|---|---|---|---|---|---|
| 2 | 2.50 | 1.58 | 4 | 9 | 0.50 | 0.50 |
| 3 | 3.75 | 1.94 | 4 | 18 | 0.12 | 0.88 |
| 4 | 5.0 | 2.24 | 4 | 27 | 0.03 | 0.97 |

De la tabla anterior se observa que $n$ = 4 réplicas producen un nivel del 97% de rechazar la hipótesis nula, si la diferencia en el nivel medio del porcentaje de encogimiento de la tela teñida para cualquier para de efectos de interacción sea a lo sumo igual a 10. Por lo tanto, se concluye que cuatro réplicas son suficientes para proporcionar el nivel deseado de sensibilidad siempre que no exista un error grave en la estimación para la desviación estándar del porcentaje de encogimiento de la tela teñida.

## 8. PROBLEMAS RESUELTOS

**PROBLEMA 1 (Diseño Bifactorial)**

El máximo voltaje de salida de un tipo particular de almacenaje de baterías, se cree que sea influenciado por el material usado en la cubierta de la batería por ciertas variaciones externas de temperatura. Cuatro réplicas de un experimento factorial son corridas en el laboratorio para tres materiales y tres temperaturas; los cuales fueron obtenidos aleatoriamente de un gran número de tipos de materiales y un gran número de temperaturas. Los resultados son mostrados en la siguiente tabla.

| Tipos de Material (Factorial A) | Temperaturas (Factor B) | | | | | |
|---|---|---|---|---|---|---|
| | 50ºF | | 65ºF | | 80ºF | |
| 1 | 130 | 155 | 34 | 40 | 20 | 70 |
| | 74 | 180 | 80 | 75 | 82 | 58 |
| 2 | 150 | 188 | 136 | 122 | 25 | 70 |
| | 159 | 126 | 106 | 115 | 58 | 45 |
| 3 | 138 | 110 | 174 | 120 | 96 | 104 |
| | 168 | 160 | 150 | 139 | 82 | 60 |

Pruebe que la influencia de el material usado y ciertas variaciones externas de la temperatura afectan el máximo voltaje de salida de un tipo particular de almacenaje de baterías.

**Solución**

En este ejemplo el análisis se hará como un Modelo de Efectos Aleatorios; ya que el investigador obtiene los tipos de materiales y las temperaturas de un conjunto de forma aleatoria.

**Variable Respuesta:** Máximo voltaje de salida de las baterías.

**Planteamiento de las Hipótesis a probar:**

a) $H_0 : \sigma_\tau^2 = 0$ (No existe variabilidad en los tipos de materiales).

$H_1 : \sigma_\tau^2 > 0$ (Existe variabilidad en los tipos de materiales).

b) $H_0 : \sigma_\tau^2 = 0$ (No existe variabilidad en los niveles de temperatura).

$H_1 : \sigma_\tau^2 > 0$ (No existe variabilidad en los niveles de temperatura).

c) $H_0 : \sigma^2_\tau = 0$ (No existe variabilidad en la interacción de los tipos de materiales y los niveles de temperatura).

$H_1 : \sigma^2_\tau > 0$ (Existe variabilidad en la interacción de los tipos de materiales y los niveles de temperaturas).

**Forma verbal de las Hipótesis**

a) $H_0$: La variabilidad de los tipos de materiales no influyen significativamente en el máximo voltaje de salida de las baterías.

$H_1$: La variabilidad de los tipos de materiales influyen significativamente en el máximo voltaje de salida de las baterías.

b) $H_0$: La variabilidad de los niveles de temperatura no influyen significativamente en el máximo voltaje de salida de las baterías.

$H_1$: La variabilidad en los niveles de temperatura influyen significativamente en el máximo voltaje de salida de las baterías.

c) $H_0$: La variabilidad de los tipos de materiales y los niveles de temperatura no influyen significativamente en el máximo voltaje de salida de las baterías.

$H_1$: La variabilidad de los tipos de materiales y niveles de temperatura influyen significativamente en el máximo voltaje de salida de las baterías.

**Datos.**

$a$ = 3 , $b$ = 3 , $n$ = 4 , $N = abn$ = 3x3x4 = 36

**Cálculos Matemáticos.**

**Totales de Celdas**

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk}$$

$y_{11.}$ = 539 , $y_{12.}$ = 229 , $y_{13.}$ = 230

$y_{21.}$ = 623 , $y_{22.}$ = 479 , $y_{23.}$ = 198

$y_{31.}$ = 576 , $y_{32.}$ = 583 , $y_{33.}$ = 342

A continuación se presenta la tabla de datos con sus respectivos totales por celda que se calcularon anteriormente:

| Tipos de Material (Factorial A) | Temperaturas (Factor B) | | | $y_{i..}$ |
|---|---|---|---|---|
| | 50ºF | 65ºF | 80ºF | |
| 1 | 539 | 229 | 230 | 998 |
| 2 | 623 | 479 | 198 | 1300 |
| 3 | 576 | 583 | 342 | 1501 |
| $y_{.j.}$ | 1738 | 1291 | 770 | $y_{...}$= 3799 |

**Sumas de Cuadrados**

$$SS_T = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{4} y_{ijk}^2 - \frac{y_{...}^2}{(3)(3)(4)}$$

$$= [(130)^2 + (155)^2 + (74)^2 + (180)^2 + (150)^2 + \ldots + (82)^2 + (60)^2] - \frac{(3799)^2}{36}$$

$$= 478547 - 400900.02$$

$$SS_T = 77646.96$$

**Sumas de Cuadrados para los Efectos Principales.**

$$SS_{\text{Tipo de Material}} = \sum_{i=1}^{3} \frac{y_{i..}^2}{(3)(4)} - \frac{(3799)^2}{36}$$

$$= \frac{\left[(998)^2 + (1300)^2 + (1501)^2\right]}{12} - \frac{(3799)^2}{36}$$

$$= 411583.75 - 400900.02$$

$$SS_{\text{Tipo de Material}} = 10683.73$$

$$SS_{\text{Temperatura}} = \sum_{j=1}^{3} \frac{y_{.j.}^2}{(3)(4)} - \frac{(3799)^2}{36}$$

$$= \frac{\left[(1738)^2 + (1291)^2 + (770)^2\right]}{12} - \frac{(3799)^2}{36}$$

$$= 440018.75 - 400900.02$$

$$SS_{\text{Temperatura}} = 39118.73$$

**Cálculo de subtotales**

$$SS_{Subtotales} = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{y_{ij.}^2}{4} - \frac{(3799)^2}{36}$$

$$= \frac{\left[(539)^2 + (229)^2 + (230)^2 + \ldots\ldots + (342)^2\right]}{4} - \frac{(3799)^2}{36}$$

= 460316.25 – 400900.02

$SS_{Subtotales}$ = 59416.23

**Cálculo de Interacción**

$SS_{Interacción} = SS_{Subtotales} - SS_{Tipo\ de\ Material} - SS_{Temperatura}$

= 59416.23 – 10683.73 – 39118.72

$SS_{Interacción}$ = 9613.78

**Cálculo de la suma de cuadrados del error**

$SS_E = SS_T - SS_{Subtotales}$

= 77646.98 – 59416.23

$SS_E$ = 18230.75

**Cálculo de las Medias de Cuadrados.**

$$MS_A = \frac{SS_A}{a-1} = \frac{10683.73}{3-1} = \frac{10683.73}{2} = 5341.86$$

$$MS_B = \frac{SS_B}{b-1} = \frac{39118.72}{3-1} = \frac{39118.72}{2} = 19559.36$$

$$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} = \frac{9613.78}{(3-1)(3-1)} = \frac{9613.78}{4} = 2403.44$$

$$MS_E = \frac{SS_E}{ab(n-1)} = \frac{18230.75}{3x3(4-1)} = \frac{18230.75}{27} = 675.21$$

**Cálculo de Estadísticos**

$F_0 = \frac{MS_A}{MS_{AB}} = \frac{5341.86}{2403.44}$ = 2.22 (**Tipo de Material**)

$F_0 = \frac{MS_B}{MS_{AB}} = \frac{19559.36}{2403.44}$ = 18.14 (**Temperatura**)

$F_0 = \frac{MS_{AB}}{MS_E} = \frac{2403.44}{675.21}$ = 3.56 (**Tipo de Material y Temperatura**)

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Medias de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Tipo de Material | 10683.72 | 2 | 5341.86 | 2.22 |
| Temperatura | 39118.72 | 2 | 19559.36 | 8.14 |
| Interacción | 9613.77 | 4 | 2403.44 | 3.56 |
| Error | 18230.75 | 27 | 675.21 | |
| Total | 77646.96 | 35 | | |

Tomando $\alpha$ = 0.05, encontrando para cada hipótesis a probar sus respectivos $F_{Tablas}$, se tiene:

a) $F_{\alpha,a-1,(a-1)(b-1)} = F_{0.05,3-1,(3-1)(3-1)} = F_{0.05,2,4}$ = 6.94

b) $F_{\alpha,b-1,(a-1)(b-1)} = F_{0.05,3-1(3-1)(3-1)} = F_{0.05,2,4}$ = 6.94

c) $F_{\alpha,(a-1)(b-1),ab(n-1)} = F_{0.05,(3-1)(3-1),3*3(4-1)} = F_{0.05,4,27}$ = 2.73

- **Conclusiones**

Se tiene:

**<u>Respecto a la hipótesis 1 (Factor A(Tipo de Material)</u>**

Se observa de la tabla de Análisis de Varianza que el $F_0 < F_{Tablas}$ (2.22 < 6.94); por lo tanto, se acepta $H_0$; es decir, que la variabilidad de los tipos de materiales no influyen significativamente en el máximo voltaje de salida de las baterías.

**<u>Respecto a la hipótesis 2 (Factor B(Temperatura))</u>**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (8.14 > 6.94); por lo tanto, se rechaza $H_0$; es decir, que la variabilidad de los niveles de temperatura influyen significativamente en el máximo voltaje de salida de las baterías.

**<u>Respecto a la hipótesis 3 (Interacción(Tipo de Material y Temperatura))</u>**

Se observa de la tabla de Análisis de Varianza que el $F_0 > F_{Tablas}$ (3.56 > 2.73 ); por lo tanto, se rechaza $H_0$; es decir, la variabilidad de los tipos de materiales y niveles de temperatura influyen significativamente en el máximo voltaje de salida de las baterías.

**PROBLEMA 2**

Para el problema 1, encontrar la estimación de los componentes de varianza.

**Datos**

$a$ = 3 , $b$ = 3 , $n$ = 4 , $N$ = 36

$MS_E$ = 675.21 , $MS_A$ = 5341.86 , $MS_B$ = 19559.36 , $MS_{AB}$ = 2403.44

**Solución**

Como se trata de un Modelo de Efectos Aleatorios las estimación que se deben de encontrar son las siguientes:

$$\hat{\sigma}^2 = MS_E = 675.21$$

$$\hat{\sigma}_\tau^2 = \frac{MS_A - MS_{AB}}{bn} = \frac{5341.86 - 2403.44}{12} = \frac{2938.42}{12} = 244.87$$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - MS_{AB}}{an} = \frac{19559.36 - 2403.44}{12} = \frac{17155.92}{12} = 1429.66$$

$$\hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n} = \frac{2403.44 - 675.21}{4} = \frac{1728.23}{4} = 432.06$$

**PROBLEMA 3**

Para el problema 1, encontrar un intervalo de confianza para $\sigma^2$ con un nivel de confianza del 95%.

**Datos**

$a$ = 3 , $b$ = 3 , $n$ = 4 , $N$ = 36 , $\alpha$ = 0.05 , $MS_E$ = 675.21

**Solución**

**Calculando los valores de chi-cuadrado.**

Como el valor de v = 27 (grados de libertad) no se encuentra en la tabla se tomaran los valores de v = 25 y v = 30 para ambos casos y se obtendrá el promedio.

| | |
|---|---|
| $\chi^2_{0.025,25}$ = 40.65 | $\chi^2_{0.975,25}$ = 13.12 |
| $\chi^2_{0.025,30}$ = 46.98 | $\chi^2_{0.975,30}$ = 16.79 |
| El promedio es: 43.81 | El promedio es: 14.95 |

Por lo tanto:

$$\chi^2_{\frac{\alpha}{2},ab(n-1)} = \chi^2_{\frac{0.05}{2},3x3(4-1)} = \chi^2_{0.025,27} = 43.81$$

$$\chi^2_{1-(\frac{\alpha}{2}),ab(n-1)} = \chi^2_{1-(\frac{0.05}{2}),3x3(4-1)} = \chi^2_{0.975,27} = 14.95$$

**Intervalo de confianza para $\sigma^2$.**

$$\frac{ab(n-1)MS_E}{\chi^2_{\frac{\alpha}{2},ab(n-1)}} \leq \sigma^2 \leq \frac{ab(n-1)MS_E}{\chi^2_{1-\frac{\alpha}{2},ab(n-1)}}$$

**Sustituyendo**

$$\frac{3x3(4-1)\,(675.21)}{\chi^2_{\frac{0.05}{2},3x3(4-1)}} \leq \sigma^2 \leq \frac{3x3(4-1)(675.21)}{\chi^2_{1-(\frac{0.05}{2}),3x3(4-1)}}$$

$$\frac{3x3x3(675.21)}{\chi^2_{0.025,27}} \leq \sigma^2 \leq \frac{3x3x3(675.21)}{\chi^2_{0..975,27}}$$

$$\frac{(27)(675.21)}{\chi^2_{0.025,27}} \leq \sigma^2 \leq \frac{(27)(675.21)}{\chi^2_{0..975,27}}$$

$$\frac{18230.67}{43.81} \leq \sigma^2 \leq \frac{18230.67}{14.95}$$

$$416.13 \leq \sigma^2 \leq 1219.44$$

Significa que la variabilidad del efecto total de los tipos de materiales, temperatura e interacción se encuentra entre [416.13 , 1219.44].

**PROBLEMA 4**

Encontrar las Medias de Cuadrados Esperados para un Modelo de Efectos Aleatorios de tres factores utilizando las Reglas estudiadas en el tema 4.2.

**Solución.**

En el modelo de Efectos Aleatorios los tres factores son aleatorios; es decir el factor A, B y C son aleatorios.

Aplicando las Reglas se obtiene la siguiente tabla:

| Factor | Factor A | Factor B | Factor C | |
|---|---|---|---|---|
| | A<br>$a$<br>$i$ | A<br>$b$<br>$j$ | A<br>$c$<br>$k$ | R<br>$n$<br>$l$ |
| $\iota_i$ | 1 | $b$ | $c$ | $n$ |
| $\beta_j$ | $a$ | 1 | $c$ | $n$ |
| $\gamma_k$ | $a$ | $b$ | 1 | $n$ |
| $(\iota\beta)_{ij}$ | 1 | 1 | $c$ | $n$ |
| $(\iota\gamma)_{ik}$ | 1 | $b$ | 1 | $n$ |
| $(\beta\gamma)_{jk}$ | $a$ | 1 | 1 | $n$ |
| $(\iota\beta\gamma)_{ijk}$ | 1 | 1 | 1 | $n$ |
| $\varepsilon_{(ijk)l}$ | 1 | 1 | 1 | 1 |

Las aplicaciones de los literales de la regla dos, para el llenado de la tabla anterior se detalla a continuación:

a) Son los unos de las primeras tres columnas de la última fila, ya que los subíndices $i,j$ y $k$; en esa fila son pasivos.

b) Son los unos en todas las filas menos los unos descritos en el numeral anterior.

c) Los niveles de cada factor ($abcn$).

d) Para encontrar el valor esperados de la media de cuadrados del factor A, se debe cubrir la columna donde aparece $i$, luego multiplicar los valores de las columnas por el componente de varianza o efecto, según la regla uno; esto se debe hacer para cada fila donde aparezca un subíndice "$i$" como activo y luego sumarse; agregándole además el $\sigma^2$ que es el componente de varianza de los errores.

De esta forma se obtienen los valores esperados, teniendo cuidado de cubrir la columna o número de columnas de los valores esperados que se desean encontrar.
Por lo tanto, de esta tabla se obtiene los siguientes resultados.

E($MS_A$) $= \sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma} + cn\sigma^2_{\tau\beta} + bcn\sigma^2_{\tau}$
E($MS_B$) $= \sigma^2 + cn\sigma^2_{\tau\beta} + an\sigma^2_{\beta\gamma} + n\sigma^2_{\tau\beta\gamma} + acn\sigma^2_{\beta}$
E($MS_C$) $= \sigma^2 + bn\sigma^2_{\tau\gamma} + an\sigma^2_{\beta\gamma} + n\sigma^2_{\tau\beta\gamma} + abn\sigma^2_{\gamma}$
E($MS_{AB}$) $= \sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta}$
E($MS_{AC}$) $= \sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma}$
E($MS_{BC}$) $= \sigma^2 + n\sigma^2_{\tau\beta\gamma} + an\sigma^2_{\beta\gamma}$
E($MS_{ABC}$) $= \sigma^2 + n\sigma^2_{\tau\beta\gamma}$
E($MS_E$) $= \sigma^2$

**PROBLEMA 5**

Utilizando las Reglas para encontrar las Sumas de Cuadrados estudiadas en el tema 4.1. Encontrar la fórmula de la Suma de Cuadrados para $(\iota\beta\gamma)_{ijk}$; es decir la fórmula para encontrar la Suma de Cuadrados de $SS_{ABC}$.

**Solución**

Los grados de libertad son $(a\text{-}1)(b\text{-}1)(c\text{-}1) = abc - ac - bc + c - ab + a + b - 1$

Por lo tanto, la Suma de Cuadrados, se obtiene de la siguiente forma.

| $Abc$ | $-ac$ | $-bc$ | $+c$ | $-ab$ | $a$ | $b$ | -1 |
|---|---|---|---|---|---|---|---|
| $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}$ | $\frac{y_{....}^2}{abcn}$ (a) |
| $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\sum_{i=1}^{a}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{k=1}^{c}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{i=1}^{a}\left(\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\sum_{j=1}^{b}\left(\sum_{i=1}^{s}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}\right)$ | $\frac{y_{....}^2}{abcn}$ (b) |
| $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(y_{ijk.}\right)$ | $\sum_{i=1}^{a}\sum_{k=1}^{c}\left(y_{i.k.}\right)$ | $\sum_{j=1}^{b}\sum_{k=1}^{c}\left(y_{.jk.}\right)$ | $\sum_{k=1}^{c}\left(y_{..k.}\right)$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij..}\right)$ | $\sum_{i=1}^{a}\left(y_{i...}\right)$ | $\sum_{j=1}^{b}\left(y_{.j..}\right)$ | $\frac{y_{....}^2}{abcn}$ (c) |
| $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y_{yijk}^2}{n}$ | $\sum_{i=1}^{a}\sum_{k=1}^{c}\frac{y_{i.k.}^2}{bn}$ | $\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y_{.jk.}^2}{an}$ | $\sum_{k=1}^{c}\frac{y_{..k.}^2}{abn}$ | $\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij..}^2}{cn}$ | $\sum_{i=1}^{a}\frac{y_{i...}^2}{bcn}$ | $\sum_{j=1}^{b}\frac{y_{.j..}^2}{acn}$ | $\frac{y_{....}^2}{abcn}$ (d) |
| $SS_{ABC} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y_{yijk}^2}{n} - \sum_{i=1}^{a}\sum_{k=1}^{c}\frac{y_{i.k.}^2}{bn} - \sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y_{.jk.}^2}{an} + \sum_{k=1}^{c}\frac{y_{..k.}^2}{abn} - \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij..}^2}{cn} + \sum_{i=1}^{a}\frac{y_{i...}^2}{bcn} + \sum_{j=1}^{b}\frac{y_{.j..}^2}{acn} - \frac{y_{....}^2}{abcn}$ | | | | | | | |

**Notas**

- $SS_{ABC}$ se obtiene de la combinación de las sumas de cuadrados no corregida en el último renglón, de acuerdo con los signos en la parte superior de cada columna.
- Esta igualdad se observa claramente si se sustituyen las fórmulas de las Sumas de Cuadrados en la igualdad que aparece abajo.

Entonces la Suma de Cuadrados es: $SS_{ABC} = SS_{subtotales(ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SS_{BC} - SS_{AC}$

# UNIDAD PROGRAMÁTICA V:

# " DISEÑOS FACTORIALES $2^K$ "

# DISEÑO FACTORIAL $2^k$

## 1. DESCRIPCIÓN

En el área industrial, la experimentación requiere frecuentemente conocer el efecto sobre una variable respuesta de un número grande de factores. En estos casos, el Diseño Factorial puede necesitar un tamaño tan grande de réplicas que lo haga irrealizable. Por ejemplo, un Diseño con seis factores cada uno con cinco niveles, en el cual se deben realizar $5^6 = 3125$ réplicas, para reducir el número de réplicas se requiere o bien eliminar factores, o bien disminuir los niveles a los que se aplican; es decir que el estudio de los Diseños Factoriales se complican cuando existe un número considerable de factores y un número bastante grande de niveles de cada factor.

Sin embargo, los Diseños Factoriales son ampliamente utilizados en Experimentos para estudiar el efecto conjunto de éstos sobre la variable respuesta. Aunque el tamaño del Diseño resultante sea aceptable, si sólo uno o dos factores influyen sobre la variable respuesta, la mayoría de los datos medirán únicamente el error experimental, por lo que el Diseño será poco eficiente. Si se tiene un Diseño de cinco factores con cuatro niveles en donde tres factores no influyen significativamente sobre la variable respuesta; es decir, son inertes, proporcionan casi el mismo resultado si se hiciera el estudio de un Diseño Factorial, para los dos factores que influyen sobre la variable respuesta (activos).

Para minimizar el riesgo de realizar Experimentos con factores inertes, se debe utilizar una estrategia secuencial; donde los factores a considerar en el Experimento en cada etapa se eligen en función de los resultados de las etapas anteriores. La ventaja principal de los Diseños Factoriales, es la posibilidad de determinar si los factores actúan de manera independiente o interaccionan entre ellos para afectar las unidades experimentales.

Existen varios casos especiales del Diseño Factorial general que resultan importantes en la práctica y se usan ampliamente en los estudios de investigación; y además constituyen la base para otros diseños de gran valor práctico.

El primer caso especial más importante de estos ocurre cuando se tienen *k* factores, fijándoles el número de niveles en dos. Estos niveles pueden ser cuantitativos como por ejemplo, dos valores de Temperatura, dos de presión o dos de tiempo, etc; también pueden ser cualitativos como dos máquinas, dos operadores, etc. Estos dos niveles podrían ser considerados como "superior" e "inferior", "alto" o "bajo", "ausencia" o "presencia" del factor.

Por lo tanto, en un Experimento en el que intervienen *k* factores y cada factor contiene dos niveles se conoce con el nombre de Experimento Factorial $2^k$.

En una réplica completa de tal Diseño se necesita que se recopilen $2x2x…x2 = 2^k$ observaciones.

Cuando es probable que hayan muchos factores por investigar, son útiles los Diseños $2^k$ en las primeras fases del trabajo experimental; ya que se hace menor número de corridas con los cuales pueden estudiarse *k* factores en un Diseño Factorial Completo, debido a que sólo hay dos niveles para cada factor.

El segundo caso especial es el Diseño Factorial con *k* factores con tres niveles cada uno; el cual se conoce como **Diseño Factorial $3^k$**.

A continuación se presentan métodos especiales para el análisis de los tipos de Diseños Factoriales $2^2$, $2^3$ y una generalización de estos ($2^k$), suponiendo que los Factores son Fijos, los Diseños son Completamente Aleatorizados, y que satisfacen la suposición usual de normalidad.

## 2. DISEÑO FACTORIAL $2^2$.

Este Diseño Factorial es el primero de la serie $2^k$, en el cual sólo existen dos factores en estudio, los cuales pueden ser A y B, cada uno con dos niveles; que se suelen considerar como "alto" y "bajo". El nivel alto del Factor A y Factor B se denota por el signo mas "+" y el nivel bajo con el signo menos "-".

En este Diseño habrán cuatro combinaciones de tratamiento que son: bajoA-bajoB, altoA-bajoB, bajoA-altoB y altoA-altoB. En general una combinación de tratamiento se representa por una serie de letras minúsculas, de la siguiente forma:

$a$ : alto A-bajo B

$b$ : bajo A-alto B

$ab$ : alto A-alto B

(1) : bajo A-bajo B

**Las cuales verbalmente significan:**

$a$ : Representa la combinación de tratamiento, en la que el factor A se encuentra en el nivel alto y el factor B en el nivel bajo.

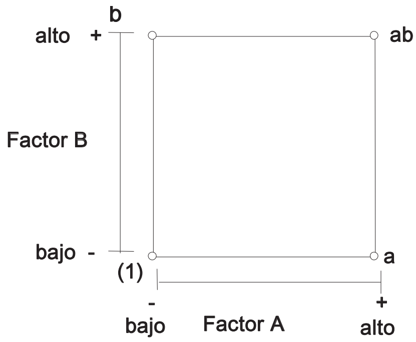
$b$ : Representa la combinación de tratamiento, en la que el factor A se encuentra en el nivel bajo y el factor B en el nivel alto.

$ab$ : Representa la combinación de tratamiento, en la que el factor A se encuentra en el nivel alto y el factor B en el nivel alto.

(1) : Representa la combinación de tratamiento, en la que ambos factores (A y B) se encuentra en el nivel bajo.

Por lo tanto, el nivel "alto" de cualquier factor de una combinación de tratamiento está representado por la presencia de la letra minúscula correspondiente; mientras que la ausencia de está representa el nivel "bajo" del factor. Por ejemplo, si hay cuatro factores (A,B,C y D) la combinación de tratamiento $cd$, correspondería a los niveles altos de los factores C y D , bajos de los factores A y B.

Las letras minúsculas *(1)*, $a$, $b$ y $ab$ también se usaran para representar los totales de las *n* réplicas de las combinaciones de tratamientos correspondientes.
Las combinaciones de tratamiento se representan geométricamente de la siguiente forma:




El uso de las palabras alto y bajo, es arbitrario y no tiene un significado particular en el análisis. Tratándose de un factor cuantitativo, un nivel tendrá un valor más alto y el otro un valor más bajo; coincidiendo con la simbología adoptada, pero si se trata de un factor cualitativo no habrá esta ordenación entre los niveles de los factores, y podría señalarse arbitrariamente a uno como "alto" y al otro como "bajo".

Las letras mayúsculas A,B y AB se utilizaran para denotar el efecto de un factor; por lo tanto:

"A" : Se refiere al efecto del factor A.
"B" : Se refiere al efecto del factor B.
"AB" : Se refiere al efecto de la interacción del factor A y el factor B.

## 2.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

Como en los Diseños Factoriales $2^2$, el número de factores son dos con dos niveles cada uno y en cada nivel existen "$n$" réplicas del Experimento; entonces la representación de los datos observados para esté tipo de Diseño Factorial será de la siguiente forma:

| Factor A | Factor B | |
|---|---|---|
| | 1 | 2 |
| 1 | $Y_{111}, y_{112}, \ldots, y_{11n}$ | $Y_{121}, y_{122}, \ldots, y_{12n}$ |
| 2 | $Y_{211}, y_{212}, \ldots, y_{21n}$ | $Y_{221}, y_{222}, \ldots, y_{22n}$ |

Como puede observarse esta tabla es sólo una parte de la Tabla de la Representación Simbólica de un Diseño de dos factores con "$a$" niveles del factor A y "$b$" del factor B con "$n$" réplicas en cada nivel.

Cualquier observación de tabla se interpreta de la misma forma que el Diseño de dos factores.

El número total de observaciones en el Experimento es de $N$ = 2x2x$n$ = 4$n$; ya que se realizan "$n$" réplicas.

**Nota:** Lo que se refiere al Modelo Estadístico que representan los datos descritos anteriormente será igual al del Diseño de dos factores, con la diferencia de que los subíndices de $i$ y $j$ tomaran valores de 1 hasta 2.

## 2.2 ESTIMACIÓN DE LOS EFECTOS DE LOS FACTORES

En general, el efecto de un factor vendrá dado por el promedio del efecto del mismo factor en los niveles bajo y alto del otro factor, siendo "$n$" el número de réplicas.

Para encontrar los efectos de los factores y de la interacción, existen dos métodos en los cuales se utiliza la representación geométrica, los cuales se explicaran a continuación:

**<u>Primer Método.</u>**

**Efecto promedio del factor A**

De la representación geométrica de las combinaciones de tratamientos; tomando de derecha a izquierda se observa que:

El efecto del Factor A en el nivel bajo del Factor B es: $\frac{(a-(1))}{n}$

El efecto del Factor A en el nivel alto del Factor B es: $\frac{(ab-b)}{n}$

**Por lo tanto:**

$$A = \frac{1}{2}\left[\frac{(a-(1))}{n} + \frac{(ab-b)}{n}\right]$$

$$= \frac{1}{2n}\left(a-(1)+ab-b\right)$$

$$A = \frac{1}{2n}\left(ab+a-b-(1)\right)$$

**Efecto promedio del factor B**

De la representación geométrica de las combinaciones de tratamientos; tomando de arriba hacia abajo se observa que:

El efecto del Factor B en el nivel bajo del Factor A es: $\frac{(b-(1))}{n}$

El efecto del Factor B en el nivel alto del Factor A es: $\frac{(ab-a)}{n}$

**Por lo tanto:**

$$B = \frac{1}{2}\left[\frac{(b-(1))}{n} + \frac{(ab-a)}{n}\right]$$

$$= \frac{1}{2n}\left(b-(1)+ab-a\right)$$

$$B = \frac{1}{2n}\left(ab+b-a-(1)\right)$$

**Efecto promedio de la interacción AB.**

Se define como la diferencia promedio entre el efecto del factor A en el nivel alto del Factor B, que es $\left[\frac{(ab-b)}{n}\right]$; y su efecto promedio en el nivel bajo del Factor B, que es $\left[\frac{(a-(1))}{n}\right]$.

**Por lo tanto:**

$$AB = \frac{1}{2}\left[\frac{(ab-b)}{n} - \frac{(a-(1))}{n}\right]$$

$$= \frac{1}{2n}\left(ab-b-a+(1)\right)$$

$$AB = \frac{1}{2n}\left(ab+(1)-a-b\right)$$

También se puede definir como la diferencia promedio entre el efecto del Factor B en el nivel alto del Factor A, que es $\left[\frac{(ab-a)}{n}\right]$; y el efecto promedio del Factor B en el nivel bajo del Factor A, que es $\left[\frac{(b-(1))}{n}\right]$.

**Por lo tanto:**

$$AB = \frac{1}{2}\left[\frac{(ab-a)}{n} - \frac{(b-(1))}{n}\right]$$

$$= \frac{1}{2n}\left[ab-a-b+(1)\right]$$

$$AB = \frac{1}{2n}\left(ab+(1)-a-b\right)$$

**<u>Segundo Método.</u>**

Sea:

$\overline{Y}_{F+}$ : La respuesta promedio para las combinaciones de tratamientos a las que el factor (F) se encuentra en el nivel alto.

$\overline{Y}_{F-}$ : La respuesta promedio para las combinaciones de tratamientos a las que el factor (F) se encuentra en el nivel bajo.

Por lo tanto, el efecto del factor será la diferencia en la respuesta promedio de $\overline{Y}_{F+}$ y $\overline{Y}_{F-}$.

**Efecto promedio del factor A.**

$\overline{Y}_{A+}$ : Respuesta promedio para las combinaciones de tratamiento a las que el factor A se encuentra en el nivel alto.

$$\overline{Y}_{A+} = \frac{(ab+a)}{n}$$

$\overline{Y}_{A-}$ : Respuesta promedio para las combinaciones de tratamiento a las que el factor A se encuentra en el nivel bajo.

$$\overline{Y}_{A-} = \frac{(b+(1))}{n}$$

Encontrando el efecto promedio se tiene:

A = $\overline{Y}_{A+}$ - $\overline{Y}_{A-}$

$$= \frac{1}{2}\left[\frac{(ab+a)}{n} - \frac{(b+(1))}{n}\right]$$

$$= \frac{1}{2n}\left[ab+a-b-(1)\right]$$

**Efecto promedio del factor B.**

$\overline{Y}_{B+}$ : Respuesta promedio para las combinaciones de tratamiento a las que el factor B se encuentra en el nivel alto.

$$\overline{Y}_{B+} = \frac{(ab+b)}{n}$$

$\overline{Y}_{B-}$ : Respuesta promedio para las combinaciones de tratamiento a las que el factor B se encuentra en el nivel bajo.

$$\overline{Y}_{B-} = \frac{(a+(1))}{n}$$

Encontrando el efecto promedio se tiene:

B = $\overline{Y}_{B+}$ - $\overline{Y}_{B-}$

$$= \frac{1}{2}\left[\frac{(ab+b)}{n} - \frac{(a+(1))}{n}\right]$$

$$= \frac{1}{2n}\left[ab+b-a-(1)\right]$$

**Efecto promedio de la interacción AB**

Se define como la diferencia del promedio de las combinaciones de tratamientos en la diagonal de derecha a izquierda: $\left[\frac{ab+(1)}{n}\right]$ y el promedio de las combinaciones de tratamientos en la diagonal de izquierda a derecha: $\left[\frac{b+a}{n}\right]$.

**Por lo tanto:**

$$AB = \frac{1}{2}\left[\frac{(ab+(1))}{n} - \frac{(b+a)}{n}\right]$$

$$= \frac{1}{2n}(ab+(1)-a-b)$$

El efecto principal de un factor es el efecto promedio que se produce a consecuencia de cambiarlo del nivel bajo al nivel alto de ese factor.
La interacción entre dos factores se da cuando el efecto de uno de ellos es diferente para los niveles alto y bajo del otro factor.

## 2.3 SUMAS DE CUADRADOS

El Análisis de Varianza en los Diseños $2^k$ pueden utilizarse para examinar la magnitud y la dirección de los efectos de los factores, para determinar cuáles variables es probable que sean importantes.

A continuación, se presenta un método rápido para llevar a cabo los cálculos del Análisis de Varianza.

Si se observa, en los efectos encontrados para A, B y AB anteriormente están involucrados o inmersos sus respectivos contrastes, que son los que se encuentran entre paréntesis en dichos efectos, como se muestra en la siguiente tabla:

| Factor | Efecto | Contraste |
|---|---|---|
| A | $\frac{1}{2n}\left[ab+a-b-(1)\right]$ | $ab+a-b-(1)$ |
| B | $\frac{1}{2n}\left[ab+b-a-(1)\right]$ | $ab+b-a-(1)$ |
| AB | $\frac{1}{2n}\left[ab+(1)-a-b\right]$ | $ab+(1)-a-b$ |

Por lo tanto, de la tabla anterior se tiene que:

$\text{Contraste}_A = ab + a - b - (1)$

$\text{Contraste}_B = ab + b - a - (1)$

$\text{Contraste}_{AB} = ab + (1) - a - b$

A estos contraste se suelen denominar efecto total de A, B y AB respectivamente. Estos tres contrastes son ortogonales. La suma de cuadrados para cada uno (como se estudio en el Unidad II) puede calcularse como el contraste elevado al cuadrado, entre el producto del número de observaciones de cada total del contraste y la suma de los cuadrados de los coeficientes del mismo. Es decir que $SS_C = \dfrac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{n\sum_{i=1}^{a} c_i^2}$

Por lo tanto, las sumas de cuadrados de los efectos principales y la interacción puede obtenerse por el contraste respectivos elevado al cuadrado entre el número total de observaciones; tal como:

$$SS = \frac{(contraste)^2}{4n}.$$

Entonces la suma de cuadrados de A, B y AB es:

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n}$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{4n}$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n}$$ respectivamente.

La suma total de cuadrados y la suma de cuadrados del error se obtiene de forma usual.

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{4n}$$

$SS_E = SS_T - SS_A - SS_B - SS_{AB}$

donde

$SS_T$ : Tiene $4n$-1 grados de libertad; ya que hay un total de $N = 4n$

$SS_E$ : Tiene $4(n$-1) grados de libertad; porque, cada uno de los efectos principales y doble tiene un grado de libertad.

**Ejemplo 1**

Un ingeniero en alimentos esta interesado en determinar los efectos de la Temperatura y el tiempo de almacenamiento en la conservación de las manzanas. La respuesta en este estudio es el número de manzanas que se pudren.

Para realizar este Experimento se escogieron 2 lotes de 120 manzanas cada uno, dividiendo cada lote en cuatro porciones de igual tamaño y asignando al azar los tratamientos a estas porciones. Para realizar este estudio se tomaron 2 niveles de tiempo de almacenamiento de 20 y 25 días (Factor A) y 2 niveles de Temperatura de 13°C y 17°C (Factor B). A continuación se presentan los resultados del número de manzanas podridas de cada 30 manzanas.

| Tiempo (Factor A) | Temperatura (Factor B) | |
|---|---|---|
| | 13°C | 17°C |
| 20 días | 5<br>3 | 12<br>11 |
| 25 días | 9<br>8 | 12<br>14 |

**Solución**

**Variable Respuesta:** Número de manzanas podridas.

**Hipótesis a probar**

**Efectos Principales**

$H_0$ : El tiempo de almacenamiento no influye significativamente en el número de manzanas que se pudren.

$H_1$ : El tiempo de almacenamiento influye significativamente en el número de manzanas que se pudren.

$H_0$ : El nivel de Temperatura no influye significativamente en el número de manzanas que se pudren.

$H_1$ : El nivel de Temperatura influye significativamente en el número de manzanas que se pudren.

**Interacción**

$H_0$ : El tiempo de almacenamiento y el nivel de Temperatura no influyen significativamente en el número de manzanas que se pudren.

$H_1$ : El tiempo de almacenamiento y el nivel de Temperatura influyen significativamente en el número de manzanas que se pudren.

Como puede observarse este Experimento es un Diseño Factorial $2^2$, donde existen dos factores, factor A (Tiempo) y factor B (Temperatura) ambos a dos niveles.

El factor A con nivel bajo que es 20 días y el nivel alto de 25 días.

El factor B con nivel bajo que es 13°C y el nivel alto de 17°C

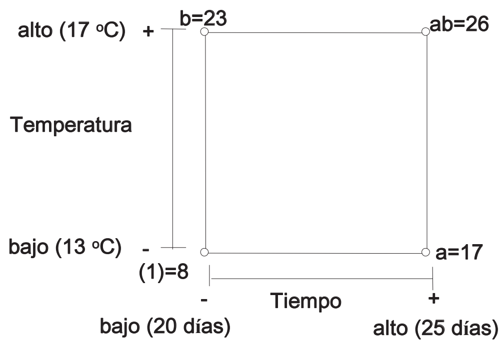
Realizándose este Experimento dos veces.

De la tabla de datos es posible obtener una tabla de combinaciones de tratamientos, como la que se muestra a continuación:

| Combinación de Tratamientos | Réplicas | | Total | Simbología |
|---|---|---|---|---|
| | I | II | | |
| A bajo, B bajo | 5 | 3 | 8 | (1) |
| A alto , B bajo | 9 | 8 | 17 | $a$ |
| A bajo, B alto | 12 | 11 | 23 | $b$ |
| A alto, B alto | 12 | 14 | 26 | $ab$ |

$y_{...} = 74$

Estas combinaciones de tratamientos Geométricamente se representa de la siguiente manera:




$n$ : el número de réplicas ($n$=2)

Utilizando el primer método para la estimación de los efectos promedios de los factores principales e interacción, se tiene:

$$A = \frac{1}{2n}\left[ab + a - b - (1)\right] = \frac{1}{2x2}\left(26 + 17 - 23 - 8\right) = \frac{1}{4}\left(12\right) = 3$$

$$B = \frac{1}{2n}\left[ab + b - a - (1)\right] = \frac{1}{2x2}\left(26 + 23 - 17 - 8\right) = \frac{1}{4}\left(24\right) = 6$$

$$AB = \frac{1}{2n}\left[ab + (1) - a - b\right] = \frac{1}{2x2}\left(26 + 8 - 17 - 23\right) = \frac{1}{4}\left(-6\right) = -1.5$$

**Los valores encontrados significan:**

a) El valor del efecto del factor A (Tiempo) es positivo e igual a 3; que significa que se espera un aumento de 3 manzanas podridas de cada 30 manzanas al aumentar el tiempo de almacenamiento de 20 días a 30 días.

b) El valor del efecto del factor B (Temperatura) es positivo e igual a 6; lo que significa que se espera un aumento de 6 manzanas podridas de cada 30 manzanas al aumentar la Temperatura de 13°C a 17°C; casi sin importar el tiempo de almacenamiento.

c) El valor del efecto de la interacción del Factor A y Factor B es negativo e igual a -1.5; lo que significa que el efecto de ambos factores no afectan en el número de manzanas podridas por cada 30 manzanas.

**A continuación se encuentran los contrastes:**

$Contraste_A$ = $ab + a - b - (1)$ = 26 + 17 - 23 - 8 = 12

$Contraste_B$ = $ab + b - a - (1)$ = 26+23 - 17 - 8 = 24

$Contraste_{AB}$ = $ab + (1) - a - b$ = 26 + 8 - 17 - 23 = -6

**Sumas de cuadrados.**

$$SS_A = \frac{(Contraste_A)^2}{4n} = \frac{(12)^2}{4x2} = \frac{144}{8} = 18$$

$$SS_B = \frac{(Contraste_B)^2}{4n} = \frac{(24)^2}{4x2} = \frac{576}{8} = 72$$

$$SS_{AB} = \frac{(Contraste_{AB})^2}{4n} = \frac{(-6)^2}{4x2} = \frac{36}{8} = 4.5$$

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} y_{ijk}^2 - \frac{y_{...}^2}{4n} = 784 - \frac{(74)^2}{4x2} = 784 - 684.5 = 99.5$$

$SS_E = SS_T - SS_A - SS_B - SS_{AB}$

$= 99.5 - 18 - 72 - 4.5$

$= 5.0$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Tiempo | 18.0 | 1 | 18.0 | 14.4 |
| Temperatura | 72.0 | 1 | 72.0 | 57.6 |
| TiempoxTemperatura | 4.5 | 1 | 4.5 | 3.6 |
| Error | 5.0 | 4 | 1.25 | |
| Total | 99.5 | 7 | | |

Tomando $\alpha$ = 0.05 para obtener el F de tablas con un grado de libertad del numerador y 4(n-1)= 4(2-1) = 4 en el denominador, se tiene que:

$F_{\alpha,1,4(n-1)} = F_{0.05,1,4} = 7.71$

**Conclusiones**

- Para el caso del factor Tiempo el $F_0 > F_{Tablas}$ (14.4>7.71); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que el tiempo de almacenamiento influye significativamente en el número de manzanas que se pudren.
- Para el caso del factor Temperatura el $F_0 > F_{Tablas}$ (57.6>7.71); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que el nivel de Temperatura influye significativamente en el número de manzanas que se pudren.

- Para el caso de la interacción Tiempo y Temperatura el $F_0 < F_{Tablas}$ (3.6<7.71); por lo tanto, se acepta $H_0$ y se rechaza $H_1$; es decir, que el tiempo de almacenamiento y el nivel de Temperatura no influyen significativamente en el número de manzanas que se pudren.

## 3. ORDEN ESTÁNDAR.

Las combinaciones de tratamientos es posible escribirlas en el orden *(1), a, b* y *ab*; el cual se conoce como **Orden Estándar**, el cual está formada por los coeficientes de los contrastes usados para estimar los efectos, y siempre son +1 ó -1, como se presenta en la siguiente tabla:

| Efectos | (1) | *a* | *b* | *ab* |
|---|---|---|---|---|
| A | -1 | +1 | -1 | +1 |
| B | -1 | -1 | +1 | +1 |
| AB | +1 | -1 | -1 | +1 |

De la tabla anterior se puede elaborar una tabla de signos algebraicos (positivos y negativos), en la cual en el encabezado de las columnas se encuentran los efectos principales (A y B), la interacción (AB), e I, que representa el total o el promedio de todo el Experimento.

| Combinación de Tratamientos | Efecto Factorial | | | |
|---|---|---|---|---|
| | I | A | B | AB |
| (1) | + | - | - | + |
| *a* | + | + | - | - |
| *b* | + | - | + | - |
| ab | + | + | + | + |

En la tabla anterior se puede observar que la columna encabezada por I, que representa el toral o el promedio de todo el experimento, esta formada sólo por signos positivos; los signos de la columna AB se obtiene como el producto de los signos de los dos efectos principales (A y B), es decir multiplicando los signos de las columnas encabezadas por A y B.

Esta tabla puede ser utilizada para encontrar los coeficientes de los contrastes, con el fin de estimar cualquier efecto; simplemente multiplicando los signos de la columna apropiada por la correspondiente combinación de tratamientos, y luego sumándolos.
Por lo tanto, los contrastes para estimar A, B y AB se obtienen:

$Contraste_A = -(1) + a - b + ab$

$Contraste_B = -(1) - a + b + ab$

$Contraste_{AB} = (1) - a - b + ab$

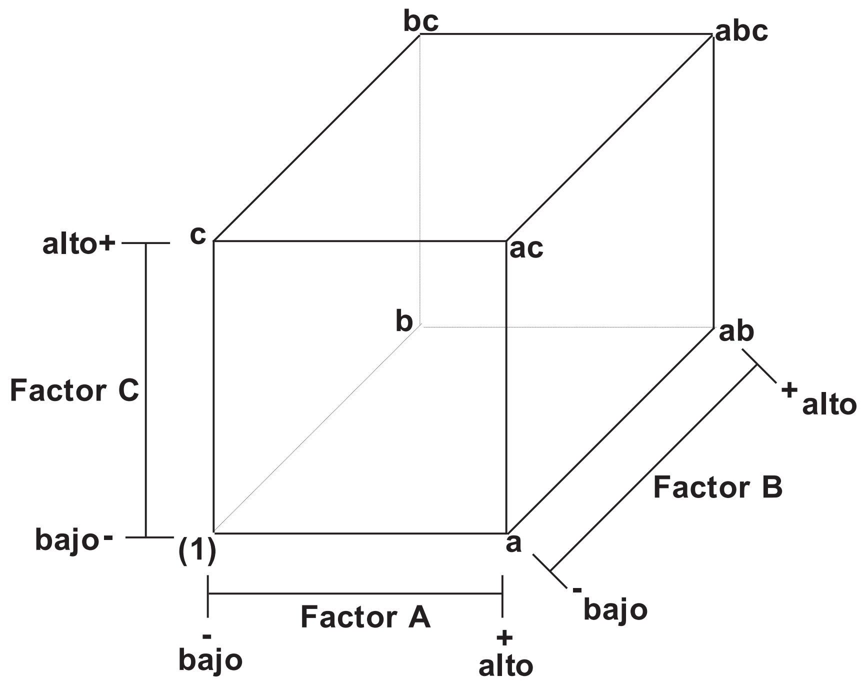
Como puede observarse estos contrastes coinciden con los que se obtuvieron anteriormente.

## 4. DISEÑO FACTORIAL $2^3$

Si en un Experimento se estudian tres factores A, B y C, cada uno con dos niveles. Este Diseño se conoce como Diseño Factorial $2^3$.

Para el estudio de este Diseño Factorial se retomará la notación definida en el Diseños Factorial $2^2$, que fue planteada anteriormente.

En este Diseño habrán ocho combinaciones de tratamiento ($2^3 = 8$) que son *(1), a, b, c, ab, ac, bc, abc*, las cuales se pueden representar geométricamente mediante un cubo de la siguiente manera:



Estas combinaciones de tratamientos representan los niveles a los que se encuentran dichos factores y significan:

(1) : el factor A, B y C se encuentra en el nivel bajo

$a$ : alto A-bajo B-bajo C

$b$ : bajo A-alto B-bajo C

$c$ : bajo A-bajo B-alto C

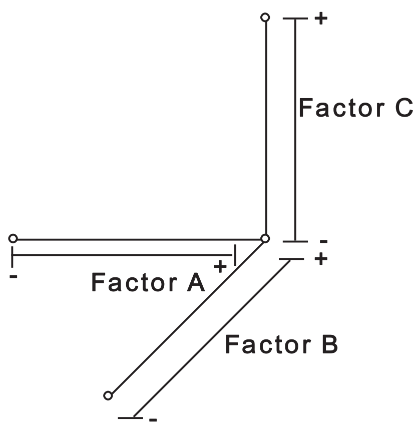
$ab$ : alto A-alto B-bajo C

$ac$ : alto A-bajo B-alto C

$bc$ : bajo A-alto B-alto C

$abc$ : alto A-alto B-alto C

Las representaciones anteriores pueden ser fácilmente observadas trasladando los ejes de los niveles de cada factor a cada uno de los vértices que representa cada combinación de tratamientos. Por ejemplo, para encontrar la combinación de tratamientos $ab$, geométricamente los ejes se trasladan de las siguiente manera:




Ya que *ab* representa que el factor A esta en nivel alto, el factor B en nivel alto y el factor C en nivel bajo.

Entre estas ocho combinaciones de tratamiento existen siete grados de libertad; de los cuales tres se asocian con los efectos principales (A, B y C); y los otros cuatro se asocian con las interacciones (AB, AC, BC y ABC), cada una con un grado de libertad.

## 4.1 REPRESENTACIÓN SIMBÓLICA DE LOS DATOS

Ya que en los Diseños Factoriales $2^3$, el número de factores son Tres (A, B y C) con dos niveles cada uno y en cada nivel existen "$n$" réplicas del Experimento; entonces la representación de los datos observados para esté tipo de Diseño Factorial será de la siguiente forma:

| Factor A | Factor B | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| | Factor C | | Factor C | |
| | 1 | 2 | 1 | 2 |
| 1 | $y_{1111},y_{1112},\ldots,y_{111n}$ | $y_{1121},y_{1122},\ldots,y_{112n}$ | $y_{1211},y_{1212},\ldots,y_{121n}$ | $y_{1221},y_{1222},\ldots,y_{122n}$ |
| 2 | $y_{2111},y_{2112},\ldots,y_{211n}$ | $y_{2121},y_{2122},\ldots,y_{212n}$ | $y_{2211},y_{2212},\ldots,y_{221n}$ | $y_{2221},y_{2222},\ldots,y_{222n}$ |

Como puede observarse esta tabla es sólo una parte de la tabla de la representación simbólica de un Diseño de Tres factores con "$a$" niveles del factor A, "$b$" del factor B y "$c$" niveles del factor C; con "$n$" réplicas cada nivel.

Cualquier observación de tabla se interpreta de la misma forma que el Diseño de Tres Factores.

El número total de observaciones en el Experimento es $N$ = 2x2x2xn = 8n; ya que se realizaron *n* réplicas.

**Nota:** Lo que se refiere al Modelo Estadístico que representan los datos descritos anteriormente será igual al Diseño de Tres factores, con la diferencia de que los subíndices de $i$, $j$ y $k$ tomaran valores de 1 hasta 2.

## 4.2 ESTIMACIÓN DE LOS EFECTOS DE LOS FACTORES

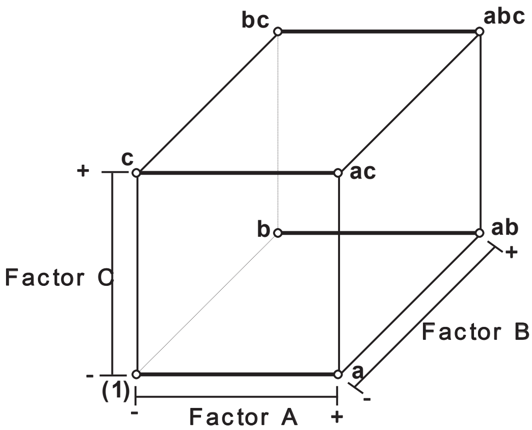
La estimación de los efectos principales y las interacciones; igual que el Diseño $2^2$ se pueden calcular por medio de dos métodos, observando la representación geométrica.

**Primer Método**

**Efecto del Factor A**

El efecto promedio del factor A vendrá dado como el promedio de la suma de los efectos del factor A de las diferentes combinaciones de los niveles de los factores B y C.

La forma de visualizarlo en el cubo es desplazando el factor A a las diferentes combinaciones de los niveles alto y bajo de los factores B y C; tomando las combinaciones de derecha a izquierda. Es decir, los vértices de los segmentos grueso del cubo siguiente:




De la figura anterior se obtiene la siguiente tabla:

| Factor B | Factor C | Efecto del Factor A |
|---|---|---|
| Nivel bajo | Nivel bajo | $\frac{(a-(1))}{n}$ |
| Nivel alto | Nivel bajo | $\frac{(ab-b)}{n}$ |
| Nivel bajo | Nivel alto | $\frac{(ac-c)}{n}$ |
| Nivel alto | Nivel alto | $\frac{(abc-bc)}{n}$ |

**El Efecto promedio del factor A es:**

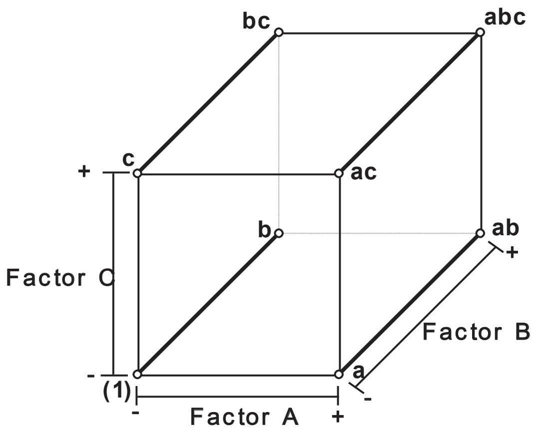
$$A = \frac{1}{4}\left[\frac{(a-(1))}{n} + \frac{(ab-b)}{n} + \frac{(ac-c)}{n} + \frac{(abc-bc)}{n}\right]$$

$$= \frac{1}{4n}\left[a-(1)+ab-b+ac-c+abc-bc\right]$$

**Efecto del Factor B**

El efecto promedio del factor B vendrá dado como el promedio de la suma de los efectos del factor B de las diferentes combinaciones de los niveles de los factores A y C.

La forma de visualizarlo en el cubo es desplazando el factor B a las diferentes combinaciones de los niveles alto y bajo de los factores A y C; tomando las combinaciones de tratamientos en diagonal del fondo hacia adelante. Es decir, los vértices del los segmentos grueso del cubo siguiente:




De la figura anterior se obtiene la siguiente tabla:

| Factor A | Factor C | Efecto del Factor B |
|---|---|---|
| Nivel bajo | Nivel bajo | $\frac{(b-(1))}{n}$ |
| Nivel alto | Nivel bajo | $\frac{(ab-a)}{n}$ |
| Nivel bajo | Nivel alto | $\frac{(bc-c)}{n}$ |
| Nivel alto | Nivel alto | $\frac{(abc-ac)}{n}$ |

**El efecto promedio del factor B es:**

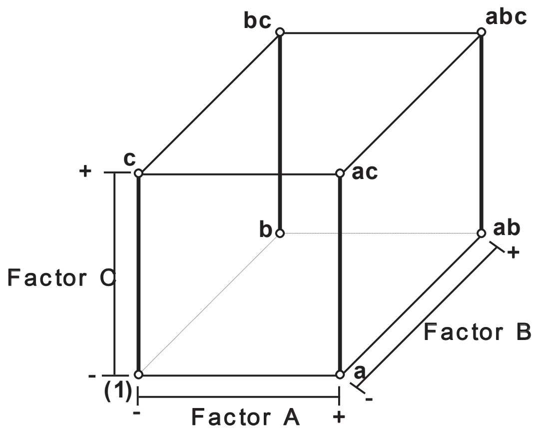
$$B = \frac{1}{4}\left[\frac{(b-(1))}{n}+\frac{(ab-a)}{n}+\frac{(bc-c)}{n}+\frac{(abc-ac)}{n}\right]$$

$$= \frac{1}{4n}\left[b+ab+bc+abc-(1)-a-c-ac\right]$$

**Efecto del Factor C**

El efecto promedio del factor C vendrá dado como el promedio de la suma de los efectos del factor C de las diferentes combinaciones de los niveles de los factores A y B.

La forma de visualizarlo en el cubo es desplazando el factor C a las diferentes combinaciones de los niveles alto y bajo de los factores A y B; tomando las combinaciones de tratamiento de arriba hacia abajo. Es decir, los vértices del los segmentos grueso del cubo siguiente:




De la figura anterior se obtiene la siguiente tabla:

| Factor A | Factor B | Efecto del Factor C |
|---|---|---|
| Nivel bajo | Nivel bajo | $\frac{(c-(1))}{n}$ |
| Nivel alto | Nivel bajo | $\frac{(ac-a)}{n}$ |
| Nivel bajo | Nivel alto | $\frac{(bc-b)}{n}$ |
| Nivel alto | Nivel alto | $\frac{(abc-ab)}{n}$ |

**El efecto promedio del factor C es:**

$$C = \frac{1}{4}\left[\frac{(c-(1))}{n}+\frac{(ac-a)}{n}+\frac{(bc-b)}{n}+\frac{(abc-ab)}{n}\right]$$

$$= \frac{1}{4n}\left[c+ac+bc+abc-(1)-a-b-ab\right]$$

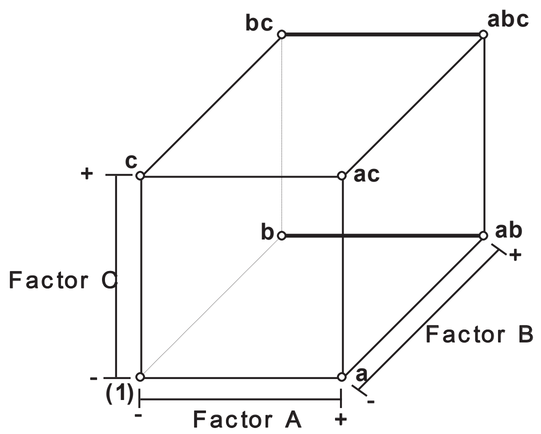
## Efectos de las Interacciones Dobles

**Efecto de AB**

El efecto promedio de la interacción AB vendrá dada como la diferencia del promedio de la suma de los efectos del factor A, a las diferentes combinaciones de los niveles del factor B, cuando el factor C se encuentra en los niveles alto y bajo. Como se muestra a continuación.

**El factor B a nivel alto**

La forma de visualizarlo en el cubo es desplazando el factor A al nivel alto de B, y luego observando los vértices para los niveles alto y bajo del factor C; tomando las combinaciones de tratamiento de derecha a izquierda. Es decir, los vértices del de los dos segmentos gruesos del cubo siguiente:

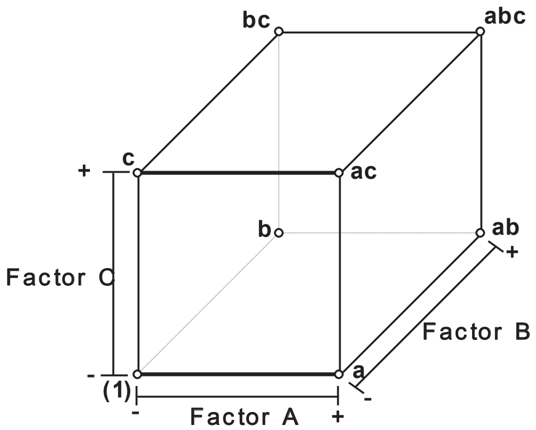



De la figura anterior se obtiene la siguiente tabla:

| Efecto de A para el nivel alto del factor B | |
|---|---|
| Nivel alto del factor C | Nivel bajo del factor C |
| $\frac{(abc-bc)}{n}$ | $\frac{(ab-b)}{n}$ |

**El factor B a nivel bajo**

La forma de visualizarlo en el cubo es desplazando el factor A al nivel bajo de B, y luego observando los vértices para los niveles alto y bajo del factor C; tomando las combinaciones de tratamiento de derecha a izquierda. Es decir, los vértices de los dos segmentos gruesos del cubo siguiente:




De la figura anterior se obtiene la siguiente tabla:

| Efecto de A para el nivel bajo del factor B | |
|---|---|
| Nivel alto del factor C | Nivel bajo del factor C |
| $\frac{(ac-c)}{n}$ | $\frac{(a-(1))}{n}$ |

**Por lo tanto, el efecto promedio de AB es:**

$$AB = \frac{1}{4}\left\{\frac{(abc-bc)}{n}+\frac{(ab-b)}{n}-\left[\frac{(ac-c)}{n}+\frac{(a-(1))}{n}\right]\right\}$$

$$= \frac{1}{4n}\{(abc-bc)+(ab-b)-[(ac-c)+(a-(1)]\}$$

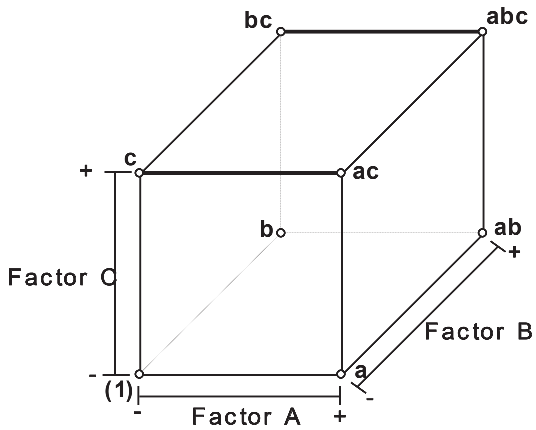
$$= \frac{1}{4n}[(1)-a-b+ab+c-ac-bc+abc]$$

**Efecto de AC**

El efecto promedio de la interacción AC vendrá dada como la diferencia del promedio de la suma de los efectos del factor A, a las diferentes combinaciones de los niveles del factor C, cuando el factor B se encuentra en los niveles alto y bajo. Como se muestra a continuación

**El factor C a nivel alto**

La forma de visualizarlo en el cubo es desplazando el factor A al nivel alto de C, y luego observando los vértices para los niveles alto y bajo del factor B; tomando las combinaciones de tratamiento de derecha a izquierda. Es decir, los vértices de los dos segmentos gruesos del cubo siguiente:

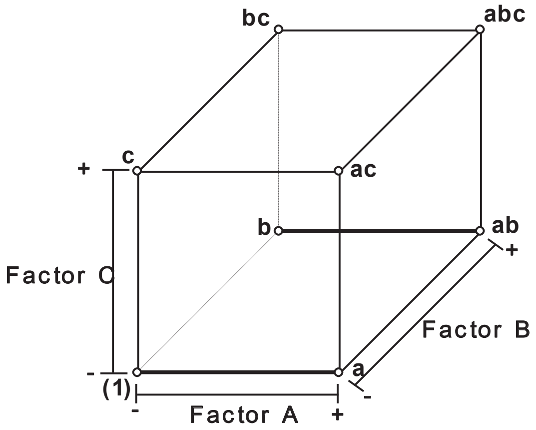



De la figura anterior se obtiene la siguiente tabla:

| Efecto de A para el nivel alto del factor C | |
|---|---|
| Nivel alto del factor B | Nivel bajo del factor B |
| $\frac{(abc - bc)}{n}$ | $\frac{(ac - c)}{n}$ |

**El factor C a nivel bajo**

La forma de visualizarlo en el cubo es desplazando el factor A al nivel bajo de C, y luego observando los vértices para los niveles alto y bajo del factor B; tomando las combinaciones de tratamiento de derecha a izquierda. Es decir, los vértices de los dos segmentos gruesos del cubo siguiente:




De la figura anterior se obtiene la siguiente tabla:

| Efecto de A para el nivel bajo del factor C | |
|---|---|
| Nivel alto del factor B | Nivel bajo del factor B |
| $\frac{(ab-b)}{n}$ | $\frac{(a-(1))}{n}$ |

Por lo tanto, el efecto promedio de AC es:

$$AC = \frac{1}{4}\left\{\frac{(abc-bc)}{n}+\frac{(ac-c)}{n}-\left[\frac{(ab-b)}{n}+\frac{(a-(1))}{n}\right]\right\}$$

$$= \frac{1}{4n}\{(abc-bc)+(ac-c)-[(ab-b)+(a-(1)]\}$$

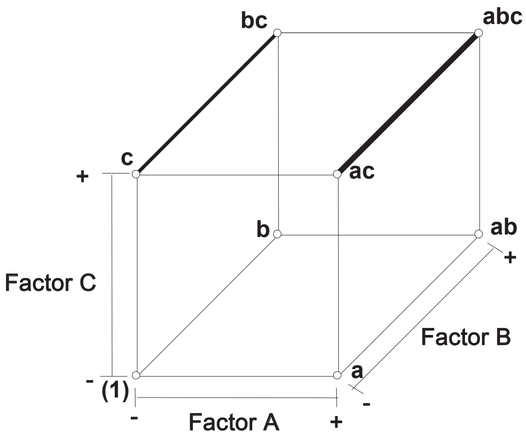
$$= \frac{1}{4n}[(1)-a+b-ab-c+ac-bc+abc]$$

**Efecto de BC**

El efecto promedio de la interacción BC vendrá dada como la diferencia del promedio de la suma de los efectos del factor B, a las diferentes combinaciones de los niveles del factor C, cuando el factor A se encuentra en los niveles alto y bajo. Como se muestra a continuación:

**El factor C a nivel alto**

La forma de visualizarlo en el cubo es desplazando el factor B al nivel alto de C, y luego observando los vértices para los niveles alto y bajo del factor A; tomando las combinaciones de tratamiento de derecha a izquierda. Es decir, los vértices de los dos segmentos gruesos del cubo siguiente:

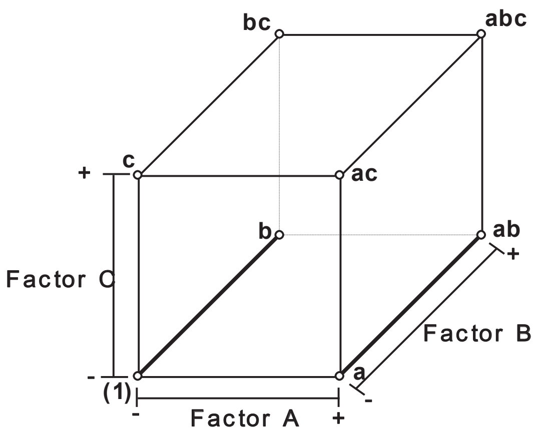



De la figura anterior se obtiene la siguiente tabla:

| Efecto de B para el nivel alto del factor C | |
|---|---|
| Nivel alto del factor A | Nivel bajo del factor A |
| $\frac{(abc - ac)}{n}$ | $\frac{(bc - c)}{n}$ |

**El factor C a nivel bajo**

La forma de visualizarlo en el cubo es desplazando el factor B al nivel bajo de C, y luego observando los vértices para los niveles alto y bajo del factor A; tomando las combinaciones de tratamiento en diagonal de adentro hacia enfrente. Es decir, los vértices de los dos segmentos gruesos del cubo siguiente:



De la figura anterior se obtiene la siguiente tabla:

| Efecto de A para el nivel bajo del factor C | |
|---|---|
| Nivel alto del factor A | Nivel bajo del factor A |
| $\frac{(ab-a)}{n}$ | $\frac{(b-(1))}{n}$ |

Por lo tanto, el efecto promedio de BC es:

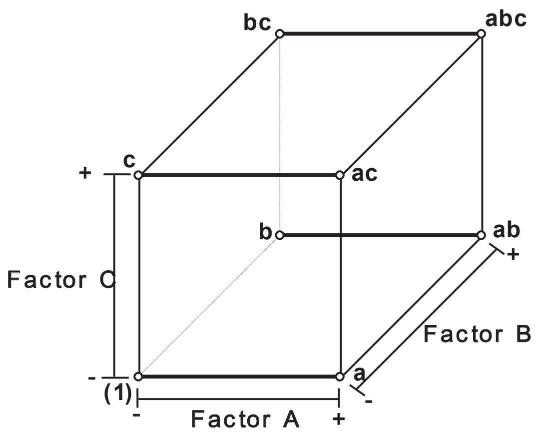
$$BC = \frac{1}{4}\left\{\frac{(abc-ac)}{n}+\frac{(bc-c)}{n}-\left[\frac{(ab-a)}{n}+\frac{(b-(1))}{n}\right]\right\}$$

$$= \frac{1}{4n}\{(abc-ac)+(bc-c)-[(ab-a)+(b-(1)]\}$$

$$= \frac{1}{4n}\left[(1)+a-b-ab-c-ac+bc+abc\right]$$

**Efecto de ABC**

Este efecto se define como la diferencia promedio entre la interacción AB para el nivel bajo y alto del factor C. Es decir, los vértices de los segmentos gruesos del cubo siguiente:




De la figura anterior se obtiene la siguiente tabla:

| Efecto de la interacción AB | |
|---|---|
| Nivel alto del factor C | Nivel bajo del factor C |
| $\frac{(abc-bc)-(ac-c)}{4n}$ | $\frac{(ab-b)-(a-(1))}{n}$ |

Por lo tanto, el efecto promedio de ABC es:

$$ABC = \left\{\frac{(abc-bc)-(ac-c)}{4n} - \left[\frac{(ab-b)-(a-(1))}{4n}\right]\right\}$$

$$= \frac{1}{4n}\left\{(abc-bc)-(ac-c)-\left[(ab-b)-(a-(1)\right]\right\}$$

$$= \frac{1}{4n}\left\{(abc-bc)-(ac-c)-(ab-b)+(a-(1))\right\}$$

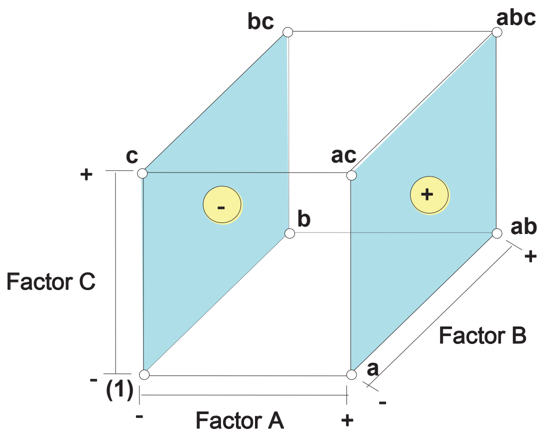
$$= \frac{1}{4n}\left[abc-bc-ac+c-ab+b+a-(1)\right]$$

**Segundo Método**

**Efecto Promedio de A.**

Se obtiene promediando las cuatro combinaciones de tratamientos en el lado derecho del cubo donde el factor A está en el nivel alto $(\bar{Y}_{A+})$, menos el promedio de las cuatro combinaciones de tratamiento en el lado izquierdo del cubo, donde el factor A está en el nivel bajo $(\bar{Y}_{A-})$. Es decir, A = $\bar{Y}_{A+}$ - $\bar{Y}_{A-}$ .

Mediante el cubo se puede observar de la siguiente manera:




**Por lo tanto:**

$\overline{Y}_{A+} = \dfrac{a + ab + ac + abc}{4n}$ $\qquad$ $\overline{Y}_{A-} = \dfrac{(1) + b + bc + c}{4n}$

**Entonces:**

A = $\overline{Y}_{A+}$ - $\overline{Y}_{A-}$

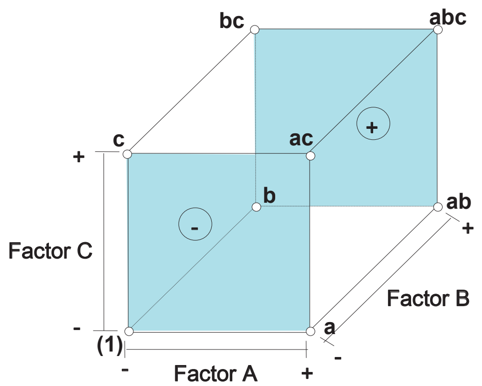
$$= \left[\frac{a + ab + ac + abc}{4n}\right] - \left[\frac{(1) + b + bc + c}{4n}\right]$$

$$= \frac{1}{4n}\left[a + ab + ac + abc - (1) - b - c - bc\right]$$

**Efecto promedio de B.**

Se obtiene por la diferencia promedio de las cuatro combinaciones de tratamientos de la cara posterior del cubo, donde el factor B se encuentra en el nivel alto $(\overline{Y}_{B+})$, y las cuatro combinaciones de tratamiento de la cara del frente del cubo, donde el factor B se encuentra en el nivel bajo $(\overline{Y}_{B-})$. Es decir, B = $\overline{Y}_{B+}$ - $\overline{Y}_{B-}$ .

Mediante el cubo se puede observar de la siguiente manera:




**Por lo tanto:**

$\overline{Y}_{B+} = \dfrac{b + ab + bc + abc}{4n}$ $\qquad$ $\overline{Y}_{B-} = \dfrac{(1) + a + ac + c}{4n}$

**Entonces:**

B = $\overline{Y}_{B+}$ - $\overline{Y}_{B-}$

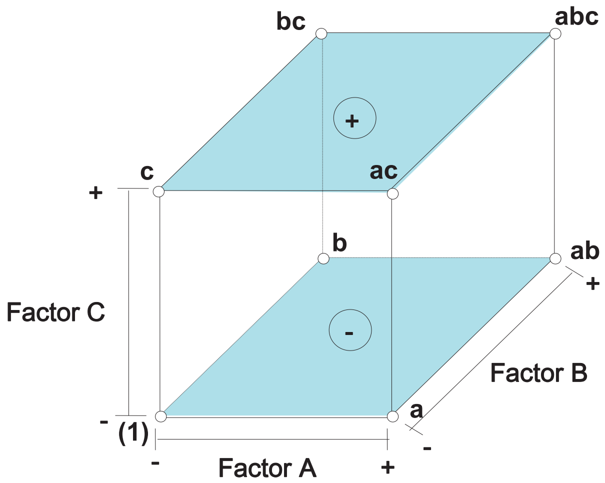
$$= \left[\frac{b+ab+bc+abc}{4n}\right] - \left[\frac{(1)+a+ac+c}{4n}\right]$$

$$= \frac{1}{4n}\left[b+ab+bc+abc-(1)-a-ac-c\right]$$

**Efecto Promedio de C**

Se define como el promedio de las cuatro combinaciones de tratamientos de la cara superior del cubo, donde el factor C se encuentra en el nivel alto $(\overline{Y}_{C+})$, menos el promedio de las cuatro combinaciones de tratamiento de la cara inferior del cubo, donde el factor C se encuentra en el nivel bajo $(\overline{Y}_{C-})$. Es decir, C = $\overline{Y}_{C+}$ - $\overline{Y}_{C-}$ .

Mediante el cubo se puede observar de la siguiente manera:




**Por lo tanto:**

$$\overline{Y}_{C+} = \frac{c+ac+bc+abc}{4n} \qquad \overline{Y}_{C-} = \frac{(1)+a+b+ab}{4n}$$

**Entonces:**

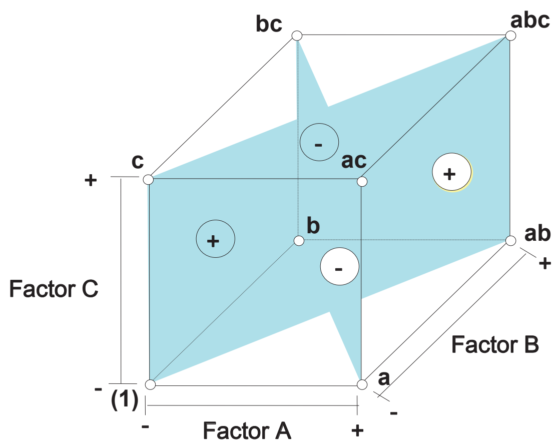
$$C = \overline{Y}_{C+} - \overline{Y}_{C-}$$
$$= \left[\frac{c+ac+bc+abc}{4n}\right] - \left[\frac{(1)+a+b+ab}{4n}\right]$$
$$= \frac{1}{4n}\left[c+ac+bc+abc-(1)-a-b-ab\right]$$

**Efecto promedio de AB.**

Se obtiene por la diferencia promedio entre los efectos promedios del factor A a los dos niveles del factor B; es decir, cuando el factor C está en el nivel bajo, AB es justo la diferencia promedio en el efecto del factor A en los dos niveles del factor B. La iteración AB esta dada por la diferencia de promedios entre las combinaciones en los dos planos diagonales del cubo. Es decir:

$$AB(C\text{ bajo}) = \frac{1}{2n}(ab-b) - \frac{1}{2n}(a-(1))$$
$$AB(C\text{ alto}) = \frac{1}{2n}(abc-bc) - \frac{1}{2n}(ac-c)$$

Mediante el cubo se puede observar que son las siguientes caras del cubo:




El efecto promedio de la interacción AB es exactamente el promedio de estas dos componentes.

**Entonces**

$$AB = \frac{1}{2}\left[\frac{1}{2n}(ab-b)-\frac{1}{2n}(a-(1))+\frac{1}{2n}(abc-bc)-\frac{1}{2n}(ac-c)\right]$$

$$= \frac{1}{4n}(ab-b-a+(1)+abc-bc-ac+c)$$

$$= \frac{1}{4n}(abc-bc+ab-b-ac+c-a+(1))$$

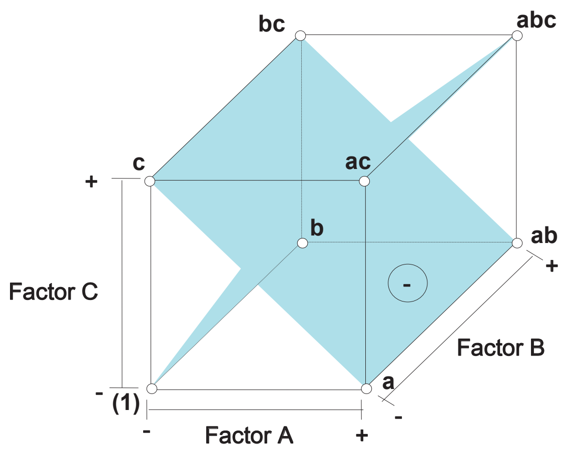
**Efecto promedio de AC.**

Se obtiene por la diferencia promedio entre los efectos promedios del factor A a los dos niveles del factor C; es decir, cuando el factor B está en el nivel bajo, AC es justo la diferencia promedio en el efecto del factor A en los dos niveles del factor C. Es decir:

$$AC(B\ bajo) = \frac{1}{2n}(ac-c)-\frac{1}{2n}(a-(1))$$

$$AC(B\ alto) = \frac{1}{2n}(abc-bc)-\frac{1}{2n}(ab-b)$$

Mediante el cubo se puede observar que son las siguientes caras del cubo:




El efecto promedio de la interacción AC es la suma del promedio de estas dos componentes.

**Entonces:**

$$AC = \frac{1}{2}\left[\frac{1}{2n}(ac-c)-\frac{1}{2n}(a-(1))+\frac{1}{2n}(abc-bc)-\frac{1}{2n}(ab-b)\right]$$

$$= \frac{1}{4n}(ac-c-a+(1)+abc-bc-ab+b)$$

$$= \frac{1}{4n}((1)-a+b-ab-c+ac-bc+abc)$$

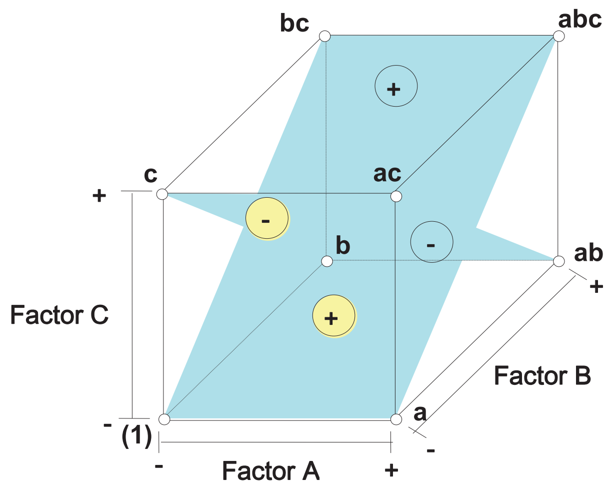
**Efecto promedio de BC**

Se obtiene por la diferencia promedio ente los efectos promedios del factor B a los dos niveles del factor C; es decir, cuando el factor A está en el nivel bajo o alto, BC es la diferencia promedio en el efecto del factor B en los dos niveles del factor C. Es decir:

$$BC(A\ bajo) = \frac{1}{2n}(bc-c)-\frac{1}{2n}(b-(1))$$

$$BC(A\ alto) = \frac{1}{2n}(abc-ac)-\frac{1}{2n}(ab-a)$$

Mediante el cubo se puede observar que son las siguientes caras del cubo:




El efecto promedio de la interacción BC es la suma del promedio de estas dos componentes.

**Entonces:**

$$BC = \frac{1}{2}\left[\frac{1}{2n}(bc-c)-\frac{1}{2n}(b-(1))+\frac{1}{2n}(abc-ac)-\frac{1}{2n}(ab-a)\right]$$

$$= \frac{1}{4n}(bc-c-b+(1)+abc-ac-ab+a)$$

$$= \frac{1}{4n}((1)+a-b-ab-c-ac+bc+abc)$$

**Efecto promedio de ABC.**

Esta se define como la diferencia promedio entre la interacción AB para los dos niveles del factor C. Es decir:

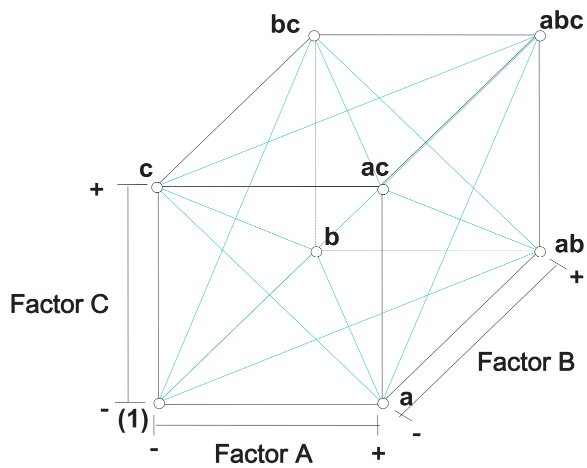
**Entonces:**

$$ABC = \frac{1}{4n}[(abc-bc)-(ac-c)-(ab-b)+(a-(1))]$$

$$= \frac{1}{4n}(abc-bc-ac+c-ab+b+a-(1))$$

La interacción ABC se puede considerar como la diferencia de dos promedios. Si las combinaciones de los dos promedios se aíslan, definen los vértices de los dos tetraedros que componen el cubo.

Mediante el cubo se puede observar de la siguiente forma:




## 4.3 ANÁLISIS DE VARIANZA

Para llevar a cabo el Análisis de Varianza se deben obtener los contrastes de los efectos encontrados anteriormente; que no son más que las cantidades entre paréntesis de los efectos. Es decir:

$\text{Contraste}_A = a + ab + ac + abc - (1) - b - c - bc$

$\text{Contraste}_B = b + ab + bc + abc - (1) - a - c - ac$

$\text{Contraste}_C = c + ac + bc + abc - (1) - a - b - ab$

$\text{Contraste}_{AB} = abc - bc + ab - b - ac + c - a + (1)$

$\text{Contraste}_{AC} = (1) - a + b - ab - c + ac - bc + abc$

$\text{Contraste}_{BC} = (1) + a - b - ab - c - ac + bc + abc$

$\text{Contraste}_{ABC} = abc - bc - ac + c - ab + b + a - (1)$

Las sumas de cuadrados de los efectos pueden ser obtenidas fácilmente; ya que a cada una le corresponde un contraste y un sólo grado de libertad. Por lo tanto, la suma de cuadrados para cualquier efecto de un Diseño $2^3$ con "$n$" réplicas, vendrá dada por el contraste correspondiente al cuadrado entre el total de las observaciones.

$$SS = \frac{(contraste)^2}{8n}$$

Entonces la Suma de Cuadrados para los efectos principales e interacciones son las siguientes:

$$SS_A = \frac{(contraste_A)^2}{8n} \qquad SS_B = \frac{(contraste_B)^2}{8n} \qquad SS_C = \frac{(contraste_C)^2}{8n}$$

$$SS_{AB} = \frac{(contraste_{AB})^2}{8n} \qquad SS_{AC} = \frac{(contraste_{AC})^2}{8n} \qquad SS_{BC} = \frac{(contraste_{BC})^2}{8n}$$

$$SS_{ABC} = \frac{(contraste_{ABC})^2}{8n}$$

La suma de cuadrado total y la suma de cuadrados del error se calcula de forma usual.

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{8n}$$

$$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC}$$

donde

$SS_T$ : Tiene $8n$-1 grados de libertad; ya que existe un total de $N = 8n$ observaciones.

$SS_E$ : Tiene $8(n-1)$ grados de libertad; porque cada suma de los efectos principales e interacciones tienen un grado de libertad cada una.

**Ejemplo 2**

Retomando el ejemplo desarrollado en el Diseño Trifactorial, se tiene:

En un Experimento para investigar las propiedades de resistencia a la compresión de mezcla de Cemento y Tierra, se utilizaron dos períodos (A) diferentes de curado en combinación con dos Temperaturas (B) diferentes de curado y dos tierras (C) diferentes. Se hicieron dos réplicas para cada combinación de niveles de los tres factores, los datos obtenidos se presentan a continuación:

| Edad (Factor A) | Temperatura (Factor B) | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| | Tierra (Factor C) | | Tierra (Factor C) | |
| | 1 | 2 | 1 | 2 |

| | | | | |
|---|---|---|---|---|
| 1 | 471<br>413 | 385<br>434 | 485<br>552 | 530<br>593 |
| 2 | 712<br>637 | 770<br>705 | 712<br>789 | 741<br>806 |

**Solución**

**Datos**

$a = 2$ , $b = 2$ , $c = 2$, $n = 2$ , $i = 1,2$ , $j = 1,2$ , $k = 1,2$

A continuación se presenta la tabla de datos codificados (divididos por 100).

| Edad (Factor A) | Temperatura (Factor B) | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| | Tierra (Factor C) | | Tierra (Factor C) | |
| | 1 | 2 | 1 | 2 |
| 1 | 4.71<br>4.13 | 3.85<br>4.34 | 4.85<br>5.52 | 5.30<br>5.93 |
| 2 | 7.12<br>6.37 | 7.70<br>7.05 | 7.12<br>7.89 | 7.41<br>8.06 |

Las hipótesis a probar serán las mismas que se plantearon cuando se analizó anteriormente este ejemplo.

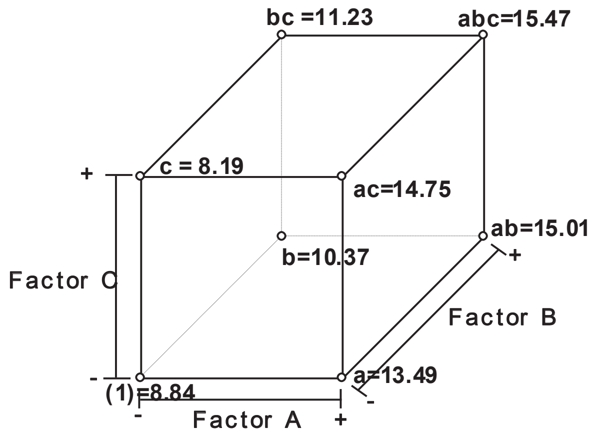
Como puede observarse este ejemplo corresponde a un Diseño $2^3$, en el cual existen tres factores Factor A (Edad), Factor B (Temperatura) y Factor C (Tierra) todos a dos niveles (bajo y alto); y el Experimento se realiza dos veces.

La tabla de combinaciones de tratamiento para este Diseño es la siguiente:

| Combinación de Tratamiento | Réplicas | | Total | Simbología |
|---|---|---|---|---|
| | I | II | | |
| A bajo, B bajo, C bajo | 4.71 | 4.13 | 8.84 | (1) |
| A alto, B bajo, C bajo | 7.12 | 6.37 | 13.49 | $a$ |
| A bajo, B alto, C bajo | 4.85 | 5.52 | 10.37 | $b$ |
| A alto, B alto, C bajo | 7.12 | 7.89 | 15.01 | $ab$ |
| A bajo, B bajo, C alto | 3.85 | 4.34 | 8.19 | $c$ |

| | | | | |
|---|---|---|---|---|
| A alto, B bajo, C alto | 7.70 | 7.05 | 14.75 | $ac$ |
| A bajo, B alto, C alto | 5.30 | 5.93 | 11.23 | $bc$ |
| A alto, B alto, C alto | 7.41 | 8.06 | 15.47 | $abc$ |

Geométricamente las combinaciones de tratamientos que se encuentran en la tabla anterior, se representan de la siguiente forma:




Para la estimación de los efectos promedios de los factores principales e interacciones se utilizará el primer método detallado anteriormente:

**Efecto de la Edad**

A = $\frac{1}{4n}\left[a-(1)+ab-b+ac-c+abc-bc\right]$

= $\frac{1}{4x2}\left[13.49-8.84+15.01-10.37+14.75-8.19+15.47-11.23\right]$

= $\frac{1}{8}(20.09) = 2.51125 \approx 2.51$

**Efecto de la Temperatura**

B = $\frac{1}{4n}\left[b+ab+bc+abc-(1)-a-c-ac\right]$

B = $\frac{1}{4x2}\left[10.37+15.01+11.23+15.47-8.84-13.49-8.19-14.75\right]$

= $\frac{1}{8}(6.81)$ = $0.85125 \approx 0.85$

**Efecto de la Tierra**

C = $\frac{1}{4n}\left[c + ac + bc + abc - (1) - a - b - ab\right]$

= $\frac{1}{4x2}\left[8.19 + 14.75 + 11.23 + 15.47 - 8.84 - 13.49 - 10.37 - 15.01\right]$

= $\frac{1}{8}(1.93)$ = 0.24125≈0.24

**Efecto de la Edad y Temperatura**

AB = $\frac{1}{4n}\left[abc - bc + ab - b - ac + c - a + (1)\right]$

= $\frac{1}{4x2}\left[15.47 - 11.23 + 15.01 - 10.37 - 14.75 + 8.19 - 13.49 + 8.84\right]$

= $\frac{1}{8}(-2.33)$ = - 0.29125 ≈ -0.30

**Efecto de Edad y Tierra**

AC = $\frac{1}{4n}\left[(1) - a + b - ab - c + ac - bc + abc\right]$

= $\frac{1}{4x2}\left[8.84 - 13.49 + 10.37 - 15.01 - 8.19 + 14.75 - 11.23 + 15.47\right]$

= $\frac{1}{8}(1.51)$ = 0.18875 ≈ 0.189

**Efecto de la Temperatura y Tierra**

BC = $\frac{1}{4n}\left[(1) + a - b - ab - c - ac + bc + abc\right]$

= $\frac{1}{4x2}\left[8.84 + 13.49 - 10.37 - 15.01 - 8.19 - 14.75 + 11.23 + 15.47\right]$

= $\frac{1}{8}(0.71)$ = 0.08875 ≈ 0.089

**Efecto de la Edad, Temperatura y Tierra**

ABC = $\frac{1}{4n}\left[abc - bc - ac + c - ab + b + a - (1)\right]$

= $\frac{1}{4x2}\left[15.47 - 11.23 - 14.75 + 8.19 - 15.01 + 10.37 + 13.49 - 8.84\right]$

= $\frac{1}{8}(-2.31)$ = -0.28875 ≈ -0.29

Los valores encontrados para cada uno de los efectos significan:

- El efecto de la Edad es positivo e igual a 2.51; significa que se espera un aumento de 2.51 en la resistencia a la compresión de la mezcla de Cemento y Tierra al aumentar la Edad del nivel bajo(1) al nivel alto(2).
- El efecto de la Temperatura es positivo e igual a 0.85; significa que se espera un aumento de 0.85 en la resistencia a la compresión de la mezcla de Cemento y Tierra al aumentar la Temperatura del nivel bajo(1) al nivel alto(2).
- El efecto de la Tierra es positivo e igual a 0.24; significa que se espera un aumento de 0.24 en la resistencia a la compresión de la mezcla de Cemento y Tierra al aumentar la Tierra del nivel bajo(1) al nivel alto(2).

- El efecto de la Edad y Temperatura es negativo (-0.30); significa que el efecto de ambos factores no afectan la resistencia a la compresión de la mezcla de Cemento y Tierra.
- El efecto de la Edad y Tierra es positivo e igual a 0.189; significa que se espera que la combinación de ambos, produzca un aumento en la resistencia a la compresión de la mezcla de Cemento y Tierra de 0.189.
- El efecto de la Temperatura y Tierra es positivo e igual a 0.089; significa que se espera que la combinación de ambos, produzca un aumento en la resistencia a la compresión de la mezcla de Cemento y Tierra de 0.089.
- El efecto de la Edad, Temperatura y Tierra es negativo (-0.29); significa que la combinación de los tres factores no afecta la resistencia a la compresión de la mezcla de Cemento y Tierra.

**Encontrando los contrastes para los efectos principales e interacciones**

Ya que los contrastes son el resultado de lo que se encuentra entre paréntesis de los efectos; se tiene:

$\text{Contraste}_A = 20.09$ , $\text{Contraste}_B = 6.81$ , $\text{Contraste}_C = 1.93$

$\text{Contraste}_{AB} = -2.33$ , $\text{Contraste}_{AC} = 1.51$ , $\text{Contraste}_{BC} = 0.71$

$\text{Contraste}_{ABC} = -2.31$

**Sumas de Cuadrados**

$$SS_A = \frac{(Contraste_A)^2}{8n} = \frac{(20.09)^2}{8x2} = 25.2255 \approx 25.23$$

$$SS_B = \frac{(Contraste_B)^2}{8n} = \frac{(6.81)^2}{8x2} = 2.8985 \approx 2.9$$

$$SS_C = \frac{(Contraste_C)^2}{8n} = \frac{(1.93)^2}{8x2} = 0.2328 \approx 0.23$$

$SS_{AB} = \dfrac{(Contraste_{AB})^2}{8n} = \dfrac{(-2.33)^2}{8x2} = 0.33930625 \approx 0.34$

$SS_{AC} = \dfrac{(Contraste_{AC})^2}{8n} = \dfrac{(1.51)^2}{8x2} = 0.1425 \approx 0.14$

$SS_{BC} = \dfrac{(Contraste_{AB})^2}{8n} = \dfrac{(0.71)^2}{8x2} = 0.03150 \approx 0.03$

$SS_{ABC} = \dfrac{(Contraste_{ABC})^2}{8n} = \dfrac{(-2.31)^2}{8x2} = 0.3335 \approx 0.33$

**Suma de cuadrados totales:**

$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}^2 - \dfrac{y_{....}^2}{8x2} = 623.2289 - \dfrac{(97.35)^2}{16} = 623.2289 - 592.3139 = 30.91499$

**Suma de cuadrados del Error**

$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC}$

$= 30.91 - 25.23 - 2.90 - 0.23 - 0.34 - 0.14 - 0.03 - 0.33$

$= 1.71$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Edad | 25.23 | 1 | 25.23 | 120.14 |
| Temperatura | 2.90 | 1 | 2.90 | 13.81 |
| Tierra | 0.23 | 1 | 0.23 | 1.09 |
| EdadxTemperatura | 0.34 | 1 | 0.34 | 1.62 |
| EdadxTierra | 0.14 | 1 | 0.14 | 0.66 |
| TemperaturaxTierra | 0.03 | 1 | 0.03 | 0.14 |
| EdadxTemperaturaxTierra | 0.33 | 1 | 0.33 | 1.57 |
| Error | 1.71 | 8 | 0.21 | |
| Total | 30.91 | 15 | | |

**Conclusión**

Como puede observarse los valores obtenidos en la Tabla de Análisis de Varianza son bastante aproximados a los que se obtuvieron en la Tabla de Análisis de Varianza, cuando se utilizaron las fórmulas de un Diseño de Tres Factores (Unidad IV).

## 5. OTRA FORMA DE ENCONTRAR LOS CONTRASTES

Una forma más sencilla de obtener los contrastes de los efectos es utilizando una tabla de signos positivos y negativos (más y menos) para un Diseño Factorial $2^3$, la cual estará formada de la siguiente manera:

| Combinaciones de tratamiento | Efecto Factorial | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | I | A | B | AB | C | AC | BC | ABC |
| (1) | + | - | - | + | - | + | + | - |
| $a$ | + | + | - | - | - | - | + | + |
| $b$ | + | - | + | - | - | + | - | + |
| $ab$ | + | + | + | + | - | - | - | - |
| $c$ | + | - | - | + | + | - | - | + |
| $ac$ | + | + | - | - | + | + | - | - |
| $bc$ | + | - | + | - | + | - | + | - |
| $abc$ | + | + | + | + | + | + | + | + |

Los signos de los efectos principales (columna A,B y C) se determinan asignando un signo positivo al nivel superior y uno negativo al nivel inferior. Es decir, signo positivo si en la combinación de tratamiento aparece el efecto Factorial y negativo si no aparece. Una vez que se han establecido los signos para los efectos principales, los de las columnas restantes o

iteraciones se encuentran multiplicando los signos de renglón por renglón de las columnas correspondientes a la interacción que se desea obtener. Por ejemplo, los signos de la columna BC, se obtienen por medio del producto de los signos de la columnas B y C de cada uno de los renglones. Los signos de la columna de la interacción triple ABC, se obtienen multiplicando los signos de cualquier conjunto de columnas cuyo efecto es ABC. Es decir, los signos para ABC se pueden obtener de cualquier producto de signos de las columnas AxBxC, ABxC, ACxB o BCxA.
En esta tabla se pueden observar algunas propiedades interesantes como:

a) Cada columna tiene un número igual de signos positivos y negativos, con excepción de la columna I.
b) La suma de los productos de los signos de cualquier par de columnas siempre es cero; ya que las columnas en la tabla son ortogonales.

c) El producto de cualquier columna con la columna I, no produce ningún cambio en sus signos de dicha columna; ya que la columna I es el elemento identidad.
d) El producto de cualquier par de columnas siempre produce otra columna que se encuentra en la tabla. Ejemplo, AxC = AC, ACxC = $AC^2$ = A.

**Obtención de los Coeficientes de los Contrastes.**

Los coeficientes de los contrastes para cada efecto se obtienen de la tabla anterior simplemente multiplicando los signos de la columna apropiada por la correspondiente combinación de tratamientos, y luego sumándolos.

Por lo tanto, los contraste para un Diseño $2^3$ obtenidos por medio de la tabla anterior son:

$\text{Contraste}_A = -(1) + a - b + ab - c + ac - bc + abc$
$\text{Contraste}_B = -(1) - a + b + ab - c - ac + bc + abc$
$\text{Contraste}_C = -(1) - a - b - ab + c + ac + bc + abc$
$\text{Contraste}_{AB} = (1) - a - b + ab + c - ac - bc + abc$
$\text{Contraste}_{AC} = (1) - a + b - ab - c + ac - bc + abc$
$\text{Contraste}_{BC} = (1) + a - b - ab - c - ac + bc + abc$
$\text{Contraste}_{ABC} = -(1) + a + b - ab + c - ac - bc + abc$

Se puede observar que son los mismo resultados que se encontraron anteriormente.

La estimación de cualquier efecto o interacción se obtiene multiplicando las combinaciones de tratamientos por los signos de la columna correspondiente de interacción o efecto que se desea obtener su efecto, sumando el resultado para producir un contraste y dividiendo luego el doble del contraste entre el número total de ejecuciones en el Experimento.

$$\text{Efecto} = \frac{2(contraste)}{8n}$$

## 6. DISEÑO FACTORIAL GENERAL $2^k$

Cuando se considera un Diseño Factorial $2^k$, se esta generalizando los Diseños Factoriales de dos niveles que tienen $k$ factores. En general el Modelo Estadístico del Diseño Factorial $2^k$ tiene $k$ factores principales, $\binom{k}{2}$ interacciones de dos factores, $\binom{k}{3}$ interacciones de tres factores, ..., y una interacción de $k$ factores; es decir, que el Modelo completo tiene $2^k - 1$ efectos.

Las combinaciones de tratamientos pueden escribirse en el orden estándar, introduciendo los factores de uno en uno, combinado en forma sucesiva cada nuevo factor con aquellos introducidos anteriormente. Por ejemplo, el orden estándar para un Diseño $2^4$ es : *(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, y abcd.* En el cual *cd* representa la combinación de tratamientos en la que el Factor C y Factor D se encuentran en el nivel superior, mientras que A y B están en el nivel inferior.

Para estimar cualquier efecto o calcular su correspondiente Suma de Cuadrado es necesario determinar el contraste de cada uno de los efectos; los cuales se pueden obtener utilizando una tabla de signos positivos y negativos; pero para un número grande de factores (k grande), resulta demasiado tedioso o complicado; y por lo tanto, es preferible utilizar otro método para encontrar los contrastes.

A continuación se presenta una forma o método más general y sencillo para obtener los contraste de cada uno de los Efectos del Diseño Factorial.

En general, el contraste para el efecto AB....K, se obtiene desarrollando el segundo miembro de la siguientes ecuación.

$\text{Contraste}_{AB....K} = (a \pm 1)(b \pm 1)....(K \pm 1)$

Para encontrar un contraste determinado, en la ecuación anterior del conjunto de paréntesis se debe usar el signo negativo si el factor esta presente en el efecto (para todas aquellas letras que aparecen en el contraste) y positivo en caso contrario.

Al terminar de desarrollar los factores de la ecuación, se debe reemplazar el "1" por (1) en la expresión final.

Luego de haber obtenido los contrastes para los efectos, la estimación de los efectos y la Suma de Cuadrado de cada uno de ellos, se puede determinar mediante las siguientes expresiones generales.

La forma general para obtener los efectos, viene dada por:

$AB....K = \frac{2}{n2^k}\left(contraste_{AB....K}\right)$ ó $AB....K = \frac{1}{n2^{k-1}}\left(contraste_{AB....K}\right)$

Por lo tanto, la forma general de obtener las sumas de cuadrados, vendrá dada por: $SS_{AB....K} = \frac{1}{n2^k}\left(contraste_{AB....K}\right)^2$ , en las ecuaciones anteriores *n* es el número de réplicas.

**Ejemplo 3**

1) En un Diseño Factorial $2^3$, el contraste para la interacción BC, utilizando la ecuación anterior se obtiene de la siguiente manera:

   como $k$=3:

$$
\begin{aligned}
\text{Contraste}_{BC} &= (a + 1)(b - 1)(c - 1) \\
&= (ab - a + b - 1)(c - 1) \\
&= abc - ac + bc - c - ab + a - b + 1 \\
&= abc + bc + a + (1) - ab - b - ac - c
\end{aligned}
$$

En un Diseño Factorial de $2^5$ el contraste para la interacción BCDE, utilizando la ecuación anterior se obtiene de la siguiente manera:

$$
\begin{aligned}
\text{Contraste}_{BCDE} &= (a+1)(b-1)(c-1)(d-1)(e-1) \\
&= abcde + bcde + ade + de + ace + ce + abe + be + acd + dc + abd + bd + \\
&\quad abc + bc + a + (1) - acde - cde - abde + bde - abce - bce - ae - e - abce
\end{aligned}
$$

$$- bcd - ad - d - ac - c - ab - b$$

2) El efecto de BC de un Diseño Factorial $2^3$ se obtiene utilizando la expresión general anterior de la siguiente manera:

Como $k$=3, entonces

$$BC = \frac{2}{2x2^3}\left(Contraste_{BC}\right)$$

$$= \frac{2}{16}\left(abc + bc + a + (1) - ab - b - ac - c\right)$$

El efecto de BCDE en un Diseño Factorial $2^5$ se obtiene de la siguiente manera:

Como $k$=5, entonces

$$BCDE = \frac{2}{2x2^5}\left(Contraste_{BCDE}\right)$$

$$= \frac{2}{64}\begin{pmatrix} abcde + bcde + ade + de + ace + ce + abe + be + acd \\ + dc + abd + bd + abc + bc + a + (1) - acde - cde - abde \\ + bde - abce - bce - ae - e - abce - bcd - ad - d - ac - c - ab - b \end{pmatrix}$$

3) La Suma de cuadrados de BC en un Diseño Factorial $2^3$, se obtiene utilizando la fórmula general planteada anteriormente de la siguiente forma: Como $k$=3

$$SS_{BC} = \frac{1}{2x2^3}\left(contraste_{BC}\right)^2 = \frac{1}{16}\left(abc + bc + a + (1) - ab - b - ac - c\right)^2$$

La Suma de Cuadrado del efecto de BCDE en un Diseño Factorial $2^5$ se obtiene de la siguiente manera:

$$SS_{BCDE} = \frac{1}{64}\begin{pmatrix} abcde + bcde + ade + de + ace + ce + abe + be + acd \\ + dc + abd + bd + abc + bc + a + (1) - acde - cde - abde \\ + bde - abce - bce - ae - e - abce - bcd - ad - d - ac - c - ab - b \end{pmatrix}^2$$

**Tabla de Análisis de Varianza para el Diseño $2^k$**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad |
|---|---|---|
| K efectos principales | | |
| A | $SS_A$ | 1 |
| B | $SS_B$ | 1 |
| C | $SS_C$ | 1 |
| . | . | . |
| . | . | . |
| . | . | . |
| K | $SS_K$ | 1 |
| $\binom{k}{2}$ interacciones de dos factores | | |
| AB | $SS_{AB}$ | 1 |
| AC | $SS_{AC}$ | 1 |
| . | . | . |
| . | . | . |
| . | . | . |
| JK | $SS_{JK}$ | 1 |
| $\binom{k}{3}$ interacciones de Tres factores | | |
| ABC | $SS_{ABC}$ | 1 |
| ABD | $SS_{ABD}$ | 1 |
| . | . | . |
| . | . | . |
| . | . | . |
| IJK | $SS_{IJK}$ | 1 |
| . | . | . |
| . | . | . |
| . | . | . |

| | | |
|---|---|---|
| $\binom{k}{k}$ interacciones de $k$ factores | | |
| ABC…K | $SS_{ABC\ldots K}$ | 1 |
| Error | $SS_E$ | $2^k(n-1)$ |
| Total | $SS_T$ | $n2^k-1$ |

## 7. ALGORITMO DE YATES PARA UN DISEÑO $2^K$.

Como se puede observar que para encontrar los contrastes y sumas de cuadrados de los efectos, los métodos vistos anteriormente resultan muy tediosos cuando $k$ crece, incluyendo la tabla de signos.

Una técnica eficiente para calcular la estimación de los efectos y las correspondientes Sumas de Cuadrados en un Diseño Factorial $2^k$ fue propuesta por Yates (1937), el cual se presenta a continuación.

Se elabora una tabla de la siguiente manera:

1) La primera columna esta compuesta por las combinaciones de los tratamientos escritos en orden estándar.
2) Luego se coloca una segunda columna llamada "**Respuesta**" que contiene las observaciones (o total de observaciones) correspondientes a las combinaciones de tratamientos del renglón.
3) Se calcula la **columna (1)**, en la cual la primera mitad de ella, se obtiene sumando los valores de la columna respuesta por pares adyacente (dos a dos) y la segunda mitad cambiando el signo del primer valor de cada par de la columna Respuesta y sumando los pares adyacentes.
4) Se crea una **columna (2)**, la cual se obtiene a partir de la columna (1) en la misma forma como la columna (1) se obtuvo de la columna respuesta.

   Y así sucesivamente se van creando más columnas hasta el número de factores en estudio.

   En general para un Diseño Factorial $2^k$ deben construirse $k$ columnas de este tipo.

Por lo tanto, la columna $k$ es el contraste del efecto representado por las letras minúsculas al comienzo del renglón.

5) Para obtener la estimación del efecto se dividen los valores de la columna $k$ por $n2^{k-1}$ y se crea esta columna.
6) Se obtiene la columna de la Suma de Cuadrado de los efectos elevando al cuadrado los valores de la columna $k$, y dividiendo por $n2^k$.

**Nota:** Para la prueba parcial de los cálculos, se deben tomar en cuenta lo siguiente:

a) El primer valor de la columna $k$, siempre es igual a la suma de todas las observaciones.
b) La suma de los cuadrados de los elementos de la j-ésima columna, es igual a $2^j$ veces la suma de los cuadrados de los elementos de la columna de Respuesta.

**Ejemplo 4**

Considérese el Experimento para investigar las propiedades a la compresión de Mezcla de Cemento y Tierra; desarrollado en el Diseño Factorial $2^3$. Tomando los datos codificados.

**Solución**

**Datos.**

$k = 3 \quad , \quad n = 2$

**Pasos para la elaboración de la Tabla**

*1)* El orden estándar de las combinaciones de tratamiento de los Diseños Factorial $2^3$ es: *(1), a , b , ab , c , ac , bc , abc.*

2) Por lo tanto, con los datos del Experimento se elabora la columna "**Respuesta**" como se describe en la siguiente tabla.

| Combinación de Tratamiento | Réplicas | | Respuesta |
|---|---|---|---|
| | I | II | |
| (1) | 4.71 | 4.13 | 8.84 |
| *a* | 7.12 | 6.37 | 13.49 |
| *b* | 4.85 | 5.52 | 10.37 |
| *ab* | 7.12 | 7.89 | 15.01 |
| *c* | 3.85 | 4.34 | 8.19 |
| *ac* | 7.70 | 7.05 | 14.75 |

| | | | |
|---|---|---|---|
| $bc$ | 5.30 | 3.93 | 11.23 |
| $abc$ | 7.41 | 8.06 | 15.47 |

3) Forma de elaborar la columna (1)

| **Primera Mitad** | **Segunda Mitad** |
|---|---|
| 8.84 + 13.49 = 22.33 | -8.84 + 13.49 = 4.65 |
| 10.37 + 15.01= 25.38 | -10.37 + 15.01 = 4.64 |
| 8.19 + 14.75 = 22.94 | -8.19 + 14.75 = 6.56 |
| 11.23 + 15.47= 26.70 | -11.23 + 15.47 = 4.24 |

4) Forma de elaborar la columna (2).

| **Primera Mitad** | **Segunda Mitad** |
|---|---|
| 22.33 + 25.38 = 47.71 | -22.33 + 25.38 = 3.05 |
| 22.94 + 26.70 = 49.64 | -22.94 + 26.70 = 3.76 |
| 4.65 + 4.64 = 9.29 | -4.65 + 4.64 = -0.01 |
| 6.56 + 4.24 = 10.80 | -6.56 + 4.24 = -2.32 |

5) Forma de elaborar la columna (3)

| **Primera Mitad** | **Segunda Mitad** |
|---|---|
| 47.71 + 49.64 = 97.35 | -47.71 + 49.64 = 1.93 |
| 9.29 + 10.80 = 20.09 | -9.29 + 10.80 = 1.51 |
| 3.05 + 3.76 = 6.81 | -3.05 + 3.76 = 0.71 |
| -0.01 - 2.32 = -2.33 | 0.01 - 2.32 = -2.31 |

Como $k$ = 3, sólo serán tres columnas de este tipo.

6) Forma de elaborar la columna de la Estimación de los Efectos; esta columna se obtiene dividiendo los valores de la columna (3) por $n2^{k-1} = 2x2^2 = 2^3 = 8$.

A = $\frac{20.09}{8}$ = 2.51 B = $\frac{6.81}{8}$ = 0.85 AB = $\frac{-2.33}{8}$ = -0.29

C = $\frac{1.93}{8}$ = 0.24 AC = $\frac{1.51}{8}$ = 0.19 BC = $\frac{0.71}{8}$ = 0.09

ABC = $\frac{-2.31}{8}$ = -0.29

7) Forma de elaborar la columna de las Sumas de Cuadrados; esta columna se obtiene elevando al cuadrado los valores de la columna (3) luego dividirlos por $n2^k = 2x2^3 = 2^4 = 16$.

$SS_A = \frac{(20.09)^2}{16} = 25.22$ $SS_B = \frac{(6.81)^2}{16} = 2.90$ $SS_{AB} = \frac{(-2.33)^2}{16} = 0.34$

$SS_C = \frac{(1.93)^2}{16} = 0.23$ $SS_{AC} = \frac{(1.51)^2}{16} = 0.14$ $SS_{BC} = \frac{(0.71)^2}{16} = 0.03$

$SS_{ABC} = \frac{(-2.31)^2}{16} = 0.33$

A continuación se presenta la Tabla completa del Algoritmo de Yates que resulto de haber realizado los pasos anteriores:

| Combinación de Tratamientos | Respuesta | (1) | (2) | (3) | Efecto | Estimación del Efecto $(3)\div2x2^2 = (3)\div8$ | Sumas de Cuadrados $(3)^2\div2x2^3 = (3)^2\div16$ |
|---|---|---|---|---|---|---|---|
| (1) | 8.84 | 22.33 | 47.71 | 97.35 | I | ——— | ——— |
| $a$ | 13.49 | 25.38 | 49.64 | 20.09 | A | 2.51 | 25.22 |
| $b$ | 10.37 | 22.94 | 9.29 | 6.81 | B | 0.85 | 2.90 |
| $ab$ | 15.01 | 26.70 | 10.80 | -2.33 | AB | -0.29 | 0.34 |
| $c$ | 8.19 | 4.65 | 3.05 | 1.93 | C | 0.24 | 0.23 |
| $ac$ | 14.75 | 4.64 | 3.76 | 1.51 | AC | 0.19 | 0.14 |
| $bc$ | 11.23 | 6.56 | -0.01 | 0.71 | BC | 0.09 | 0.03 |
| $abc$ | 15.47 | 4.24 | -2.32 | -2.31 | ABC | -0.29 | 0.33 |

**Conclusiones**

- La suma total de observaciones es $N$ = 97.35; por lo tanto coincide con el primer valor de la columna (3).

- Los valores de los efectos y las Sumas de Cuadros coinciden con los valores encontrados cuando se desarrollo en los ejemplos anteriores.
- Teniendo ya estos valores se puede formar la Tabla de Análisis de Varianza.

## 8. PROBLEMAS RESUELTOS

**PROBLEMA 1**

Un ingeniero industrial que trabaja en una embotelladora está interesado en el efecto de dos tipos de botellas de 32 onzas sobre el tiempo de reparto de cajas de 12 botellas de este producto. Los dos tipos de botella son de plástico y de vidrio, y se utilizan dos repartidores para realizar una tarea que consiste en mover 40 cajas del producto una distancia de 50 pies sobre un carrito repartidor, y acomodarlas. Se realizan cuatro réplicas de un Diseño Factorial $2^2$; y los tiempos observados se muestran en la siguiente tabla.

| Tipo de botella | Operario | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| Vidrio | 5.12 | 4.89 | 6.65 | 6.24 |
| | 4.98 | 5.00 | 5.49 | 5.55 |
| Plástico | 4.95 | 4.43 | 5.28 | 4.91 |
| | 4.27 | 4.25 | 4.75 | 4.71 |

Analice los datos y determine las conclusiones apropiadas.

**Solución**

**Variable Respuesta:** Tiempo de reparto de cajas de 12 botellas.

**Hipótesis a probar**

**Efectos Principales**

$H_0$ : El tipo de botella no influye significativamente en el tiempo de reparto de las cajas de 12 botellas.

$H_1$ : El tipo de botella influye significativamente en el tiempo de reparto de las cajas de 12 botellas.

$H_0$ : El operario no influye significativamente en el tiempo de reparto de las cajas de 12 botellas.

$H_1$ : El operario influye significativamente en el tiempo de reparto de las cajas de 12 botellas.

**Interacción**

$H_0$ : El tipo de botella y el operario no influyen significativamente en el tiempo de reparto de las cajas de 12 botellas.

$H_1$ : El tipo de botella y el operario influyen significativamente en el tiempo de reparto de las cajas de 12 botellas.

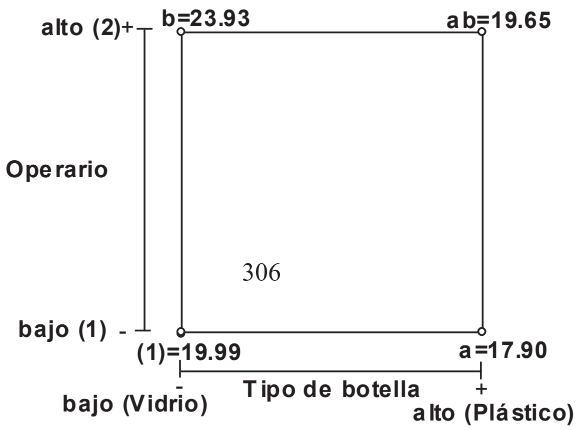
Como puede observarse este Experimento es un Diseño Factorial $2^2$, donde existen dos factores, factor A (Tipo de botella) y factor B (Operario) ambos a dos niveles.
El factor A con nivel bajo que es Vidrio y el nivel alto Plástico.
El factor B con nivel bajo que es "1" y el nivel alto es "2".
Realizándose este Experimento cuatro veces.

De la tabla de datos es posible obtener una tabla de combinaciones de tratamientos, como la que se muestra a continuación:

| Combinación de Tratamientos | Réplicas | | | | Total | Simbología |
|---|---|---|---|---|---|---|
| | I | II | III | IV | | |
| A bajo, B bajo | 5.12 | 4.98 | 4.89 | 5.00 | 19.99 | (1) |
| A alto , B bajo | 4.95 | 4.27 | 4.43 | 4.25 | 17.90 | a |
| A bajo, B alto | 6.65 | 5.49 | 6.24 | 5.55 | 23.93 | b |
| A alto, B alto | 5.28 | 4.75 | 4.91 | 4.71 | 19.65 | ab |

$y_{\dots} = 81.47$

Estas combinaciones de tratamientos Geométricamente se representa de la siguiente



manera:

$n$ : el número de réplicas ($n$=4)

**A continuación se encuentran los contrastes:**

$Contraste_A$ = $ab + a - b - (1)$ = 19.65 + 17.90 – 23.93 – 19.99 = -6.37

$Contraste_B$ = $ab + b - a - (1)$ = 19.65 + 23.93 – 17.90 – 19.99 = 5.69

$Contraste_{AB}$ = $ab + (1) - a - b$ = 19.65 + 19.99 – 17.90 – 23.93 = -2.19

**Sumas de cuadrados.**

$$SS_A = \frac{(Contraste_A)^2}{4n} = \frac{(-6.37)^2}{4x4} = \frac{40.5769}{16} = 2.54$$

$$SS_B = \frac{(Contraste_B)^2}{4n} = \frac{(5.69)^2}{4x4} = \frac{32.3761}{16} = 2.02$$

$$SS_{AB} = \frac{(Contraste_{AB})^2}{4n} = \frac{(-2.19)^2}{4x4} = \frac{4.7961}{16} = 0.30$$

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{4} y_{ijk}^2 - \frac{y_{...}^2}{4n} = 421.1855 - \frac{(81.47)^2}{4x4} = 421.1855 - 414.83505 = 6.35$$

$SS_E$ = $SS_T$ - $SS_A$ - $SS_B$ - $SS_{AB}$

= 6.35 – 2.54 – 2.02 – 0.30

= 1.49

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Tipo de botella | 2.54 | 1 | 2.54 | 21.16 |
| Operario | 2.02 | 1 | 2.02 | 16.83 |
| Tipo de BotellaxOperario | 0.30 | 1 | 0.30 | 2.50 |
| Error | 1.49 | 12 | 0.12 | |
| Total | 6.35 | 15 | | |

Tomando $\alpha$ = 0.05 para obtener el F de tablas con un grado de libertad del numerador y 4($n$-1)= 4(4-1) = 12 en el denominador, se tiene que:

$F_{\alpha,1,4(n-1)}$ = $F_{0.05,1,12}$ = 4.75

**Conclusiones**

- Para el caso del factor Tipo de botella el $F_0 > F_{Tablas}$ (21.16>4.75); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que el tipo de botella influye significativamente en el tiempo de reparto de las cajas de 12 botellas.
- Para el caso del factor Operario el $F_0 > F_{Tablas}$ (16.83>4.75); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que el operario influye significativamente en el tiempo de reparto de las cajas de 12 botellas.
- Para el caso de la interacción Tipo de botella y Operario el $F_0 < F_{Tablas}$ (2.50<4.75); por lo tanto, se acepta $H_0$ y se rechaza $H_1$; es decir, que el tipo de botella y el operario no influyen significativamente en el tiempo de reparto de las cajas de 12 botellas.

**<u>PROBLEMA 2</u>**

En un hospital se lleva a cabo un experimento para investigar como influyen tres factores en la tensión arterial de un grupo de personas. Los factores fueron: A: Nicotina(fuma, no fuma), B: Edad(menor de 40 años , mayor de 40 años), C: Sexo(hombre, mujer).

Los datos obtenidos se presentan en la siguiente tabla:

| Nicotina (Factor A) | Edad (Factor B) | | | |
|---|---|---|---|---|
| | Menor de 40 años | | Mayor de 40 años | |
| | Sexo (Factor C) | | Sexo (Factor C) | |
| | Hombre | Mujer | Hombre | Mujer |
| Fuma | 110<br>110 | 120<br>120 | 145<br>150 | 120<br>120 |
| No fuma | 110<br>115 | 90<br>95 | 130<br>140 | 130<br>130 |

Analice los datos y determine las conclusiones apropiadas.

**Solución.**

Este experimento es un Diseño Factorial $2^3$; ya que se analizan tres factores y cada uno tiene dos niveles (alto y bajo).

Para el factor A el nivel bajo Fuma y el alto No fuma.

Para el factor B el nivel bajo Menor de 40 años y el alto Mayor de 40 anos.

Para el factor C el nivel bajo Hombre y el alto Mujer.

**Datos**

$a = b = c = 2$ , $n = 2$

**Forma verbal de las Hipótesis a probar:**

a) $H_0$: La nicotina no influye significativamente en la tensión arterial de las personas.

$H_1$: La nicotina influye significativamente en la tensión arterial de las personas.

b) $H_0$: La edad no influye significativamente en la tensión arterial de las personas.
$H_1$: La edad influye significativamente en la tensión arterial de las personas.

c) $H_0$: El sexo no influye significativamente en la tensión arterial de las personas.
$H_1$: El sexo influye significativamente en la tensión arterial de las personas.

d) $H_0$: La nicotina y la edad no influyen significativamente en la tensión arterial de las personas.
$H_1$: La nicotina y la edad influyen significativamente en la tensión arterial de las personas.

e) $H_0$: La nicotina y el sexo no influyen significativamente en la tensión arterial de las personas.
$H_1$: La nicotina y el sexo influyen significativamente en la tensión arterial de las personas.

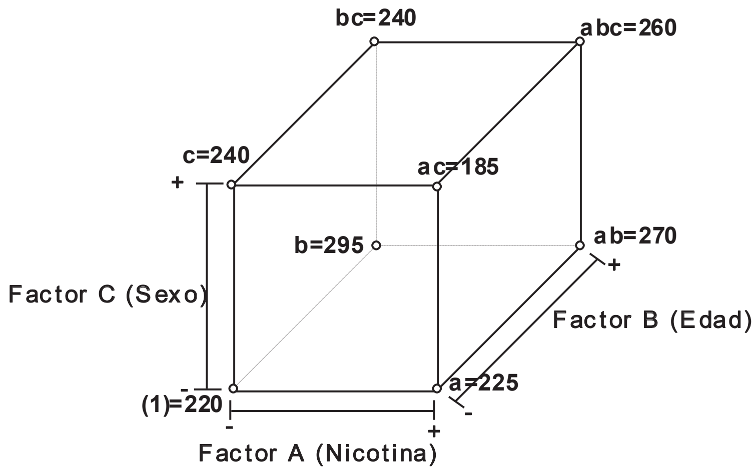
f) $H_0$: La edad y el sexo no influyen significativamente en la tensión arterial de las personas.
$H_1$: La edad y el sexo influyen significativamente en la tensión arterial de las personas.

g) $H_0$: La nicotina, la edad y el sexo no influyen significativamente en la tensión arterial de las personas.
$H_1$: La nicotina, la edad y el sexto influyen significativamente en la tensión arterial de las personas.

La tabla de combinaciones de tratamientos para este Diseño Factorial es la siguiente:

| Combinación de Tratamiento | Réplicas | | Total | Simbología |
|---|---|---|---|---|
| | I | II | | |
| A bajo, B bajo, C bajo | 110 | 110 | 220 | (1) |
| A alto, B bajo, C bajo | 110 | 115 | 225 | $a$ |
| A bajo, B alto, C bajo | 145 | 150 | 295 | $b$ |
| A alto, B alto, C bajo | 130 | 140 | 270 | $ab$ |
| A bajo, B bajo, C alto | 120 | 120 | 240 | $c$ |
| A alto, B bajo, C alto | 90 | 95 | 185 | $ac$ |
| A bajo, B alto, C alto | 120 | 120 | 240 | $bc$ |
| A alto, B alto, C alto | 130 | 130 | 260 | $abc$ |

Geométricamente las combinaciones de tratamientos que se encuentran en la tabla anterior, se representan de la siguiente forma:




Para la estimación de los efectos promedios de los factores principales e interacciones se utilizará el primer método detallado anteriormente:

**Efecto de la Nicotina**

A = $\frac{1}{4n}\left[a-(1)+ab-b+ac-c+abc-bc\right]$

= $\frac{1}{4x2}\left[225-220+270-295+185-240+260-240\right]$

= $\frac{1}{8}(-55)$ = -6.875

**Efecto de la Edad**

B = $\frac{1}{4n}\left[b + ab + bc + abc - (1) - a - c - ac\right]$

B = $\frac{1}{4x2}\left[295 + 270 + 240 + 260 - 220 - 225 - 240 - 185\right]$

= $\frac{1}{8}(195)$ = 24.375

**Efecto del Sexo**

C = $\frac{1}{4n}\left[c + ac + bc + abc - (1) - a - b - ab\right]$

= $\frac{1}{4x2}\left[240 + 185 + 240 + 260 - 220 - 225 - 295 - 270\right]$

= $\frac{1}{8}(-85)$ = -10.625

**Efecto Nicotina y Edad**

AB = $\frac{1}{4n}\left[abc - bc + ab - b - ac + c - a + (1)\right]$

= $\frac{1}{4x2}\left[260 - 240 + 270 - 295 - 185 + 240 - 225 + 220\right]$

= $\frac{1}{8}(45)$ = 5.625

**Efecto Nicotina y Sexo**

AC = $\frac{1}{4n}\left[(1) - a + b - ab - c + ac - bc + abc\right]$

= $\frac{1}{4x2}\left[220 - 225 + 295 - 270 - 240 + 185 - 240 + 260\right]$

= $\frac{1}{8}(-15)$ = -1.875

**Efecto Edad y Sexo**

BC = $\frac{1}{4n}\left[(1) + a - b - ab - c - ac + bc + abc\right]$

= $\frac{1}{4x2}\left[220 + 225 - 295 - 270 - 240 - 185 + 240 + 260\right]$

= $\frac{1}{8}(-45)$ = -5.625

**Efecto Nicotina, Edad y Sexo**

ABC = $\frac{1}{4n}\left[abc - bc - ac + c - ab + b + a - (1)\right]$

= $\frac{1}{4x2}\left[260 - 240 - 185 + 240 - 270 + 295 + 225 - 220\right]$

$= \frac{1}{8}(105) = 13.125$

Los valores encontrados para cada uno de los efectos significan:

- El efecto de la Nicotina es negativo (–6.875); significa que el efecto de la Nicotina no afecta la tensión arterial de las personas.
- El efecto de la Edad es positivo e igual a 24.375; significa que se espera un aumento de 24.375 en la tensión de las personas al tener una edad de mas de 40 años.
- El efecto del Sexo es negativo (-10.625); significa que el efecto del sexo no afecta la tensión arterial de las personas.
- El efecto de la Nicotina y Edad es positivo e igual a 5.625; significa que se espera que la combinación de ambos, produzca un aumento de 5.625 en la tensión arterial de las personas.
- El efecto de la Nicotina y el Sexo es negativo (-1.875); significa que la combinación de ambos no afecta la tensión arterial de las personas.
- El efecto de la Edad y el Sexo es negativo (-5.625); significa que la combinación de ambos no afecta la tensión arterial de las personas.
- El efecto de la Nicotina, Edad y el Sexo es positivo e igual a 13.125; significa que se espera que la combinación de los tres factores, produzca un aumento en la tensión arterial de las personas en 13.125.

**Encontrando los contrastes para los efectos principales e interacciones:**

Ya que los contrastes es el resultado de lo que se encuentra entre paréntesis de los efectos; se tiene:

$Contraste_A$ = - 55 , $Contraste_B$ = 195 , $Contraste_C$ = -85

$Contraste_{AB}$ = 45 , $Contraste_{AC}$ = -15 , $Contraste_{BC}$ = -45

$Contraste_{ABC}$ = 105

**Sumas de Cuadrados**

$$SS_A = \frac{(Contraste_A)^2}{8n} = \frac{(-55)^2}{8x2} = \frac{3025}{16} = 189.06$$

$$SS_B = \frac{(Contraste_B)^2}{8n} = \frac{(195)^2}{8x2} = \frac{38025}{16} = 2376.562$$

$$SS_C = \frac{(Contraste_C)^2}{8n} = \frac{(-85)^2}{8x2} = \frac{7225}{16} = 451.56$$

$$SS_{AB} = \frac{(Contraste_{AB})^2}{8n} = \frac{(45)^2}{8x2} = \frac{2025}{16} = 126.56$$

$SS_{AC} = \dfrac{(Contraste_{AC})^2}{8n} = \dfrac{(-15)^2}{8x2} = \dfrac{225}{16} = 14.06$

$SS_{BC} = \dfrac{(Contraste_{AB})^2}{8n} = \dfrac{(-45)^2}{8x2} = \dfrac{2025}{16} = 126.56$

$SS_{ABC} = \dfrac{(Contraste_{ABC})^2}{8n} = \dfrac{(105)^2}{8x2} = \dfrac{11025}{16} = 689.06$

**Suma de cuadrado totales:**

$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} y_{ijkl}^2 - \dfrac{y_{....}^2}{8x2} = 238075 - \dfrac{(1935)^2}{16} = 238075 - 234014.06 = 4060.94$

**Suma de cuadrado del Error**

$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC}$

$= 4060.94 - 189.06 - 2376.562 - 451.56 - 126.56 - 14.06 - 126.56 - 689.06$

$= 87.518$

**Tabla de Análisis de Varianza**

| Fuente de Variación | Suma de Cuadrados | Grados de Libertad | Media de Cuadrados | $F_0$ |
|---|---|---|---|---|
| Nicotina | 189.06 | 1 | 189.06 | 17.28 |
| Edad | 2376.562 | 1 | 2376.562 | 217.24 |
| Sexo | 451.56 | 1 | 451.56 | 41.28 |
| NicotinaxEdad | 126.56 | 1 | 126.56 | 11.57 |
| NicotinaxSexo | 14.06 | 1 | 14.06 | 1.28 |
| EdadxSexo | 126.56 | 1 | 126.56 | 11.57 |
| NicotinaxEdadxSexo | 689.06 | 1 | 689.06 | 62.98 |
| Error | 87.518 | 8 | 10.94 | |
| Total | 4060.94 | 15 | | |

Tomando $\alpha = 0.05$ para obtener el F de tablas con un grado de libertad del numerador y $8(n\text{-}1)= 8(2\text{-}1) = 8$ en el denominador, se tiene que:

$F_{\alpha,1,8(n-1)} = F_{0.05,1,8} = 5.32$

**Conclusiones**

a) Para el caso del factor A (Nicotina) el $F_0 > F_{Tablas}$ (17.28>5.32); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que la nicotina influye significativamente en la tensión arterial de las personas.

b) Para el caso del factor B (Edad) el $F_0 > F_{Tablas}$ (217.24>5.32); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que la Edad influye significativamente en la tensión arterial de las personas.

c) Para el caso del factor C (Sexo) el $F_0 > F_{Tablas}$ (41.28>5.32); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que el Sexo influye significativamente en la tensión arterial de las personas.

d) Para el caso de la interacción del Factor A y el factor B (Nicotina*x*Edad) el $F_0 > F_{Tablas}$ (11.57>5.32); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que la Nicotina y la Edad influyen significativamente en la tensión arterial de las personas.

e) Para el caso de la interacción del Factor A y el factor C (Nicotina*x*Sexo) el $F_0 < F_{Tablas}$ (1.28<5.32); por lo tanto, no se rechaza $H_0$; es decir, que la Nicotina y el Sexo no influyen significativamente en la tensión arterial de las personas.

f) Para el caso de la interacción del Factor B y el factor C (Edad*x*Sexo) el $F_0 > F_{Tablas}$ (11.57>5.32); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que la Edad y el Sexo influyen significativamente en la tensión arterial de las personas.

g) Para el caso de la interacción del Factor A, factor B y el factor C (Nicotina*x*Edad*x*Sexo) el $F_0 > F_{Tablas}$ (62.98>5.32); por lo tanto, se rechaza $H_0$ y se acepta $H_1$; es decir, que la Nicotina, la Edad y el Sexo influyen significativamente en la tensión arterial de las personas.

**PROBLEMA 3**

Para el Diseño Factorial planteado en el problema 1; encontrar la estimación de los efectos promedios de los factores principales e interacción y analice los resultados que se obtenidos.

**Datos**

$n = 4$ , $a = 17.90$ , $b = 23.93$ , $ab = 19.65$ , $(1) = 19.99$

**Solución**

**Estimación del Efecto de los Tipo de botella (Factor A)**

$$A = \frac{1}{2n}[ab + a - b - (1)] = \frac{1}{2x4}(19.65 + 17.90 - 23.93 - 19.99) = \frac{1}{8}(-6.37) = -0.80$$

**Estimación del Efecto de los Operarios (Factor B)**

$$B = \frac{1}{2n}[ab + b - a - (1)] = \frac{1}{2x4}(19.65 + 23.93 - 17.90 - 19.99) = \frac{1}{8}(5.69) = 0.71$$

**Estimación del Efecto del Tipo de botella y Operario (Factor A y B)**

$$AB = \frac{1}{2n}[ab + (1) - a - b] = \frac{1}{2x4}(19.65 + 19.99 - 17.90 - 23.93) = \frac{1}{8}(-2.19) = -0.27$$

**Los valores encontrados significan:**

- El valor del efecto del factor A (Tipo de botella) es negativo e igual a −0.80; lo que significa que el tipo de botella no afecta al tiempo de reparto de las cajas de 12 botellas.
- El valor del efecto del factor B (Operario) es positivo e igual a 0.71; lo que significa que se espera un aumento de 0.71 en el tiempo de reparto de las cajas de 12 botellas.
- El valor del efecto de la interacción del Factor A y el Factor B es negativo e igual a −0.27; lo que significa que la combinación de los factores no afectan en el tiempo de reparto de las cajas de 12 botellas.

**PROBLEMA 4**

Para el Diseño Factorial planteado en el problema 2; encontrar las estimaciones de los Efectos y las Sumas de Cuadrados utilizando el Algoritmo de Yates.

**Solución**

**Datos.**

$k = 3$ , $n = 2$

**Pasos para la elaboración de la Tabla**

1) El orden estándar de las combinaciones de tratamientos de los Diseños $2^3$ es:*(1)*, $a$ , $b$ , $ab$ , $c$ , $ac$ , $bc$ , $abc$.

2) Por lo tanto, con los datos del Experimento se elabora la columna "Respuesta" como se describe en la siguiente tabla.

| Combinación de Tratamiento | Réplicas | | Respuesta |
|---|---|---|---|
| | I | II | |
| (1) | 110 | 110 | 220 |
| a | 110 | 115 | 225 |
| $b$ | 145 | 150 | 295 |
| $ab$ | 130 | 140 | 270 |
| $c$ | 120 | 120 | 240 |

| | | | |
|---|---|---|---|
| $ac$ | 90 | 95 | 185 |
| $bc$ | 120 | 120 | 240 |
| $abc$ | 130 | 130 | 260 |

3) Forma de elaborar la columna (1)

| **Primera Mitad** | **Segunda Mitad** |
|---|---|
| 220 + 225 = 445 | -220 + 225 = 5 |
| 295 + 270 = 565 | -295 + 270 = - 25 |
| 240 + 185 = 425 | -240 + 185 = - 55 |
| 240 + 260 = 500 | -240 + 260 = 20 |

4) Forma de elaborar la columna (2).

| **Primera Mitad** | **Segunda Mitad** |
|---|---|
| 445 + 565 = 1010 | -445 + 565 = 120 |
| 425 + 500 = 925 | -425 + 500 = 75 |
| 5 + (-25) = -20 | -5 + (-25) = -30 |
| -55 + 20 = -35 | -(-55) + 20 =75 |

5) Forma de elaborar la columna (3)

| **Primera Mitad** | **Segunda Mitad** |
|---|---|
| 1010 + 925 = 1935 | -1010 + 925 = -85 |
| (-20) + (-35) = -55 | -(-20) + (-35) = -15 |
| 120 + 75 = 195 | -120 + 75 = -45 |
| (-30) + 75 = 45 | -(-30) + 75 =105 |

como $k$ = 3, sólo serán tres columnas de este tipo.

6) Forma de elaborar la columna de la estimación de los efectos; esta columna se obtiene dividiendo los valores de la columna (3) por $n2^{k-1} = 2x2^2 = 8$.

A = $\frac{-55}{8}$ = -6.875 B = $\frac{195}{8}$ = 24.375 AB = $\frac{45}{8}$ = 5.625

C = $\frac{-85}{8}$ = - 10.625 AC = $\frac{-15}{8}$ = -1.875 BC = $\frac{-45}{8}$ = -5.625

$$ABC = \frac{105}{8} = 13.125$$

7) Forma de elaborar la columna de las Sumas de Cuadrados; esta columna se obtiene elevando al cuadrado los valores de la columna (3) luego dividirlos por $n2^k = 2x2^3 = 16$.

$SS_A = \frac{(-55)^2}{16} = 189.06$ $SS_B = \frac{(195)^2}{16} = 2376.56$ $SS_{AB} = \frac{(45)^2}{16} = 126.56$

$SS_C = \frac{(-85)^2}{16} = 451.56$ $SS_{AC} = \frac{(-15)^2}{16} = 14.06$ $SS_{BC} = \frac{(-45)^2}{16} = 126.56$

$SS_{ABC} = \frac{(105)^2}{16} = 689.06$

A continuación se presenta la Tabla completa del Algoritmo de Yates que resultó de haber realizado los pasos anteriores:

| Combinación de Tratamientos | Respuesta | (1) | (2) | (3) | Efecto | Estimación del Efecto $(3) \div 2x2^2 = (3) \div 8$ | Sumas de Cuadrados $(3)^2 \div 2x2^3 = (3)^2 \div 16$ |
|---|---|---|---|---|---|---|---|
| (1) | 220 | 445 | 1010 | 1935 | I | ― | ― |
| *a* | 225 | 565 | 925 | -55 | A | -6.875 | 189.06 |
| *b* | 295 | 425 | -20 | 195 | B | 24.375 | 2376.56 |
| *ab* | 270 | 500 | -35 | 45 | AB | 5.625 | 126.56 |
| *c* | 240 | 5 | 120 | -85 | C | -10.625 | 451.56 |
| *ac* | 185 | -25 | 75 | -15 | AC | -1.875 | 14.06 |
| *bc* | 240 | -55 | -30 | -45 | BC | -5.625 | 126.56 |
| *abc* | 260 | 20 | 75 | 105 | ABC | 13.125 | 689.06 |

**Conclusiones**

- La suma total de observaciones es $N = 1935$; por lo tanto coincide con el primer valor de la columna (3).

- Los valores de los efectos y las Sumas de Cuadros coinciden con los valores encontrados cuando se desarrollo este Diseño Factorial en el problema 2.
- Teniendo ya estos valores se puede formar la Tabla de Análisis de Varianza.

## PROBLEMA 5

Encontrar los contrastes de un Diseño Factorial $2^4$, utilizando la tabla de signos positivos y negativos.

### Solución

Como en un Diseño Factorial $2^4$ existen cuatro factores (A,B,C y D) con dos niveles cada uno; entonces las combinaciones de tratamientos escritas en orden estándar son:$(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd$.

La tabla de signos positivos y negativos de estas combinaciones de tratamientos se presenta a continuación.

| Combinaciones de Tratamientos | Efecto Factorial | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | A | B | AB | C | AC | BC | ABC | D | AD | BD | ABD | CD | ACD | BCD | ABCD |
| (1) | + | - | - | + | - | + | + | - | - | + | + | - | + | - | - | + |
| *a* | + | + | - | - | - | - | + | + | - | - | + | + | + | + | - | - |
| *b* | + | - | + | - | - | + | - | + | - | + | - | + | + | - | + | - |
| *ab* | + | + | + | + | - | - | - | - | - | - | - | - | + | + | + | + |
| *c* | + | - | - | + | + | - | - | + | - | + | + | - | - | + | + | - |
| *ac* | + | + | - | - | + | + | - | - | - | - | + | + | - | - | + | + |
| *bc* | + | - | + | - | + | - | + | - | - | + | - | + | - | + | - | + |
| *abc* | + | + | + | + | + | + | + | + | - | - | - | - | - | - | - | - |
| *d* | + | - | - | + | - | + | + | - | + | - | - | + | - | + | + | - |
| *ad* | + | + | - | - | - | - | + | + | + | + | - | - | - | - | + | + |
| *bd* | + | - | + | - | - | + | - | + | + | - | + | - | - | + | - | + |
| *abd* | + | + | + | + | - | - | - | - | + | + | + | + | - | - | - | - |
| *cd* | + | - | - | + | + | - | - | + | + | - | - | + | + | - | - | + |
| *acd* | + | + | - | - | + | + | - | - | + | + | - | - | + | + | - | - |
| *bcd* | + | - | + | - | + | - | + | - | + | - | + | - | + | - | + | - |
| *abcd* | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |

Los signos de los efectos principales (columnas A, B, C y D) se determinan asignando un signo **positivo**, si en la combinación de tratamiento aparece el efecto factorial y **negativo** si no aparece; y los signos de las columnas restantes o iteraciones se encuentran multiplicando los signos de renglón por renglón de las columnas correspondientes a la interacción que se desea obtener.

**A continuación se obtienen los Coeficientes de los Contrastes de la siguiente manera:**

Los coeficientes de los contrastes para cada efecto se obtienen de la tabla anterior simplemente multiplicando los signos de la columna apropiada por la correspondiente combinación de tratamientos, y luego sumándolos.

Por lo tanto, los contraste para este Diseño Factorial $2^3$ obtenidos por medio de la tabla anterior son:

$\text{Contraste}_{A}= -(1) + a - b+ ab - c+ ac - bc + abc - d + ad - bd + abd- cd + acd- bcd + abcd$

$\text{Contraste}_{B}= -(1) - a + b + ab - c - ac + bc + abc - d - ad + bd + abd- cdd - acd+ bcd + abcd$

$\text{Contraste}_{AB} = (1) - a - b + ab + c - ac - bc + abc + d - ad - bd + abd + cd - acd - bcd + abcd$

$\text{Contraste}_{C} = -(1) - a - b - ab + c + ac + bc + abc - d - ad - bd - abd + cd + acd + bcd + abcd$

$\text{Contraste}_{AC} = (1) - a + b - ab - c + ac - bc + abc + d - ad + bd - abd - cd + acd - bcd + abcd$

$\text{Contraste}_{BC} = (1) + a - b - ab - c - ac + bc + abc + d + ad - bd - abd - cd - acd + bcd + abcd$

$\text{Contraste}_{ABC} = -(1) + a + b - ab + c- ac- bc + abc - d + ad + bd - abd + cd - acd- bcd+ abcd$

$\text{Contraste}_{D} = -(1) - a - b - ab - c - ac - bc - abc + d + ad + bd + abd + cd+ acd + bcd + abcd$

$\text{Contraste}_{AD}= (1) - a + b - ab + c - ac + bc - abc - d+ ad - bd + abd - cd + acd - bcd + abcd$

$\text{Contraste}_{BD}= (1) + a - b - ab + c + ac - bc - abc - d - ad + bd + abd - cd - acd + bcd + abcd$

$\text{Contraste}_{ABD}= -(1) + a + b - ab - c + ac+ bc - abc + d - ad - bd + abd + cd- acd - bcd + abcd$

$\text{Contraste}_{CD}= (1) + a + b + ab - c - ac - bc - abc - d - ad - bd - abd + cd+ acd + bcd + abcd$

$\text{Contraste}_{ACD}= -(1) + a - b + ab + c - ac + bc - abc + d - ad + bd - abd - cd+ acd - bcd+ abcd$

$\text{Contraste}_{BCD}= -(1) - a + b + ab + c + ac- bc - abc + d+ ad - bd - abd - cd- acd+ bcd+ abcd$

$\text{Contraste}_{ABCD} = (1) - a - b + ab - c + ac + bc - abc - d + ad + bd - abd + cd - acd - bcd + abcd$

# A N E X O S

# APÉNDICE DE LA UNIDAD PROGRAMÁTICA II
# ( DISEÑOS UNIFACTORIALES )

## (1) Descomposición de la suma total de cuadrados.

### Caso Balanceado.

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

**Se puede escribir como:**

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left[(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})\right]^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n}\left[(\bar{y}_{i.} - \bar{y}_{..})^2 + 2\,(\bar{y}_{i.} - \bar{y}_{..})\,(y_{ij} - \bar{y}_{i.}) + (\bar{y}_{ij} - \bar{y}_{i.})^2\right]$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) + \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{ij} - \bar{y}_{i.})^2$$

$$= n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

**el segundo término de esta expresión es cero ya que:**

$$\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.}) = \sum_{j=1}^{n} y_{ij} - \sum_{j=1}^{n}\bar{y}_{i.} = y_{i.} - n\,\bar{y}_{i.} = y_{i.} - n\left(\frac{y_{i.}}{n}\right) = y_{i.} - y_{i.} = 0$$

**por lo tanto, nos queda:**

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

**L.q.q.d**

## (2) Deducción de las fórmulas Matemáticas de las Sumas de Cuadrados

### Suma de Cuadrado Total

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}^2 - 2y_{ij}\,\bar{y}_{..} + \bar{y}_{..}^{\,2})$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \sum_{i=1}^{a}\sum_{j=1}^{n} 2y_{ij}\,\bar{y}_{..} + \sum_{i=1}^{a}\sum_{j=1}^{n}\bar{y}_{..}^{\,2}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - 2\bar{y}_{..}\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} + an\,\bar{y}_{..}^{\,2}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - 2\,\bar{y}_{..}\,y_{..} + an\,\bar{y}_{..}^{\,2} \qquad , \text{ con } \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} = y_{..}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - 2\bar{y}_{..}(N\bar{y}_{..}) + N\bar{y}_{..}^2 \quad , \quad \text{con } N\bar{y}_{..} = y_{..} \quad y \quad N = an$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - 2N\bar{y}_{..}^2 + N\bar{y}_{..}^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - N\bar{y}_{..}^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - N\left(\frac{y_{..}^2}{N^2}\right) \quad , \text{con } \bar{y}_{..}^2 = \frac{y_{..}^2}{N^2}$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$$

**L.q.q.d**

**(3) Suma de Cuadrados de Tratamientos**

**Tomando el primer término de la descomposición de las sumas de Cuadrados; se tiene:**

$$SS_{Tratamientos} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= n\sum_{i=1}^{a}(\bar{y}_{i.} - 2\bar{y}_{i.}\bar{y}_{..} + \bar{y}_{..}^2)$$

$$= n\left[\sum_{i=1}^{a}\bar{y}_{i.}^2 - 2\bar{y}_{..}\sum_{i=1}^{a}\bar{y}_{i.} + \sum_{i=1}^{a}\bar{y}_{..}^2\right] \quad , \quad \text{con } \bar{y}_{i.} = \frac{y_{i.}}{n}$$

$$= n\left[\sum_{i=1}^{a}\bar{y}_{i.}^2 - 2\bar{y}_{..}\sum_{i=1}^{a}\frac{y_{i.}}{n} + a\,\bar{y}_{..}^2\right]$$

$$= n\left[\sum_{i=1}^{a}\bar{y}_{i.}^2 - 2\frac{\bar{y}_{..}}{n}\sum_{i=1}^{a}y_{i.} + a\,\bar{y}_{..}^2\right]$$

$$= n\left[\sum_{i=1}^{a}\bar{y}_{i.}^2 - 2\frac{\bar{y}_{..}}{n}y_{..} + a\,\bar{y}_{..}^2\right] \quad , \quad \text{con } y_{..} = \sum_{i=1}^{a}y_{i.}$$

$$= n\left[\sum_{i=1}^{a}\frac{y_{i.}^2}{n^2} - \frac{2}{n}\frac{y_{..}}{an}y_{..} + a\frac{y_{..}^2}{a^2n^2}\right] \quad , \quad \text{con } \bar{y}_{..} = \frac{y_{..}}{an}, \ \bar{y}_{i.} = \frac{y_{i.}}{n}$$

$$= \sum_{i=1}^{a}\frac{y_{i.}^2}{n} - 2\frac{y_{..}^2}{an} + \frac{y_{..}^2}{an}$$

$$= \sum_{i=1}^{a}\frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$$

$$SS_{\text{Tratamientos}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$

**L.q.q.d**

## (4) Deducción del Intervalo de confianza para $\sigma^2$

$$P\left( \chi^2_{1-(\frac{\alpha}{2}),N-a} \leq \frac{(N-a)MS_E}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},N-a} \right) = 1 - \alpha$$

$$\frac{1}{\chi^2_{1-(\frac{\alpha}{2}),N-a}} \geq \frac{1}{\frac{(N-a)MS_E}{\sigma^2}} \geq \frac{1}{\chi^2_{\frac{\alpha}{2},N-a}} \quad \textbf{inverso}$$

$$\frac{1}{\chi^2_{1-(\frac{\alpha}{2}),N-a}} \geq \frac{\sigma^2}{(N-a)MS_E} \geq \frac{1}{\chi^2_{\frac{\alpha}{2},N-a}}$$

$$\frac{(N-a)MS_E}{\chi^2_{1-(\frac{\alpha}{2}),N-a}} \geq \frac{((N-a)MS_E)\sigma^2}{(N-a)MS_E} \geq \frac{(N-a)MS_E}{\chi^2_{\frac{\alpha}{2},N-a}}$$

$$\frac{(N-a)MS_E}{\chi^2_{\frac{\alpha}{2},N-a}} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi^2_{1-(\frac{\alpha}{2}),N-a}}$$

**L.q.q.d**

## (5) Deducción del Intervalo de confianza para $\frac{\sigma_\tau^2}{\sigma^2}$

$$P\left( F_{1-\frac{\alpha}{2},a-1,N-a} \leq \frac{MS_{Tratamiento}}{MS_E} \frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2} \leq F_{\frac{\alpha}{2},a-1,N-a} \right) = 1 - \alpha$$

**Inverso**

$$\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} \leq \frac{1}{\frac{MS_{Tratamiento}}{MS_E} \frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2}} \leq \frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}$$

$$\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} \leq \frac{MS_E}{MS_{Tratamiento}} \frac{n\sigma_\tau^2 + \sigma^2}{\sigma^2} \leq \frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}$$

$$\frac{MS_{Tratamiento}}{MS_E} \frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} \leq \frac{MS_{Tratamiento}}{MS_E} \frac{MS_E}{MS_{Tratamiento}} \frac{n\sigma_\tau^2 + \sigma^2}{\sigma^2} \leq \frac{MS_{Tratamiento}}{MS_E} \frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}$$

$$\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} \le \frac{n\sigma_\tau^2+\sigma^2}{\sigma^2} \le \frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}$$

$$\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} \le \frac{n\sigma_\tau^2}{\sigma^2}+\frac{\sigma^2}{\sigma^2} \le \frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}$$

$$\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} \le \frac{n\sigma_\tau^2}{\sigma^2}+1 \le \frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}$$

$$\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}} - 1 \le \frac{n\sigma_\tau^2}{\sigma^2} \le \frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}} - 1$$

**Multiplicando por** $\frac{1}{n}$

$$\frac{1}{n}\left(\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}}-1\right) \le \frac{n\sigma_\tau^2}{n\sigma^2} \le \frac{1}{n}\left(\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}-1\right)$$

$$\frac{1}{n}\left(\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}}-1\right) \le \frac{\sigma_\tau^2}{\sigma^2} \le \frac{1}{n}\left(\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}}-1\right)$$

$$\mathrm{L} \le \frac{\sigma_\tau^2}{\sigma^2} \le \mathrm{U}$$

**donde:**

$$\mathrm{L} = \frac{1}{n}\left(\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,N-a}}-1\right) \qquad \mathrm{U} = \frac{1}{n}\left(\frac{MS_{Tratamiento}}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,N-a}} - 1\right)$$

**L.q.q.d**

**(6) Deducción de intervalo de confianza de:** $\dfrac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$

**Ya que**

$$L \leq \frac{\sigma_\tau^2}{\sigma^2} \leq U$$

$$\frac{1}{U} \leq \frac{1}{\frac{\sigma_\tau^2}{\sigma^2}} \leq \frac{1}{L} \qquad \textbf{Inverso}$$

$$\frac{1}{U} \leq \frac{\sigma^2}{\sigma_\tau^2} \leq \frac{1}{L}$$

$$\frac{1}{U} + \frac{\sigma_\tau^2}{\sigma_\tau^2} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2} + \frac{\sigma^2}{\sigma_\tau^2} \leq \frac{1}{L} + \frac{\sigma_\tau^2}{\sigma_\tau^2}$$

$$\frac{1}{U} + 1 \leq \frac{\sigma_\tau^2+\sigma^2}{\sigma_\tau^2} \leq \frac{1}{L} + 1$$

$$\frac{1+U}{U} \leq \frac{\sigma_\tau^2+\sigma^2}{\sigma_\tau^2} \leq \frac{1+L}{L}$$

$$\frac{1}{\frac{1+U}{U}} \geq \frac{1}{\frac{\sigma_\tau^2+\sigma^2}{\sigma_\tau^2}} \geq \frac{1}{\frac{1+L}{L}} \qquad \textbf{Inverso}$$

$$\frac{U}{U+1} \geq \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \geq \frac{L}{L+1}$$

$$\frac{L}{L+1} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leq \frac{U}{U+1}$$

**L.q.q.d**

## (7) Especificación de un incremento en la Desviación Estándar (Efectos Fijos)

$$\frac{\sqrt{\sigma^2 + \sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\sigma} = 1 + 0.01p$$

$$\left(\frac{\sqrt{\sigma^2 + \sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\sigma}\right)^2 = (1 + 0.01p)^2$$

$$\frac{\sigma^2 + \sum_{i=1}^{a} \frac{\tau_i^2}{a}}{\sigma^2} = (1 + 0.01p)^2$$

$$\frac{\sigma^2}{\sigma^2} + \frac{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}{\sigma^2} = (1 + 0.01p)^2$$

$$1 + \frac{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}{\sigma^2} = (1 + 0.01p)^2$$

$$\frac{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}{\sigma^2} = (1 + 0.01p)^2 - 1$$

$$\sqrt{\frac{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}{\sigma^2}} = \sqrt{(1+0.01p)^2 - 1}$$

$$\frac{\sqrt{\sum_{i=1}^{a} \frac{\tau_i^2}{a}}}{\sigma} = \sqrt{(1+0.01p)^2 - 1}$$

**L.q.q.d**

**(8) Especificación de un incremento en la Desviación Estándar. (Efectos Aleatorios)**

$$\frac{\sqrt{\sigma^2+\sigma_\tau^2}}{\sigma} = 1 + 0.01p$$

$$\left(\frac{\sqrt{\sigma^2+\sigma_\tau^2}}{\sigma}\right)^2 = (1 + 0.01p)^2$$

$$\frac{\sigma^2+\sigma_\tau^2}{\sigma^2} = (1 + 0.01p)^2$$

$$\frac{\sigma^2}{\sigma^2}+\frac{\sigma_\tau^2}{\sigma^2} = (1 + 0.01p)^2$$

$$1 + \frac{\sigma_\tau^2}{\sigma^2} = (1 + 0.01p)^2$$

$$\frac{\sigma_\tau^2}{\sigma^2} = (1 + 0.01p)^2 - 1$$

**L.q.q.d**

# APÉNDICE DE LA UNIDAD PROGRAMÁTICA III ( DISEÑO POR BLOQUES COMPLETOS )

## Deducción de las fórmulas Matemáticas de las sumas de cuadrados

**La descomposición de las sumas de Cuadrados quedo:**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + a\sum_{i=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j}-\bar{y}_{i.}+\bar{y}_{..})^2$$

**(1) Suma de Cuadrado Total**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}^2-2y_{ij}\,\bar{y}_{..}+\bar{y}_{..}^{\,2})$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}(2y_{ij}\,\bar{y}_{..})+\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{..}^{\,2}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ -2} \bar{y}_{..} \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} + ab\,\bar{y}_{..}^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ -2} \bar{y}_{..}\, y_{..} + ab\,\bar{y}_{..}^2 \quad , \text{ con } y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ -2} \bar{y}_{..} (N\bar{y}_{..}) + ab\,\bar{y}_{..}^2 \quad , \text{ con } N\bar{y}_{..} = y_{..}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ -2 } ab\,\bar{y}_{..}^2 + ab\,\bar{y}_{..}^2 \quad , \text{ con } N = ab$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ - } ab\,\bar{y}_{..}^2 \quad \text{con } \bar{y}_{..}^2 = \frac{y_{..}^2}{(ab)^2}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ - } ab\left(\frac{y_{..}^2}{(ab)^2}\right)$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ - } \frac{y_{..}^2}{ab} \quad , \text{ con } N = ab$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \text{ - } \frac{y_{..}^2}{N}$$

**L.q.q.d**

## (2) Suma de Cuadrados de Tratamiento

**Considerando el primer término de la descomposición de las sumas de Cuadrados; se tiene:**

$$SS_{\text{Tratamientos}} = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 = b\left(\sum_{i=1}^{a}(\bar{y}_{i.}^2 - 2\bar{y}_{i.}\bar{y}_{..} + \bar{y}_{..}^2\right)$$

$$= b\left(\sum_{i=1}^{a}\bar{y}_{i.}^2 - 2\bar{y}_{..}\sum_{i=1}^{a}\bar{y}_{i.} + \sum_{i=1}^{a}\bar{y}_{..}^2\right)$$

$$= b\left(\sum_{i=1}^{a}\left(\frac{y_{i.}}{b}\right)^2 - 2\bar{y}_{..}\sum_{i=1}^{a}\frac{y_{i.}}{b} + a\bar{y}_{..}^2\right) \quad , \text{ con } \bar{y}_{i.} = \frac{y_{i.}}{b}$$

$$= b\left(\sum_{i=1}^{a}\frac{y_{i.}^2}{b^2} - 2\frac{y_{..}}{N}\sum_{i=1}^{a}\frac{y_{i.}}{b} + a\frac{y_{..}^2}{N^2}\right) \quad , \text{ con } \bar{y}_{..} = \frac{y_{..}}{N}$$

$$= b\left(\sum_{i=1}^{a}\frac{y_{i.}^2}{b^2} - 2\frac{y_{..}}{ab}\frac{y_{..}}{b} + a\frac{y_{..}^2}{a^2b^2}\right) \quad , \text{ con } N = ab \quad y \quad \sum_{i=1}^{a} y_{i.} = y_{..}$$

$$= b\left(\sum_{i=1}^{a}\frac{y_{i.}^2}{b^2} - 2\frac{y_{..}^2}{ab^2} + \frac{y_{..}^2}{ab^2}\right) = b\left(\sum_{i=1}^{a}\frac{y_{i.}^2}{b^2} - \frac{y_{..}^2}{ab^2}\right) = \sum_{i=1}^{a}\frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$$

$$SS_{\text{Tratamientos}} = \sum_{i=1}^{a}\frac{y_{i.}^2}{b} - \frac{y_{..}^2}{N}$$

**L.q.q.d**

## (3) Suma de Cuadrados de Bloques

**Considerando el Segundo Término de la descomposición de las sumas de Cuadrados; se tiene:**

$$SS_{\text{Bloques}} = a \sum_{j=1}^{b} (\bar{y}_{.j} - \bar{y}_{..})^2 = a \left( \sum_{j=1}^{b} (\bar{y}_{.j}^2 - 2\bar{y}_{.j}\bar{y}_{..} + \bar{y}_{..}^2 \right)$$

$$= a \left( \sum_{j=1}^{b} \bar{y}_{.j}^2 - 2\bar{y}_{..} \sum_{j=1}^{b} \bar{y}_{.j} + \sum_{j=1}^{b} \bar{y}_{..}^2 \right)$$

$$= a \left( \sum_{j=1}^{b} \left( \frac{y.j}{a} \right)^2 - 2\bar{y}_{..} \sum_{j=1}^{b} \frac{y_{.j}}{a} + b\,\bar{y}_{..}^2 \right) \quad , \text{ con } \bar{y}_{.j} = \frac{y_{.j}}{a}$$

$$= a \left( \sum_{j=1}^{b} \frac{y_{.j}^2}{a^2} - 2\frac{y_{..}}{N} \sum_{j=1}^{b} \frac{y_{.j}}{a} + b\,\frac{y_{..}^2}{N^2} \right) \quad , \text{ con } \bar{y}_{..} = \frac{y_{..}}{N}$$

$$= a \left( \sum_{j=1}^{b} \frac{y_{.j}^2}{a^2} - 2\frac{y_{..}y_{..}}{Na} + b\,\frac{y_{..}^2}{N^2} \right) \quad , \text{ con } \sum_{j=1}^{b} y_{.j} = y_{..}$$

$$= a \left( \sum_{j=1}^{b} \frac{y_{.j}^2}{a^2} - 2\frac{y_{..}^2}{(ab)a} + b\,\frac{y_{..}^2}{(ab)^2} \right) \quad , \text{ con } N = ab$$

$$= a \left( \sum_{j=1}^{b} \frac{y_{.j}^2}{a^2} - 2\frac{y_{..}^2}{a^2 b} + b\,\frac{y_{..}^2}{a^2 b^2} \right) \quad \text{distribuyendo “}a\text{”}$$

$$= \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - 2\frac{y_{..}^2}{ab} + \frac{y_{..}^2}{ab} \quad , \text{ con } N = ab$$

$$SS_{\text{Bloques}} = \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{N}$$

**L.q.q.d**

# APÉNDICE DE LA UNIDAD PROGRAMATICA IV
# ( DISEÑOS FACTORIALES )

## DISEÑO DE DOS FACTORES

### Deducción de las fórmulas Matemáticas de las sumas de cuadrados

**La descomposición de las sumas de Cuadrados quedo:**

$$SS_T = bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{...})^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$$
$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij.})^2$$

**(1) Suma de Cuadrado Total**

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}^2-2y_{ijk}\bar{y}_{...}+\bar{y}_{...}^2)$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}2y_{ijk}\bar{y}_{...} + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\bar{y}_{...}^2$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - 2\bar{y}_{...}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk} + abn\,\bar{y}_{...}^2$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - 2\bar{y}_{...}y_{...} + abn\,\bar{y}_{...}^2 \quad , \text{ con } y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - 2\left(\frac{y_{...}}{abn}\right)y_{...} + abn\frac{y_{...}^2}{(abn)^2} \quad , \text{ con } \bar{y}_{...} = \frac{y_{...}}{abn}$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - 2\frac{y_{...}^2}{abn} + \frac{y_{...}^2}{abn}$$
$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

**L.q.q.d**

## (2) Suma de Cuadrados del Factor A

**Considerando el primer término de la descomposición de las sumas de Cuadrados; se tiene:**

$$SS_A = bn \sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2 = bn\left( \sum_{i=1}^{a} \bar{y}_{i..}^{2} - 2\bar{y}_{...} \sum_{i=1}^{a} \bar{y}_{i..} + \sum_{i=1}^{a} \bar{y}_{...}^{2} \right)$$

$$= bn\left( \sum_{i=1}^{a} \bar{y}_{i..}^{2} - 2\bar{y}_{..} \sum_{i=1}^{a} \frac{y_{i..}}{bn} + a\bar{y}_{...}^{2} \right) \quad , \quad \text{con } \bar{y}_{i..} = \frac{y_{i..}}{bn}$$

$$= bn\left( \sum_{i=1}^{a} \frac{y_{i..}^{2}}{b^2n^2} - 2\frac{y_{...}}{abn} \sum_{i=1}^{a} \frac{y_{i.}}{bn} + a\frac{y_{...}^{2}}{(abn)^2} \right) \quad , \quad \text{con } \bar{y}_{i..}^{2} = \frac{y_{i..}^{2}}{b^2n^2} \text{ y } \bar{y}_{...}^{2} = \frac{y_{...}^{2}}{(abn)^2}$$

$$= bn\left( \sum_{i=1}^{a} \frac{y_{i..}^{2}}{b^2n^2} - 2\frac{y_{...}}{abn}\frac{y_{...}}{bn} + a\frac{y_{...}^{2}}{a^2b^2n^2} \right) \quad , \quad \text{con } \sum_{i=1}^{a} y_{i..} = y_{...}$$

$$= bn\left( \sum_{i=1}^{a} \frac{y_{i..}^{2}}{b^2n^2} - 2\frac{y_{...}^{2}}{ab^2n^2} + \frac{y_{...}^{2}}{ab^2n^2} \right) = bn\left( \sum_{i=1}^{a} \frac{y_{i..}^{2}}{b^2n^2} - \frac{y_{...}^{2}}{ab^2n^2} \right), \text{ Distribuyendo } bn$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i..}^{2}}{bn} - \frac{y_{...}^{2}}{abn}$$

**L.q.q.d**

## (3) Suma de Cuadrados del Factor B

**Considerando el Segundo Término de la descomposición de las sumas de Cuadrados; se tiene:**

$$SS_A = an \sum_{i=1}^{b} (\bar{y}_{.j.} - \bar{y}_{...})^2 = an\left( \sum_{j=1}^{b} (\bar{y}_{.j.}^{2} - 2\bar{y}_{.j.}\bar{y}_{...} + \bar{y}_{...}^{2}) \right)$$

$$= an\left( \sum_{j=1}^{b} \bar{y}_{.j.}^{2} - 2\bar{y}_{...} \sum_{j=1}^{b} \bar{y}_{.j.} + \sum_{j=1}^{b} \bar{y}_{...}^{2} \right)$$

$$= an\left( \sum_{j=1}^{b} \bar{y}_{.j.}^{2} - 2\bar{y}_{...} \sum_{j=1}^{b} \frac{y_{.j.}}{an} + b\bar{y}_{...} \right) \quad , \text{ con } \bar{y}_{.j.} = \frac{y_{.j.}}{an}$$

$$= an\left( \sum_{j=1}^{b} \frac{y_{.j.}^2}{a^2 n^2} - 2\,\frac{y_{...}}{N} \sum_{j=1}^{b} \frac{y_{.j.}}{an} + b\,\frac{y_{...}^2}{N^2} \right)$$ , con $\overline{y}_{.j.}^2 = \frac{y_{.j.}^2}{a^2 n^2}$ y $\overline{y}_{...} = \frac{y_{...}}{N}$

$$= an\left( \sum_{j=1}^{b} \frac{y_{.j.}^2}{a^2 n^2} - 2\,\frac{y_{...}y_{...}}{Nan} + b\,\frac{y_{...}^2}{N^2} \right)$$ , con $\sum_{j=1}^{b} y_{.j.} = y_{...}$

$$= an\left( \sum_{j=1}^{b} \frac{y_{.j.}^2}{a^2 n^2} - 2\,\frac{y_{...}^2}{(abn)an} + b\,\frac{y_{...}^2}{(abn)^2} \right)$$ , con $N = abn$

$$= an\left( \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - 2\,\frac{y_{...}^2}{abn} + \frac{y_{...}^2}{abn} \right)$$ , distribuyendo $an$

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

**L.q.q.d**

# APÉNDICE: USO DEL SOFTWARE ESTADÍSTICO EN EL DISEÑO EXPERIMENTAL.

## INTRODUCCIÓN

El uso del Software en la Estadística es muy importante; ya que en la mayoría de los casos nos facilita los cálculos Matemáticos y la elaboración de gráficos y muchas otras cosas.

Existen muchos software que se pueden utilizar como herramienta para el Análisis Estadístico, desde los más sencillos hasta los más complicados, entre los cuales se pueden mencionar el **SPSS, STATIST, MINITAD, STATGRAPHICS, EPI-INFO, INFOSTAT** y otros mas; los cuales en su mayoría se iniciaron con versiones de MS-DOS y en la actualidad existen versiones para Windows y redes de comunicaciones.

Estos software contienen en su mayoría todas las aplicaciones Estadísticas que se necesitan para hacer trabajos de Investigación Estadística.

También existen software para áreas especificas de la Estadística, que son creados por especialistas de esas áreas, tal es el caso del área de Diseño de Experimentos; los cuales contienen aplicaciones especificas para el estudios de los Diseños Experimentales.

En el Análisis de los Diseños Experimentales resulta de gran importancia la utilización de un software, ya sea de uso general ó de uso específico; ya que facilita la obtención de la Tabla de Análisis de Varianza, gráficos importantes, etc.

Para realizar la tabla de Análisis de Varianza u otro Análisis por medio de un software estadístico, lo más importante es saber definir la Variable Dependiente y las Variables Independientes que serán introducidas en el computador a partir de los datos obtenidos en el experimento; ya que para cada Diseño Experimental cambia el número de Variables Independientes y también es importante saber que información se le va ha asignar a cada una de las Variables y lo demás la computadora lo hace sola.

En todo Diseño Experimental existe una Variable Dependiente y muchas Variables Independientes dependiendo el tipo de Diseño que se esta estudiando.

Tomando en cuenta todo lo anterior a continuación se presentará una guía general para la creación de cada una de las Variables y la información que deben contener; para llevar a cabo el Análisis de Varianza de los diferentes Diseños Experimentales que se han estudiado en cada una de las Unidades Programáticas, colocando además el ejemplo desarrollado y Analizado en cada una de las Unidades Programáticas.

## DISEÑO UNIFACTORIAL

En este tipo de Diseño Experimental sólo se definen dos Variables que son:

i) **Variable Dependiente (Variable Respuesta):** Es aquella variable a la que se le asignan los valores de los datos u observaciones obtenidas en el experimento y es de tipo numérica.

ii) **Variable Independiente:** Será la que contendrá los niveles del factor (Tratamientos) en forma numérica ó Carácter, colocándose consecutivamente tantas veces como observaciones tenga dicho Tratamiento.

**Ejemplo 1**

Retomando el ejemplo de los programas de estudio, desarrollado en la Unidad Programática II, si se lleva a cabo el análisis de varianza por medio de un software estadístico se debe realizar de la siguiente manera:

**a) Definición de Variables:**

Variable Dependiente : HAB-LECT (Tipo numérica)

Variable Independiente: PROG (Tipo numérica)

**b) Introducción de datos de las Variables en el Software Estadístico.**

| PROG | HAB-LECT |
|---|---|
| 1 | 20 |
| 1 | 18 |
| 1 | 18 |
| 1 | 23 |
| 1 | 22 |
| 1 | 17 |
| 1 | 15 |
| 1 | 13 |
| 1 | 21 |
| 2 | 15 |
| 2 | 20 |
| 2 | 13 |
| 2 | 12 |
| 2 | 16 |
| 2 | 17 |
| 2 | 21 |
| 2 | 15 |
| 2 | 13 |
| 3 | 12 |
| 3 | 15 |
| 3 | 18 |
| 3 | 20 |
| 3 | 18 |
| 3 | 17 |
| 3 | 10 |
| 3 | 24 |
| 3 | 16 |

**Nota:** La Variable PROG, se pudo definir como A: programa 1, B: programa 2 y C: programa 3; y ver colocado como valores de dicha Variable A nueve veces, B Nueve veces y C también nueve veces. (Igual como se definió de tipo Numérica).

**c) Seleccionar la opción que realiza el Análisis de Varianza y luego definir las Variables como Dependiente (**HAB-LECT**) e Independiente (**PROG**).**

**d) Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S D E V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrado | F | P>F |
|---|---|---|---|---|---|
| TRATAMIENTOS | 36.222168 | 2 | 18.111084 | 1.4404 | 0.256 |
| RESIDUAL | 301.777832 | 24 | 12.574077 | | |
| TOTAL | 338.000000 | 26 | | | |

## DISEÑO POR BLOQUES COMPLETOS

En este tipo de Diseño Experimental se deben definir Tres Variables que son:

i) **Variable Dependiente (Variable Respuesta):** Es aquella variable que se le asignan los valores de los datos u observaciones obtenidas en el experimento y debe ser de tipo numérica.

ii) **Variable de Tratamientos:** Será la que contendrá los niveles del factor (Tratamientos) en forma numérica ó carácter, colocándose consecutivamente tantas veces como observaciones tenga dicho tratamiento.

iii) **Variable de Bloques:** Será la que contendrá el número de bloques en forma numérica ó carácter; colocándose consecutivamente tantas veces como bloques se tengan.

**Nota:** Hay que tomar en cuenta al introducir los datos en la Variable Respuesta a que tratamiento y a que bloque pertenece dicho valor, ya que le corresponde un número específico de tratamiento y bloque.

**Ejemplo 2**

Retomando el ejemplo de las Raciones para Novillos; tomando los datos sin codificar desarrollado en la Unidad Programática III, si se lleva a cabo el Análisis de Varianza por medio de un software Estadístico se debe realizar lo siguiente:

**a) Definición de Variables:**

Variable Dependiente : **PESOS-NOVILLOS (Tipo Numérica)**

Variable de Tratamientos: **RACIONES (Tipo Carácter) (Tratamientos)**

Variable de Bloques : **NOVILLOS (Tipo Numérica) (Bloques)**

**b) Introducción de datos de las Variables en el Software Estadístico.**

| NOVILLOS | RACIONES | PESOS-NOVILLOS |
|---|---|---|
| 1 | A | 0.9 |
| 1 | B | 3.6 |
| 1 | C | 0.5 |
| 1 | D | 3.6 |
| 1 | E | 1.8 |
| 2 | A | 1.4 |
| 2 | B | 3.2 |
| 2 | C | 0.9 |
| 2 | D | 3.6 |
| 2 | E | 1.8 |
| 3 | A | 1.4 |
| 3 | B | 4.5 |
| 3 | C | 0.5 |
| 3 | D | 3.2 |
| 3 | E | 0.9 |
| 4 | A | 2.3 |
| 4 | B | 4.1 |
| 4 | C | 0.9 |
| 4 | D | 3.6 |
| 4 | E | 1.4 |

**c) Seleccionar la opción que realiza el Análisis de Varianza y luego definir la Variable como Dependiente (**PESOS-NOVILLOS**) y las otras dos Variables como Independientes (**RACIONES Y NOVILLOS**).**

**d) Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S D E V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrados | F | P>F |
|---|---|---|---|---|---|
| RACIONES | 30.71 | 4 | 7.68 | 39.11 | 0.0000 |
| NOVILLOS | 0.46 | 3 | 0.15 | 0.78 | 0.5257 |
| RESIDUAL | 2.36 | 12 | 0.20 | | |
| TOTAL | 33.53 | 19 | | | |

## DISEÑO DE CUADRADO LATINO

En este tipo de Diseño Experimental se deben definir Cuatro Variables que son:

**i)** **Variable Dependiente (Variable Respuesta):** Es aquella variable que se le asignan los valores de los datos u observaciones obtenidas en el experimento y debe ser de tipo numérica.

**ii)** **Variable de Fila:** Será la que contendrá los números de las filas en forma numérica ó Carácter (de preferencia numérica para que no se confunda con la Letra Latina), colocándose consecutivamente tantas veces como observaciones tenga dicha Fila.

**iii)** **Variable de Columna:** Será la que contendrá los números de las columnas en forma numérica ó de carácter (de preferencia numérica para que no se confunda con la Letra Latina), colocándose consecutivamente tantas veces como observaciones tenga cada Columna.

**iv)** **Variable de Letra Latina:** Será la contendrá las diferentes letras Latinas que corresponden a los tratamientos que están presentes en el experimento en forma de carácter; colándose consecutivamente tantas veces como estén involucradas en el experimento.

**Nota:** El valor de la Variable Respuesta va a corresponder a un valor de la variable fila, columna y Letra Latina; es decir hay que tener cuidado al asignar estos valores, es de observar dicho valor en tres direcciones.

**Ejemplo 3**

Se tomará como ejemplo, el Cuadrado Latino Analizado en la Unidad Programática III; que se refiere a la capacidad de producción, si se realiza el Análisis de Varianza por medio de un software estadístico se debe realizar de la siguiente manera:

**a) Definición de Variables:**

Variable Dependiente: **CAPACIDAD (Tipo Numérica)**

Variable Fila : **PERÍODO (Tipo Numérica)**

Variable Columna : **OPERADORES (Tipo Numérica )**

Variable Letra Latina : **MÁQUINAS (Letra Latina, Tipo carácter)**

**b) Introducción de datos de las Variables en el software estadístico.**

| PERÍODO | OPERADORES | MÁQUINAS | CAPACIDAD |
|---|---|---|---|
| 1 | 1 | C | 31 |
| 1 | 2 | D | 43 |
| 1 | 3 | A | 67 |
| 1 | 4 | B | 36 |
| 2 | 1 | D | 39 |
| 2 | 2 | A | 96 |
| 2 | 3 | B | 40 |
| 2 | 4 | C | 48 |
| 3 | 1 | B | 57 |
| 3 | 2 | C | 33 |
| 3 | 3 | D | 40 |
| 3 | 4 | A | 84 |
| 4 | 1 | A | 85 |
| 4 | 2 | B | 46 |
| 4 | 3 | C | 48 |
| 4 | 4 | D | 50 |

c) **Seleccionar la opción que realiza el Análisis de Varianza y luego definir la Variable como Dependiente (**CAPACIDAD**) y las otras tres Variables como Independientes (**MÁQUINAS, PERÍODO Y OPERADORES**).**

d) **Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S D E V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrados | F | P>F |
|---|---|---|---|---|---|
| MÁQUINAS | 4946.6875 | 3 | 1640.8958 | 19.178 | 0.0018 |
| PERÍODO | 408.1875 | 3 | 136.0625 | 1.583 | 0.2890 |
| OPERADORES | 88.6875 | 3 | 29.5625 | 0.344 | 0.7952 |
| RESIDUAL | 515.87500 | 6 | 85.979167 | | |
| TOTAL | 5959.4375 | 15 | | | |

## DISEÑO DE CUADRADO GRECO-LATINO

En este tipo de Diseño Experimental se deben definir Cinco Variables que son:

**i)** **Variable Dependiente (Variable Respuesta):** Es aquella variable que se le asignan los datos u observaciones obtenidas en el experimento y debe ser de tipo numérica.

**ii)** **Variable de Fila:** Será la que contendrá los números de las filas en forma numérica ó Carácter (de preferencia numérica para que no se confunda con la Letra Latina), colocándose consecutivamente tantas veces como observaciones tenga dicha Fila.

**iii)** **Variable de Columna:** Será la que contendrá los números de las columnas en forma numérica ó de carácter (de preferencia numérica para que no se confunda con la Letra Latina), colocándose consecutivamente tantas veces como observaciones tenga dicha Columna.

**iv)** **Variable de Letra Latina:** Será la que contendrá las diferentes Letras Latinas que están presentes en el experimento, que corresponden a los tratamientos asignados a la Letras Latinas en forma de carácter; colándose en el orden que aparecen en el experimento.

**v)** **Variable de Letra Griega:** Será la que contendrá las diferentes Letras Griegas que están presentes en el experimento, que corresponden a los tratamientos asignados a las Letras Griegas en forma de carácter; colándose en el orden en que aparecen en el experimento. Para esta variable se podría optar como una opción inmediata utilizar los caracteres Romanos (I, II, III,…) para representar cada letra griega; ya que será difícil poder incorporar las letras griegas en el editor de el Software Estadístico.

**Nota:** En este caso cada valor de la Variable Respuesta le corresponden cuatro direcciones; es de observar cuidadosamente a que fila, columna, Letra Latina y Letra Griega corresponde dicho valor.

**Ejemplo 4**

Se tomará como ejemplo, el Cuadrado Greco-latino Analizado en la Unidad Programática III; que se refiere a la fabricación de ladrillos de concreto, si se lleva a cabo el análisis de varianza por medio de un software estadístico se debe realizar de la siguiente manera:

**a) Definición de Variables:**

Variable Dependiente: **LADRILLO (Tipo Numérica)**

Variable Fila : **MEZCLAS (Tipo Numérica )**

Variable Columna : **DÍAS (Tipo Numérica)**

Variable Letra Latina : **MÁQUINAS (Letra Latina, Tipo Carácter)**

Variable Letra Griega : **PROCED (Letra Griega, Tipo Carácter).** Se tomaran las Letras Griegas de la siguiente forma: $\alpha$ = I, $\beta$ = II, $\gamma$ = III, $\delta$ = IV, $\varepsilon$ = V

**b) Introducción de datos de las Variables en el Software Estadístico.**

| MEZCLAS | DÍAS | MÁQUINAS | PROCED | LADRILLO |
|---|---|---|---|---|
| 1 | 1 | A | I | 1 |
| 1 | 2 | B | III | 0 |
| 1 | 3 | C | V | 4 |
| 1 | 4 | D | II | 0 |
| 1 | 5 | E | IV | 1 |
| 2 | 1 | B | II | 3 |
| 2 | 2 | C | IV | 4 |
| 2 | 3 | D | I | 1 |
| 2 | 4 | E | III | 5 |
| 2 | 5 | A | V | 3 |
| 3 | 1 | C | III | 2 |
| 3 | 2 | D | V | 5 |
| 3 | 3 | E | II | 0 |
| 3 | 4 | A | IV | 0 |
| 3 | 5 | B | I | -1 |
| 4 | 1 | D | IV | -1 |
| 4 | 2 | E | I | 2 |
| 4 | 3 | A | III | -1 |
| 4 | 4 | B | V | 1 |
| 4 | 5 | C | II | 4 |
| 5 | 1 | E | V | 0 |
| 5 | 2 | A | II | 1 |
| 5 | 3 | B | IV | -2 |
| 5 | 4 | C | I | -3 |
| 5 | 5 | D | III | 1 |

c) **Seleccionar la opción que realiza el Análisis de Varianza y luego definir la Variable como Dependiente (**LADRILLO**) y las otras cuatro Variables como Independientes (**MEZCLAS, DÍAS, MÁQUINAS Y PROCED**).**

d) **Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S D E V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrados | F | P>F |
|---|---|---|---|---|---|
| MÁQUINAS | 11.6000 | 4 | 2.9000 | 0.841 | 0.5366 |
| PROCED | 21.2000 | 4 | 5.3000 | 1.536 | 0.2801 |
| MEZCLAS | 36.4000 | 4 | 9.1000 | 2.638 | 0.1133 |
| DÍAS | 13.2000 | 4 | 3.3000 | 0.957 | 0.4804 |
| RESIDUAL | 27.6000 | 8 | 3.4500 | | |
| TOTAL | 110.0000 | 24 | | | |

## DISEÑO POR BLOQUES INCOMPLETOS BALANCEADOS

En este tipo de Diseño Experimental se deben definir Tres Variables al igual que el Diseño por Bloques Completos; que son:

- **i)** **Variable Dependiente (Variable Respuesta):** Es aquella variable que se le asignan los valores de los datos u observaciones obtenidas en el experimento y debe ser de tipo numérica.
- **ii)** **Variable de Tratamientos:** Será la que contendrá los niveles del factor (Tratamientos) en forma numérica ó carácter, colocándose consecutivamente tantas veces como observaciones tenga dicho tratamiento.
- **iii)** **Variable de Bloques:** Será la que contendrá el número de bloques en forma numérica ó carácter; colocándose consecutivamente tantas veces como bloques se tengan.

**Nota:** Hay que tomar en cuenta al introducir los datos en la Variable Respuesta a que tratamiento y a que bloque pertenece dicho valor; al que le corresponden a un número específico de tratamiento y bloque; en este tipo de Diseño no van ha existir valores para algunas combinaciones de las variables de Tratamientos y Bloques; ya que faltaran algunos valores en la tabla de datos.

**Ejemplo 5**

Retomando el ejemplo de la Germinación de semillas de tomate; desarrollado en la Unidad Programática III, si se lleva a cabo el Análisis de Varianza por medio de un software Estadístico se debe realizar lo siguiente:

**a) Definición de Variables:**

Variable Dependiente : **GERMINACIÓN (Tipo Numérica)**

Variable de Tratamientos: **TEMPERATURA (Tipo Numérica) (Tratamientos)**

Variable de Bloques : **CORRIDAS (Tipo Numérica) (Bloques)**

**b) Introducción de datos de las Variables en el Software Estadístico.**

| TEMPERATURA | CORRIDAS | GERMINACIÓN |
|---|---|---|
| 25 | 1 | 24.65 |
| 25 | 3 | 29.17 |
| 25 | 5 | 28.90 |
| 30 | 2 | 24.38 |
| 30 | 3 | 21.25 |
| 30 | 6 | 25.53 |
| 35 | 4 | 5.90 |
| 35 | 5 | 18.27 |
| 35 | 6 | 8.42 |
| 40 | 1 | 1.34 |
| 40 | 2 | 1.24 |
| 40 | 4 | 1.83 |

**c) Seleccionar la opción que realiza el Análisis de Varianza y luego definir la Variable como Dependiente (**GERMINACIÓN**) y las otras dos Variables como Independientes (**TEMPERATURA Y CORRIDAS**).**

**d) Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S D E V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrados | F | P>F |
|---|---|---|---|---|---|
| TEMPERATURAS(Corregida) | 718.29 | 3 | 239.43 | 17.52 | 0.0210 |
| CORRIDAS(Corregido) | 67.54 | 5 | 13.508 | 0.988 | 0.5393 |
| RESIDUAL | 40.99 | 3 | 13.665 | | |
| TOTAL | 1372.949 | 11 | | | |

Como puede observarse en la tabla de Análisis de Varianza que se obtiene, el valor de la Suma de Cuadrado de los Bloques (CORRIDAS) son los corregidos; por lo tanto, la Suma de Cuadrado de los tratamientos, los bloques y el error no será igual a la Suma Total de Cuadrados. Ya que en un Diseño por Bloques Incompletos Balaceados la Suma Total de cuadrados es igual a la Suma de Tratamientos(Corregidos) más la Suma de los Bloques(no corregidos) más la Suma de Cuadrado del Error.

## DISEÑO DE DOS FACTORES

En este tipo de Diseño Experimental se deben definir Tres Variables que son:

i) **Variable Dependiente (Variable Respuesta):** Es aquella variable que se le asignan los datos u observaciones obtenidas en el experimento y debe ser de tipo numérica.

ii) **Variable del Factor A:** Será la que contendrá los niveles del factor A en forma numérica ó Carácter, colocándose consecutivamente tantas veces como niveles tenga dicho Factor.

iii) **Variable del Factor B**: Será la que contendrá los niveles del factor B en forma numérica ó Carácter, colocándose consecutivamente tantas veces como niveles tenga dicho Factor.

**Nota:** Una observación de la Variable Respuesta corresponde a un nivel específico del Factor A y Factor B; por lo tanto, se debe observar en dos direcciones.

**Ejemplo 6**

Se tomara como ejemplo el desarrollado en la Unidad Programática IV; que se refiere al porcentaje de encogimiento de tela teñida; si se lleva a cabo el Análisis de Varianza por medio de un software estadístico se debe realizar de la siguiente manera:

**a) Definición de Variables:**

Variable Dependiente: **PORCENTAJE (Tipo Numérica)**

Variable del Factor A : **TIPOS (Tipo Numérica)**

Variable del Factor B : **TEMPERATURA (Tipo Numérica)**

**b) Introducción de datos de las Variables en el software estadístico.**

| TIPOS | TEMPERATURA | PORCENTAJE |
|---|---|---|
| 1 | 210 | 1.8 |
| 1 | 210 | 2.1 |
| 1 | 215 | 2.0 |
| 1 | 215 | 2.1 |
| 1 | 220 | 4.6 |
| 1 | 220 | 5.0 |
| 1 | 225 | 7.5 |
| 1 | 225 | 7.9 |
| 2 | 210 | 2.2 |
| 2 | 210 | 2.4 |
| 2 | 215 | 4.2 |
| 2 | 215 | 4.0 |
| 2 | 220 | 5.4 |
| 2 | 220 | 5.6 |
| 2 | 225 | 9.8 |
| 2 | 225 | 9.2 |
| 3 | 210 | 2.8 |
| 3 | 210 | 3.2 |
| 3 | 215 | 4.4 |
| 3 | 215 | 4.8 |
| 3 | 220 | 8.7 |
| 3 | 220 | 8.4 |
| 3 | 225 | 13.2 |
| 3 | 225 | 13.0 |
| 4 | 210 | 3.2 |
| 4 | 210 | 3.6 |
| 4 | 215 | 3.3 |
| 4 | 215 | 3.5 |
| 4 | 220 | 5.7 |
| 4 | 220 | 5.8 |
| 4 | 225 | 10.9 |
| 4 | 225 | 11.1 |

c) **Seleccionar la opción que realiza el Análisis de Varianza y luego definir la Variable como Dependiente (**PORCENTAJE**) y las otras dos Variables como Independientes (**TIPOS Y TEMPERATURA**). Como en este tipo de Diseño interesa obtener efectos de interacción doble entonces cada software estadístico trae la opción donde se le especifica que se deben calcular tales interacciones. Como por ejemplo, en el STATGRAPHIP se le debe colocar en la ventana donde se definen las Variables en la opción "Ignore Interactions Higher than orden:" el valor de 2, que significa que se calculen las interacciones dobles.**

d) **Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S D E V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrados | F | P>F |
|---|---|---|---|---|---|
| TIPOS | 41.8762 | 3 | 13.9587 | 279.17 | 0.0000 |
| TEMPERATURA | 283.9362 | 3 | 94.6454 | 1892.91 | 0.0000 |
| INTERACICÓN | | | | | |
| AB | 15.8562 | 9 | 1.7618 | 35.23 | 0.0000 |
| RESIDUAL | 0.8000 | 16 | 0.0500 | | |
| TOTAL | 342.4687 | 31 | | | |

## DISEÑO DE TRES FACTORES

En este tipo de Diseño Experimental se deben definir Cuatro Variables que son:

i) **Variable Dependiente (Variable Respuesta):** Es aquella variable que se le asignan los datos u observaciones obtenidas en el experimento y debe ser de tipo numérica.

ii) **Variable del Factor A:** Será la que contendrá los niveles del factor A en forma numérica ó Carácter, colocándose consecutivamente tantas veces como niveles tenga dicho Factor.

iii) **Variable del Factor B**: Será la que contendrá los niveles del factor B en forma numérica ó Carácter, colocándose consecutivamente tantas veces como niveles tenga dicho Factor.

iv) **Variable del Factor C**: Será la que contendrá los niveles del factor C en forma numérica ó Carácter, colocándose consecutivamente tantas veces como niveles tenga dicho Factor.

**Nota:** Una observación de la Variable Respuesta corresponde a un nivel específico del Factor A, Factor B y Factor C; por lo tanto, se debe observar en tres direcciones.

**Ejemplo 7**

Se tomará como ejemplo el desarrollado en la Unidad Programática IV; que se refiere a las propiedades de Resistencia a la Compresión de Mezclas de Cemento y Tierra (Datos codificados); si se lleva a cabo el Análisis de Varianza por medio de un software estadístico se debe realizar de la siguiente manera:

**a) Definición de Variables:**

Variable Dependiente: **COMPRESIÓN (Tipo Numérica)**

Variable del Factor A : **EDAD (Tipo Numérica)**

Variable del Factor B : **TEMPERATURA (Tipo Numérica)**

Variable del Factor C : **TIERRA (Tipo Numérica)**

**c) Introducción de datos de las Variables en el software estadístico.**

| EDAD | TEMPERATURA | TIERRA | COMPRESIÓN |
|---|---|---|---|
| 1 | 1 | 1 | 4.71 |
| 1 | 1 | 1 | 4.13 |
| 1 | 1 | 2 | 3.85 |
| 1 | 1 | 2 | 4.34 |
| 1 | 2 | 1 | 4.85 |
| 1 | 2 | 1 | 5.52 |
| 1 | 2 | 2 | 5.30 |
| 1 | 2 | 2 | 5.93 |
| 2 | 1 | 1 | 7.12 |
| 2 | 1 | 1 | 6.37 |
| 2 | 1 | 2 | 7.70 |
| 2 | 1 | 2 | 7.05 |
| 2 | 2 | 1 | 7.12 |
| 2 | 2 | 1 | 7.89 |
| 2 | 2 | 2 | 7.41 |
| 2 | 2 | 2 | 8.06 |

**c) Seleccionar la opción que realiza el Análisis de Varianza y luego definir la Variable como Dependiente (**COMPRESIÓN**) y las otras tres Variables como Independientes (**EDAD, TEMPERATURA Y TIERRA**). Como en este tipo de Diseño interesa obtener efectos de interacciones dobles y triple; entonces cada Software Estadístico trae la opción donde se le especifica que se deben calcular tales interacciones. Como por ejemplo, en el STATGRAPHIP se le debe colocar en la ventana donde se definen las Variables en la opción "Ignore Interactions Higher than orden:" el valor de 3 que significa que se calculen las interacciones dobles y triple.**

**e) Resultado del Análisis de Varianza que proporciona la computadora.**

**A N Á L I S I S  D E  V A R I A N Z A**

| Fuente de Variación | Suma de Cuadrado | Grados de Libertad | Medias de Cuadrados | F | P>F |
|---|---|---|---|---|---|
| EDAD (A) | 25.2255 | 1 | 25.2255 | 117.92 | 0.0000 |
| TEMPERATURA(B) | 2.8985 | 1 | 2.8985 | 13.55 | 0.0062 |
| TIERRA (C) | 0.2328 | 1 | 0.2328 | 1.08 | 0.3273 |
| INTERACCIÓN | | | | | |
| AB | 0.3393 | 1 | 0.3393 | 1.58 | 0.2434 |
| AC | 0.1425 | 1 | 0.1425 | 0.66 | 0.4465 |
| BC | 0.3150 | 1 | 0.3150 | 0.14 | 0.7151 |
| ABC | 0.3335 | 1 | 0.335 | 1.55 | 0.2471 |
| RESIDUAL | 1.7113 | 8 | 0.2139 | | |
| TOTAL | 30.9149 | 15 | | | |

# BIBLIOGRAFÍA


- DISEÑO DE EXPERIMENTOS
  Principios Estadísticos de Diseño y Análisis de Investigación
  Autor: Robert O. Kuehl
  2ª. Edición

- ESTADÍSTICA
  MODELOS Y METODOS
  Autor: Daniel Peña Sánchez de Rivera
  2ª. Edición
  Alianza Universidad Textos.

- PROBABILIDAD Y ESTADÍSTICA
  Para Ingeniería y Ciencias
  Autor: Jay L. Devore
  4ª. Edición
  International Thomson Editores

- PROBABILIDAD Y ESTADÍSTICA
  Para Ingeniería y Administración
  Autores: William W. Hines, Douglas C. Montgomery
  3ª. Edición
  Compañía Editorial Continental, S.A de C.V.

- DISEÑO Y ANALISIS DE EXPERIMENTOS
  Autor: Douglas C. Montgomery
  Grupo Editorial Iberoamérica S.A de C.V.

- ESTRATEGIAS EXPERIMENTALES PARA LA MEJORA DE LA CALIDAD EN LA INDUSTRIA.
  Autores: John Lawson, Jose L. Madrigal , John Erjavec
  Grupo Editorial Iberoamérica.

- DISEÑO Y ANALISIS DE EXPERIMENTOS
  Modelos y Problemas
  Autor: Prof. Ing. Victor Manuel Avilés
  Universidad Centroamericana José Simeón Cañas
  Impreso en El Salvador por Talleres Gráficos UCA.

- ESTRATEGIAS EXPERIMENTALES PARA EL MEJORAMIENTO DE LA CALIDAD EN LA INDUSTRIA.
  Autores: John Lawson/José L. Madrigal/John Erjevec
  Grupo Editorial Iberoamérica

- MODELOS ESTADISTICOS CON STATGRAPHICS
  Autores: Andés Gonzáles Carmona, José Fernando Vera Vera, Pedro Antonio García López,Diego Terrecilla de Amo, Juan Antonio Maldonado Jurado,María del Mar Rueda García, Antonio Arcos Cebran, Jorge Ollero Hinojosa, Francisco Torres Ruiz. Granada, 1994

- DISEÑOS EXPERIMENTALES MULTIVARIABLES
  Autor: Jaime Arnavi Gras
  Alianza Editorial 1990

- MANUAL DE DISEÑOS EXPERIMENTALES CON APLICACIONES A LA AGRICULTURA Y GANADERIA.
  Autores: Ing.Agr. Julia Amelia Nulia de Mejía
  Ing. Agr. Marco Antonio Mejía Mejía

- DISEÑOS EXPERIMENTALES
  Autor: Cochram G. William
  Editorial Trillas

- ESTADISTICA APLICADA
  Técnica de la Estadística Moderna, Cuando y Donde Aplicarla.
  Autor: Bernard Ostle
  Editorial Limusa-Wile, S.A

- DISEÑO Y ANALISIS DE EXPERIMENTOS
  Autor: Ing.Agr. José René Alvarado Lozano

- DISEÑOS EXPERIMENTALES
  Departamento de Biometría y Estadística
  San Andres, La Libertad, El Salvador, C.A
  Autor: Ing.Agr. José René Alvarado Lozano

- GUIA TECNICA PARA ENZAYOS DE VARIEDADES DE CAMPO
  Autores: J. M. Elena Rosselló , M. Fernández de Gorostiza

- METODOS ESTADISTICOS PARA INVESTIGACIONES AGRICOLAS
  Autores: V. G. Panse , P.V. Sukhatme
  Editorial: Fondo de Cultura Económica
  Mexico-Buenos Aires

- DISEÑO DE EXPERIMENTOS Y TEORIA DE MUESTRAS, Unidad Didáctica/4
  Autor: Miguel Martín Dávila
  Primera Edición. Madrid , Junio 1991

- INFORMACIÓN INVESTIGADA EN LA RED DE INTERNET