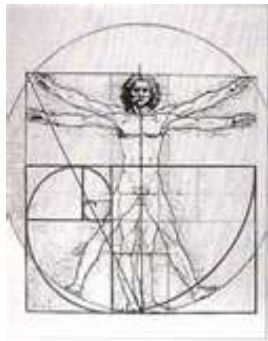


El Diseño Experimental en las Ciencias de la Salud

Miguel Ángel Rodríguez Feliciano





P r o l o g o

Una de las áreas donde empezó a desarrollarse la experimentación, fue en el sector agrícola y pecuario, por la naturaleza de los sujetos con los que se realizaban (plantas y animales). Así que de manera general los ejemplos que en la mayoría de la bibliografía encontramos sobre los diseños de experimentos, en una gran parte de los casos, se basan en ejemplos del sector antes mencionado y el industrial.

El área de la salud se encuentra en ciertos momentos de su desarrollo y a diferentes niveles apoyada por la experimentación; para solucionar o progresar en la cura o tratamiento de enfermedades. Encontrándose esta colaboración en áreas como la epidemiología, la fisiología, la farmacología (diseño de medicamentos), la microbiología, la biología molecular, la inmunología, etc. Por lo que el éxito de esta colaboración, se debe a que los modelos matemáticos de los Diseños Experimentales, permiten describir e interpretar la gran cantidad de números y variables, además de las relaciones que se encuentran presentes en los sistemas biológicos y de salud. Pudiéndose llegar a realizar predicciones sobre diversos patrones de comportamiento de enfermedades, así como la interacción de diversos factores en el desarrollo de la misma. Sin embargo es necesario contar con guías que permitan conocer la estrategia más adecuada para la recolección y análisis de los datos para la elaboración de medicamentos y tratamientos más eficaces; y es aquí donde el Diseño Experimental interviene como la herramienta auxiliar en el desarrollo de estas actividades.

Por lo que el objetivo de este libro es el de aportar una perspectiva completamente orientada al área de la salud, que facilite la comprensión del tema al profesionista que trabaja dentro de la esfera de estos servicios y requiera hacer uso de esta herramienta.

Miguel A. Rodríguez F..



Indice

Prologo

Parte I: Generalidades

Capitulo I:

Principios Básicos del Diseño de Experimentos
Métodos de Investigación
Métodos de Recolección de Datos
Métodos de Interpretación de Datos
Etapas del Método de Investigación Científica
Diseño de Experimentos
Etapas de un Experimento
Diseño Estadístico de Experimentos
Aplicaciones del Diseño de Experimentos
Principios Básicos del Diseño de Experimentos
Elementos del Diseño de Experimentos
Certidumbre de los Resultados de un Experimento
Capitulo II:
Capitulo III: Los modelos Estadísticos y el Análisis de la Varianza
Modelo Matemático Que Origina el Análisis de Varianza
Esquema de Muestreo
Supuestos de Aplicación
Conceptos de Aplicación
Desarrollo del Modelo Matemático del Análisis de Varianza
Hipótesis
Tabla de Análisis de Varianza
Problemario de ANVA

Parte II: Modelos Clásicos

Capitulo IV:

Comparaciones Poblacionales
Comparación de Poblaciones Normales.
Análisis de Dos Poblaciones Independientes
Problemario de Poblaciones Independientes
Comparación de Medias con Grupos Pareados.
Problemario de Grupos Pareados

Capitulo V:

Diseños Simples
Diseño Completamente al Azar
Análisis Estadístico



	Notación
	El Modelo
	Estimación
	Problemario de Diseño
	Completamente al Azar
	Diseño de Bloques al Azar
	Características
	Ventajas
	Desventajas
	Análisis Estadístico
	Notación
	El Modelo
	Estimación
	Eficiencia Relativa del DBA
	Problemario de Diseño de Bloques al Azar
	Diseño en Cuadro Latino
	Características
	Ventajas
	Desventajas
	Distribución de los Tratamientos
	Análisis Estadístico
	Notación
	El modelo
	Estimación
	Eficiencia Relativa del Diseño en Cuadro Latino
	Problemario de Diseño en Cuadrado Latino
Capitulo VI:	Diseños Incompletos
	Bloques al Azar con Observaciones Perdidas
	Cuando se pierde una observación:
	Cuando se pierden dos observaciones
	Cuadro Latino con Observaciones Perdidas
	Cuando se pierde una observación
	Cuando se pierden dos observaciones
Capitulo VII:	Comparación de Medias de Poblaciones o Tratamientos
	Tipos de Tratamientos
	Comparaciones Múltiples de Medias
	Diferencia Mínima Significativa (DMS)
	Método de Duncan



		Método de Student-Newman-Keuls (S-N-K)
		Método de Tukey
		Método de Scheffé
		Método de Dunnett
		Consideraciones Sobre Las Pruebas
	Capitulo VIII:	Comparaciones Ortogonales
		Comparaciones Independientes
		Contrastes Ortogonales
		Problemario de Comparaciones Ortogonales
Parte III:	Arreglos Factoriales	
	Capitulo IX:	Principios Básicos de los Experimentos Factoriales
		Experimentos Factoriales
		Ventajas
		Desventajas
		El Modelo
		Notación
		Representacion Grafica de los Diseños Factoriales
		Análisis Estadístico
		Análisis de las Interacciones
	Capitulo X:	Análisis de un Arreglo Factorial por Contrastes Ortogonales.
		Arreglo Factorial 2 x 2 (22).
		Arreglo Factorial 2x2x2 (23).
		Problemario de Arreglos Factoriales por Contrastes Ortogonales
	Capitulo XI:	Análisis de un Arreglo Factorial por Medio del Modelo Factorial.
		Arreglo Factorial 22
		Arreglo Factorial 23
		Problemario de Arreglos Factoriales por Modelos Factoriales
	Tablas	
		DISTRIBUCION "t"
		Tabla "F"
		DUNCAN
		TUKEY
		DUNNETT
	Resultados	



Parte I: Generalidades



Capítulo I: Principios Básicos del Diseño de Experimentos

Hablar del diseño experimental, pareciera ser un tema árido y difícil de entender, sobre todo en el área de la salud donde las matemáticas no son precisamente su fuerte. Sin embargo es imprescindible el apoyo de esta herramienta para poder avanzar en el conocimiento que permita combatir y prevenir enfermedades. Por lo que una explicación sencilla del diseño experimental es que es una herramienta de la investigación científica, que permite responder preguntas mediante el establecimiento de causalidad entre variables a través del control de las mismas y así poder conocer: “quienes”, “que”, “donde” y “como interaccionan”.

Los Diseños Experimentales, básicamente son arreglos de variables sujetas a estudio que permiten el estudio de un fenómeno de interés, y poder inferir causalidad. Además poseen un modelo matemático que sustenta el posterior análisis estadístico.

La fase que se conoce como Diseño, es el proceso de planear un experimento para obtener datos apropiados que puedan ser analizados mediante métodos estadísticos, con el objeto de producir conclusiones válidas y objetivas. Y la fase del Experimento, es una prueba o una serie de pruebas, en las cuales se inducen cambios deliberados en las variables de entrada de un proceso o sistema, de manera que sea posible observar e identificar las causas de los cambios en la respuesta de salida. Considerándose que los experimentos exitosos consisten en proponer preguntas que son importantes en el campo de la investigación en el que se está trabajando y en efectuar experimentos que las contesten.



Así que el diseño experimental y el análisis estadístico de los datos están estrechamente relacionados, ya que el método de análisis depende directamente del diseño empleado.

Debido a que el diseño experimental, es una herramienta de la investigación científica, es necesario conocer algunos conceptos previos antes de poder entrar de lleno en materia; por lo que antes que nada es necesario hablar sobre que es la Investigación Científica, considerándosele como la búsqueda permanente de la verdad de un hecho mediante métodos objetivos, adecuados y precisos. Por otra parte, como ya se menciono, para poderla realizar se requiere de los Métodos de Investigación, los cuales, son procesos (desarrollo de una actividad) mediante los cuales se obtienen conocimientos para acrecentar el cuerpo de una disciplina científica; pudiéndose aplicar a cualquier nivel de la investigación (Básica o Aplicada): Al mismo tiempo los procedimientos en los que se apoya son una sucesión cronológica de operaciones que se concatenan entre sí, para constituir una unidad de función que realiza una actividad o tarea específica dentro de un ámbito de aplicación.

Por otra parte un término que se usará frecuentemente es el de Variable Estadística, para comprender su significado, consideremos el hecho de que al hacer un estudio en una población, se observa una característica o propiedad de interes en los elementos o individuos que la constituyen. Así que cada una de estas características estudiadas reciben el nombre de variable estadística (edad, peso, sexo, etc).

Dependiendo de la característica podemos distinguir varios tipos de variables:

- Por su estructura
 - Simple.- Mide un solo indicador
 - Compleja.- Se requiere de 2 o más indicadores



- Por la forma de medirse
 - Cualitativas
 - Nominales (nombres).- Nombra la modalidad de una característica sin compararla con grados de intensidad.
 - Ordinales.- Expresa que elementos pueden poseer características en distintos grados o intensidades
 - Cuantitativas
 - Enteras
 - Fraccionadas
- Por dependencia
 - Independientes
 - Dependientes

Métodos de Investigación

El cuadro 1 muestra los dos tipos de métodos de investigación, pudiendo ser generales cuando la información encontrada puede ser usada en diversas aplicaciones, mientras que la específica tienen un propósito muy estrecho y va a dar solución a cosas muy concretas.

Cuadro1. Métodos de Investigación

Generales	Específicos
<ul style="list-style-type: none">• Empírico• Experimental• Estadístico*• Documental	<ul style="list-style-type: none">• Técnicas ó• Rutinas
<ul style="list-style-type: none">• Histórico	



Métodos de Recolección de Datos

Todos los sistemas requieren de una entrada y una salida por lo que en la investigación la entrada, se da en los Métodos de Recolección de Datos, que va a depender del método de investigación empleado siendo algunos de ellos:

- **Empírico.-** Que consiste en la acumulación de información en forma de experiencias, sin haber pasado por una fase de verificación científica; considerándosele la primera fase del método científico
- **Experimental*.-** Cuando se diseña y ejecuta un experimento que genere datos que den respuesta a preguntas concretas planteadas por el investigador para poder interpretar de manera correcta el fenómeno de interés.
- **Histórico.-** Son métodos de recolección retrospectivos, que se encargan de buscar hechos en el pasado de una sociedad
- **Reconocimiento.-** El investigador observa las características del entorno de la unidad de investigación. En el caso del área de salud, se basa en los factores que condicionan la aparición de una enfermedad debido a la exposición a estos; partiendo el estudio con la identificación del factor de riesgo que puede generar la enfermedad y estima el riesgo de adquirirla, es decir va de la causa (factor de riesgo) al efecto (enfermedad).
- **De Casos.-** El investigador observa las características de la unidad de investigación. En el caso del área de salud la o las características se centran en las enfermedades; por lo que el estudio de casos y controles parte individuos que tienen una enfermedad y busca cual es el factor que condiciona su aparición, es decir va desde el efecto (enfermedad) a la causa (factor de riesgo).

Es importante destacar que los métodos de recolección de datos, son el medio a través del cual el investigador se relaciona con los elementos de la población participante para obtener la información necesaria que le permita lograr los objetivos de la investigación.

* *Son los métodos de interés en este libro*



Métodos de Interpretación de Datos

Por otra parte, de nada sirven los métodos de recolección, si no cuentan con los Métodos de Interpretación de Datos, adecuados. Estos se dan en función a la disciplina donde se hace la investigación y al método de investigación empleado, sin embargo de manera general tenemos dos grandes tipos de métodos de interpretación:

- **Estadísticos*.-** Es la herramienta base que utiliza el método científico para el análisis de datos, para una interpretación correcta del fenómeno de interés. Cuyo objetivo es hacer referencia de los errores aleatorios que se presentan en la mediciones (de que manera influyen y cual es su magnitud), es decir también se encarga de analizar la validez de los resultados
- **Documental.-** Método muy usado en las ciencias sociales, encarga de relatar, describir, comparar en orden eventos actuales o pasados que no precisamente requieren de una medición como lo es en el estadístico.



Etapas del Método de Investigación Científica

1. Observación: Observación en forma crítica de un hecho o fenómeno. Colecta hechos por: a) Observación, b) Experimentación empírica. En este paso no se concluye.
2. Planteamiento del Problema: Aquí se formula el problema en forma específica y con precisión. Se fundamenta en los conceptos de la ciencia (conocimientos científicos ya elaborados).
3. Formulación de Hipótesis: Se inicia a partir del análisis del problema planteado. Tiene la finalidad de explicar los hechos conocidos y pronosticar los desconocidos. Su alcance depende de su capacidad lógica (no debe ser auto contradictoria).
En el planteamiento de una Hipótesis no debe haber una negación; además la investigación básica no lleva hipótesis.
Estas deben ser operacionales para poder comprobarlas (debe de contener variables que podamos operar). Además contendrá la información de las variables de interés para su interpretación y conclusión.
Se deben proponer hipótesis bien definidas y fundamentadas; y no ser suposiciones que no comprometan nada en concreto, ni ocurrencias sin fundamentos.
4. Verificación de la Hipótesis: Este paso nos ayuda a adquirir información relevante al problema; además de conclusiones con cierto grado de confiabilidad.
Sin embargo esto dependerá de: a) Cantidad de información, b) Forma de recolección, c) Técnicas de análisis. Aquí es donde es donde los métodos estadísticos son una herramienta indispensable
5. Conclusiones: Deben ser lo más relevantes sobre el fenómeno estudiado. Cuyo propósito es el de confirmar o en caso contrario, aumentar los conocimientos al cuerpo de la ciencia.



Diseño de Experimentos

Un experimento (proceso mediante el cual obtenemos una observación o dato) diseñado es una prueba o una serie de pruebas, en las cuales se inducen cambios deliberados en las variables de entrada de un proceso o sistema, de manera que sea posible observar e identificar las causas de los cambios en la respuesta de salida.

Por lo que cualquier tipo de Experimentación que se realice forma parte del método científico debido a que realiza operaciones ordenadas, destinadas a demostrar, comprobar o descubrir fenómenos o principios básicos de interés para el investigador

Etapas de un Experimento

- **Planear:** Identifica los elementos o variables a evaluar, del fenómeno de interés. Plantea los Objetivos y las Hipótesis.
- **Diseñar:** Aquí se selecciona y aplica el diseño más eficiente y que minimice el error experimental (forma de la distribución del experimento). Identifica los factores que afecten a las variables de estudio y determina el tamaño de la muestra.
- **Ejecución:** Seguir el diseño seleccionado; realizando todas las actividades por igual excepto las que se desean evaluar. Recolecta de manera ordenada las observaciones.
- **Análisis e Interpretación:** Se organiza la información y se aplica el método estadístico apropiado, para evaluar el diseño y probar las hipótesis planteadas.



Diseño Estadístico de Experimentos

Es el proceso de planear un experimento para obtener datos apropiados que puedan ser analizados mediante métodos estadísticos, con el objeto de producir conclusiones validas y objetivas. El diseño de experimentos y el análisis estadístico de los datos están estrechamente relacionados, ya que el método de análisis depende directamente del diseño empleado.

Para tener validez en las conclusiones, los objetivos deben de establecerse con claridad y plantearse como:

- Preguntas que se tienen que contestar,
- Hipótesis que se tienen que probar y
- Efectos que hay que estimar.

Además se debe de considerar:

- Qué variables tienen mayor influencia en la respuesta.
- Cuál es el valor de la variable independiente (X) que mejor influye en la variable dependiente (Y), de modo que “Y” tenga siempre un valor cercano al valor nominal deseado.
- Cuál es el valor de la variable independiente (X) que mejor influye en la variable dependiente (Y), de modo que “Y” sea pequeña.
- Cuál es el valor de la variable independiente (X) que mejor influye en la variable dependiente (Y), de modo que se minimicen los efectos de las variables incontrolables.

Aplicaciones del Diseño Experimental

Las aplicaciones de técnicas de diseño experimental pueden dar por resultado:

- Lograr mejorar el rendimiento de un proceso
- Obtener menor variabilidad y mayor apego a los requerimientos nominales u objetivos.
- Optimizar los tiempos de desarrollo.
- Minimizar costos globales y
- Es empleado extensamente en el desarrollo de nuevos productos o para mejorar productos ó técnicas ya existentes.



Principios Básicos del Diseño de Experimentos

Se consideran 3 aspectos fundamentales

- 1) El Experimento
 - Establecimiento y definición completa del problema.
 - Selección de las variables de respuesta o variables dependientes.
 - Selección de los factores o variables independientes que afectan las variables de respuesta.
 - Selección de los niveles de los factores y sus combinaciones
- 2) El Diseño
 - Número de observaciones o ensayos a realizar.
 - Aleatorización y orden del experimento
 - Modelo matemático que describe el experimento
- 3) El Análisis
 - Captura y procesamiento de la información.
 - Forma de hacer las pruebas estadísticas
 - Análisis e interpretación de los resultados
 - Costo del experimento
 - Conclusiones y recomendaciones

Elementos del Diseño de Experimentos

Aleatorización: Es una técnica ideada por Fisher. Y no es otra cosa que la asignación de los tratamientos a las unidades experimentales (UE). Por lo que todas las UE tienen las mismas probabilidades de recibir un tratamiento. Permite: a) Estimación insesgada del Error Experimental (EE). b) Estimación imparcial de la media de los tratamientos.

Bloque: Conjunto de UE lo más homogéneas posibles; aparecen todos los tratamientos una sola vez. Aquí el gradiente de variabilidad es perpendicular al bloque. Tiene como objetivo minimizar el EE.

Diseño de Experimentos: Ya se habló de lo que es el diseño experimental, pero ahora lo haremos en función de los elementos que lo constituyen; por lo que se considera que es el procedimiento para asignar los tratamientos a las unidades experimentales en los



bloques. Es un método aleatorio, pudiéndose decidir por sorteo o promedio de números aleatorios, que incluye: Bloques, Aleatorización, Tratamientos y Repeticiones.

Error Experimental (EE): Es debido a la variabilidad proveniente de los factores que afectan a los tratamientos. Estos factores enmascaran el efecto de los tratamientos. El EE. no es una equivocación. Mide la variabilidad por: a) Variación de la UE, b) Variaciones del manejo de las UE, c) Variaciones por efecto del ambiente. Estas variaciones no se pueden eliminar pero si se pueden reducir. Tienen como finalidad obtener una mejor estimación de los efectos de los tratamientos. Si EE es el error experimental y es uniforme para todas las UE entonces EE_i tiene una distribución normal con una media de cero y una varianza $[EE \sim N(0, \sigma^2)]$. Además está presente en la prueba de hipótesis. Existen 2 tipos de errores: a) Error tipo I: Rechazar la hipótesis verdadera, b) Error tipo II: Aceptar una hipótesis falsa. Las frecuencias de ocurrencia de ambos se dan en términos de probabilidad. El error más importante es el tipo I $[\alpha = \max P(\text{Error})]$ siendo los niveles de significancia convencionales 1% y 5%.

Espacio Muestral: Conjunto de todos los resultados posibles de un experimento aleatorio (población). Cada resultado se denomina punto muestral.

Experimento: Son de dos tipos a) Determinístico: Determina un mismo resultado bajo ciertas condiciones invariables, b) Aleatorio: No siempre se obtiene el mismo resultado aunque desarrollen bajo idénticas condiciones (uso de probabilidades)

Factor: Elemento, condicionante que contribuye a lograr un resultado

Repetición: Aplicación de un tratamiento 2 o más veces en un experimento. El objetivo es: a) Estimar el error experimental, b) Medición más precisa del efecto de los tratamientos, c) Disminuir la desviación estándar de la muestra de la media.

Testigo: Sujeto de comparación. Son los elementos base para la comparación de los tratamientos.

Tratamiento: Elemento, sujeto o ensayo de estudio en cantidad o calidad. Es cualquier variable cuyo efecto se desea medir.



Unidad Experimental (UE): Material o lugar sobre el cual se aplican los tratamientos bajo estudio. El efecto de los tratamientos se mide en la UE.

Variable Aleatoria: Función que transforma a los eventos definidos en un espacio muestral en números reales. Generan funciones de probabilidad (discretas) y de densidad (continuas). Siendo las variables: a)Discretas: Bernoulli, Binomial, Poisson, b)Continua: Normal, X^2 , 't', F

Referencias

- Andión G. M., Walterio Beller, Heinz Dleterich
Taboada. Guía De Investigación Científica. Ediciones
De Cultura Popular. 2ª. Reimpresión. México D.F.
1988.
- Box, G. E., Hunter, W., Hunter, J. S., "Estadística para
investigadores". Ed. Reverté. 1998.
- De Canales F. H., Eva Luz De Alvarado, Elia Beatriz
Pineda. Metodología De La Investigación. Manual
Para El Desarrollo De Personal De Salud. Editorial
Limusa. 2ª Impresión. México, D.F. 1988.
- Elston, R.C., W.D. Johnson. 1990. Principios De
Bioestadística. México, Ed. El Manual Moderno, S.A.
De C.V.
- Box G.E.P., W.G. Hunter, J.S. Hunter Estadística para
experimentadores. Ed. Reverté S.A. Barcelona 1989.
- Box G. E. P., William G. Hunter, J. Stuart Hunter.
Estadística Para Investigadores. Editorial Reverté. 1ª
Reimpresión. México, D.F. 1999.
- Martínez, A. Diseños Experimentales. Métodos y
Elementos de Teoría. Editorial Trillas, México. 1988.
- Ostle B. Estadística Aplicada. Limusa-Wiley, México,
1975,
- Rojas Soriano Raúl. El Proceso De La Investigación
Científica. Editorial Trillas. 4ª. Edición. México, D.F.
1981.
- Winer, B.J., Brown, D.R. and Michels, K.M. Statistical
Principles in Experimental Design. 1991.



Parte I: Generalidades



Capítulo II: Certidumbre de los Resultados de un Experimento

Una de las principales consideraciones a la hora de desarrollar una investigación apoyada en la experimentación, es el hecho de que tan confiables sean los resultados que se obtengan, por lo que para poder estar seguros sobre lo que estamos haciendo, es necesario tomar en consideración algunos aspectos que abordaremos en este capítulo.

Por principio de cuentas, es necesario considerar que los resultados de los experimentos no solo se ven afectados debido a la acción de los tratamientos bajo estudio, si no que también a causas desconocidas que encubren sus efectos, a los que se les conoce como Error Experimental, teniendo como principales fuentes a:

1. La variabilidad inherente al material experimental (UE) al cual se aplican los tratamientos.
2. La variabilidad debida a la falta de uniformidad en la conducción del experimento.

A pesar de que los Errores Experimentales pueden tener una gran influencia en los resultados, el efecto de estos se puede diferenciar de los efectos de los tratamientos.

En los resultados experimentales aún en los experimentos muy bien planeados y conducidos, no se puede estimar la exactitud en ellos, debido a que no se conoce el valor verdadero con el que se debe de comparar la aproximación del valor obtenido; pero se puede trabajar sobre la precisión de los resultados, es decir sobre la reproducibilidad de estos. Por lo que en términos de probabilidad, se puede señalar cierta confiabilidad de los resultados mientras sean aproximados unos de otros.

En la experimentación, el uso del Diseño Experimental permite obtener estimaciones insesgadas de Medias, Diferencias de Tratamientos y del Error Experimental.



Esto indica que la medición del valor de una variable aleatoria experimental, variará en mayor o menor cantidad del valor verdadero de la variable, Si se repite la medición varias veces, éstas variarán alrededor del valor verdadero; a esto se llama: Exactitud.

Exactitud: Es la cercanía con la cual la medición de una variable se aproxima al valor verdadero de esa variable

El valor verdadero de una variable aleatoria no se conocerá con exactitud, debido a los errores que se cometen en la medición y los factores que influyen sobre esa variable.

Las mediciones que se hacen tienen sesgos, más aún cuando se trabajan con muestras, ya que se calculan estimadores y no parámetros poblacionales.

Sesgo: Un sesgo del estimador es la diferencia entre el valor esperado de una variable aleatoria y el valor verdadero.

A medida que sesgo disminuye o aumenta el número de repeticiones del experimento, la exactitud aumenta. El resultado del estimador puede repetirse varias veces o no, dando como resultado la siguiente pregunta ¿Son confiables los datos que obtenemos y las recomendaciones que hacemos?

Los valores de una variable aleatoria son obtenidos de condiciones reales pero controladas, por lo que los datos reflejan la realidad pero ¿Que tan confiables son?

La respuesta a la anterior pregunta requiere la consideración de que es necesario establecer una medida, que permita en términos relativos tener confianza en los datos colectados, entonces ¿Como se debe de medir la confiabilidad?

Antes de responder debemos de comprender cuando un proceso es confiable. Por lo que “Un procedimiento es confiable, cuando en todos los casos genera el mismo resultado”

Por lo que la confianza va a variar de acuerdo a la frecuencia con la que se obtiene el resultado deseado. Esto conduce a la utilización de la teoría de las probabilidades para la medición de la confianza; siendo el Nivel de Significancia o Significación quien da la medición de la confianza

Al probar un conjunto de tratamientos, el Análisis de Varianza (ANVA) sirve para probar la Hipótesis Nula (H_0) contra la Hipótesis Alternativa (H_a) en donde:



- H_0 : El efecto de los tratamientos es el mismo (no hay diferencia)
- H_a : Existen diferencias entre el efecto de los tratamientos

La prueba de hipótesis se efectúa a través de una prueba de significancia estadística, mediante la cual se comprueba si la Hipótesis nula es “verdadera” o “falsa”, teniendo como probabilidades de error lo que se observa en el cuadro 2:

Cuadro 2.- Probabilidades de la hipótesis nula

Decisión	H_0	
	Falsa	Verdadera
Rechazar la Hipótesis	Correcta	Error tipo I (α)
No Rechazar la Hipótesis	Error tipo II (β)	Correcta

Cuando se rechaza a una hipótesis verdadera se le llama Error Tipo I y el nivel de significancia es la probabilidad de cometerlo; esto es, la probabilidad de aceptar que existen diferencias entre tratamientos cuando no las hay.

Nivel de significancia = α = p (Error Tipo I).

La elección del nivel de significancia (α) está determinado por los costos y riesgos que implica una decisión incorrecta, teniéndose en el área de salud mayor poder de decisión en la selección del nivel de significancia los riesgos, por lo que cuando son muy grandes el nivel (α) debe ser pequeño. Por lo que el estándar utilizado en la mayoría de los experimentos, con el estadístico de prueba “F” es de $\alpha=0.05$ (5%) y $\alpha=0.01$ (1%), que proporcionan una confianza adecuada, pero se pierde al utilizar pruebas no confiables.

Los factores que afectan la confianza en los resultados experimentales son:

- Conducción del experimento
- Elección del sitio experimental
- Número de repeticiones

Al no detectarse diferencias entre tratamientos, se pueden encontrar tendencias. Ya que en la mayoría de los casos se encontrarán algunos tratamientos que rinden más que otros y la probabilidad de que esa tendencia implique diferencias es α ; entonces con la Hipótesis Alternativa (H_a) se plantea la pregunta ¿De que magnitud nos interesan estas diferencias?



Para poder contestar la pregunta, se requiere definir el término precisión.

Precisión: Es la capacidad de repetición de las medidas estudiadas.

Es la magnitud de la diferencia entre dos tratamientos que un experimento es capaz de detectar.

La precisión es de gran importancia en la investigación para probar hipótesis mediante el Análisis de Varianza.

Ya que las hipótesis estadísticas, son complementarias y mutuamente excluyentes, siempre se van a tener dos hipótesis, la nula (H_0) y la alterna (H_a), que es donde realmente descansa la idea del investigador; en el caso del análisis de varianza (ANVA) para el diseño de experimentos la hipótesis nula siempre se plantea de la forma: “El efecto de los tratamientos es el mismo”, y que esto sea cierto depende de la precisión que deseamos tener en la investigación, y la precisión va a depender de los costos y problemas en la realización del experimento.

Además la precisión, se mide por el inverso de la varianza de la media:

$$\text{Precisión} = \frac{1}{\sigma^2} = \frac{n}{\sigma^2}$$

Por lo que:

- A medida que σ^2 aumenta, la precisión decrece.
- A medida que n aumenta la precisión aumenta.
- La precisión de un experimento es afectada por:
 1. La conducción del experimento
 2. Técnicas de análisis de la información
 3. Número de repeticiones



Referencias

Box, G. E., Hunter, W., Hunter, J. S., "Estadística para investigadores". Ed. Reverté. 1998.

Elston, R.C., W.D. Johnson. 1990. Principios De Bioestadística. México, Ed. El Manual Moderno, S.A. De C.V.

Box G.E.P., W.G. Hunter, J.S. Hunter Estadística para experimentadores. Ed. Reverté S.A. Barcelona 1989.

Box G. E. P., William G. Hunter, J. Stuart Hunter. Estadística Para Investigadores. Editorial Reverté. 1ª Reimpresión. México, D.F. 1999.

Martínez, A. Diseños Experimentales. Métodos y Elementos de Teoría. Editorial Trillas, México. 1988.

Ostle B. Estadística Aplicada. Limusa-Wiley, México, 1975,

Prat, A., Tort-Martorell, X., Grima, P., Pozueta, L., "Métodos estadísticos. Control y Mejora de la Calidad". Ed. UPC (1997).

Winer, B.J., Brown, D.R. and Michels, K.M. Statistical Principles in Experimental Design. 1991.



Parte I: Generalidades



Capítulo III:

Los modelos Estadísticos y el Análisis de la Varianza

Los diseños de experimentos tienen como herramienta base al análisis de varianza (ANVA), por lo que no es posible entrar de lleno en materia, sin antes analizar la forma en que trabaja el ANVA.

De manera general se considera que un MODELO, es la “representación a escala de un fenómeno de la vida real”. Y un Modelo Matemático, es el que se encarga de relacionar las variables que describen o explican el fenómeno. Dentro de los Modelos Matemáticos tenemos que estos se dividen en dos grandes grupos:

- **Determinístico:** Establecen una relación exacta entre las variables.
- **Estadístico:** La relación entre las variables depende del aspecto aleatorio de los fenómenos.

Debido a que en el área de SALUD, se trabaja con sistemas biológicos, el modelo matemático que satisface las características de los fenómenos en estudio es el estadístico.

Modelo Matemático Que Origina el Análisis de Varianza

En el proceso de recolección de datos, las observaciones (y_i) provenientes de una población normal pueden ser representados por la ecuación:

$$y_i = \mu + \varepsilon_i \quad \text{con } i = 1, 2, 3, \dots, n \quad \text{Ecuación (1)}$$

En donde:

- μ : Es una constante, la cual depende de un conjunto de factores con el mismo efecto para todos los valores de la población. Es la media de todas las observaciones (y_i) de la población



- ε_i : Es el error de muestreo para cada observación (y_i). Es decir, es la parte aleatoria, la cual depende de un conjunto de factores que influyen en diferente forma sobre el fenómeno.

En la ecuación (1), los ε_i son aleatorios y pueden tomar valores tanto positivos como negativos, debido a que los y_i también son aleatorios; y μ es una constante. Por lo que cualquier inferencia sobre μ dependerá del modelo probabilístico que se suponga para la variable ε_i

Esquema de Muestreo

En el esquema de muestreo se tiene: $y_i = \mu + \varepsilon_i$

Donde se considera que los datos deben de tener una distribución normal con una media y una varianza [$y_i \sim N(\mu, \sigma^2)$] y en consecuencia el error también debe de presentar una distribución normal, con una media de cero y una varianza [$\varepsilon_i \sim N(0, \sigma^2)$] como se observa en las siguientes figura 1.

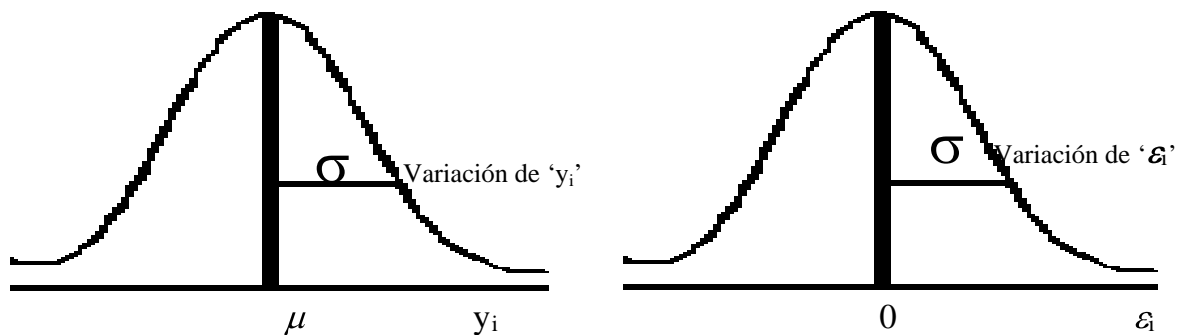


Figura 1.- Distribución de los valores de y_i y de ε_i , siendo μ una constante

Los ε_i son valores de variables aleatorias no observables, entonces la base de la inferencia sobre la media poblacional (μ), que es desconocida, son los valores (y_i). En las gráficas mostradas, lo único que se hizo fue la traslación de ejes de las abscisas, esto es, un cambio en el parámetro de localización tanto “ ε_i ” como para “ y_i ”.

La importancia de la ecuación (1), reside en la representación de la variable y_i como la suma de un parámetro “ μ ” y una variable aleatoria no observable “ ε_i ”, dando origen a una ecuación que pretende explicar el comportamiento de la variable aleatoria “ y_i ”.



Para el análisis estadístico de la ecuación (1) puede utilizarse el método del Análisis de la Varianza (ANVA), el cual es un procedimiento aritmético, que consiste en descomponer la Suma de Cuadrados Total (Variación Total) en fuentes de variación reconocidas, incluyendo la variación que no se ha podido medir, que es el ERROR EXPERIMENTAL.

Supuestos de Aplicación

En la aplicación del ANVA se supone que:

- Los efectos de los tratamientos y los ambientales son aditivos.
- El Error Experimental constituye un elemento al azar, normal e independiente, con una distribución normal con una media 0 y una varianza (σ^2).

Conceptos de Aplicación

Dos conceptos en la aplicación del ANVA.

- Grados de Libertad (G.L.): Es el número de contrastes o comparaciones ortogonales (independientes) menos el número de restricciones (en este caso, las medias de las hipótesis) impuestas que se realiza en un grupo de datos (n-1).
- Cuadrado Medio (C.M.): Es el cociente de una suma de cuadrados (SC) entre su respectivo grado de libertad. [CM = SC / GL]

Desarrollo del Modelo Matemático del Análisis de Varianza

Retomando la ecuación (1), despejando ε_i tenemos:

$$\text{Si } y_i = \mu + \varepsilon_i \text{ con } y = 1, 2, 3, \dots, n \implies \varepsilon_i = y_i - \mu \quad (2)$$

En el esquema de muestreo tenemos como referencia a la media muestral (\bar{Y}), por lo que desarrollando a la ecuación (2) tenemos:

$$y_i - \mu = (y_i - \bar{Y}) + (\bar{Y} - \mu) \quad (3)$$

La figura 2 describe gráficamente a la ecuación

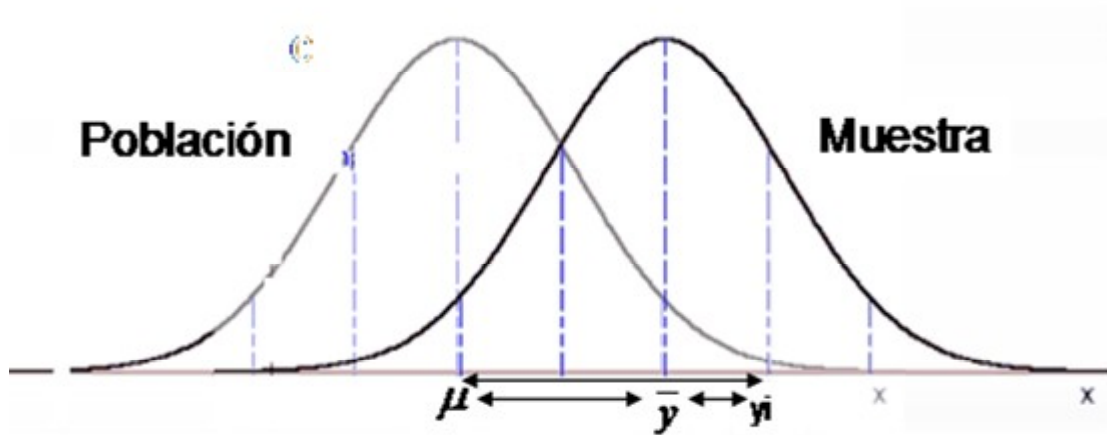


Figura 2.- Interpretación gráfica de la ecuación 3

Elevando al cuadrado a la ecuación 3 tenemos:

$$(y_i - \mu)^2 = [(y_i - \bar{y}) + (\bar{y} - \mu)]^2 \quad (4)$$

Como la ecuación (4) es cierta para todas y cada una de las y_i , se tienen que sumar todas y cada una de las diferencias:

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n [(y_i - \bar{y}) + (\bar{y} - \mu)]^2 \quad (5)$$

Desarrollando el binomio al cuadrado de la ecuación (5)

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n [(y_i - \bar{y})^2 + 2(y_i - \bar{y})(\bar{y} - \mu) + (\bar{y} - \mu)^2]$$

Eliminando los corchetes tenemos:

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n 2(y_i - \bar{y})(\bar{y} - \mu) + \sum_{i=1}^n (\bar{y} - \mu)^2 \quad (6)$$

En la ecuación (6) se tiene que la media muestral (\bar{y}) y poblacional (μ) son constantes, entonces la diferencia de las constantes ($\bar{y} - \mu$) es constante; además, tomando en consideración que la sumatoria de una constante, es igual al producto del número de veces que aparece por el valor de la misma ($\sum_{i=1}^n C = nC$) y que la sumatoria de una constante por una variable es igual al producto de la sumatoria de la constante por la sumatoria de la variable ($\sum_{i=1}^n cX_i = \sum_{i=1}^n c \sum_{i=1}^n X_i$), entonces:

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + 2(\bar{y} - \mu) \sum_{i=1}^n (y_i - \bar{y}) + n(\bar{y} - \mu)^2 \quad (7)$$



La sumatoria de las diferencias de las observaciones con la media es igual a cero $\sum_{i=1}^n (y_i - \bar{y}) = 0$, entonces:

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \quad (8)$$

La ecuación (8) descompone a la variación total [errores de muestreo (ϵ)] en dos partes, que conducen a la suma de Cuadrados (SC) de las desviaciones originales en la ecuación (3); razón por la cual se le llama a la ecuación (8), SUMA DE CUADRADOS (SC).

Es importante determinar la distribución de los componentes de la ecuación (8), los cuales generarán un procedimiento para probar hipótesis sobre la media poblacional (μ). Esto se realiza de la manera siguiente:

Debido a que la variación encontrada (errores de muestreo), es tan solo la de la muestra, y ésta representa tan solo una parte de la variación poblacional; se debe de considerar tan solo como la fracción correspondiente, por tal motivo se debe de dividir a la ecuación 8 entre la varianza poblacional (σ^2), obteniéndose la ecuación 9.

$$\frac{\sum_{i=1}^n (y_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2} \quad \text{Ecuación (9)}$$

Analizando los componentes de la ecuación (9), de manera general se observa que tienden a comportarse de manera más parecida (algunos más que otros) a la distribución Ji-cuadrada (ver figura 3). Por lo que se analizan en detalle a cada uno de los elementos que forman la ecuación

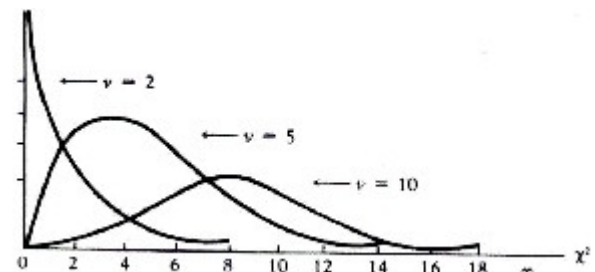


Figura 3.- Gráfica de distribución X^2



$$\frac{\sum_{i=1}^n (y_i - \mu)^2}{\sigma^2}$$

a) Distribución del componente:

Las observaciones tienen una distribución normal, con una media y una varianza poblacional $y_i \sim N(\mu, \sigma^2)$ estandarizando la ecuación se observa que tiene una distribución normal con una media de cero y varianza de uno.

$\frac{(y_i - \mu)}{\sigma} \sim N(0,1)$ por lo que esta ecuación presenta una

distribución Ji- cuadrada $\frac{(y_i - \mu)^2}{\sigma^2} \sim X^2$

Además, como las observaciones (y_i) son independientes,

entonces: $\frac{\sum_{i=1}^n (y_i - \mu)^2}{\sigma^2} \sim X^2_{(n)}$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)}$$

$$x^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$x^2 = \frac{\sum (x_i - \bar{x})^2}{\sigma^2}$$

b) Analizando el componente: $\frac{n(\bar{y} - \mu)^2}{\sigma^2}$

La media muestral tiene distribución normal con una media y una varianza poblacional $\bar{y} \sim N(\mu, \sigma^2/n)$ estandarizando la ecuación tenemos que tiene una distribución normal con una media de cero y una varianza de uno; por lo que esta ecuación presenta una distribución aproximadamente Ji-cuadrada.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)}$$

$$x^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$x^2 = \frac{\sum (x_i - \bar{x})^2}{\sigma^2}$$

$\frac{(\bar{y} - \mu)}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \implies \frac{(y_i - \mu)^2}{\sigma^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2} \sim X^2_{(1)}$

c) Distribución del componente $\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\sigma^2}$



Tomaremos como base a la ecuación del análisis de varianza de la muestra:

$$S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

despejando esta ecuación tenemos:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = (n-1)S^2$$

y si a esto lo dividimos entre la varianza poblacional (σ^2) tenemos:

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim X^2_{(n-1)}$$

que presenta una distribución Ji-cuadrada

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)}$$

$$x^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$x^2 = \frac{\sum (x_i - \bar{x})^2}{\sigma^2}$$

Por lo tanto la ecuación (9)

$$\sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2} = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\sigma^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2}$$

es equivalente a la partición de la x^2 siguiente:

$X^2(n) = X^2(n-1) + X^2(1)$ <p>(10)</p>	Ecuación
--	----------

Donde:

$X^2(n)$: Varianza total

$X^2(n-1)$: Varianza debida a la muestra (Error Experimental)

$X^2(1)$: Varianza debido a la media poblacional



Hipótesis

Una vez obtenidas las distribuciones dadas por la ecuación (9) se pueden utilizar para probar hipótesis sobre medias poblacionales (μ). La partición de la variabilidad dada por la ecuación (10), permite probar hipótesis únicamente para μ , esto es, probar el juego de hipótesis siguiente:

$$\begin{array}{llll} \text{Ho: } \mu = \mu_0 & \text{VS} & \text{Ha: } \mu \neq \mu_0 \\ \text{Ó} & & & \\ \text{Ho: } \mu - \mu_0 = 0 & \text{VS} & \text{Ha: } \mu - \mu_0 \neq 0 \end{array}$$

Siendo μ_0 el valor supuesto del parámetro desconocido y μ La media poblacional real.

El Análisis de Varianza “Estima” semejanzas o diferencias de parámetros.

Para probar las hipótesis antes descritas, se debe derivar una estadística de Prueba de hipótesis para μ , para ello se utilizan los siguientes componentes de la ecuación (9):

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2}$$

La estadística de prueba a utilizar no deberá contener a la varianza poblacional (σ^2) ya que esta es desconocida, por lo cual se derivará a partir de una razón de X^2 , dividiendo a una entre la otra de la manera siguiente:

$$\frac{\frac{n(\bar{y} - \mu)^2}{\sigma^2 (1)}}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2 (n-1)}} = \frac{\frac{n(\bar{y} - \mu)^2}{1}}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} = \frac{n(\bar{y} - \mu)^2}{S^2} = \sim F_{1, n-1}$$

El producto de esta razón nos da como resultado una ecuación cuya distribución es la del estadístico de prueba “F”. Por lo que si la hipótesis nula $\text{Ho: } \mu = \mu_0$ es cierta, entonces:

$$F_c = \frac{n(\bar{y} - \mu)^2}{S^2} \sim F_{1, n-1} \text{ Es decir la “F” calculada se aproxima a una}$$

distribución “F” teórica (de tablas). Por lo tanto se utiliza F_c para probar el juego de hipótesis planteado al inicio.



REGLA DE DECISIÓN.- La regla de decisión para la prueba con un nivel de significancia α es de:

$$\text{Rechazar } H_0 \text{ si } F_c > F_{n-1, \alpha}^1$$

Siendo F_c la calculada y la $F_{n-1, \alpha}^1$ la de tablas

Tabla de Análisis de Varianza

Todo el proceso antes descrito se resume en una tabla llamada Tabla de Análisis de Varianza (ANVA), que a continuación se detalla en el cuadro 3:

Cuadro 3.- Análisis de Varianza

Fuentes de Variación (FV)	Grados de Libertad (GL)	Sumas de Cuadrados (SC)	Cuadrado Medio (CM)	Fo
Media	1	$n(\bar{y} - \mu_0)^2$	$n(\bar{y} - \mu)^2/1$	$n(\bar{y} - \mu)^2/S^2$
Error	$n-1$	$\sum_{i=1}^n (y_i - \bar{y})^2$	$[\sum_{i=1}^n (y_i - \bar{y})^2 / n-1] = S^2$	
Total	N	$\sum_{i=1}^n (y_i - \mu_0)^2$		

El estadístico “F” calculado, bajo la hipótesis nula, tiene una distribución F_{n-1}^1 . Debemos de tener siempre presente a la regla de decisión, ya que a través de esta evaluaremos a la hipótesis nula, rechazándola si la F calculada es mayor que la F de tablas (rechazar H_0 si $F_c > F_{n-1, \alpha}^1$)

Si el propósito del ANVA es probar el juego de hipótesis $H_0: \mu=0$ VS $H_a: \mu \neq 0$, el ANVA es equivalente a la tabla anterior, únicamente haciendo la transformación de $x_i = y_i - \mu_0$, en donde las nuevas variables tendrán como centro el valor de cero, quedando el análisis de varianza como se observa en el cuadro 4.

Cuadro 4.- Análisis de Varianza cuando $H_0: \mu=0$

FV	GL	SC	CM	Fo
Media (μ)	1	$n \bar{x}^2$	$n \bar{x}^2/1$	$n \bar{x}^2/S_x^2$
Error	$n-1$	$\sum_{i=1}^n (x_i - \bar{x})^2$	$[\sum_{i=1}^n (x_i - \bar{x})^2 / n-1] = S^2$	
Total	N	$\sum_{i=1}^n x_i^2$		

La regla de decisión que se toma es la misma que para la tabla antes descrita.

**Ejemplo:**

En un estudio sobre los niveles de hemoglobina (Hb) de una comunidad ubicada 1600 m.s.n.m. en la costa del Estado de Chiapas, se cuantificaron las concentraciones de Hb de 28 varones que accedieron a participar en el trabajo, siendo los valores obtenidos:

12.72	13.38	13.94	17.34	15.74	14.60	19.03
14.11	13.01	17.53	19.25	13.72	12.26	13.29
18.92	17.65	12.13	13.90	10.41	15.03	14.44
13.62	11.49	14.75	13.68	14.81	5.21	17.03

De estudios previos se sabe que las concentraciones de hemoglobina en personas que viven en zonas con alturas similares a las de la comunidad en estudio es de 13.23 gr/dL.

- ¿Cual es la población de interés en este estudio?
- ¿Se puede considerar que la concentración de Hb de los habitantes de la comunidad en estudio es igual a la reportada en comunidades con alturas similares. Si es así a que nivel de significancia?

Solución:

- La concentración de hemoglobina

El juego de hipótesis a probar es: $H_0: \mu = \mu_0$ VS $H_a: \mu \neq \mu_0$

Como μ_0 = Hb reportada en comunidades con alturas similares: $\mu_0 = 13.23$

Entonces $H_0: \mu = 13.23$ VS $H_a: \mu \neq 13.23$

Este juego de hipótesis se probará mediante el ANVA. En donde la suma de cuadrados es:

$$\begin{aligned} \text{S.C. Media} &= n(\bar{y} - \mu_0)^2 \\ \text{S.C. Error} &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2 / n \\ \text{S.C. Total} &= \sum_{i=1}^n (y_i - \mu_0)^2 \end{aligned}$$

De los datos obtenidos tenemos:

$$\begin{aligned} n &= 28 \\ \sum y &= 402.93 \\ \sum y^2 &= 6029.9671 \\ \sum y/n &= \bar{y} = 14.39 \end{aligned}$$

Sustituyendo en las formulas tenemos:

$$\begin{aligned} \text{S.C. Media} &= n(\bar{y} - \mu_0)^2 = 28 (14.39 - 13.23)^2 = \underline{37.67} \\ \text{S.C. Error} &= \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2 / n = 6029.9671 - (402.93)^2 / 28 = \underline{231.66} \end{aligned}$$



$$\text{S.C. Total} = \sum_{i=1}^n (y_i - \mu_0)^2 = \text{S.C. Error} + \text{S.C. Media} = 231.66 + 37.67 = \underline{269.33}$$

Cuadro 4.- Análisis de Varianza del ejemplo

FV	GL	SC	CM	Fo
Media	1	$n(\bar{y} - \mu_0)^2$	$n(\bar{y} - \mu)^2/1$	$n(\bar{y} - \mu)^2/S^2$
Error	n-1	$\sum_{i=1}^n (y_i - \bar{y})^2$	$[\sum_{i=1}^n (y_i - \bar{y})^2 / n-1] = S^2$	
Total	n	$\sum_{i=1}^n (y_i - \mu_0)^2$		
FV	GL	SC	CM	Fc
Media	1	37.67	37.67	4.39*
Error	27	231.66	8.58	
Total	28	269.33		

$$F_{27,0.05}^1 = 4.17$$

$$F_o > F_t$$

$$F_{27,0.01}^1 = 7.56$$

$$F_o < F_t$$

b) Existe diferencia significativa al 5% por lo que solo a este nivel se rechaza H_0 .

Conclusión: Los niveles de Hb de la comunidad en estudio, son **DIFERENTES** a los de la población utilizada como referencia con un nivel de significancia del 5%

Con este ejemplo concluimos con la parte relacionada con el análisis de varianza, siendo esta la forma más sencilla de su uso. A partir de aquí la tabla del ANVA crecerá pero tan solo en el número de hileras debido al incremento de los factores de variación, quedando permanentemente con el mismo número de columnas

Problemario de ANVA

1. Se considera que la infertilidad masculina que existe en una comunidad, es debido a una oligospermia (bajo nivel espermático), siendo lo normal 20 millones/ml. Se hace un estudio con pacientes que presentan infertilidad de los cuales se obtuvieron los siguientes datos en millones/ml.: 16, 14, 20, 21, 15, 17, 19, 17, 16, 14 a que conclusión se llega con este estudio.
2. La frecuencia cardiaca normal de un adulto en reposo es de 70 latidos/minuto. En una población de pacientes hipertensos se evalúa la posibilidad de que tengan una frecuencia cardiaca mayor. Se toman 6



muestras dándonos los siguientes resultados: 82, 75, 80, 72, 85, 80. A que conclusión se llega con este estudio

3. La capacidad inspiratoria de un adulto sano es de 3 lt. de aire. Se desea evaluar dicha capacidad en individuos con problemas bronquiales, para lo cual se tomo una muestra de 8 personas obteniéndose los siguientes resultados: 2.8, 2.7, 2.5, 2.7, 2.6, 2.8, 2.8, 2.6. A que conclusión se llega con este estudio.
4. El promedio de hemoglobina en el ganado vacuno es de 11. se analizan 9 muestras de animales con problemas de babesiosis y se les cuantifico su hemoglobina siendo los resultados: 10, 9, 8.5, 10.5, 11, 10, 9.5, 9, 10. ¿Influye l babesiosis en los niveles de hemoglobina de los vacunos? Explique.
5. El promedio en los niveles sanguíneos de la hormona triyodo tironina (T3) en jóvenes de 10 a 15 años de edad es de 145 ng/dl en personas sanas. En la comunidad de Pochutla, Oaxaca se hace la valoración de la T3 en 10 adolescentes, dándonos los siguientes resultados: 80, 100, 120, 140, 120, 100, 95, 130, 90, 100. Podemos considerar que la baja estatura de los habitantes de esa zona se debe a las diferencias que existen en sus niveles de T3 en la adolescencia con respecto al resto de la población. Explique.

Referencias

Neter-Kutner-Nachtsheim-Li. Applied Linear Statistical Models. Mc. Graw-Hill. 5ª Ed. 2005.

Arnau Gras, J. Diseños Experimentales Multivariabes. Alianza Editorial, Madrid. 1990.

Box, G. E., Hunter, W., Hunter, J. S., "Estadística para investigadores". Ed. Reverté. 1998.

Box, G., Hunter, W.G. and Hunter, J.S. Estadística para Investigadores. Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos. Editorial Reverté. 2002.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.



Box G.E.P., W.G. Hunter, J.S. Hunter Estadística para experimentadores. Ed. Reverté S.A. Barcelona 1989.

Box G. E. P., William G. Hunter, J. Stuart Hunter. Estadística Para Investigadores. Editorial Reverté. 1ª Reimpresión. México, D.F. 1999.

Lara, A.M. Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS. Proyecto Sur de Ediciones, Granada. 2001.

Martínez, A. Diseños Experimentales. Métodos y Elementos de Teoría. Editorial Trillas, México. 1988.

Montgomery-Peck-Vining. Introduction to Linear Regression Analysis. Wiley. 3ª Ed. 2001.

Ostle B. Estadística Aplicada. Limusa-Wiley, México, 1975,

Peña D. Estadística, Modelos y Métodos. Volumen II: Modelos Lineales y Series Temporales. Alianza Universidad, Madrid. 2000.

Peña, D., Regresión y Diseño de Experimentos., Alianza Editorial. 2002.

Ríos S. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Walpole R.E., Myers R.H., Myers S.L: "Probabilidad y Estadística para Ingenieros", Ed. Prentice Hall, 6ª edición.1998,



Parte II: Modelos Clásicos



Capítulo IV: Comparaciones Poblacionales

Comparación de Poblaciones Normales.

Después de la descripción del comportamiento de una variable a través de la estadística descriptiva, se tiene que parte de la estadística inferencial son las pruebas de hipótesis, que en este caso las utilizamos para comparar las medias de dos o más poblaciones en estudio. Esto se debe a que en los procesos de investigación es común tener que efectuar comparaciones al menos entre dos medias de muestras aleatorias de una misma o diferente población bajo algún parámetro en estudio. Por ejemplo:

- El comportamiento bajo una dieta para individuos de ambos sexos en una población.
- La comparación de alguna característica fisiológica entre diferentes grupos étnicos.

Lo anterior implica la utilización de algún estadístico de prueba que nos permita definir sobre una hipótesis nula planteando que las medias poblacionales son iguales considerando que los resultados de una muestra no influyen sobre los resultados de la otra $H_0: \mu_i = \mu_j$; esto es comparar dos grupos independientes. Los cuales a su vez pueden tener varianzas iguales o diferentes

También es necesario considerar las características de las muestras aleatorias, en función de su tamaño, variación y correspondencia de los elementos de una respecto de la otra; teniendo la comparación de grupos pareados.

Los dos tipos de comparaciones antes mencionadas corresponden a la forma más sencilla de experimentación (ver figura 4).



Análisis de Dos Poblaciones Independientes

La comparación de dos poblaciones con distribución normal puede ocurrir en experimentación o en muestreo. En experimentación ocurre cuando se comparan dos tratamientos en unidades experimentales **HOMOGENEAS** (iguales). Por ejemplo:

- Comparación del control de insectos por la aplicación de insecticidas con diferente formula, etc.
- La temperatura optima de lectura para muestras de semen
- El tiempo en se alcanza el mayor crecimiento de un microorganismo

Este análisis estadístico también es usado en muestreos de poblaciones reales. Por ejemplo:

- La concentración media de fosfatasa alcalina en niños,
- La concentración de Nitrógeno en individuos con diferentes dietas ricas en proteínas (fisiculturistas, atletas),
- La estatura de los habitantes de dos regiones diferentes,

A este tipo de experimentación se le conoce como muestras independientes o grupos sorteados; en virtud de que los tratamientos son asignados de manera aleatoria a las unidades experimentales y su análisis es por medio de los estimadores en forma independiente. En el caso de poblaciones o tratamientos se dice que son independientes porque el resultado de un grupo de ninguna manera afecta a los resultados del otro grupo en comparación. Siendo el caso más simple del diseño completamente al azar (DCA), que se aborda en el capítulo V.

En general, se pueden comparar dos poblaciones que cumplan con los siguientes supuestos:

- Las poblaciones tienen una distribución normal
- Las poblaciones tienen igual varianza
- Las muestras son independientes

Además, si se toman muestras al azar de dos poblaciones con distribución normal con varianzas iguales, se tienen los parámetros que se presentan en el cuadro 5.



Cuadro 5.- Parámetros y estimadores

	Población 1		Población 2	
	Parámetro	Estimador	Parámetro	Estimador
Media	μ_1	\bar{X}_1	μ_2	\bar{X}_2
Varianza	σ^2_1	S^2_1	σ^2_2	S^2_2

Lo primero con lo que requerimos trabajar, es con la varianza, debido no solo a su importancia práctica tanto en el aspecto de los análisis estadísticos, si no también como el de uniformidad y calidad de lo evaluado.

Al calcular las varianzas puede suceder que, sean homogéneas (iguales) o diferentes como grupo; así que debemos establecer una hipótesis que nos permita despejar la duda sobre el supuesto de homogeneidad; lo que se puede plantear de la siguiente manera:

Ho: $S^2_i = S^2_j$ (Varianzas homogéneas),

Ha: $S^2_i \neq S^2_j$ (Varianzas heterogéneas)

El estadístico de prueba está basado en una F, bajo las siguientes características:

$$F_c = \frac{\text{varianza}_{\text{mayor}}(m)}{\text{varianza}_{\text{menor}}(n)} \sim F_{n-1}^{m-1} \Rightarrow F_c = \frac{S_1^2(\text{mayor})}{S_2^2(\text{menor})}$$

La regla de decisión para esta hipótesis es:

1°.- $F_{n-1}^{m-1} = F_t$ (F de tabla)

2°.- Si: $F_c > F_t \Rightarrow$ Se rechaza Ho (varianzas heterogéneas)

$F_c < F_t \Rightarrow$ No se rechaza Ho (varianzas homogéneas)

En función de lo anterior y del número de elementos contenidos en las muestras, el juego de hipótesis planteadas son:

Ho: $\mu_1 = \mu_2$ vs Ha: $\mu_1 \neq \mu_2$

Por lo que la hipótesis planteada podrá ser probada bajo los siguientes estadísticos:



Varianzas Homogéneas: Si se considera igual varianza para las dos poblaciones, entonces el indicador de la varianza es la varianza ponderada S_p^2 . Además no se requiere que tengan el mismo tamaño de muestra. Y la Hipótesis a probar es: $H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 \neq \mu_2$

Por lo tanto, si en la hipótesis nula se tiene que ($\mu_1 = \mu_2$), entonces el estadístico para probar la hipótesis es:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

El cual se distribuye como una “t” student con $n_1 + n_2 - 2$ grados de libertad. Siendo el criterio de prueba o regla de decisión:

Si $|t_c| > t(0.025, n_1 + n_2 - 2)$ se rechaza la hipótesis nula ($H_0: \mu_1 = \mu_2$)

Varianzas Heterogéneas: En el caso de que las varianzas sean heterogéneas (las unidades experimentales tienen diferentes características) y el número de elementos que componen a las muestras es igual o diferentes, el estadístico de prueba para la hipótesis $H_0: \mu_1 = \mu_2$, es:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)}}$$

En ambos casos, la regla de decisión es:

Si $|t_c| > t(0.025, n_1 + n_2 - 2)$ se rechaza la hipótesis nula ($H_0: \mu_1 = \mu_2$)

Ejemplo: En un laboratorio médico, se estudian el pH de la sangre de hombres y mujeres de una población de pacientes con insuficiencia renal, siendo las lecturas de: 5.78, 5.74, 5.80 y 5.84, para hombres y para mujeres de: 5.82, 5.87 y 5.96. Determine si existen diferencias significativas en las lecturas realizadas*.

Una vez hecha la tabla de datos, lo primero que se debe hacer es la: “Prueba de

	Hombres	Mujeres
	5.78	5.82
	5.74	5.87
	5.80	5.96
	5.84	
$\sum X_i$	23.16	17.65
$\sum X_i^2$	134.10	103.85
Media	5.79	5.88

* El ejemplo se trabajó con pocos datos por ser demostrativo (las operaciones son las mismas con muchos que con pocos)



homogeneidad de Varianzas”

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$$

Para lo cual primero hay que calcular las varianzas de “A” y “B”. Se pueden obtener de 2 formas a través de formulas o directamente de la calculadora

$$\begin{aligned} S^2_A &= 0.00173 \\ S^2_B &= 0.00503 \end{aligned} \quad Fc = \frac{S_1^2(mayor)}{S_2^2(menor)} = \frac{0.00503}{0.00173} = 2.9 \quad Ft^2_{3,0.01} = 30.8 \quad Fc < Ft.$$

En base a la regla de decisión, se acepta H_0 , por lo que hay igualdad de varianzas.

Así que, una vez que se ha comprobado que las varianzas son homogéneas, se continúa con la prueba de hipótesis.

La prueba de hipótesis es la siguiente: $H_0: \mu_1 = \mu_2$ VS $H_a: \mu_1 \neq \mu_2$

El estadístico de prueba es:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{(4-1)(0.00173) + (3-1)(0.00503)}{4+3-2} = 0.00305$$

$$t = \frac{5.79 - 5.88}{\sqrt{0.00305 \left(\frac{1}{4} + \frac{1}{3} \right)}} = 2.134$$

$$t_{\alpha/2, n_1+n_2-2} ==> t_{0.025,5} = 3.163$$

$|t_c| < t_{0.025,5}$ Por lo que la H_0 , no se rechaza.

El pH de la sangre de ambos sexos es igual.



Problemario de Poblaciones Independientes

6. Se evalúa la cantidad de semen eyaculado, en varones que presentan niveles elevados de Antígeno Prostático Específico, siendo un grupo el del problema y el otro de control:
Problema: 5, 7, 3, 4, 5, 3, 6.
Control: 3, 3, 4, 6, 5, 7, 4.
¿Son iguales los niveles de eyaculado? Explique
7. Se efectúa un experimento para establecer la dosis letal (DL) de una droga. Para lo cual se utilizan dos especies distintas de ratones. Los resultados se obtuvieron en “gr” de la droga utilizada hasta la muerte del roedor; siendo estos:
Especie “Mus musculus”: 0.5, 0.4, 0.6, 0.4, 0.3, 0.5.
Especie “Microtus ochrogaster”: 0.8, 0.5, 0.6, 0.6, 0.7, 0.8, 0.7.
¿Influye la especie del roedor en el efecto de la droga? Explique.
8. En la población de Mapastepec, Chiapas se realiza un estudio para determinar el índice de mortalidad en la población durante los últimos 8 años. Los resultados obtenidos fueron el número de defunciones por año agrupados por sexos, siendo estos:
Hombres: 152, 143, 98, 115, 123, 137, 142, 144
Mujeres: 143, 165, 134, 148, 122, 151, 156, 155
¿Influye el sexo en la mortalidad de la gente de Mapastepec? Explique.
9. Se realiza un experimento para determinar si existe diferencia entre la temperatura que alcanzan los animales infectados con 2 tipos diferentes de hemoparásitos (Babesia y Anaplasma). El experimento se realiza con 6 animales por parásito y se midió la temperatura en °C, obteniéndose los siguientes resultados:
Babesia: 39.5, 40.2, 38.7, 40.6, 39.7, 41.0
Anaplasma: 38.5, 39.0, 38.2, 40.1, 39.8, 39.2
¿Existen diferencias en el aumento de la temperatura por efecto de la infección de los parásitos? Explique.
10. Se realiza un experimento para determinar si existe diferencia en el tiempo de sangrado en dos grupos de personas. En el experimento se utiliza un grupo control (personas sanas) y un grupo de personas con diabetes: Se procede a hacer el experimento obteniéndose los siguientes resultados:
Control: 4, 7, 6, 4, 5, 3, 4
Diabetes: 9, 9, 8, 9, 7, 8, 9, 10, 11, 7
¿Existen diferencias en los tiempos de sangrado? Explique.



Comparación de Medias con Grupos Pareados.

Cuando se usa la técnica de muestras independientes para comparar dos tratamientos y las unidades experimentales son muy HETEROGENEAS (diferentes) dentro de las muestras, la varianza dentro de ellas es muy grande, por lo que es difícil detectar las diferencias entre los tratamientos.

Por lo que resulta conveniente agrupar a las unidades experimentales formando pares, de tal forma que los elementos de un par sean lo más homogéneos posible. Las unidades experimentales que forman un par deben ser homogéneas en el sentido de que si no se aplicaran diferentes tratamientos, se esperarían observaciones similares.

Con esta forma de agrupación, la varianza entre pares no se incluye en la varianza del error experimental, incrementándose la precisión del experimento. A este diseño se le conoce como COMPARACION DE MEDIAS CON GRUPOS PAREADOS. Sin embargo es el caso más simple del diseño de bloques al azar (DBA), que se aborda en el capítulo V

Algunas comparaciones que se pueden evaluar son del tipo:

- La respuesta de un individuo a una dieta, antes y después de la aplicación de un medicamento.
- El rendimiento de un atleta antes y después de la aplicación de un tratamiento con cierta droga.
- Los niveles de estrógenos en mujeres de diferentes grupos de edades

Si $X_{11}, X_{12}, X_{13}, \dots, X_{1n}$ son n observaciones de la muestra 1 y $X_{21}, X_{22}, X_{23}, \dots, X_{2m}$ son m observaciones de la muestra 2, entonces el experimento de poblaciones dependientes arrojará los como se observan e el cuadro 6.

Cuadro 6.- Datos de poblaciones dependientes

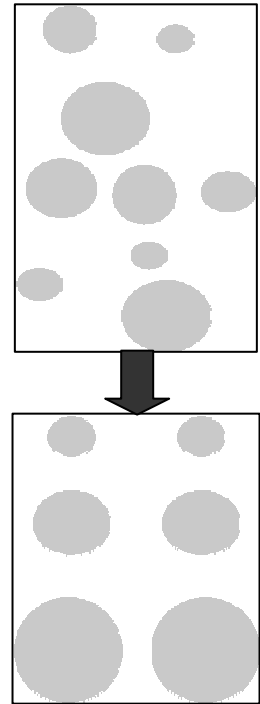
Par	Población 1	Población 2	$X_{1i} - X_{2i} = d_i$
I	X_{11}	X_{21}	$X_{11} - X_{21} = d_1$
II	X_{12}	X_{22}	$X_{12} - X_{22} = d_2$
III	X_{13}	X_{23}	$X_{13} - X_{23} = d_3$
...
N	X_{1n}	X_{im}	$X_{1n} - X_{2m} = d_n$

El juego de hipótesis a probar es:

$$H_0: \mu_1 - \mu_2 = \delta = 0$$

VS

$$H_a: \mu_1 - \mu_2 = \delta \neq 0$$





Los estimadores y parámetros se presentan en el cuadro 7.

Cuadro 7.- Estimadores y parámetros de poblaciones independientes

	Parámetro	Estimador
Media	δ	\bar{d}
Varianza	$\sigma^2\delta$	S_d^2

En donde:

$$\bar{d} = \frac{\sum di}{n} \quad s_d^2 = \frac{\sum (di - \bar{d})^2}{n-1} = \frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1}$$

Por lo tanto, para probar la hipótesis $H_0: \delta = 0$ se utiliza el estadístico de prueba:

La regla de decisión para la hipótesis planteada es
 Si $t_c > t_{n-1, \alpha}$ Existe significancia. Se rechaza H_0 .
 Si $t_c < t_{n-1, \alpha}$ No hay significancia. Se acepta H_0 .

$$t_c = \frac{\bar{d}}{\sqrt{\frac{S_d^2}{n}}}$$

Ejemplo:

Se comparan dos dietas basadas en proteínas derivadas de la soya y de maíz para medir la eficiencia alimenticia en relación a la ganancia de peso en niños menores de 6 años de una región con deficiencia nutricional. Debido a que los niños que participaron en el estudio tenían diferentes pesos (muy heterogéneos), se agruparon por pares, de tal forma que los niños de un par fueran lo más homogéneos (iguales en peso) posibles.

Par	Dieta		di	d2i
	Soya	Maíz		
1	16.8	21.6	-4.8	23.04
2	19.2	21.5	-2.3	5.29
3	20.5	24.1	-3.6	12.96
4	20.6	22.2	-1.6	2.56
5	22.0	23.3	-1.3	1.59
6	22.1	24.2	-2.1	4.41
7	23.0	23.3	-0.3	0.09
8	23.3	24.9	-1.6	2.56
	Σ		-17.6	52.6

Los pesos resultantes fueron:

Pruebe la hipótesis de que el efecto de las dietas es el mismo en el incremento de peso de las personas.

La prueba de hipótesis sobre las medias es la siguiente:

$$H_0: \mu_1 - \mu_2 = \delta = 0 \quad \text{VS} \quad H_a: \mu_1 - \mu_2 = \delta \neq 0$$

$$\bar{d} = \frac{\sum di}{n} = \frac{-17.6}{8} = -2.2 \quad s_d^2 = \frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1} = \frac{52.6 - \frac{38.72}{7}}{7} = \frac{13.88}{7} = 1.98$$



El estadístico de prueba:

$$t_c = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}} = \frac{-2.2}{\sqrt{\frac{1.98}{8}}} = \frac{-2.2}{0.49} = -4.42$$

Aplicando la regla de desición:

$t_{0.05/2,7} = 2.841$ como $t_c > t_{0.05/2,7} \implies$ Se rechaza H_0

$t_{0.01/2,7} = 4.029$ como $t_c > t_{0.01/2,7} \implies$ Se rechaza H_0

Debido a que se rechaza la hipótesis nula (H_0). El efecto de las dietas no es el mismo en el incremento de peso de los niños.

Problemario de Grupos Pareados

11. Se evalúa la función renal de pacientes de ambos sexos con diabetes, para lo cual se mide la concentración de glucosa en orina/24h en mg. Se efectúan las lecturas ordenando a las personas en base a los niveles de glucosa en sangre, obteniéndose los siguientes resultados:

Glucosa:	1	2	3	4	5	6
Hombres:	280,	300,	270,	267,	325,	277.
Mujeres:	290,	289,	267,	274,	302,	280.

¿Existen diferencias en la concentración de glucosa en orina por efecto del sexo? Explique.

12. Debido a las variaciones climáticas recientes, se han estado dando brotes de enfermedades en la costa del Estado de Chiapas, por lo que se desea saber si en la costa la temperatura es la misma, procediéndose a registrar la temperatura de 2 ciudades en los extremos de la costa de Chiapas (Tonala y Tapachula) durante una semana a las 12:00 hrs. a la sombra a 7 diferentes alturas, obteniéndose los siguientes resultados:

Atura:	1	2	3	4	5	6	7
Tonala:	32,	29,	31,	30,	29,	31,	30.
Tapachula:	29,	31,	28,	30,	29,	28,	29

¿Existen diferencias en las temperaturas ambientales de la costa? Explique.



13. Se realiza un experimento para evaluar el efecto de una nueva droga en la reducción de los niveles de glucosa en sangre de pacientes diabéticos. La prueba se realizó con pacientes que tenían hiperglucemia y se les suministró la droga evaluando los niveles de glucosa en sangre 24 hrs. después. Los resultados obtenidos fueron los siguientes:

Paciente:	1	2	3	4	5	6
Antes:	300,	280,	220,	200,	210,	230
Después:	140,	135,	120,	115,	118,	130

¿Existen diferencias en las concentraciones de glucosa por efecto de la droga? Explique.

14. Se evalúan las concentraciones de Progesterona en mujeres embarazadas de dos edades diferentes (25 y 35 aprox.). Se tomaron muestras durante los nueve meses de embarazo obteniéndose los siguientes resultados promedios:

Mes:	1	2	3	4	5	6	7	8	9
25:	8,	14,	17,	19,	22,	25,	27,	28,	31
35:	9,	12,	15,	17,	21,	24,	27,	29,	32

¿Existen diferencias en los niveles de Progesterona debido a la edad? Explique.

15. Se hacen un experimento con dos tipos de leches infantiles para observar la ganancia de peso en niños recién nacidos con problemas de peso al nacer. Se les da la leche durante un mes a dos grupos de niños distribuidos en relación a los pesos al nacer y se pesan obteniéndose por diferencia la ganancia de peso en gramos. Los resultados obtenidos son en:

Peso:	1	2	3	4	5	6
Leche A (proteína de soya):	300,	400,	600,	450,	520,	730

Leche B (prot. de leche de vaca): 380, 440, 423, 750, 600, 800

¿Existen diferencias en las ganancias de pesos por las diferentes leches? Explique.

Referencias

Box, G. E., Hunter, W., Hunter, J. S., "Estadística para investigadores". Ed. Reverté. 1998.

Box, G., Hunter, W.G. and Hunter, J.S. Estadística para Investigadores. Introducción al Diseño de



Experimentos, Análisis de Datos y Construcción de Modelos. Editorial Reverté. 2002.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

G.E.P. Box, W.G. Hunter, J.S. Hunter Estadística para experimentadores. Ed. Reverté S.A. Barcelona 1989.

Lara, A.M. Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS. Proyecto Sur de Ediciones, Granada. 2001.

Mead, R. The Design of Experiments: Statistical Principles for Practical Applications. Cambridge University Press, Cambridge. 1994.

Ostle B. Estadística Aplicada. Limusa-Wiley, México, 1975,

Pérez, C. Estadística Práctica con Statgraphics. Prentice Hall, Madrid. 2002.

Prat, A., Tort-Martorell, X., Grima, P., Pozueta, L., "Métodos estadísticos. Control y Mejora de la Calidad". Ed. UPC (1997).

S. Chatterjee - A. Hadi - B. Price., Regression Analysis by Example., (3ª Ed. 2000) WILEY.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Walpole R.E., Myers R.H., Myers S.L: "Probabilidad y Estadística para Ingenieros", Ed. Prentice Hall, 6ª edición.1998,

George E. P. Box, William G. Hunter, J. Stuart Hunter. Estadística Para Investigadores. Editorial Reverté. 1ª Reimpresión. México, D.F. 1999.



Parte II: Modelos Clásicos



Capítulo V: Diseños Simples

En este capítulo es donde abordamos de manera formal lo que se conoce como diseños de experimentos, lo anterior visto, son precursores de lo que se necesitan saber, para facilitar la comprensión de los diseños propiamente dichos.

La pregunta que responde el Análisis de Varianza de los Diseños Simples es la de: ¿El efecto de los tratamientos es el mismo? Solo pudiendo contestarnos si tienen o no el mismo efecto; así que si se desea conocer quienes tienen el mismo efecto y quienes un efecto diferente se requiere de un análisis posterior que se le conoce como “Comparación de Medias”

En principio de cuentas, al trabajar con experimentos debemos de tener en consideración dos aspectos: El arreglo geométrico de los tratamientos sobre las unidades experimentales, y las combinaciones de tratamientos cuando se ensayan dos o más factores. El primero corresponde a lo que se conoce como DISEÑO EXPERIMENTAL, el cual establece la asignación de los tratamientos a las unidades experimentales e indica la forma de obtener datos. Mientras que el segundo punto es el que se conoce como DISEÑO DE LOS TRATAMIENTOS, el cual indica la forma de combinar los niveles de los factores que interesan evaluar.

El objetivo de estudiar los datos de una población a través de un experimento es el de “realizar una prueba de hipótesis sobre el parámetro de interés (las medias) de donde proviene el experimento”, esto utilizando como herramienta al Análisis de la Varianza, cuyas características van a depender del Diseño Experimental que se haya considerado para el experimento. Y el tipo de diseño va a depender de las características de las unidades experimentales (figura 5).



Con los experimentos simples se investigan los efectos de tratamientos de un factor. Aunque, existen varios tipos de diseños experimentales, los que más se ocupan a nivel de laboratorio, pero sobre todo en el área de la salud son los siguientes:

- Comparación de dos poblaciones normales,
- Diseño completamente al Azar,
- Diseño de Bloques al Azar,
- Diseño cuadro Latino.



Diseño Completamente al Azar

Es el tipo de arreglo más sencillo; los tratamientos están asignados completamente al azar a las unidades experimentales; por lo que la variabilidad total de las observaciones del experimento se divide en:

1. Una debido al efecto de los tratamientos
2. Otra debido al error experimental

Por otra parte, el análisis de la varianza contiene únicamente dos fuentes de variación y se caracteriza por:

1. Puede aplicarse cuando se estudian dos o más tratamientos
2. Las unidades experimentales deben ser homogéneas
3. Los tratamientos deben asignarse a las unidades experimentales totalmente al azar.

Ventajas:

1. Puede utilizarse cuando las repeticiones por tratamiento son diferentes
2. Cuando sea probable que parte del experimento, ya sean unidades experimentales o tratamientos se pierdan o se rechacen por alguna razón.
3. El análisis estadístico que se desarrolla es fácil
4. En experimentos pequeños, se tiene mayor precisión, ya que contiene más grados de libertad para estimar el error experimental.

Desventajas:

1. Cuando las unidades experimentales son heterogéneas pierde precisión.
2. La variación que existe entre las unidades experimentales forma parte del error experimental.



Este diseño se utiliza cuando las unidades experimentales son esencialmente homogéneas, o sea, que no es posible identificar un factor que las afecte sistemáticamente. Por ello, es poco utilizado a nivel de campo sin embargo, puede utilizarse en invernaderos o laboratorios en los cuales debe de existir mayor control de los factores que las puedan afectar.

Análisis Estadístico

Las suposiciones que se consideran para probar la hipótesis de igualdad de efectos de tratamientos con el ANVA son:

- 1) Los valores observados de un tratamiento son una muestra aleatoria de un conjunto infinito de observaciones del tratamiento bajo las mismas condiciones del experimento.
- 2) Las variaciones en las unidades experimentales tratadas igualmente es la misma para todos los tratamientos.
- 3) Las observaciones se distribuyen normalmente.

Estas suposiciones son equivalentes a asumir que se tienen t muestras aleatorias independientes procedentes de t poblaciones normales con varianza común.

Notación

Si se tiene un experimento bajo este diseño y se probaron t tratamientos con n_i repeticiones cada uno, los datos deben clasificarse en t grupos de acuerdo a los tratamientos aplicados; la tabla de concentración de datos se observa en el cuadro 8:

Cuadro 8.- Concentración de datos del DCA

Tratamientos						
1	2	3	t
Y_{11}	Y_{21}	Y_{31}	Y_{t1}
Y_{12}	Y_{22}	Y_{32}	Y_{t2}
Y_{13}	Y_{23}	Y_{33}	Y_{t3}
...
Y_{1n1}	Y_{2n2}	Y_{3n3}	Y_{tnk}
$Y_{i.}$	$\sum_{j=1}^{n1} Y_{1j}$	$\sum_{j=1}^{n2} Y_{2j}$	$\sum_{j=1}^{n3} Y_{3j}$	$\sum_{j=1}^{nk} Y_{tj}$
$\sum_{i=1}^t \sum_{j=1}^{nk} Y_{ij} = Y_{..}$						



Si todos los tratamientos tienen el mismo número de repeticiones entonces: $n_1=n_2=n_3=.....=n_k=r$.

En la tabla, se tiene que las observaciones se denotan con una letra con dos subíndices, el primero indica el tratamiento (i) y el segundo la repetición (j).

El Modelo

Las observaciones procedentes de cada una de las unidades experimentales pueden ajustarse a un modelo estadístico. El desarrollo teórico de los diseños experimentales está basado en la Teoría de los modelos lineales. En este caso, el modelo para el Diseño completamente al Azar (DCA) es:

$$Y_{ij} = \mu + t_i + \varepsilon_{ij} \quad \text{con } i=1,2,3,...,t \quad \text{y} \quad j=1,2,3...n_k$$

Con:

Y_{ij} = Observación del tratamiento en la repetición

μ = Es la media verdadera general de todas las poblaciones, es decir, si cada observación tiene una distribución normal, con una media y una varianza; entonces la media poblacional es el resultado de las medias de tratamientos entre el número de estos:

$$\text{Si } Y_{ij} \sim N(\mu, \sigma^2) \implies \mu = \mu_i/t$$

t_i = Es el efecto verdadero del i-ésimo tratamiento, es decir: $t_i = \mu_i - \mu$.

ε_{ij} = Es el error experimental de la ij-ésima observación. Los ε_{ij} son variables aleatorias, distribuidas normal e independientemente con media cero y varianza (σ^2), esto es: $E(\varepsilon_{ij})=0$; $E(\varepsilon_{ij}^2)= \sigma^2$

Estimación

La estimación, el análisis de la varianza y los desarrollos para obtener las esperanzas de los cuadrados medios, se realizaron anteriormente; en este caso tenemos:

$$\sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - Y_{..})^2 = \sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{i.} - Y_{..})^2 + \sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - Y_{i.})^2$$

S.C. Total

S.C. Tratamiento

S.C. Error



Las fórmulas computarizadas a utilizar son:

$$S.C. \text{ Total} = \sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 - (Y_{..})^2/n.$$

$$S.C. \text{ Tratamientos} = \sum_{i=1}^t (Y_{i.})^2/n_i - (Y_{..})^2/n.$$

$$S.C. \text{ Error} = S.C. \text{ Total} - S.C. \text{ Tratamientos}$$

en donde:

$$i = 1, 2, 3, \dots, t$$

$$j = 1, 2, 3, \dots, n_i$$

$$Y_{..} = \sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}$$

$$n. = \sum_{i=1}^t n_i$$

$Y_{i.}$ = Es la suma de las i-esimas repeticiones del t-ésimo tratamiento.

La tabla del ANVA, para cuando se tiene diferente número de repeticiones por tratamiento se observa en el cuadro 9.

Cuadro 9.- ANVA del DCA con diferentes repeticiones en los tratamientos

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_{i.})^2/n_i - (Y_{..})^2/n.$	S.C. Trat./t-1	$\frac{C.M. \text{ Trat}}{C.M. \text{ Error}}$
Error	n.-t	S.C.Tot.-S.C.Tratam.	S.C. Error/n.-t	
Total	n.-1	$\sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 - (Y_{..})^2/n.$		

Si los tratamientos tienen igual número de repeticiones, entonces la tabla del ANVA es la que se muestra en el cuadro 10.

Cuadro 10.- ANVA del DCA con igual número de repeticiones en los tratamientos

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_{i.})^2/r - (Y_{..})^2/rt$	S.C. Trat./t-1	$\frac{C.M. \text{ Trat}}{C.M. \text{ Error}}$
Error	t(r-1)	S.C.Tot.-S.C.Tratam.	S.C. Error/n.-t	
Total	rt-1	$\sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 - (Y_{..})^2/rt$		

En la tabla del ANVA, Fc se distribuye como una F con t-1 y n.-t o t(r-1) grados de libertad. Por lo tanto el criterio de prueba es:

$$\text{Rechazar } H_0, \text{ si } F_c > F_{t-1, n.-t \text{ ó } t(r-1), \alpha}$$



El coeficiente de variación (CV) es una medida de la varianza en forma relativa; es debido al Error Experimental, que es independiente a las unidades en que están dadas las observaciones; se calcula:

$$C.V. = \frac{\sqrt{C.M. \text{ error}}}{Y..} * 100$$

Es la variabilidad de cada dato en forma relativa con respecto a la media.

Ejemplo:

Debido a que el síndrome de Turner es un defecto genético en niñas, las cuales entre otros factores, tienen un retraso de crecimiento. Se desarrolla una investigación para evaluar cual es el tratamiento hormonal más adecuado que pueda significar una mejoría en la talla adulta final. Por lo que a nivel de laboratorio se evaluó el incremento de tamaño de ratones recién nacidos, por efecto de tres dosis de una hormona de crecimiento (GH) biosintética.

	D1	D2	D3
1	58	56	67
2	60	54	64
3	61	60	58
4	54	58	65
5	58	52	
6		59	
7		54	
Yi.	291	393	254

Después de 30 días se midió el tamaño de los ratones desde las patas traseras a las patas delanteras, estirando al ratón sin forzarlo ni lastimarlo. El resultado de la medición del tamaño de cada ratón se hizo en mm y dio como resultado los valores que a continuación se presentan:

Pruebe la hipótesis de que los incrementos de tamaño de los ratones no se ven afectados por los diferentes niveles de T4.

Ho: $\mu_1 = \mu_2 = \mu_3$ VS. Ha: al menos un μ_i es diferente

$$n. = n_1 + n_2 + n_3 = 5 + 7 + 4 = 16$$

$$\sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij} = 938$$

$$\sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 = 55256$$

$$S.C. \text{ Total} = \sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 - (Y..)^2/n. \\ 55256 - (938)^2/16 = 55256 - 54990.25 = 265.75$$

$$S.C. \text{ Tratamientos} = \sum_{i=1}^t (Y_{i.})^2/n_i - (Y..)^2/n. \\ (291)^2/5 + (393)^2/7 + (254)^2/4 - 54990.25 \\ 55129.34 - 54990.25 = 139.09$$

$$S.C. \text{ Error} = S.C. \text{ Total} - S.C. \text{ Tratamientos} \\ 265.75 - 139.09 = 126.66$$



Como se tiene diferente número de repeticiones por tratamiento se utiliza la siguiente tabla:

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_i)^2/n_i - (Y_{..})^2/n$	S.C. Trat./t-1	<u>C.M. Trat</u> C.M. Error
Error	n.-t	S.C.Tot.-S.C.Tratam.	S.C. Error/n.-t	
Total	n.-1	$\sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 - (Y_{..})^2/n$		

FV	GL	SC	CM	Fc
Tratam.	2	139.09	68.545	7.035
Error	13	126.66	9.743	
Total	15	265.75		

$$F_t 0.05 = 3.81 \quad \Rightarrow F_c > F_t$$

$$F_t 0.01 = 6.70 \quad \Rightarrow F_c > F_t$$

Se rechaza la hipótesis nula. Si hay diferencias significativas en el crecimiento de los ratones con las tres diferentes dosis de la hormona de crecimiento biosintética.

Problemario de Diseño Completamente al Azar

16. Los niveles de estrógenos (E_4) en mujeres durante la ovulación es de 280 pg/ml en promedio. En un estudio se les determinó los niveles de E_4 a mujeres de 3 distintas edades (20, 30 y 40 años aprox.), dando como resultado los valores que en la siguiente tabla se presentan:

20:	437, 380, 400, 420, 350, 370, 400.
30:	250, 300, 320, 325, 290, 270.
40:	122, 180, 200, 150, 190.

¿Influye de alguna forma la edad en la producción de E_4 ? Explique

17. Se realiza un estudio en personas adictas al café, mayores de 40 años para evaluar el efecto de esta práctica en la concentración de calcio en sangre. Los niveles de referencia de calcio entre 8,5-10,5 (mg/100 mL).

De 5 a 10 tasas al día	5, 7, 4, 5, 5.
De 11 a 15 tasas al día	4, 4, 5, 6, 6, 4.
Más de 15 tasas al día	4, 5, 7, 6, 7.

A qué conclusión puede llegar. Explique



- 18.El ruido producido por el motor de un vehículo no debe de exceder los 30 decibeles. 4 compañías indican que sus motores son los que poseen los índices más bajos de contaminación sonora (ruido), para tal efecto se tomaron muestras de varios automóviles y los resultados se presentan en la siguiente tabla:

Compañía A	40, 45, 48, 36, 32.
Compañía B	38, 30, 47, 43.
Compañía C	48, 40, 40, 39, 35.
Compañía D	42, 41, 47, 48, 40

A que conclusión puede llegar. Explique

- 19.Se considera que la población joven de los países latinoamericanos se concentra en los lugares más urbanizados. En un estudio en la República Mexicana, se seleccionaron 3 sitios al azar siendo estos: Campeche, Guadalajara y Veracruz. De la zona centro de cada ciudad se tomaron de manera aleatoria a los individuos que se les pregunto su edad, siendo los resultados:

Guadalajara:	22, 36, 40, 18, 25.
Campeche:	32, 41, 56, 28, 37, 43.
Veracruz:	18, 25, 32, 14, 29.

A que conclusión llega con este estudio. Explique.

- 20.El promedio de la frecuencia respiratoria es de 12/min. en reposo. Se estudia a un grupo de personas para determinar si la complexión es un factor en la frecuencia respiratoria. Se tomaron muestras al azar de los 3 tipos de complexiones y los resultados fueron los siguientes:

Delgados:	12, 13, 14, 15, 14.
Medianos:	13, 12, 12, 14, 13.
Gruesos:	16, 15, 14, 15.

Cual es su conclusión. Explique.



Diseño de Bloques al Azar

Cuando las unidades experimentales son heterogéneas (existe al menos un factor que los afecta sistemáticamente) en el DCA la varianza dentro de tratamientos es muy alta, la cual está medida por el cuadrado medio del Error, que en este caso es grande, por lo que es más difícil rechazar la hipótesis de igualdad de medios de tratamientos.

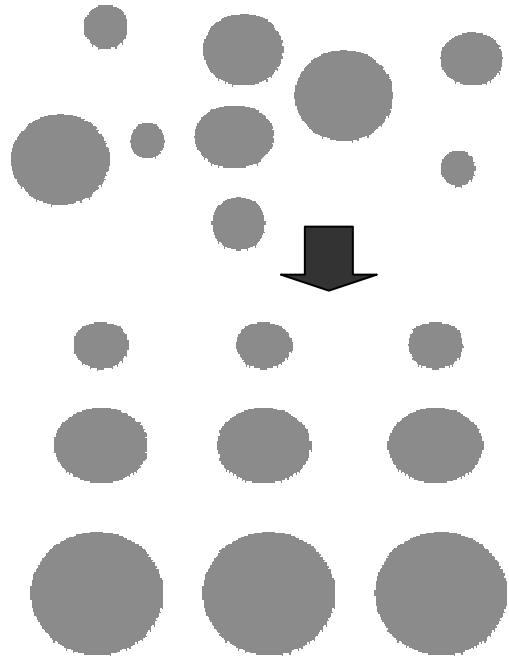
Bajo la situación antes mencionada, es conveniente dividir el material experimental en grupos (BLOQUES), de tal forma que las unidades experimentales sean homogéneas dentro de bloques y heterogéneas entre bloques.

El Diseño de Bloques al Azar se caracteriza porque todos los tratamientos se agrupan en dos o más bloques completos y compactos; cada tratamiento a ensayar está representado una sola vez en cada bloque, estos se asignan al azar a las unidades experimentales de cada bloque y posteriormente cada bloque se distribuye al azar en su establecimiento.

Este diseño es apropiado para los casos en que se observa una cierta tendencia de variación en el material experimental. Con la asignación de los tratamientos y los bloques al azar, es posible estimar la varianza entre bloques por separado y tener un Error Experimental pequeño.

Características:

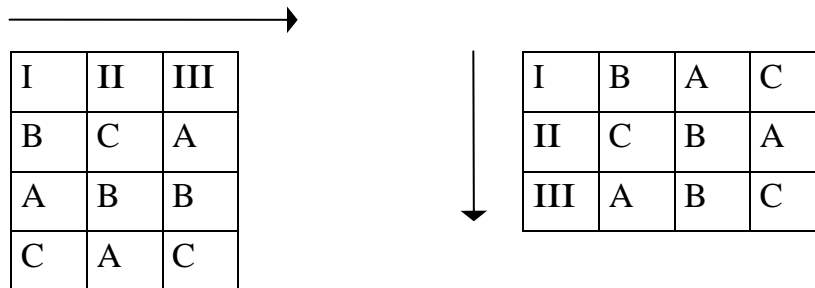
- a) Las unidades experimentales deben de ser homogéneas dentro de cada bloque, salvo por pequeñas variaciones aleatorias; puede haber cierta heterogenicidad, el propósito de los bloques es: absorber en máximo grado la variabilidad del material experimental.
- b) Es el más utilizado en trabajos experimentales, sobre todo a nivel de laboratorio
- c) Aunque no es particularmente una característica, los grados de libertad para el Error no debe ser menor de 12





- d) Es fácil de planear y el procedimiento de cálculo es fácil
- e) La disposición de tratamientos y bloques es ortogonal entre sí.
- f) El establecimiento del Diseño en Campo, los Bloques se colocan perpendicularmente al gradiente de variabilidad.

Gradiente de variación



Un bloque es un conjunto de unidades experimentales con una o varias características comunes de modo que en el modelo se les asigna un efecto común denominado efecto de bloque (beta). Una vez hechos los bloques se procede a la asignación de los tratamientos de manera aleatoria en cada bloque

Ventajas:

- 1) Es más preciso que el DCA cuando hay un factor que causa variación en las unidades experimentales
- 2) Es flexible, debido a que puede tener cualquier número de tratamientos y de bloques (mínimo dos)
- 3) Es posible estimar datos perdidos

Desventajas:

- 1) Cuando el número de tratamientos es muy grande, es difícil mantener la homogeneidad dentro de bloques, se pierde precisión.
- 2) Estima el Error Experimental con menos grados de libertad que el D.C.A.

Diseño	gl
DCA	$t(r-1)$
DBA	$(t-1)(r-1)$

**Análisis Estadístico:**

Las suposiciones consideradas para probar las hipótesis de igualdad de efectos de tratamientos, mediante el Análisis de la Varianza son las mismas que para el DCA.

Notación:

Si se tiene un experimento bajo el DBA y se prueban t tratamientos con r repeticiones cada uno, los datos se clasifican en t grupos, de acuerdo a los tratamientos aplicados y r bloques, según la tabla siguiente:

Cuadro 11.- Concentración de datos en el DBA

Bloque	Tratamientos							$Y_{.j}$
	1	2	3	t	$\sum_{j=1}^r Y_{.j}$
1	Y_{11}	Y_{21}	Y_{31}	Y_{t1}	$\sum_{j=1}^r Y_{i1}$
2	Y_{12}	Y_{22}	Y_{32}	Y_{t2}	$\sum_{j=1}^r Y_{i2}$
3	Y_{13}	Y_{23}	Y_{33}	Y_{t3}	$\sum_{j=1}^r Y_{i3}$
...
r	Y_{1r}	Y_{2r}	Y_{3r}	Y_{tr}	$\sum_{j=1}^r Y_{ir}$
$Y_{i.}$	$\sum_{j=1}^r Y_{1j}$	$\sum_{j=1}^r Y_{2j}$	$\sum_{j=1}^r Y_{3j}$	$\sum_{j=1}^r Y_{tj}$	$\sum_{i=1}^t \sum_{j=1}^r Y_{ij} = Y_{..}$

Como se puede observar en la tabla todos los tratamientos tienen el mismo número de repeticiones (r).

El Modelo:

El modelo estadístico para el DBA está dado por

$$Y_{ij} = \mu + t_i + \beta_j + \varepsilon_{ij} \quad \text{con } i=1,2,3,\dots,t \quad \text{y} \quad j=1,2,3,\dots,r$$

Y_{ij} = Observación del tratamiento (y) en el bloque (j)

μ = Es el efecto verdadero de la media general.

t_i = Es el efecto verdadero del i -ésimo tratamiento, es decir: $t_i = \mu_i - \mu$.

β_j = Es el efecto verdadero del j -ésimo bloque

ε_{ij} = Es el error experimental

Estimación:

La estimación, el análisis de la varianza y los desarrollos para obtener las esperanzas de los cuadrados medios, se realizaron anteriormente; en este caso tenemos:



$$\sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - Y_{..})^2 = \sum_{i=1}^t \sum_{j=1}^r (Y_{i.} - Y_{..})^2 + \sum_{i=1}^t \sum_{j=1}^r (Y_{.j} - Y_{..})^2 + \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - Y_{i.})^2$$

$$\text{S.C. Total} = \text{S.C. Tratamiento} + \text{S.C. Bloques} + \text{S.C. Error}$$

Las fórmulas computarizadas a utilizar son:

$$\text{S.C. Total} = \sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - (Y_{..})^2 / rt$$

$$\text{S.C. Tratamientos} = \sum_{i=1}^t (Y_{i.})^2 / r - (Y_{..})^2 / rt$$

$$\text{S.C. Bloques} = \sum_{j=1}^r (Y_{.j})^2 / t - (Y_{..})^2 / rt$$

$$\text{S.C. Error} = \text{S.C. Total} - \text{S.C. Tratamientos} - \text{S.C. Bloques}$$

En donde:

$$i = 1, 2, 3, \dots, t$$

$$j = 1, 2, 3, \dots, r$$

$$Y_{..} = \sum_{i=1}^t \sum_{j=1}^r Y_{ij}$$

$Y_{i.}$ = Es la suma de las i-esimas repeticiones del i-ésimo tratamiento.

$Y_{.j}$ = Es la suma de las j-esimas repeticiones del j-ésimo bloque.

El cuadro 12 muestra, la tabla del ANVA, de este diseño:

Cuadro 12.- Analisis de varianza para el DBA

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_{i.})^2 / r - (Y_{..})^2 / rt$	S.C. Trat./t-1	C.M. Trat/ C.M. Error
Bloques	r-1	$\sum_{j=1}^r (Y_{.j})^2 / t - (Y_{..})^2 / rt$	S.C. Bloq/r-1	C.M. Bloq./ C.M. Error
Error	(t-1)(r-1)	S.C.Tot.-S.C.Tr.-S.C.Bl.	S.C.Err./t-1 r-1	
Total	rt-1	$\sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - (Y_{..})^2 / rt$		

El criterio de decisión es:

Rechazar H_{01} , si $F_c > F_{t-1(t-1)(r-1), \alpha}$ Para tratamientos

Rechazar H_{02} , si $F_c > F_{r-1(t-1)(r-1), \alpha}$ Para Bloques



Eficiencia Relativa del DBA

La eficiencia (mayor precisión) que se tiene en el análisis de un experimento bajo un DBA, se da en forma relativa con respecto al DCA; esto a través del cálculo del índice:

$$ER (BA \rightarrow CA) = \frac{(r-1)(C.M.Bloques) + r(t-1)(C.M.error)}{(rt-1)(C.M.Error)}$$

Con: r = Número de repeticiones

t = Número de tratamientos

La regla de decisión es:

Si $ER (BA \rightarrow CA) > 1$, es más eficiente DBA

Si $ER (BA \rightarrow CA) < 1$, es más eficiente DCA

La eficiencia se expresa en porcentajes y de acuerdo con la fracción excedente de la unidad.

Ejemplo:

Se evalúa la eficacia de sulfato ferroso a 4 diferentes concentraciones 2, 2.5, 3, y 4 mg/kg/día para combatir la anemia en personas con desnutrición. Dada la variabilidad de pesos y sexos, se decidió agrupar en bloque de personas por pesos lo más semejantes posibles. Se les dio el tratamiento durante 4 meses y se procedió a determinar su volumen sanguíneo dando los siguientes resultados:

	T1	T2	T3	T4	Y.j
I	4.1	6.5	4.1	4.9	19.6
II	4.1	5.3	4.0	6.2	19.6
III	6.5	6.9	4.5	4.8	22.7
IV	4.3	6.8	4.3	4.2	19.6
V	6.0	6.5	4.1	6.9	23.5
Yi.	25	32	21	27	105

Las hipótesis a probar son:

$H_{01}: T1=T2=T3=T4$

VS H_{a1} : Al menos un tratamiento es diferente

$H_{02}: B1=B2=B3=B4=B5$

VS H_{a2} : Al menos un bloque (pesos) es diferente



Las sumas de cuadrados son:

$$\begin{aligned}
 \text{S.C. Total} &= \sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - (Y_{..})^2/rt \\
 &= (4.1)^2 + (4.1)^2 + \dots + (4.2)^2 + (6.9)^2 - (105)^2/(5*4) \\
 &= 575.9 - 551.25 = \mathbf{24.65} \\
 \text{S.C. Tratamientos} &= \sum_{i=1}^t (Y_{i.})^2/r - (Y_{..})^2/rt \\
 &= (25)^2 + \dots + (27)^2 - (105)^2/(5*4) = 563.8 - 551.25 = \mathbf{12.55} \\
 \text{S.C. Bloques} &= \sum_{j=1}^r (Y_{.j})^2/t - (Y_{..})^2/rt \\
 &= (19.6)^2 + \dots + (23.5)^2 - (105)^2/(5*4) = 555.005 - 551.25 = \mathbf{3.755} \\
 \text{S.C. Error} &= \text{S.C. Total} - \text{S.C. Tratamientos} - \text{S.C. Bloques} \\
 &= 24.65 - 12.55 - 3.755 = \mathbf{8.345}
 \end{aligned}$$

La tabla del ANVA, es la siguiente:

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_{i.})^2/r - (Y_{..})^2/rt$	S.C. Trat./t-1	C.M. Trat/ C.M. Error
Bloques	r-1	$\sum_{j=1}^r (Y_{.j})^2/t - (Y_{..})^2/rt$	S.C. Bloq/r-1	C.M. Bloq./ C.M. Error
Error	(t-1)(r-1)	S.C.Tot.-S.C.Tr.-S.C.Bl.	S.C.Err./t-1 r-1	
Total	rt-1	$\sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - (Y_{..})^2/rt$		
FV	GL	SC	CM	Fc
Tratam.	3	12.550	4.1830	6.015**
Bloques	4	3.755	0.9387	1.35ns
Error	12	8.345	0.6954	
Total	19	24.650		

Para Tratamientos $F_c > F_{t0.05}$ y $F_c > F_{t0.01}$ Se rechaza H_0 . Hay diferencias altamente significativas entre los tratamientos. Algunas dosis funcionan mejor que otras.

Para Bloques $F_c < F_{t0.05}$ y $F_c < F_{t0.01}$ Se acepta H_0 . Son iguales los efectos de los bloques (Pesos).



Problemario de Diseño de Bloques al Azar

21. Tres empresas (A, B y C) concursan para venderle a la policía un equipo para la determinación de alcohol en el aliento. Los equipos son de diferentes precios, por lo que se hace una prueba con un equipo de cada compañía para determinar si existe diferencia entre estos. La prueba consiste en la medición de metanol (Valor de Referencia < 0.80 p.p.m.) al personal de oficina de la policía de diferentes edades. Los resultados obtenidos fueron:

Edades	Empresas		
	A	B	C
I	0.90	0.80	0.70
II	0.70	0.70	0.60
III	0.70	0.60	0.80
IV	0.60	0.70	0.70

- ¿Existen diferencias en las mediciones de metanol con los equipos?
- ¿Existen diferencias en las mediciones de metanol a diferentes edades?

22. El nivel máximo en la atmósfera de dióxido de azufre (SO_2) recomendado por la OMS es de $40 \mu\text{gr}/\text{m}^3$. Se considera al Distrito Federal la ciudad más contaminada del país; para comprobarlo se toman muestras de 4 de las ciudades con mayor número de habitantes del país (Distrito Federal, Monterrey, Guadalajara, Tijuana) durante una semana y se obtuvieron los siguientes resultados:

Días	Ciudades			
	D.F.	MTY	GDA	TJA
1	82	70	67	61
2	90	74	49	53
3	71	82	54	48
4	60	68	72	45
5	75	75	94	67
6	100	90	98	62
7	110	88	85	49

- En base a los resultados obtenidos concluya.

23. La presión arterial promedio de un individuo es de 100 (80-120). Se toman muestras de las tres zonas del estado (Altos, Centro y Costa) a individuos seleccionados al azar de diversas edades y se les mide la presión arterial para determinar si la altitud a la que viven influye en esta presión. Los resultados obtenidos son:

Edades	Zonas		
	ALT	CEN	COS
I	110	95	100
II	115	98	98



III	105	102	94
IV	108	104	99
V	104	100	92

- ¿Es la altitud un factor que influya en la presión arterial de los individuos?
- ¿Es la edad un factor que influya en la presión arterial de los individuos?

24. En la Ciudad de México el número de defunciones por mes atribuidas a la contaminación ambiental son de aproximadamente 1000. Se hace un estudio en tres ciudades del país (Mérida, Oaxaca y Tijuana) para determinar si el índice de mortalidad por efecto de la contaminación es el mismo entre las diferentes regiones del país. Para tal efecto se registraron el número de defunciones (por contaminación) por mes durante cinco meses, obteniéndose los siguientes datos:

Meses	Ciudades		
	MDA	OAX	TJA
I	400	357	800
II	300	421	725
III	600	480	1005
IV	457	300	930
V	525	500	915

- ¿Fuera de la Ciudad de México los índices de mortandad son parecidos entre las diferentes regiones de la República? Explique y concluya.

25. Un experimento sobre adicciones se realizó en una población de adultos farmacodependientes con diferentes niveles de adicción y de diferentes edades a los cuales se les retiró la droga para establecer el número de días que lograban aguantar sin el consumo de esta. Los resultados fueron los siguientes:

Edades	Niveles de Adicción		
	Alt	Med	Baj
I	12	14	15
II	10	8	13
III	12	10	10
IV	8	8	9
V	4	5	7

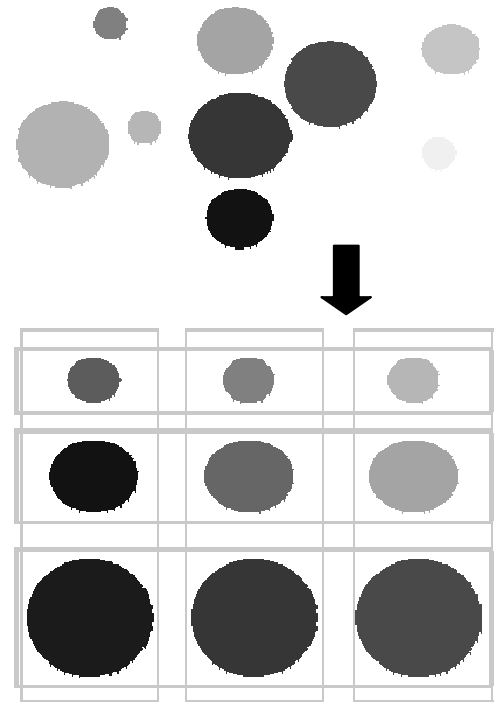
- ¿Existe diferencia para aguantar la abstinencia del consumo de droga de acuerdo a los niveles de adicción? Explique.
- ¿Existe diferencia para aguantar la abstinencia del consumo de droga de acuerdo a las diferentes edades considerando que se tienen agrupadas de menor a mayor edad? Explique



Diseño en Cuadro Latino

Este diseño se utiliza cuando hay dos factores que afectan sistemáticamente a las unidades experimentales. Para evitar que el efecto de los factores se acumule en el error experimental se hace un doble bloqueo.

Los diferentes sistemas de distribución de las unidades experimentales adoptados y que llevan los distintos tratamientos a ensayar satisfacen en mayor o menor grado dichas condiciones; para ello, se tienen los diseños que utilizan el BLOQUEO, y una variante es aquel que se da bajo dos factores en forma ortogonal (independiente), dando origen al DISEÑO EN CUADRO LATINO denominado así por Fisher, quien utilizó letras latinas para denotar los tratamientos.



Esta asignación de ortogonalidad (perpendicular) a los efectos de los tratamientos respecto a los dos factores de variabilidad asociados con el experimento, requiere que cada tratamiento aparezca una vez en cada una de esas variantes, denominados: HILERAS Y COLUMNAS.

Características:

1. Se genera cuando los tratamientos se agrupan en bloques homogéneos en dos direcciones: hileras y columnas quienes constituyen una repetición completa de los tratamientos.
2. Cualquier tratamiento aparece una sola vez en la misma columna o en la misma hilera.
3. Se impone a las UNIDADES EXPERIMENTALES una restricción de doble bloqueo.
4. El número de repeticiones es igual al número de tratamientos a evaluar. Esto es, si hay t tratamientos, el total de UNIDADES EXPERIMENTALES será t^2 . t es mayor de 2.



Ventajas:

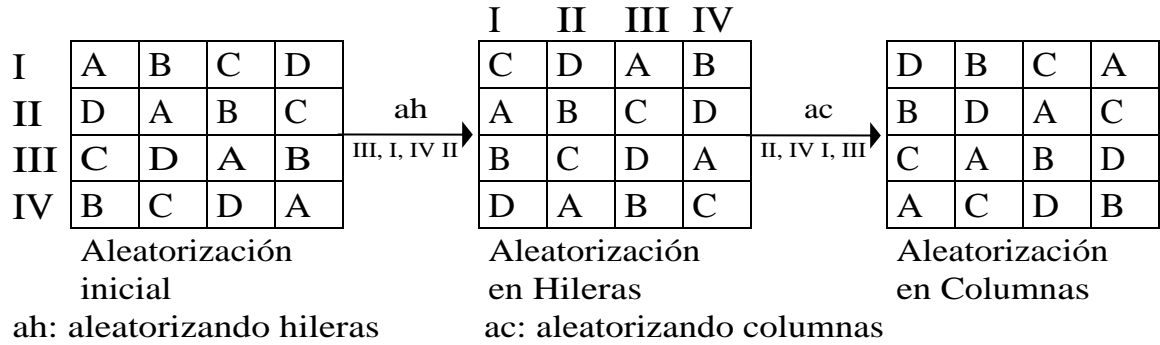
1. Reduce el error experimental al introducir el doble bloqueo.
2. El análisis estadístico es simple, ligeramente más complicado que DBA.
3. Puede utilizarse cuando las UNIDADES EXPERIMENTALES forman una línea continua.
4. Proporciona una comparación más precisa de los efectos de los tratamientos.
5. Aún con datos perdidos, el análisis estadístico es simple.
6. Es más preciso que el DBA.

Desventajas:

1. Es poco flexible, ya que el No. de Hileras o Columnas dependen del No. de tratamientos.
2. No se pueden comparar muchos tratamientos, el rango es 4-10
3. Con pocos tratamientos, se tienen demasiados parámetros en el modelo con pocas observaciones, siendo ineficiente la estimación de la varianza del error.

Distribución de los Tratamientos:

1. Dividir el lote experimental en un número de unidades experimentales igual al cuadrado del número de tratamientos.
2. Formar hileras y columnas de unidades experimentales iguales al número de tratamientos.
3. Distribuir los tratamientos, de tal manera que ningún tratamiento se repita en la hilera ni en la columna. Esto se realiza asignando aleatoriamente las hileras y las columnas de un modelo inicial, por permutaciones.



Análisis Estadístico:

Al igual que para los DCA y DBA, las suposiciones de distribución de las observaciones son las mismas.

Notación: En el cuadro 13 se observa que la distribución de los tratamientos están evaluados en todas las hileras y en todas las columnas, entonces se tienen T^2 variables de respuesta a ensayar, debido al doble bloqueo, por lo que todas las observaciones serán denotados como Y_{ijk} , los que se organizan de la siguiente forma:

Cuadro 13.- Concentración de datos en el DCL

Hile- ras	Columnas				$\sum_{j=1}^4 Y_{.j} = Y_{.j}$
	I	II	III	IV	$Y_{.j}$
I	Y_{411}	Y_{212}	Y_{313}	Y_{114}	$Y_{.1}$
II	Y_{221}	Y_{422}	Y_{123}	Y_{324}	$Y_{.2}$
III	Y_{331}	Y_{132}	Y_{233}	Y_{434}	$Y_{.3}$
IV	Y_{141}	Y_{342}	Y_{443}	Y_{244}	$Y_{.4}$
$Y_{..K}$	$Y_{..1}$	$Y_{..2}$	$Y_{..3}$	$Y_{..4}$	$Y_{...}$

$$\sum_{i=1}^4 \sum_{j=1}^4 \sum_{k=1}^4 Y_{ijk} = Y_{...}$$

Totales de tratamientos: $\sum_{i=1}^4 Y_{i..}$

$$T1 = Y_{1..} = Y_{114} + Y_{123} + Y_{132} + Y_{141}$$

$$T2 = Y_{2..} = Y_{212} + Y_{221} + Y_{233} + Y_{244}$$

$$T3 = Y_{3..} = Y_{313} + Y_{324} + Y_{331} + Y_{342}$$

$$T4 = Y_{4..} = Y_{411} + Y_{422} + Y_{434} + Y_{443}$$



El modelo:

La suposición lógica de un D.C.L. con una observación por cada unidad experimental es que las observaciones pueden representarse mediante el modelo estadístico:

$$Y_{ijk} = \mu + T_i + H_j + C_k + E_{ijk}$$

Con $i=1,2,3,\dots,t$, $j=1,2,3,\dots,t$ y $k=1,2,3,\dots,t$

en donde:

Y_{ijk} : Es la observación del i -ésimo tratamiento en la j -ésima hilera de la k -ésima columna.

μ : Es el efecto verdadero de la media general

T_i : Es el efecto del i -ésimo tratamiento

H_j : Es el efecto de la j -ésima hilera

C_k : Es el efecto de la k -ésima columna

E_{ijk} : Es el error experimental

Además, se supone:

- Los errores (E_{ijk}) se distribuyen normal e independientemente con $\mu=0$ y Varianza σ^2 .
- No hay interacción entre tratamientos hileras y columnas.

Estimación: Los componentes de la suma de cuadrados para el DCL son:

$$\begin{aligned} \sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - Y_{..})^2 = & t \sum_{i=1}^t (Y_{i..} - Y_{...})^2 + \\ & t \sum_{j=1}^t (Y_{.j.} - Y_{...})^2 + \\ & \tau \sum_{k=1}^t (Y_{..k} - Y_{...})^2 + \\ & \sum_{i=1}^t \sum_{j=1}^t \sum_{k=1}^t (Y_{ijk} - Y_{i..} - Y_{.j.} - Y_{..k} + 2Y_{...})^2 \end{aligned}$$

S.C. Total = S.C. Tratamiento - S.C. Hileras - S.C. Columnas - S.C. Error



Las fórmulas computarizadas a utilizar son:

$$S.C. \text{ Total} = \sum_{i=1}^t \sum_{j=1}^t \sum_{k=1}^t Y_{ijk}^2 - (Y_{...})^2/t^2$$

$$S.C. \text{ Tratamientos} = \sum_{i=1}^t (Y_{i..})^2/t - (Y_{...})^2/t^2$$

$$S.C. \text{ Hileras} = \sum_{j=1}^t (Y_{.j.})^2/t - (Y_{...})^2/t^2$$

$$S.C. \text{ Columnas} = \sum_{k=1}^t (Y_{..k})^2/t - (Y_{...})^2/t^2$$

$$S.C. \text{ Error} = S.C. \text{ Total} - S.C. \text{ Tratam.} - S.C. \text{ Hil.} - S.C. \text{ Col.}$$

en donde: $i = 1, 2, 3, \dots, t$ $j = 1, 2, 3, \dots, t$ $k = 1, 2, 3, \dots, t$

$$Y_{...} = \sum_{i=1}^t \sum_{j=1}^t \sum_{k=1}^t Y_{ijk}$$

$Y_{i..}$ = Es la suma de las i-esimas repeticiones del i-ésimo Tratamiento.

$Y_{.j.}$ = Es la suma de las j-esimas repeticiones de la j-ésimo Hilera.

$Y_{..k}$ = Es la suma de las k-esimas repeticiones de la k-ésimo Columna.

En el análisis de la varianza hay una fuente de variación para cada uno de los factores: Tratamientos, Hileras, Columnas y Error Experimental; calculándose la contribución de ks sumas de cuadrados de cada componente, para probar los juegos de hipótesis siguientes:

$H_{01}: t_1 = t_2 = \dots = t_i$ VS H_{a1} : al menos un t_i es diferente ($t_i \neq t_i'$)

$H_{02}: H_1 = H_2 = \dots = H_i$ VS H_{a2} : al menos un H_i es diferente ($H_i \neq H_i'$)

$H_{03}: C_1 = C_2 = \dots = C_i$ VS H_{a3} : al menos un C_i es diferente ($C_i \neq C_i'$)

La tabla del ANVA, se observa en el cuadro 14:

Cuadro 14.- Analisis de varianza para el DCL

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_{i..})^2/t - (Y_{...})^2/t^2$	S.C. Trat./t-1	C.M. Trat C.M. Error
Hileras	t-1	$\sum_{j=1}^t (Y_{.j.})^2/t - (Y_{...})^2/t^2$	S.C. Hil./t-1	C.M. Hil. C.M. Error
Cols.	t-1	$\sum_{k=1}^t (Y_{..k})^2/t - (Y_{...})^2/t^2$	S.C. Col./t-1	C.M. Col. C.M. Error
Error	(t-1) (t-2)	S.C.Tot.-S.C.T.-S.C.H-S.C.C.	S.C.Err. t-1 t-2	
Total	t2-1	$\sum_{i=1}^t \sum_{j=1}^t \sum_{k=1}^t Y_{ijk}^2 - (Y_{...})^2/t^2$		



El criterio de decisión es:

Rechazar H_{01} , si $F_c > F_{t-1(t-1)(t-2), \alpha}$ Para tratamientos

Rechazar H_{02} , si $F_c > F_{r-1(t-1)(t-2), \alpha}$ Para Hileras

Rechazar H_{03} , si $F_c > F_{r-1(t-1)(t-2), \alpha}$ Para Columnas

Eficiencia Relativa del Diseño en Cuadro Latino

La eficiencia relativa del DCL con respecto a los diseños antes descritos se determina de la manera siguiente:

1) Respecto de un DBA

$$ER (CL \rightarrow BA) = \frac{CM(Hileras) + (t-1)CM(Error)}{(t)(CM(Error))}$$

2) Respecto de un DCA

$$ER (CL \rightarrow CA) = \frac{CM(Hil.)+CM(Col.)+(t-1)CM(Error)}{(t+1)(CM(Error))}$$

La regla de decisión en ambos casos, es:

Si $ER > 1$, es mejor el diseño utilizado

Si $ER < 1$, es mejor el diseño que se compara

La $ER-1$, se multiplica por 100 y se expresa en porcentaje la eficiencia total.

Ejemplo:

Se desea evaluar el funcionamiento hepático en perros con diferentes patologías hepáticas que se encuentran con tratamiento de vitaminas B a 4 diferentes concentraciones. Por lo que se hace un doble arreglo en base a las patologías y razas, y se procedió a la medición de la Transaminasa Glutámico Pirúvica (TGP).



		Patologías hepáticas							
R		I Isquemia		II Tumor		III Hepatitis viral		IV Hemocromatosis	
A	I.- Cocker Spaniel	t1	10	t2	8	t3	14	t4	14
Z	II.- Pastor Aleman	t4	7	t1	12	t2	12	t3	11
A	III.- Boxer	t3	14	t4	12	t1	16	t2	16
S	IV.- Schnauzer	t2	13	t3	13	t4	10	t1	15

De la tabla de datos se construye el croquis de la siguiente forma:

1. Se ordenan la unidades experimentales en hileras y columnas
2. Se asignan los tratamientos de tal forma que los tratamientos ocurran en todas las hileras y todas las columnas.
3. Se aleatorizan las hileras
4. Se aleatorizan las columnas

	I	II	III	IV	Y.j.
I	t2 12	t1 15	t3 13	t4 7	47
II	t1 16	t3 11	t4 12	t2 13	52
III	t3 14	t4 14	t2 8	t1 10	46
IV	t4 10	t2 16	t1 12	t3 14	52
Y..k	52	56	45	44	197

Para evaluar los tratamientos se utiliza como variable principal la concentración de TGP.

Ho1:t1=t2 = t3 VS Ha1:al menos un ti es diferente (ti≠ti')

Ho2:H1=H2 H3 VS Ha2:al menos un Hi es diferente (Hi≠Hi')

Ho3:C1=C2 =C3 VS Ha3:al menos un Ci es diferente (Ci≠Ci')

Totales de tratamientos: $\sum ti=1Yi..$

$$T1= Y_{1..} = Y_{112} + Y_{121} + Y_{134} + Y_{143} = 15+16+10+12=53$$

$$T2= Y_{2..} = Y_{212} + Y_{224} + Y_{233} + Y_{242} = 12+13+08+16=49$$

$$T3= Y_{3..} = Y_{313} + Y_{322} + Y_{331} + Y_{344} = 13+11+14+14=52$$

$$T4= Y_{4..} = Y_{414} + Y_{423} + Y_{432} + Y_{441} = 07+12+14+10=43$$



Las fórmulas computarizadas a utilizar son:

$$\begin{aligned}
 \text{S.C. Total} &= \sum_{i=1}^t \sum_{j=1}^t \sum_{k=1}^t Y_{ijk}^2 - (Y_{...})^2/t^2 \\
 &= 2529 - (38809/(4)^2) = 103.44 \\
 \text{S.C. Tratamientos} &= \sum_{i=1}^t (Y_{i..})^2/t - (Y_{...})^2/t^2 \\
 &= 2440.75 - 2425.56 = 15.19 \\
 \text{S.C. Hileras} &= \sum_{j=1}^t (Y_{.j.})^2/t - (Y_{...})^2/t^2 \\
 &= 2433.25 - 2425.56 = 7.69 \\
 \text{S.C. Columnas} &= \sum_{k=1}^t (Y_{..k})^2/t - (Y_{...})^2/t^2 \\
 &= 2450.25 - 2425.56 = 24.69 \\
 \text{S.C. Error} &= \text{S.C. Total} - \text{S.C. Tratam.} - \text{S.C. Hil.} - \text{S.C. Col.} \\
 &= 103.44 - 15.19 - 7.69 - 24.69 = 55.87
 \end{aligned}$$

La tabla del ANVA, es la siguiente:

FV	GL	SC	CM	Fc
Tratam.	t-1	$\sum_{i=1}^t (Y_{i..})^2/t - (Y_{...})^2/t^2$	S.C. Trat./t-1	C.M. Trat C.M. Error
Hileras	t-1	$\sum_{j=1}^t (Y_{.j.})^2/t - (Y_{...})^2/t^2$	S.C. Hil./t-1	C.M. Hil. C.M. Error
Cols.	t-1	$\sum_{k=1}^t (Y_{..k})^2/t - (Y_{...})^2/t^2$	S.C. Col./t-1	C.M. Col. C.M. Error
Error	(t-1) (t-2)	S.C.Tot.-S.C.T.- S.C.H-S.C.C.	S.C.Err. t-1 t-2	
Total	t2-1	$\sum_{i=1}^t \sum_{j=1}^t \sum_{k=1}^t Y_{ijk}^2 - (Y_{...})^2/t^2$		

FV	GL	SC	CM	Fc	F36,0.05	F36,0.01
Tratamientos	3	15.19	5.0633	0.5438n.s.	4.76	9.78
Hileras	3	7.69	2.5633	0.2752n.s.	4.76	9.78
Columnas.	3	24.69	8.2300	0.8838n.s.	4.76	9.78
Error	6	55.87	9.3117			
Total	15	103.44				

El criterio de decisión es:

Rechazar Ho1, si $F_c > F_{t-1(t-1)(t-2), \alpha}$ Para tratamientos

Rechazar Ho2, si $F_c > F_{r-1(t-1)(t-2), \alpha}$ Para Hileras

Rechazar Ho3, si $F_c > F_{r-1(t-1)(t-2), \alpha}$ Para Columnas

Los niveles de TGP son iguales en las diferentes razas, patologías y concentración de vitaminas.



Problemario de Diseño en Cuadrado Latino

26. Se evalúa un nuevo tratamiento con Selenio para reducir la carga viral de VIH probando 4 diferentes dosis (200, 300, 400 y 500 µg) en pacientes con VIH de diferentes edades (20, 30, 40 y 50 años) con periodos de incubación de (8, 9, 10 y 11 años): Se dio el tratamiento durante 6 meses y se procedió a medir la carga viral arrojando los siguientes resultados en copias/ml:	Periodo de incubación	Edades			
		20	30	40	50
	8	t2-6000	t3-6500	t4-7000	t1-5500
	9	t1-5000	t2-7000	t3-7000	t4-7500
	10	t4-8200	t1-4500	t2-4800	t3-6100
	11	t3-6900	t4-7100	t1-5100	t2-6300

Se dio el tratamiento durante 6 meses y se procedió a medir la carga viral arrojando los siguientes resultados en copias/ml:

- ¿Existen diferencias entre las diferentes edades?
- ¿Existen diferencias entre las diferentes periodos de incubación?
- ¿Existen diferencias entre los diferentes dosis de Selenio?

27. Se evalúan cuatro vacunas derivadas de la cepa RB51 (t1, t2, t3 y t4) contra la brucelosis en ganado bovino de diferentes razas de ganado cebú (Guzerat, Nelore, Sardo Negro e IndoBrasil) y edades (1, 2, 3 y 4 años). Se manejaron lotes de 25 animales por raza y edad. Se evaluaron al año de la aplicación de las vacunas, habiendo recibido un manejo Sanitario y Zootécnico normal. Se procedió a cuantificar el número de animales vacunados que contrajeron la enfermedad, obteniéndose los siguientes resultados

Edad	Razas			
	A	B	C	D
I	t1-18	t2-16	t3-12	t4-21
II	t4-22	t1-17	t2-18	t3-19
III	t3-11	t4-17	t1-19	t2-17
IV	t2-19	t3-18	t4-19	t1-20

- ¿Existen diferencias entre las diferentes Vacunas? Explique.
- ¿Existen diferencias entre las diferentes Razas? Explique.
- ¿Existen diferencias entre las diferentes Edades? Explique.

28. Se evalúan 4 diferentes tipos de insecticidas (piretroide, organoclorado, organofosforado, carbamato) durante 4 semanas (I, II, III y IV) en 4 especies de Triatomos (Triatoma barberi, T. brasiliensis, T. pallidipennis y T. dimidiata). Los resultados obtenidos se dieron en relación al número de triatomos encontrados muertos, siendo estos:



Semanas	Especies			
	A	B	C	D
I	t1-12	t2-9	t3-6	t4-7
II	t4-6	t1-14	t2-10	t3-4
III	t3-5	t4-7	t1-11	t2-8
IV	t2-9	t3-5	t4-7	t1-12

- ¿Existen diferencias entre los diferentes Insecticidas? Explique.
- ¿Existen diferencias entre las diferentes Especies? Explique.
- ¿Existen diferencias entre las diferentes Semanas? Explique.

29. Se evalúan 4 medidas contra la contaminación ambiental:

t1: Uso de gas en vehículos públicos,

t2: Eliminación de carros de diesel,

t3: Un día sin circulación y

t4: No permitir correr por arriba de 80Km/hr,

En diversas ciudades del país (Cd. De México, Guadalajara, Monterrey y Puebla), durante las 4 semanas de Julio (I, II, III y IV). El parámetro a medir fue el SO_2 en $\mu\text{g}/\text{m}^3$, (Índice recomendado $40\mu\text{g}/\text{m}^3$) los resultados obtenidos fueron los siguientes:

Semanas	Ciudad			
	A	B	C	D
I	t1-80	t2-69	t3-71	t4-
II	t4-93	t1-83	t2-72	t3-
III	t3-85	t4-88	t1-69	t2-
IV	t2-xx	t3-	t4-	t1-

- ¿Existen diferencias entre las diferentes Medidas contra la contaminación? Explique.
- ¿Existen diferencias entre las diferentes Ciudades? Explique.
- ¿Existen diferencias entre las diferentes Semanas? Explique.

30. El tratamiento con la hormona progesterona puede ayudar a prevenir el parto prematuro entre las mujeres de alto riesgo. Por lo que se realiza un estudio para evaluar los niveles de progesterona en mujeres embarazadas que han tenido abortos previos, las edades de las participantes son de: 20, 25, 30, 35 y 40 años, y se evalúan los niveles durante los meses de gestación (3°, 4°, 5°, 6° y 7°). Después de evaluar que no tuvieran brucelosis ni toxoplasmosis, se les suministro un derivado de la progesterona en dosis de (100, 150, 200, 250 y 300 mg)

Mes	Edades				
	20	25	30	35	40
3°	t5-90	t4-88	t3-85	t4-82	t1-80
4°	t1-88	t5-86	t4-83	t3-80	t2-78
5°	t2-85	t1-83	t5-79	t4-77	t3-75
6°	t3-81	t2-80	t1-76	t5-74	t4-72
7°	t4-77	t3-76	t2-73	t1-70	t5-69



- ¿Existen diferencias entre los diferentes meses de gestación?
- ¿Existen diferencias entre las diferentes Edades?.
- ¿Existen diferencias entre las diferentes dosis de progesterona?

Referencias

Arnau Gras, J.. Diseños Experimentales Multivariables. Alianza Editorial, Madrid. 1990.

Box, G., Hunter, W.G. and Hunter, J.S. Estadística para Investigadores. Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos. Editorial Reverté. 2002.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Cox, D.R. Planning of Experiments. Wiley, New York. 1992.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C.: "Análisis Multivariante", Ed. Prentice Hall, 5ª edición. 1999,

Hinkelmann, K. Design and analysis of experiments. John Wiley & Sons, N.Y. (2005).

Jobson, J.D. Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design. Springer-Verlag, New York. 1991.

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lara, A.M. Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS. Proyecto Sur de Ediciones, Granada. 2001.

Lindman, H.R. Analysis of Variance in Experimental Design. Springer-Verlag, N.Y. 1992.

Montgomery, D.C. Diseño y Análisis de Experimentos. Limusa, México. 2003.



Peña Sánchez de Rivera, D. Estadística, Modelos y Métodos. Volumen II: Modelos Lineales y Series Temporales. Alianza Universidad, Madrid. 2000.

Peña, D., Regresión y Diseño de Experimentos., Alianza Editorial. 2002.

Rencher, A.C. Multivariate Statistical Inference and Applications. 1998.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Walpole R.E., Myers R.H., Myers S.L: "Probabilidad y Estadística para Ingenieros", Ed. Prentice Hall, 6ª edición.1998,

Wiley. N.Y. Toutenburg, H. Statistical Analysis of Designed Experiments. 2a edición. Springer-Verlag, New York. 2002.



Parte II: Modelos Clásicos



Capítulo VI: Diseños Incompletos

En la experimentación es frecuente, por diversas causas, que se pierdan las observaciones de una o varias unidades experimentales, ocasionando, que el experimento pierda simetría, complicando el método de análisis.

El caso más simple es cuando se pierde sólo la observación de una unidad experimental en los diseños simples; en este caso el valor perdido se puede estimar con una fórmula sencilla; el valor estimado se inserta en el lugar que le corresponde a la unidad experimental donde se perdió la observación y se completa el análisis con la rutina usual de cálculo, restando un grado de libertad al término de error, debido a la estimación del valor perdido. Es un procedimiento no muy exacto, pero suficientemente confiable.

Cuando se tienen dos o más unidades experimentales pérdidas, puede recurrirse a varios procedimientos para resolverlos:

1. Realizar el análisis del experimento empleando el método general de análisis basado en el modelo de la clasificación de doble entrada, con números desiguales en las subclases y sin interacción.
2. Estimar los valores perdidos a través de la técnica de la covarianza múltiple.
3. Mediante el uso de procedimientos dados en paquetes computacionales.

El cálculo de la estimación de unidades experimentales pérdidas en los DBA y DCL, requieren de fórmulas establecidas, los cuales se han obtenido a través de análisis por diversos estudiosos de la materia.



Bloques al Azar con Observaciones Perdidas

Cuando se usa el DCA, la pérdida de una unidad experimental no afecta el procedimiento de cálculo del ANVA, por esa razón no se requiere estimar el valor perdido.

Cuando por alguna razón en un experimento bajo un DBA se pierde una unidad experimental, entonces no es posible analizar el experimento, ya que el diseño requiere, para ello igual número de repeticiones por tratamiento; es necesario hacer una estimación de la observación faltante, antes de aplicar el ANVA. Así mismo, será para cuando se pierden dos observaciones; los procedimientos a seguir son:

1) Cuando se pierde una observación:

El valor de la observación pérdida y_{ij} será estimada por

$$\hat{Y}_{ij} = \frac{t T_i + r B_j - G}{(t-1)(r-1)}$$

En donde:

\hat{Y}_{ij}

= Valor estimado de y_{ij}

T_i = Total del tratamiento que contiene la parcela pérdida

B_j = Total del Bloque que contiene la parcela pérdida

G = Total de todas las observaciones antes de estimar y_{ij}

t = Número de tratamientos

r = Número de Bloques

Al estimar este valor, se coloca en la parcela faltante y se analiza el experimento normalmente, sólo que los grados de libertad del Error y Total se disminuyen en uno.

El ANVA resultante se muestra en el cuadro 15



Cuadro 15.- Análisis de varianza para el DBA con un observación perdida

FV	GL	SC	CM	Fc
Trat.	t-1	$\sum_{i=1}^t (Y_{i.})^2/r - (Y_{..})^2/rt$	S.C. Trat./t-1	C.M. Trat
				C.M. Error
Bloq.	r-1	$\sum_{i=1}^t (Y_{.j})^2/r - (Y_{..})^2/rt$	S.C. Bloq/r-1	C.M. Bloq.
				C.M. Error
Error	(t-1) (r-1)-1	S.C.Tot.-S.C.Tr.-S.C.Bl.	S.C.Err./t-1 r-1	
Total	(rt-1)-1	$\sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - (Y_{..})^2/rt$		

Este ANVA presenta a la Sumatoria de Cuadrados de Tratamientos sesgada hacia un valor grande debido a que el valor estimado de la parcela pérdida está calculado de tal forma que la Sumatoria de Cuadrados del Error es mínima; siendo necesario construir otro análisis de la varianza para hacer una prueba exacta para tratamientos (cuando existan más de 10 tratamientos).

Debido a lo anterior se calcula el sesgo que presenta la Sumatoria de Cuadrados de Tratamientos mediante la formula:

$$\text{Sesgo} = \Delta = \frac{\sum [B_j - (t-1) Y_{ij}]^2}{t(t-1)}$$

Con:

B_j = Total del bloque que contiene la parcela pérdida antes de estimar y_{ij} .

\wedge

Y_{ij} = Valor estimado de la parcela pérdida

t = Número de tratamientos

La Sumatoria de Cuadrados de Tratamientos insesgada, también llamada Sumatoria de Cuadrados de Tratamientos ajustadas por bloques, se calcula mediante:

$$SC(\text{Trat/Bloques}) = SC(\text{Trat}) - \Delta$$



Posteriormente se calculan las Sumatoria de Cuadrados Total y Sumatoria de Cuadrados de Bloques con todas las observaciones originales, excepto Y_{ij} , la Sumatoria de Cuadrados Error se toma del ANVA antes obtenido.

Otra forma de calcular SC(Trat/Bloques) es:

$$SC(\text{Trat/Bloques}) = SC(\text{Total}) - SC(\text{Bloques}) - SC(\text{Error})$$

En este caso la Sumatoria de Cuadrados Total es calculada con las observaciones originales y las Sumatoria de Cuadrados de Bloques y Sumatoria de Cuadrados Error cuando ya se estimó la observación perdida (Y_{ij}).

2) Cuando se pierden dos observaciones

La pérdida de dos observaciones puede ser:

- Del mismo tratamiento
- Del mismo bloque
- De diferente bloque y diferente tratamiento

a) Del mismo tratamiento

$$x_1 = \frac{t T_{12} + (r-1)B_1 + B_2 - G}{(t-1)(r-2)}$$

$$x_2 = \frac{t T_{12} + B_1 + (r-1)B_2 - G}{(t-1)(r-2)}$$

donde:

X_1 y X_2 = Valores estimados de las observaciones pérdidas

T_{12} = Total de las observaciones presentes en el tratamiento donde falta X_1 y X_2 .

B_1 = Total de las observaciones presentes en el bloque donde falta X_1 .

B_2 = Total de las observaciones presentes en el bloque donde falta X_2 .

G = Total de todas las observaciones antes de estimar X_1 y X_2



b) Del mismo bloque

$$x_1 = \frac{t B_{12} + (r-1)T_1 + T_2 - G}{(r-1)(t-2)}$$

$$x_2 = \frac{t B_{12} + T_1 + (r-1)T_2 - G}{(r-1)(t-2)}$$

donde:

X_1 y X_2 Valores estimados de las observaciones pérdidas
=

B_{12} = Total de las observaciones presentes en el bloque donde falta X_1 y X_2 .

T_1 = Total de las observaciones presentes en el tratamiento donde falta X_1 .

T_2 = Total de las observaciones presentes en el tratamiento donde falta X_2 .

G = Total de todas las observaciones antes de estimar X_1 y X_2

C) De diferente bloque y diferente tratamiento

$$x_1 = \frac{(t-1)(r-1)(t T_1 + r B_1 - G) - (t T_2 + r B_2 - G)}{[(t-1)(r-1)]^2 - 1}$$

$$x_2 = \frac{(t-1)(r-1)(t T_2 + r B_2 - G) - (t T_1 + r B_1 - G)}{[(t-1)(r-1)]^2 - 1}$$

Una vez calculados los datos en cualesquiera de las posibilidades, se sustituyen en el cuadro de concentración de datos, se obtiene el ANVA reduciendo en dos unidades (-2) los grados de libertad en el error.



Ejemplo:

Tabla con parcelas perdidas

Bloque	Tratamientos				$\bar{y}_{.j}$
	1	2	3	4	
I	7.3	x?	6.7	5.7	19.7
II	7.2	5.5	x?	6.9	19.6
III	7.6	6.8	6.8	6.4	27.6
IV	7.2	6.5	7.4	6.1	27.2
V	7.5	6.8	7.5	6.4	28.2
VI	7.6	7.1	6.3	6.3	27.3
$\sum y_{i.}$	44.4	32.7	34.7	37.8	149.6

Como las parcelas perdidas son de diferente tratamiento y bloque.

$$x1 = \frac{(t-1)(r-1)(tT1 + rB1 - G) - (tT2 + rB2 - G)}{[(t-1)(r-1)]^2 - 1}$$

$$x1 = \frac{(4-1)(6-1)[(4)(32.7) + (6)(19.7) - 149.6] - [(4)(34.7) + (6)(19.6) - 149.6]}{[(4-1)(6-1)]^2 - 1} = 6.18$$

$$x2 = \frac{(t-1)(r-1)(tT2 + rB2 - G) - (tT1 + rB1 - G)}{[(t-1)(r-1)]^2 - 1}$$

$$x2 = \frac{(4-1)(6-1)[(4)(34.7) + (6)(19.6) - 149.6] - [(4)(32.7) + (6)(19.7) - 149.6]}{[(4-1)(6-1)]^2 - 1} = 6.70$$

Los valores estimados se introducen en las parcelas perdidas y se recalculan los totales afectados:

Bloque	Tratamientos				$\bar{y}_{.j}$
	1	2	3	4	
I	7.3	6.18	6.7	5.7	25.88
II	7.2	5.5	6.70	6.9	26.3
III	7.6	6.8	6.8	6.4	27.6
IV	7.2	6.5	7.4	6.1	27.2
V	7.5	6.8	7.5	6.4	28.2
VI	7.6	7.1	6.3	6.3	27.3
$\sum y_{i.}$	44.4	38.9	41.4	37.8	162.5

El experimento se analiza normalmente, sólo que los grados de libertad del Error y Total se disminuyen en uno (-2).



Cuadro Latino con Observaciones Perdidas

Cuando se pierden las observaciones en un DCL pueden estimarse mediante fórmulas ya establecidas al igual que en el DBA; para ello tenemos los siguientes tres casos:

1) Cuando se pierde una observación

El valor de la observación perdida Y_{ijk} se estima con:

$$\hat{Y}_{ijk} = \frac{t(H + C + T) - 2G}{(t-1)(t-2)}$$

Donde:

\hat{Y}_{ijk}

= Valor estimado de Y_{ijk}

H= Total de la hilera en donde falta la observación

C= Total de la columna en donde falta la observación

T= Total del tratamiento en donde falta la observación (Σ)

G= Total de las observaciones antes de estimar Y_{ijk}

t= Número de tratamientos

Una vez estimado el Y_{ijk} faltante se coloca en la posición que le corresponde en la tabla de datos y se analiza en forma normal para obtener el ANVA correspondiente, únicamente que se le resta un grado de libertad, tanto al Error como al total.

Al igual que el DBA, la Sumatoria de Cuadrados de los Tratamientos se subestima y cuando el número de tratamientos sea mayor de 10, se resta el término:

$$\text{Sesgo} = \Delta = \frac{G - H - C - (t-1) T^2}{[(t-1)(t-2)]^2}$$

Con H,C,T,G y t definidos para este diseño

	C1	C2	C3	C4
H1	T1	T2	T3	T4
H2	T4	T1	T2	T3
H3	T3	T4	T1	T2
H4	T2	T3	T4	T1



2) Cuando se pierden dos observaciones

En este caso se tienen las siguientes posibilidades

- En el mismo Tratamiento pero en Hileras y Columnas diferentes
- En la misma Hilera, pero en Columna y Tratamientos diferentes
- En la misma Columna, pero en Hileras y Tratamientos diferentes
- En diferentes Tratamientos, Hilera y Columna

a) En el mismo Tratamiento pero en Hileras y Columnas diferentes

$$x_1 = \frac{(t-1)Q_1 + Q_2}{t(t-2)^2} \quad x_2 = \frac{(t-1)Q_2 + Q_1}{t(t-2)^2}$$

$$Q_1 = t(H_1 + C_1 + T_{12}) - 2G \quad Q_2 = t(H_2 + C_2 + T_{12}) - 2G$$

	C1	C2	C3	C4
H1	T1	T2	T3	T4
H2	T4	T1	T2	T3
H3	T3	T4	T1	T2
H4	T2	T3	T4	T1

además:

H_1 y C_1 = Totales de la Hilera y la Columna donde falta X_1
 H_2 y C_2 = Totales de la Hilera y la Columna donde falta X_2
 T_{12} = Total del tratamiento donde faltan X_1 y X_2
 G = Total de las observaciones antes de estimar a X_1 y X_2
 X_1 y X_2 = observaciones faltantes
 t = Número de tratamientos

b) En la misma Hilera, pero en Columna y Tratamientos diferentes

Se utilizan las mismas fórmulas de estimación antes descritas, con las siguientes variantes:

$$Q_1 = t(T_1 + C_1 + H_{12}) - 2G \quad Q_2 = t(T_2 + C_2 + H_{12}) - 2G$$

	C1	C2	C3	C4
H1	T1	T2	T3	T4
H2	T4	T1	T2	T3
H3	T3	T4	T1	T2
H4	T2	T3	T4	T1

además:

T_1 y C_1 = Totales de tratamiento y columna donde falta X_1
 T_2 y C_2 = Totales de tratamiento y columna donde falta X_2
 H_{12} = Total de la Hilera donde faltan X_1 y X_2
 G = Total de las observaciones antes de estimar a X_1 y X_2
 t = Número de tratamientos



c) En la misma Columna, pero en Hileras y Tratamientos diferentes

$$x1 = \frac{(t-1) Q_1 + Q_2}{t(t-1)} \quad x2 = \frac{(t-1) Q_2 + Q_1}{t(t-1)}$$

$$Q_1 = t(T_1 + H_1 + C_{12}) - 2G \quad Q_2 = t(T_2 + H_2 + C_{12}) - 2G$$

	C1	C2	C3	C4
H1	T1	T2	T3	T4
H2	T4	T1	T2	T3
H3	T3	T4	T1	T2
H4	T2	T3	T4	T1

Además: T_1, T_2, H_1 y H_2 definidos anteriormente

C_{12} = Total de la columna donde faltan X_1 y X_2

d) En diferentes Tratamientos, Hilera y Columna

$$x1 = \frac{f Q_1 - 2Q_2}{f^2 - 4} \quad x2 = \frac{f Q_2 - 2Q_1}{f^2 - 4}$$

$$Q_1 = t(T_1 + H_1 + C_1) - 2G \quad Q_2 = t(T_2 + H_2 + C_2) - 2G$$

$$f = (t-1)(t-2)$$

	C1	C2	C3	C4
H1	T1	T2	T3	T4
H2	T4	T1	T2	T3
H3	T3	T4	T1	T2
H4	T2	T3	T4	T1

Además: T_i, H_j, C_k, t y G definidos anteriormente

Una vez calculados los valores faltantes, se colocan en los lugares correspondientes en la tabla de datos, luego se analizan en forma normal, en el cuadro del ANVA se reducen en dos unidades (-2) los grados de libertad del error y del total.

Ejemplo:

H i l e r a s	Columnas										Y _{.j.}
		I		II		III		IV		V	
I	t3	10	t5	8	t2	14	t4	13	t1	17	62
II	t5	7	t2	12	t4	?	t1	17	t3	16	52
III	t2	13	t4	12	t1	16	t3	15	t5	11	67
IV	t1	14	t3	13	t5	10	t2	16	t4	16	69
V	t4	14	t1	16	t3	15	t5	12	t2	18	75
Y _{..k}	58		61		55		73		78		325

$$t4 = 14 + 12 + 13 + 16 = 55$$



Como solo se perdió una observación el valor de la observación perdida Y_{ijk} se estima con:

$$\hat{Y}_{ijk} = \frac{t(H+C+T)-2G}{(t-1)(t-2)}$$

$$= \frac{5(52+55+55)-2(325)}{(5-1)(5-2)} = 13.33$$

Se introduce el valor estimado en la parcela perdida y se recalculan los totales afectados

		Columnas										
H i l e r a s		I		II		III		IV		V		Y. _{j.}
	I	t3	10	t5	8	t2	14	t4	13	t1	17	62
	II	t5	7	t2	12	t4	13.3	t1	17	t3	16	65.33
	III	t2	13	t4	12	t1	16	t3	15	t5	11	67
	IV	t1	14	t3	13	t5	10	t2	16	t4	16	69
	V	t4	14	t1	16	t3	15	t5	12	t2	18	75
	Y.. _k	58		61		68.3		73		78		338.3

$$t4 = 14+12+13.3+13+16 = 68.3$$

El experimento se analiza en forma normal para obtener el ANVA correspondiente, restándosele un grado de libertad, al Error y al total.

Referencias

Box, G., Hunter, W.G. and Hunter, J.S. Estadística para Investigadores. Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos. Editorial Reverté. 2002.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.



Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lara, A.M. Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS. Proyecto Sur de Ediciones, Granada. 2001.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Wiley. N.Y. Toutenburg, H. Statistical Analysis of Designed Experiments. 2a edición. Springer-Verlag, New York. 2002.



Parte II: Modelos Clásicos



Capítulo VII: Comparación de Medias de Poblaciones o Tratamientos

Debido a que el análisis de la varianza de los Diseños Simples únicamente nos indica que algunos tratamientos son diferentes, la siguiente pregunta obligada es ¿Quiénes son diferentes?, por lo que el problema ahora es como hacerle para poder conocer quienes son diferentes entre sí. Para poder establecer las diferencias tenemos que utilizar el recurso estadístico de las comparaciones de medias. Sin embargo estas comparaciones solo pueden ser utilizadas hasta que se ha probado con la prueba de F, el juego de hipótesis:

Ho: $\mu_1 = \mu_2 = \dots = \mu_k$

VS

Ha: al menos un μ_i es diferente

Ho: El efecto de los tratamientos es el mismo VS

Ha: Al menos un tratamiento tiene un efecto diferente

Si la prueba de F es significativa, indica que la variabilidad entre tratamientos no se debe al azar, sino a un efecto debido a los tratamientos; sin embargo, no señala cuales medias son iguales o diferentes.

Entonces se requiere del uso de pruebas estadísticas de comparación de medias de acuerdo a la naturaleza de los tratamientos en estudio.

Tipos de Tratamientos

La metodología apropiada para la comparación de tratamientos depende de la naturaleza de los mismos; los cuales se pueden clasificar como se observa en el cuadro 16.

Cuadro 16.- Clasificación de los tratamientos

	Sin Estructura	no es posible asignar una escala numérica con
--	----------------	---



Tratamientos	Cualitativos		significado diferente al nominativo
		Con Estructura	Se puede agrupar con sentido a alguna(s) característica(s)
	Cuantitativos	Es posible asignar una escala numérica con significado (niveles de un factor de estudio)	

a) Tratamientos Cualitativos sin Estructura: Las comparaciones de medias que se pueden utilizar son los métodos de comparaciones múltiples siguientes:

- Diferencia Mínima Significativa (DMS)
- Método de Duncan
- Método de Student-Newman-Keuls (S-N-K)
- Método de Tukey
- Método de Scheffé
- Método de Dunnett

Al encontrar diferencias significativas en un conjunto de tratamientos, necesariamente se manifiesta entre el tratamiento con la media menor y el tratamiento con la mayor media.

b) Tratamientos Cualitativos con Estructura: Se puede realizar por comparación ortogonales o contrastes ortogonales en el primer caso se puede utilizar la DMS.

c) Tratamientos Cuantitativos: El procedimiento es fraccionar la suma de cuadrados de tratamientos por medio de polinomios ortogonales, es utilizar los modelos lineales en los parámetros de estudio.

La metodología apropiada es ajustar una superficie de respuesta por medio de la regresión.



Comparaciones Múltiples de Medias

Como se puede observar para los tratamientos cualitativos sin estructura, existen una gran variedad de métodos de comparación, sin embargo es necesario mencionar que la selección de unos de estos depende de las características que el experimento tenga, es decir va a estar en función del número de tratamientos en estudio, si tienen o no el mismo número de repeticiones, si es un experimentos con observaciones perdidas, etc..

Por lo que antes de seleccionar uno de estos métodos hay que observar las condiciones que se requieren para que pueda ser usado, de otra forma estamos cometiendo errores en el análisis estadístico del experimento

Así que el objetivo de la experimentación es conocer cual de los tratamientos es el que da los mejores resultados (mayores rendimientos o menores efectos nocivos), es decir, se desea conocer si uno o más tratamientos son superiores sobre otros.

Por lo que el mecanismo para esta determinación se conoce como método de comparaciones múltiples, en la metodología de análisis se utilizan los datos del análisis de la varianza, realizando las pruebas de significación o comparaciones entre las medias de los tratamientos; esto es probar el juego de hipótesis:

$$H_0: \tau_i = \tau_{i'} \quad \text{V.S.} \quad H_a: \tau_i \neq \tau_{i'} \text{ con } i \neq i'$$

Pudiéndose hacer $t(t-1)/2$ combinaciones posibles de medias. Las comparaciones no son todas independientes (ortogonales). A continuación se describen los métodos más comunes para este tipo de comparaciones.



Diferencia Mínima Significativa (DMS)

Es un método ideado por Fisher en 1935; también se le conoce como prueba de “t”, debido a que se basa en la distribución de “t” de Student.

Se aplica solo cuando la prueba de F es significativa sólo debe emplearse para comparar medias adyacentes en un arreglo ordenado de las mismas.

Es una técnica no muy precisa para hacer comparaciones múltiples; sin embargo, es válida para comparaciones planeadas entre dos medias; esto es, comparar un tratamiento estándar con otros tratamientos y cuando las comparaciones entre medias son independientes.

La ventaja de esta prueba es que es muy fácil de realizar y emplea un estimador único para efectuar todas las comparaciones.

El análisis consiste en probar el juego de hipótesis:

$$H_0: \tau_i = \tau_{i'} \quad \text{V.S.} \quad H_a: \tau_i \neq \tau_{i'} \text{ con } i \neq i'$$

Mediante el estimador:

$$DMS_{ii'} = (\tau_{\alpha, GL \text{ Error}}) (S(\bar{Y}_i - \bar{Y}_{i'}))$$

En donde: $DMS_{ii'} = \text{Estimador}$

$\tau_{\alpha, GL \text{ Error}} = \text{Valor de “t” en la tabla}$

$$S(\bar{y}_i - \bar{y}_{i'}) = \sqrt{CMError \left\langle \frac{1}{ni} + \frac{1}{ni'} \right\rangle}$$

La formula general del estimador queda de la siguiente forma:

$$DMS_{ii'} = t_{(\alpha, GL_E)} \sqrt{CMError \left\langle \frac{1}{ni} + \frac{1}{ni'} \right\rangle}$$

El procedimiento a seguir es:

- 1) Se ordenan las medias de los tratamientos
- 2) Se calcula el estimador $DMS_{ii'}$
- 3) Se comparan las medias de tal forma que si la diferencia de dos medias es mayor que el valor de $DMS_{ii'}$, entonces se rechaza $H_0: \tau_i = \tau_{i'}$, en caso contrario no se rechaza $H_0: \tau_i = \tau_{i'}$, esto es:
Rechazar $H_0: \tau_i = \tau_{i'}$ Si $|\bar{Y}_i - \bar{Y}_{i'}| > DMS_{ii'}$



- 4) Se presentan los resultados en forma tabular, identificando con la misma letra las medias estadísticamente iguales, o sea, las que no presentan diferencias significativas entre ellas.

Ejemplo: Se toman los datos de la tabla del ejemplo del DCA

	D1	D2	D3
1	58	56	67
2	60	54	64
3	61	60	58
4	54	58	65
5	58	52	
6		59	
7		54	
Yi.	291	393	254
\bar{y}	58.20	56.14	63.50

$$GL_E = 13$$

$$CM_E = 9.743$$

$$t_{0.01,13} = 3.012$$

1.- Se ordenan las medias.

$$T3 = 63.50$$

$$T1 = 58.20$$

$$T2 = 56.14$$

2.- Se calculan los estimadores $DMS_{ii'}$

$$DMS_{12} = (3.012) \sqrt{9.743 \left(\frac{1}{5} + \frac{1}{7} \right)} = 5.5$$

$$DMS_{13} = (3.012) \sqrt{9.743 \left(\frac{1}{5} + \frac{1}{4} \right)} = 6.3$$

$$DMS_{23} = (3.012) \sqrt{9.743 \left(\frac{1}{7} + \frac{1}{4} \right)} = 5.9$$

3.- Comparar las medias y aplicar la regla de decisión

$$|T1 - T2| = |58.20 - 56.14| = 2.06 < 5.5 \text{ No se rechaza } H_0. \text{ Son iguales.}$$

$$|T1 - T3| = |58.20 - 63.50| = 5.30 < 6.3 \text{ No se rechaza } H_0. \text{ Son iguales.}$$

$$|T2 - T3| = |56.14 - 63.50| = 7.36 > 5.9 \text{ Se rechaza } H_0. \text{ No son iguales.}$$

4.- Se presentan los datos ordenados identificando con la misma letra a los que son iguales.

Cuadro de Medias

Trat.	Media	Clasif.
T3	63.85	a
T1	58.20	ab
T2	56.14	b



Método de Duncan

Este método fue desarrollado por Duncan en 1951, tratando de encontrar métodos de comparaciones múltiples más eficientes que el de la DMS. Ha tenido poca aceptación, debido a que es un método aproximado que tiene poco control del error tipo I.

También es conocido como prueba de “t” modificada. La prueba de rango múltiple de Duncan; utiliza una tabla de “t” modificada por Duncan.

Se calculan valores estudentizados para cada una de las comparaciones de medias adyacentes, así como para aquellas que están separadas por otras (se puede compara un tratamiento contra todos los demás); es de tipo secuencial. Es necesario que los tratamientos tengan el mismo número de repeticiones.

La ventaja de esta prueba es que no se necesita que el valor de F sea significativo para poder usarla. Permite hacer comparaciones múltiples de todas las medias de tratamientos, sin restricciones.

El análisis consiste en probar el juego de hipótesis:

$$H_0: \tau_i = \tau_{i'} \quad \text{V.S.} \quad H_a: \tau_i \neq \tau_{i'} \text{ con } i \neq i'$$

Mediante el estimador: $RME_{ii'} = S_y(\tau_{\alpha, p, GL \text{ Error}})$

En donde:

$RME_{ii'}$ (Rango Mínimo de Error) = Estimador

P = No. de medias en el rango de la prueba

$$S_y = \sqrt{\frac{CMError}{r}}$$

El procedimiento es:

- 1) Se ordenan las medias de los tratamientos
- 2) Se calcula el estimador $RME_{ii'}$, con los valores de “t” en la tabla de rangos estudentizados para el número de medias en el rango.
- 3) Se comparan las medias de tal forma que si la diferencia de dos medias es mayor que el valor de $RME_{ii'}$, entonces se rechaza $H_0: \tau_i = \tau_{i'}$, en caso contrario no se rechaza $H_0: \tau_i = \tau_{i'}$, esto es:
Rechazar $H_0: \tau_i = \tau_{i'}$ Si $|Y_i - Y_{i'}| > RME_{ii'}$
- 4) Se presentan los resultados en forma tabular, identificando con la misma letra las medias estadísticamente iguales, o sea, las que no presentan diferencias significativas entre ellas.



Ejemplo: Retomando el ejemplo del DBA

	T1	T2	T3	T4	Y.j
I	4.1	6.5	4.1	4.9	19.6
II	4.1	5.3	4.0	6.2	19.6
III	6.5	6.9	4.5	4.8	22.7
IV	4.3	6.8	4.3	4.2	19.6
V	6.0	6.5	4.1	6.9	23.5
Yi.	25	32	21	27	105
\bar{y}	5.0	6.4	4.2	5.4	

F.V.	G.L.	S.C.	C.M.	F
Tratam	3	12.550	4.1830	6.015*
Bloques	4	3.7557	0.9387	1.35ns
Error	12	8.3454	0.6954	
Total	19	24.650		

Trat. Media
 T2= 6.4 a
 T4= 5.4 a
 T1= 5.0 a b
 T3= 4.2 b

Los datos extraídos de la tabla del análisis de varianza son:

CMError = 0.6954
 G.L.Error = 12
 r = 5

$$Sy = \sqrt{\frac{CMError}{r}} = \sqrt{\frac{0.6954}{5}} = 0.3729$$

α	=	0.05		
p	2	3	4	
t	3.08	3.23	3.33	
Sy	0.3729	0.3729	0.3729	
RME	1.15	1.20	1.24	



Método de Student-Newman-Keuls (S-N-K)

Es una prueba similar a la de Duncan; se considera menos potente y conservador en el número de diferencias significativas que el de Duncan.

También es de tipo secuencial, utiliza un único valor estandarizado para cada una de las comparaciones de medias adyacentes.

Con este método se prueba el juego de hipótesis:

$$H_0: \tau_i = \tau_{i'} \quad \text{V.S.} \quad H_a: \tau_i \neq \tau_{i'} \text{ con } i \neq i'$$

Mediante el estimador: $RME_p = S_{\bar{y}} q(\alpha, p, GL \text{ Error})$

En donde:

$RME_p = \text{Estimador}$

$$S_{\bar{y}} = \sqrt{\frac{CMError}{r}}$$

$P = \text{No. de medias en el rango de la prueba}$

$q(\alpha, p, GL \text{ Error}) = \text{valor "t" Modificado}$

El procedimiento es:

- 1) Se ordenan las medias de los tratamientos
- 2) Se calcula el estimador RME_p , con los valores de "t" en la tabla de rangos estudentizados para el número de medias en el rango.
- 3) Se comparan las medias de tal forma que si la diferencia de dos medias es mayor que el valor de RME_p , entonces se rechaza $H_0: \tau_i = \tau_{i'}$, en caso contrario no se rechaza $H_0: \tau_i = \tau_{i'}$, esto es:

$$\text{Rechazar } H_0: \tau_i = \tau_{i'} \quad \text{Si } |Y_i - Y_{i'}| > RME_p$$

- 4) Se presentan los resultados en forma tabular, identificando con la misma letra las medias estadísticamente iguales, o sea, las que no presentan diferencias significativas entre ellas.



Ejemplo: Retomando el ejemplo del DBA:

	T1	T2	T3	T4	Y.j
I	4.1	6.5	4.1	4.9	19.6
II	4.1	5.3	4.0	6.2	19.6
III	6.5	6.9	4.5	4.8	22.7
IV	4.3	6.8	4.3	4.2	19.6
V	6.0	6.5	4.1	6.9	23.5
Yi.	25	32	21	27	105
\bar{y}	5.0	6.4	4.2	5.4	

F.V.	G.L.	S.C.	C.M.	F
Tratam	3	12.550	4.1830	6.015*
Bloques	4	3.7557	0.9387	1.35ns
Error	12	8.3454	0.6954	
Total	19	24.650		
Trat. Media T2= 6.4 a T4= 5.4 a T1= 5.0 a b T3= 4.2 b				

Los datos extraídos de la tabla del análisis de varianza son:

$$\begin{aligned} CMError &= 0.6954 \\ G.L.Error &= 12 \\ r &= 5 \end{aligned}$$

$$S_y = \sqrt{\frac{CMError}{r}} = \sqrt{\frac{0.6954}{5}} = 0.3729$$

$$\begin{aligned} t &= 3.33 \\ \alpha &= 0.05 \end{aligned}$$

$$RMEp = S_{\bar{y}} q(\alpha, p, GL Error) =$$

$$(0.3729)(3.33) = 1.24$$



Método de Tukey

Es una prueba que se emplea para hacer todas las comparaciones múltiples que son posibles con t tratamientos; basada en la distribución del rango estandarizado.

Se le conoce con el nombre de Diferencia Significativa Honesta (DSH), en virtud de que mantiene el nivel de significancia sin alterar.

El número de comparaciones múltiples posibles son $t(t-1)/2$. Esta prueba no es secuencial, y considera un valor teórico común para las comparaciones posibles.

Es exacta cuando los tratamientos tienen el mismo número de repeticiones, y para medias que no han sido ajustadas por covarianza (diseños incompletos).

Con este método se prueba el juego de hipótesis:

$$H_0: \tau_i = \tau_{i'} \quad \text{V.S.} \quad H_a: \tau_i \neq \tau_{i'} \text{ con } i \neq i'$$

Mediante el estimador: $DHS = S_{\bar{y}} Q_{(\alpha, p, GL \text{ Error})}$

$$S_{\bar{y}} = \sqrt{\frac{CMError}{r}}$$

En donde:

DHS = Estimador

$Q_{(\alpha, p, GL \text{ Error})}$ = Valor en la tabla de rangos estandarizados

p = Número de tratamientos

El procedimiento es:

- 1) Se ordenan las medias de los tratamientos
- 2) Se calcula el estimador DSH, con los valores de Q en la tabla de rangos estandarizados para el número de medias en el rango.
- 3) Se comparan las medias de tal forma que si la diferencia de dos medias es mayor que el valor de DSH, entonces se rechaza $H_0: \tau_i = \tau_{i'}$, en caso contrario no se rechaza $H_0: \tau_i = \tau_{i'}$, esto es:



Rechazar $H_0: \mu = \mu'$

Si $|Y_i - Y_i'| > DHS$

- 4) Se presentan los resultados en forma tabular, identificando con la misma letra las medias estadísticamente iguales, o sea, las que no presentan diferencias entre ellas.

Nota:

Cuando el experimento conste de 8 o menos tratamientos, se puede usar DMS ó DUNCAN, pero cuando se tengan arriba de 10 ó 12 tratamientos, es más recomendable usar Tukey

Ejemplo: Retomando el ejemplo del DBA:

	T1	T2	T3	T4	Y.j
I	4.1	6.5	4.1	4.9	19.6
II	4.1	5.3	4.0	6.2	19.6
III	6.5	6.9	4.5	4.8	22.7
IV	4.3	6.8	4.3	4.2	19.6
V	6.0	6.5	4.1	6.9	23.5
Yi.	25	32	21	27	105
\bar{y}	5.0	6.4	4.2	5.4	

F.V.	G.L.	S.C.	C.M.	F
Tratam.	3	12.550	4.1830	6.015**
Bloques	4	3.755	0.9387	1.35ns
Error	12	8.345	0.6954	
Total	19	24.650		

Trat. Media

T2= 6.4 a

T4= 5.4 a b

T1= 5.0 a b

T3= 4.2 b

Los datos extraídos de la tabla del análisis de varianza son:

$$CMError = 0.6954$$

$$G.L.Error = 12$$

$$r = 5$$

$$S_y = \sqrt{\frac{CMError}{r}} = \sqrt{\frac{0.6954}{5}} = 0.3729$$

$$Q(\alpha, p, GL Error) = Q(0.05, 5, 12) = 4.20$$

$$\alpha = 0.05$$

$$DHS = S_{\bar{y}} Q(\alpha, t, GL Error) =$$

$$(0.3729)(4.20) = 1.56$$



Método de Scheffé

Es un método más general, se basa en la distribución F; sin embargo, se considera más estricta que las demás, por lo que en la práctica es frecuente usarla al 10% de significancia.

Es más exacto que las comparaciones múltiples basadas en el rango estandarizado, ya que se puede utilizar para comparar medias provenientes de tratamientos con igual o desigual tamaño ($n_i \neq n_{i'}$) y para medias que han sido ajustadas por covarianza.

También puede emplearse para probar la hipótesis general de que una función lineal de las medias poblacionales es igual a cero.

El juego de hipótesis a probar es:

$$H_0: \tau_i = \tau_{i'} \quad \text{V.S.} \quad H_a: \tau_i \neq \tau_{i'} \text{ con } i \neq i'$$

Mediante el estimador:

Si las repeticiones de los tratamientos son diferentes:

$$SCHEFFE = \sqrt{(t-1)(F_{\alpha, t-1, GLError}) \left[CMError \left(\frac{1}{n_i} + \frac{1}{n_{i'}} \right) \right]}$$

Si las repeticiones de los tratamientos son iguales:

$$SCHEFFE = \sqrt{(t-1)(F_{\alpha, t-1, GLError}) \left[\frac{2CMError}{r} \right]}$$

Donde:

Scheffé = Estimador

n_i = Número de repeticiones del tratamiento i

$n_{i'}$ = Número de repeticiones del tratamiento i' ,

t = Número de tratamientos



El procedimiento es:

- 1) Se ordenan las medias de los tratamientos
- 2) Se calcula el estimador Scheffé, con los valores de Q en la tabla de la distribución F, con el CMError, el número de tratamientos y el nivel de significancia α .
- 3) Se comparan las medias de tal forma que si la diferencia de dos medias es mayor que el valor de Scheffé, entonces se rechaza H_0 : $\mu = \mu'$, en caso contrario no se rechaza H_0 : $\mu = \mu'$, esto es:

Rechazar H_0 : $\mu = \mu'$

Si $|Y_i - Y_{i'}| > SCHEFFE$

- 4) Se presentan los resultados en forma tabular, identificando con la misma letra las medias estadísticamente iguales, o sea, las que no presentan diferencias significativas entre ellas.

Ejemplo: Retomando el ejemplo del DBA:

	T1	T2	T3	T4	Y.j
I	4.1	6.5	4.1	4.9	19.6
II	4.1	5.3	4.0	6.2	19.6
III	6.5	6.9	4.5	4.8	22.7
IV	4.3	6.8	4.3	4.2	19.6
V	6.0	6.5	4.1	6.9	23.5
Yi.	25	32	21	27	105
\bar{y}	5.0	6.4	4.2	5.4	

F.V.	G.L.	S.C.	C.M.	F
Tratam.	3	12.550	4.1830	6.015**
Bloques	4	3.755	0.9387	1.35ns
Error	12	8.345	0.6954	
Total	19	24.650		

Trat. Media

T2= 6.4 a

T4= 5.4 a b

T1= 5.0 a b

T3= 4.2 b

Los datos extraídos de la tabla del análisis de varianza son:

CMError = 0.6954

G.L.Error = 12

r = 5

α = 0.05

$$SCHEFFE = \sqrt{(t-1) \left(F_{\alpha, t-1, GLErr} \right) \left[\frac{2CMError}{r} \right]} = \sqrt{(4-1) (3.49) \left[\frac{2(0.6954)}{5} \right]} = 1.70$$



Método de Dunnett

Este método es recomendable cuando el interés en un experimento es comparar un grupo de tratamientos con un testigo o tratamiento control.

Con este método se prueba el juego de hipótesis:

$$H_0: \pi_i = \pi_1 \quad \text{V.S.} \quad H_a: \pi_i \neq \pi_1 \text{ con } i = 1, 2, 3, \dots, t$$

Mediante el estimador: $F_D = S_{(Y_1 - Y_i)} t_D$

En donde:

F_D = Estimador, Factor Dunnet

$$t_D = t_{(\alpha, p, GL_{Error})}$$

$$S_{y_t - y_i} = \sqrt{\frac{2(CM_{Error})}{r}}$$

$$P = t - 1$$

$t_{(\alpha, p, GL_{Error})}$ = Valor en la tabla “t” de Dunnet

t = Número de tratamientos

El procedimiento es:

- 1) Se calcula el estimador F_D , con lo valores de t_D en la tabla de “t” de Dunnett.
- 2) Se comparan las medias de tal forma que si la diferencia de la media del testigo y cualquier otra media es mayor que el valor de F_D , entonces se rechaza $H_0: \pi = \pi_1$, en caso contrario no se rechaza $H_0: \pi = \pi_1$, esto es:

$$\text{Rechazar } H_0: \pi = \pi_1 \quad \text{Si } |Y_t - Y_i| > F_D$$

- 3) Se presentan los resultados en forma tabular, identificando con la misma letra las medias estadísticamente iguales al testigo.



Consideraciones Sobre Las Pruebas

Todas las pruebas son iguales cuando se comparan únicamente dos tratamientos en el experimento; a medida que el número de tratamientos aumenta, las diferencias en los resultados de las pruebas son mayores.

Con dos tratamientos, el método de la DMS prueba la hipótesis de igualdad de medias con un error tipo I mayor al que se declara ($\approx 13\%$).

Un orden, en las pruebas, en cuanto a la capacidad de declarar diferencias significativas es:

1. DMS
2. DUNCAN
3. S-N-K
4. DUNNETT
5. TUKEY
6. SCHEFFÉ

Referencias

Box, G., Hunter, W.G. and Hunter, J.S. Estadística para Investigadores. Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos. Editorial Reverté. 2002.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Cox, D.R. Planning of Experiments. Wiley, New York. 1992.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C.: "Análisis Multivariante", Ed. Prentice Hall, 5ª edición. 1999,



Hinkelmann, K. Design and analysis of experiments. John Wiley & Sons, N.Y. (2005).

Jobson, J.D. Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design. Springer-Verlag, New York. 1991.

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lara, A.M. Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS. Proyecto Sur de Ediciones, Granada. 2001.

Lindman, H.R. Analysis of Variance in Experimental Design. Springer-Verlag, N.Y. 1992.

Mason, R.L., Gunst, R.F. and Hess, J.L. Statistical Design and Analysis of Experiments. Wiley, N.Y. 2003.

Montgomery, D.C. Diseño y Análisis de Experimentos. Limusa, México. 2003.

Peña Sánchez de Rivera, D. Estadística, Modelos y Métodos. Volumen II: Modelos Lineales y Series Temporales. Alianza Universidad, Madrid. 2000.

Peña, D., Regresión y Diseño de Experimentos., Alianza Editorial. 2002.

Rencher, A.C. Multivariate Statistical Inference and Applications. 1998.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Wiley. N.Y. Toutenburg, H. Statistical Analysis of Designed Experiments. 2a edición. Springer-Verlag, New York. 2002.



Parte II: Modelos Clásicos



Capítulo VIII: Comparaciones Ortogonales

En el capítulo anterior se observan los diferentes tipos de tratamientos. Correspondiendo las comparaciones de medias a los tratamientos cualitativos sin estructura; sin embargo en algunos casos los tratamientos poseen estructura, es decir, pueden ser agrupados con sentido a alguna(s) característica(s), por lo que el análisis correspondiente es a través de comparaciones independientes o contrastes ortogonales, utilizándose en el primer caso el estimador DMS.

Estos métodos son aplicables cuando el investigador planea con anterioridad las comparaciones que le van a dar las respuestas a PREGUNTAS ESPECÍFICAS.

El número de comparaciones independientes o contrastes ortogonales es igual a $t-1$ (G.L. de tratamientos); pudiéndose planear varios juegos de comparaciones o contrastes.



Comparaciones Independientes

También se les conoce como comparaciones planeadas u ortogonales; se diseñan según las necesidades de cada investigador.

Las comparaciones son independientes cuando cada media de tratamiento aparece únicamente en una comparación, aunque puede aparecer varias veces en un conjunto como miembro de dicho conjunto.

En el caso de tratamientos cualitativos sin estructura el método más sensible para detectar diferencias, es el que se realiza a través del estimador DMS; el cual consiste en calcular un valor de DMS, para cada par de medias de tratamientos en comparación; por lo que debido a estas características del estimador también es posible utilizarlo para tratamientos cualitativos con estructura.

El juego de hipótesis a probar es:

$$H_0: \mu_i = 0 \quad \text{V.S.} \quad H_a: \mu_i \neq 0$$

El estimador para cada comparación se calcula por medio:

$$DMS = t_{\alpha} \sqrt{\frac{CMError}{n_1} + \frac{CMError}{n_2}}$$

En donde:

t_{α} = Valor tabular de “t” de Student, con G.L. del Error y α

n_1 = No. de observaciones del primer componente en la comparación.

n_2 = No. de observaciones del segundo componente en la comparación.

El procedimiento de prueba es:

- 1) Se formulan las hipótesis que permitan dar respuesta a preguntas iniciales.
- 2) Se planean las comparaciones en base a las hipótesis formuladas que permitan dar respuesta a preguntas iniciales.
- 3) Se calcula la DMS para cada comparación, con los valores de la distribución de “t” de Student.



- 4) Se analizan las diferencias que resultan de las comparaciones independientes con los valores de la DMS respectiva; si este valor es mayor que la DMS, entonces se rechaza la $H_0: C_i=0$, y se dice que los tratamientos comparados difieren estadísticamente en caso contrario no se rechaza.

Estas comparaciones trabajan con las medias de los tratamientos y antes que nada hay que plantear las hipótesis que darán origen a las Comparaciones Independientes. Y de las ecuaciones planteadas en las hipótesis se despejan igualando a cero, dándonos la comparación a evaluar

Ejemplo 1: Si se tienen 5 tratamientos, A, B, C, D, E, y un tratamiento testigo; entonces algunas comparaciones que el investigador puede hacer, son:

$H_{01}: T_a + T_b + T_c + T_d = 4T_t$	VS	$H_{a1}: T_a + T_b + T_c + T_d \neq 4T_t$
$H_{02}: T_a + T_b + T_c = 3T_d$	VS	$H_{a2}: T_a + T_b + T_c \neq 3T_d$
$H_{03}: T_a + T_b = 2T_c$	VS	$H_{a3}: T_a + T_b \neq 2T_c$
$H_{03}: T_a = T_b$	VS	$H_{a3}: T_a \neq T_b$

Despejando tenemos las siguientes comparaciones:

$$c_1: \frac{T_a + T_b + T_c + T_d}{4} - T_t$$

$$c_2: \frac{T_a + T_b + T_c}{3} - T_d$$

$$c_3: \frac{T_a + T_b}{2} - T_c$$

$$c_4: T_a - T_b$$



Ejemplo 2: Debido a los altos niveles de humedad en la zona, se estudia a las Dermatomicosis causadas por Scytalidium dimidiatum y S. hyalinum ya que estos hongos son agentes etiológicos de infecciones crónicas cutáneas que afectan a la piel altamente queratinizada de palmas, plantas, uñas y espacios interdigitales. Considerando que clínicamente se parecen a tinea unguium, tinea pedis y tinea manum producidas por hongos dermatofitos, es importante establecer el diagnóstico etiológico ya que el tratamiento es diferente según se trate de una tiña o de una dermatomicosis por Scytalidium.

Por lo anterior se desarrollo un experimento con 7 antimicóticos para determinar la eficacia de estos en el tratamiento de la dermatomicosis causadas por Scytalidium dimidiatum en uñas de los pies. Se trabajaron con grupos de 30 personas infectadas a las cuales se les dio tratamiento tópico durante 25 días al término de los cuales se observo el número de pacientes que se curaron. Los tratamientos aplicados fueron los siguientes:

T₁ y T₂: Antimicóticos con base de Mercurio.
T₃ y T₄: Antimicóticos no Mercúricos (Anilidas)
T₅, T₆ y T₇: Antimicóticos no Mercúricos (Imidazoles)
T₈: Testigo (sin Antimicóticos)
T₆ y T₇: Nuevas formulas de T₅

	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	Y _j
I	16	24	28	19	16	27	30	8	168
II	19	28	18	16	25	16	23	10	155
III	16	20	23	16	22	17	25	18	157
IV	19	21	20	17	22	18	24	19	160
Y _i .	70	93	89	68	85	78	102	55	
\bar{Y} .	17.5	23.25	22.5	17	21.25	19.5	25.5	13.75	



El investigador se formula varias preguntas, dentro de ellas tenemos:

1. ¿Son efectivos los antimicóticos empleados?
2. ¿Fueron más efectivos los antimicóticos mercúricos de los no mercúricos?
3. ¿Funcionan igual los antimicóticos mercúricos?
4. ¿Los antimicóticos anilidas protegen mejor que los imidazoles?
5. ¿Los antimicóticos de las anilidas funcionan igual?
6. ¿Funcionan igual la vieja y las nuevas formulas de los antimicóticos de los imidazoles?
7. ¿Funcionan igual las nuevas formulaciones del T5?

Para cada una de las preguntas se desarrolla la comparación de medias de tratamientos correspondientes:

$$c_1: \frac{\bar{T}_1 + \bar{T}_2 + \bar{T}_3 + \bar{T}_4 + \bar{T}_5 + \bar{T}_6 + \bar{T}_7}{7} - \bar{T}_8$$

$$c_2: \frac{\bar{T}_1 + \bar{T}_2}{2} - \frac{\bar{T}_3 + \bar{T}_4 + \bar{T}_5 + \bar{T}_6 + \bar{T}_7}{5}$$

$$c_3: \bar{T}_1 - \bar{T}_2$$

$$c_4: \frac{\bar{T}_3 + \bar{T}_4}{2} - \frac{\bar{T}_5 + \bar{T}_6 + \bar{T}_7}{3}$$

$$c_5: \bar{T}_3 - \bar{T}_4$$

$$c_6: \bar{T}_5 - \frac{\bar{T}_6 + \bar{T}_7}{2}$$

$$c_7: \bar{T}_6 - \bar{T}_7$$



Antes que nada se resuelva mediante el diseño simple seleccionado, siendo el diseño empleado el DCA. Y la hipótesis a probar es:

Ho: $T_1=T_2=T_3=T_4=T_5=T_6=T_7=T_8$ VS Ha: al menos un T_i es diferente

$$\begin{aligned} \text{S.C. Total} &= \sum_{i=1}^t \sum_{j=1}^{n_i} Y_{ij}^2 - (Y_{..})^2/n. \\ &13560 - (640)^2/32 = 13560 - 12800 = 760 \end{aligned}$$

$$\begin{aligned} \text{S.C. Tratamientos} &= \sum_{i=1}^t (Y_{i.})^2/n_i - (Y_{..})^2/n. \\ &[(70)^2 + \dots + (55)^2]/4 - 12800 = 13208 - 12800 = 408 \end{aligned}$$

$$\begin{aligned} \text{S.C. Error} &= \text{S.C. Total} - \text{S.C. Tratamientos} \\ &760 - 408 = 352 \end{aligned}$$

FV	GL	SC	CM	Fc	Ft0.05	Ft0.01
Tratam.	7	408	58.286	3.974**	2.43	3.70
Error	24	352	14.66			
Total	31	760				

Se rechaza Ho. Existen diferencias altamente significativas entre los tratamientos.

Posterior al ANVA, se tienen que resolver las preguntas planteadas.

1.- Son efectivos los antimicóticos empleados.

$$\begin{aligned} c_1: & \frac{\bar{T}_1 + \bar{T}_2 + \bar{T}_3 + \bar{T}_4 + \bar{T}_5 + \bar{T}_6 + \bar{T}_7}{7} - \bar{T}_8 \\ c_1: & \frac{17.5 + 23.25 + 22.25 + 17 + 21.25 + 19.5 + 25.5}{7} - 13.75 \end{aligned}$$

$$c_1: \frac{146.25}{7} - 13.75 = 20.893 - 13.75 = 7.142$$

$$DMS = t_{\alpha} \sqrt{\frac{CMError}{n_1} + \frac{CMError}{n_2}}$$

$$t = 2.797$$



$$DMS = 2.797 \sqrt{14.667 \left(\frac{1}{28} + \frac{1}{4} \right)} = 5.73$$

$$C1 > DMS \quad \text{-----} > \quad 7.142 > 5.73$$

Como la diferencia entre los tratamientos es mayor que la DMS, los tratamientos comparados $((T_1+T_2+T_3+T_4+T_5+T_6+T_7)/7 \text{ VS } T_8)$ son diferentes entre si. Lo cual indica que los antimicóticos mercúricos y no mercúricos. Son efectivos en el tratamiento de la infección por hongos en las uñas de los pies.

Hasta aquí solo se ha resuelto la primera pregunta, teniéndose que resolver las 6 restantes. Sin embargo debido a que el procedimiento es exactamente el mismo omitimos la resolución de los restantes; viéndose la solución completa más adelante.



Contrastes Ortogonales

Son prueba de comparación de tratamientos que el investigador debe saber antes de iniciar su experimento; debiendo conocer que comparaciones de tratamientos son los que le van a dar la información deseada.

El método consiste en descomponer los grados de libertad y la suma de cuadrados para tratamientos en comparaciones componentes; pudiendo consistir en comparaciones de clase o de tendencia de respuesta.

Existen $t-1$ contrastes ortogonales; si son bien seleccionados pueden responder tantas preguntas como grados de libertad de tratamientos existan, conduciendo a afirmaciones de probabilidad bien definidas.

Es un método poco usado por los investigadores, debido a que en muchas de las ocasiones se piensa que es muy difícil y complicado, pero en realidad tiene ventajas, como son:

- 1) Permite responder importantes preguntas específicas sobre los efectos de los tratamientos.
- 2) Los cálculos son simples.
- 3) Suministra un útil control sobre la suma de cuadrados de tratamientos.

A diferencia de las Comparaciones Independientes que utilizan a la media de los tratamientos, en los Contrastes Ortogonales se trabaja con los totales de los tratamientos

Con este método se prueban hipótesis de la forma:

$$\text{I) } H_{01}: 3t_1 = t_2 + t_3 + t_4 \quad \text{VS} \quad H_{a1}: 3t_1 \neq t_2 + t_3 + t_4$$

Comparación del efecto del tratamiento 1 contra los totales de los efectos de los tratamientos 2, 3, y 4.

$$\text{II) } H_{02}: t_1 + t_2 = t_3 + t_4 \quad \text{VS} \quad H_{a2}: t_1 + t_2 \neq t_3 + t_4$$

Comparación del efecto conjunto de los tratamientos 1 y 2 contra el efecto conjunto de los tratamientos 3 y 4.

$$\text{III) } H_{03}: t_1 + t_2 = 2t_3 \quad \text{VS} \quad H_{a3}: t_1 + t_2 \neq 2t_3$$

Comparación del efecto conjunto de los tratamientos 1 y 2 contra el efecto del tratamiento 3.



Los contrastes se plantean a partir de las hipótesis formuladas, de las que se despeja la ecuación, igualándola a cero pero conservando cada tratamiento su coeficiente

Hipótesis		Contraste
$H_{01}: 3t_1 = t_2 + t_3 + t_4$	→	$C_1: 3t_1 - t_2 - t_3 - t_4 = 0$
$H_{02}: t_1 + t_2 = t_3 + t_4$	→	$C_2: t_1 + t_2 - t_3 - t_4 = 0$
$H_{03}: t_1 + t_2 = 2t_3$	→	$C_3: t_1 + t_2 - 2t_3 = 0$

Una vez planteados los contrastes, se procede a la construcción del cuadro de los coeficientes de los contrastes (ver cuadro 17), los cuales la sumatoria deberá ser igual a cero; en caso contrario el contraste y por consecuencia la hipótesis fue mal planteada.

Cuadro 17.- Coeficientes de los contrastes

	T1	T2	T3	T4	
C1	3	-1	-1	-1	= 0
C2	1	1	-1	-1	= 0
C3	1	1	-2	0	= 0

Como se puede observar en la tabla de los coeficientes las sumatorias son iguales a cero, por lo que los contrastes y las hipótesis están bien planteados; sin embargo aún nos falta evaluar la ortogonalidad de los contrastes por lo que se procede a dicha evaluación

Así que se considera que dos o más comparaciones se llaman ortogonales, si la suma del producto de sus coeficientes correspondientes es igual a cero (ver cuadro 18).

Cuadro 18.- Producto de los coeficientes

C1y C2	3*1	-1*1	-1*-1	-1*-1	= 4
C1y C3	3*1	-1*1	-1*-2	-1*0	= 4
C2 y C3	1*1	1*1	-1*-2	-1*0	= 4

Como se puede observar en este ejemplo aunque están bien planteadas las hipótesis no son ortogonales. Por lo que no procede el análisis.



En general con este método se prueban juegos de hipótesis de la forma:

$$\mathbf{H_0: \sum_{i=1}^t \lambda_i \tau_i = 0 \quad VS \quad H_a: \sum_{i=1}^t \lambda_i \tau_i \neq 0}$$

A la combinación lineal de tratamientos $\lambda_i \tau_i$ se le llama contraste, en donde los λ_i son los coeficientes de los efectos de los tratamientos, en el contraste.

Esta prueba se realiza conforme a las siguientes reglas:

- 1) Cualquier función de la forma $\lambda_1 \tau_1 + \lambda_2 \tau_2 + \lambda_3 \tau_3 + \dots + \lambda_t \tau_t = 0$ se llama comparación o contraste si la suma de sus coeficientes es cero, esto es:

$$\sum_{i=1}^t \lambda_i = 0$$

- 2) Si C es un contraste entre los posibles, entonces la cantidad:

$$SC_{Ci} = \frac{C_j^2}{r \sum_{i=1}^t \lambda_{ij}^2}$$

Es una componente de la suma de cuadrados de tratamientos y representa un contraste con un grado de libertad. Además:

$$C_j = \sum_{i=1}^t \lambda_{ij} \tau_i = 0$$

Con: λ_{ij} Coeficiente del tratamiento i en el contraste j
 τ_i Total del tratamiento i.

- 3) Dos o más comparaciones se llaman ortogonales, si la suma del producto de sus coeficientes correspondientes es igual a cero; esto es:

$$\sum_{i=1}^t \lambda_{i1} \lambda_{i2} = 0$$

Los contrastes son:

$$C_1 = \lambda_{11} \tau_1 + \lambda_{12} \tau_2 + \lambda_{13} \tau_3 + \dots + \lambda_{1t} \tau_t$$

$$C_2 = \lambda_{21} \tau_1 + \lambda_{22} \tau_2 + \lambda_{23} \tau_3 + \dots + \lambda_{2t} \tau_t$$

Al comprobar la ortogonalidad, algunos coeficientes de los tratamientos no intervienen en el contraste, dichos coeficientes se toman como cero.

- 4) Si en t tratamientos, t-1 contrastes son ortogonales, entonces la suma de cuadrados de todos los t-1 contrastes es igual a la suma de cuadrados de los tratamientos; esto es:



$$SC_{TRAT} = \frac{C_1^2}{r \sum_{i=1}^t \lambda_{i1}^2} + \frac{C_2^2}{r \sum_{i=1}^t \lambda_{i2}^2} + \frac{C_3^2}{r \sum_{i=1}^t \lambda_{i3}^2} + \dots + \frac{C_{(t-1)}^2}{r \sum_{i=1}^t \lambda_{i(t-1)}^2}$$

Después de haber obtenido la suma de cuadrados para cada contraste se procede a construir el ANVA descompuesta la suma de cuadrados de tratamientos y se prueban las hipótesis planteadas con la distribución F.

Ejemplo: Se retomará el ejemplo 2 presentado en el tema “Comparación de Medias”

1. ¿Son efectivos los antimicóticos empleados?
2. ¿Fueron más efectivos los antimicóticos mercurícos de los no mercurícos?
3. ¿Funcionan igual los antimicóticos mercurícos?
4. ¿Los antimicóticos anilidas protegen mejor que los imidazoles?
5. ¿Los antimicóticos de las anilidas funcionan igual?
6. ¿Funcionan igual la vieja y las nuevas formulas de los antimicóticos de los imidazoles?
7. ¿Funcionan igual las nuevas formulaciones del T5?

Hipótesis planteadas:

$$H_{01}: T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_7 = 7T_8$$

$$H_{02}: 5(T_1 + T_2) = 2(T_3 + T_4 + T_5 + T_6 + T_7)$$

$$H_{03}: T_1 = T_2$$

$$H_{04}: 3(T_3 + T_4) = 2(T_5 + T_6 + T_7)$$

$$H_{05}: T_3 = T_4$$

$$H_{06}: 2T_5 = T_6 + T_7$$

$$H_{07}: T_6 = T_7$$

Contrastes de las Hipótesis:

$$C_1: T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_7 - 7T_8 = 0$$

$$C_2: 5T_1 + 5T_2 - 2T_3 - 2T_4 - 2T_5 - 2T_6 - 2T_7 = 0$$

$$C_3: T_1 - T_2 = 0$$

$$C_4: 3T_3 + 3T_4 - 2T_5 - 2T_6 - 2T_7 = 0$$

$$C_5: T_3 - T_4 = 0$$

$$C_6: 2T_5 - T_6 - T_7 = 0$$

$$C_7: T_6 - T_7 = 0$$



Tabla de los coeficientes de los contrastes

	T1	T2	T3	T4	T5	T6	T7	T8
C1	1	1	1	1	1	1	1	-7
C2	5	5	-2	-2	-2	-2	-2	0
C3	1	-1	0	0	0	0	0	0
C4	0	0	3	3	-2	-2	-2	0
C5	0	0	1	-1	0	0	0	0
C6	0	0	0	0	2	-1	-1	0
C7	0	0	0	0	0	1	-1	0

Si los contrastes pasan la prueba de ortogonalidad, se procede a calcular la Suma de Cuadrados (SC) de los contrastes mediante la formula:

$$SC_{Ci} = \frac{C_j^2}{r \sum_{i=1}^t \lambda_{ij}^2}$$

$$SC_{C1} = \frac{(T1 + T2 + T3 + T4 + T5 + T6 + T7 - 7T8)^2}{4 \sum_1^8 (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (-7)^2}$$

$$SC_{C1} = \frac{(70 + 93 + 89 + 68 + 85 + 78 + 102 - 7(55))^2}{(4)(1 + 1 + 1 + 1 + 1 + 1 + 1 + 49)} = \frac{40000}{224} = 178.57$$

$$SC_{C2} = \frac{(5(70) + 5(93) - 2(89) - 2(68) - 2(85) - 2(78) - 2(102))^2}{(4)(25 + 25 + 4 + 4 + 4 + 4 + 4)} = \frac{841}{280} = 3.004$$

$$SC_{C3} = \frac{(70 - 93)^2}{(4)(1 + 1)} = \frac{529}{8} = 66.125$$

$$SC_{C4} = \frac{(3(89) + 3(68) - 2(85) - 2(78) - 2(102))^2}{(4)(9 + 9 + 4 + 4 + 4)} = \frac{3481}{120} = 29.008$$

$$SC_{C5} = \frac{(89 - 68)^2}{(4)(1 + 1)} = \frac{441}{8} = 55.125$$



$$SC_{C6} = \frac{(2(85) - 78 - 102)^2}{(4)(4 + 1 + 1)} = \frac{100}{24} = 4.166$$

$$SC_{C7} = \frac{(78 - 102)^2}{(4)(1 + 1)} = \frac{576}{8} = 72$$

$$SC_{\text{trat}} = \sum SC_{Cj} = 178.571 + 3.004 + 66.125 + 29.008 + 55.125 + 4.166 + 72 = 408$$

$$S.C. \text{ Total} = \sum_{i=1}^t \sum_{j=1}^{n_i} Y_{2ij}^2 - (Y_{..})^2/n.$$

$$S.C. \text{ Error} = S.C. \text{ Total} - S.C. \text{ Tratamientos}$$

Una vez calculados las SC de los contrastes se procede a realizar el ANVA.

F.V.	G.L.	S.C.	C.M.	Fc	Fto.o5	Fto.o1
Trat.	t-1=7	408	58.286	3.97 ^{n.s}	4.35	8.01
C1	1	178.571	178.571	12.175 ^{**}	4.35	8.01
C2	1	3.004	3.004	0.205 ^{n.s}	4.35	8.01
C3	1	66.125	66.125	4.508 [*]	4.35	8.01
C4	1	29.008	29.008	1.978 ^{n.s.}	4.35	8.01
C5	1	55.125	55.125	3.758 ^{n.s.}	4.35	8.01
C6	1	4.166	4.166	0.284 ^{n.s.}	4.35	8.01
C7	1	72	72	4.909 [*]	4.35	8.01
Error	t(r-1)=24	352	14.667			
Total	tr-1=31					

** Diferencias altamente significativas

* Diferencias significativas

n.s. No Significancia

C F Respuesta a las preguntas

C1 12.175^{**} Si demuestran efectividad

C2 0.205^{n.s} Son igualmente efectivos

C3 4.508^{*} Uno de los antimicóticos mercúricos funciona mejor

C4 1.978^{n.s.} Protegen igual

C5 3.758^{n.s.} Protegen igual

C6 0.284^{n.s.} Protegen igual

C7 4.909^{*} Uno de los antimicóticos de las nuevas formulaciones de T5 funciona mejor



Problemario de Comparaciones Ortogonales

31. Se realizó un experimento para combatir al agente etiológico de la enfermedad de Chagas o Tripanosomiasis, el Trypanosoma cruzi; para el estudio se utilizaron dos grupos de medicamentos para este tipo de parásito, se usaron nifurtimox (T1: 15 mg/kg/día y T2: 20 mg/kg/día), benznidazol (T3: 3 mg/kg/día, T4: 6 mg/kg/día) utilizando como testigo a un placebo. Se definieron 4 grupos terapéuticos y un grupo testigo con un estudio doble ciego; durante 90 días se les suministró 3 veces al día las dosis de los tratamientos, trabajándose con grupos de 35 adultos enfermos de ambos sexos, y registrándose el número de pacientes que no presentaron mejoría.

T1	T2	T3	T4	T5
19	15	26	27	35
20	17	30	26	33
16	12	25	30	32
17	14	29	28	34

T1 y T2: Nifurtomox

T2 Y T4: Benznidazol

T5: Placebo

Solo se pueden formular 4 preguntas debido a la restricción de t-1 contrastes. Formule y evalúe, resuelva y conteste todas las preguntas que se pueden hacer. Sin olvidar los análisis primarios.

32. Se desea saber en que caldo crecen más las 4 especies probables de Salmonella a encontrar en heces. Para esto se tienen 4 cepas del Genero Salmonella. Una vez inoculadas en el caldo de tetratationato, se les adiciona el inhidor de lugol en el porcentaje requerido y se espera 12 horas a 37°C y se siembran en el medio de cultivo de Sulfito de bismuto, y se deja crecer 24 horas a 37°C y se leen el número de colonias de salmonella que son capaces de crecer.

T1	T2	T3	T4	
8	11	14	15	S. typhi
10	13	8	11	S. cholerasuis
12	17	16	10	S. enteritidis
9	5	10	10	S. paratyphi

T1: Caldo de Tetratationato marca "A" Modificado

T2: Caldo de Tetratationato marca "B" Altamente selectivo.

T3 : Caldo de Tetratationato marca "B" Bajamente selectivo.

T4 : Con extracto de Levadura y Fosfato Disodico



Solo se pueden formular 3 preguntas debido a la restricción de t-1 contrastes. Formule y evalúe, resuelva y conteste todas las preguntas que se pueden hacer. Sin olvidar los análisis primarios.

33. Se realizó un experimento con personas con bajo nivel de HEMOGLOBINA probándose 4 tratamientos (T1: ácido fólico (500 µg, T2: sulfato ferroso a dosis alta (300 mg), T3: sulfato ferroso a dosis baja (15 mg), T4: sulfato ferroso (200 mg) adicionado con ac. Fólico (400µg)), esto con el fin de observar que tiempo tardan en aumentar los niveles de hemoglobina. Los resultados obtenidos fueron los niveles de hemoglobina de las personas participantes en el estudio:

T1	T2	T3	T4
18	12	16	15
14	16	13	13
16	14	12	14
13	18	15	12

Solo se pueden formular 3 preguntas debido a la restricción de t-1 contrastes. Formule y evalúe, resuelva y conteste todas las preguntas que se pueden hacer. Sin olvidar los análisis primarios.

34. Se realizó un estudio para determinar la concentración de coliformes que inciden sobre el río Huixtla. Para tal estudio se agruparon en base a la altura con 2 puntos de recolección situados antes de que el río entre a la ciudad y 2 puntos después de atravesar la ciudad, las cuales fueron seleccionadas al azar. Tomándose lecturas durante 5 días consecutivos, y los resultados se reportaron en UFC/100 ml. Para la identificación de coliformes totales se utilizó el medio de cultivo de Agar Endo.17 En la realización de pruebas presuntivas y confirmatorias se empleó el caldo lactosado y el caldo verde brillante, respectivamente. La temperatura de incubación empleada fue de $37 \text{ oC} \pm 0.5 \text{ oC}$ durante 24-30 horas.

DÍAS	I	II	III	IV
1	6	7	10	10
2	3	3	16	11
3	2	2	8	9
4	3	9	10	4
5	4	5	6	5

- T1: Punto 1 de recolección antes de entrar a la ciudad
T2: Punto 2 de recolección antes de entrar a la ciudad
T3 : Punto 3 de recolección después de atravesar la ciudad
T4 : Punto 4 de recolección después de atravesar la ciudad

Solo se pueden formular 3 preguntas debido a la restricción de t-1 contrastes. Formule y evalúe, resuelva y conteste todas las preguntas que se pueden hacer. Sin olvidar los análisis primarios.



35. Por las condiciones climáticas de la ciudad, la falta de asesoría médica y frecuentes hábitos de automedicación se hicieron 30 exudados faríngeos a alumnos del COBACH plantel 29 de Unión Juárez, Chiapas, México, con cuadro clínico de faringoamigdalitis; para la búsqueda intencionada de staphylococcus coagulasa positiva y para determinar la sensibilidad antimicrobiana obteniéndose los siguientes resultados en milímetros del halo de inhibición:

De los 30 cultivos 21 fueron positivos a staphylococcus coagulasa positivo. Y de los 12 antibióticos probados solo a 5 presentaron sensibilidad. Por lo que solo se toman a los pacientes con Staphylococcus coagulasa positivo y los antibióticos que causaron sensibilidad:

T1	T2	T3	T4	T5
CTX (mm)	PEF (mm)	TE (mm)	GE (mm)	CAZ (mm)
0	290	280	280	0
0	250	300	280	260
160	260	270	280	0
0	250	0	250	150
0	250	240	270	160
0	280	300	290	0
140	260	300	280	0
0	270	270	290	170
0	260	290	250	240
0	250	300	290	240
0	230	300	280	200
0	230	270	290	220
160	190	260	250	0
130	240	270	290	210
220	290	270	290	200
0	160	290	250	0
0	230	230	250	0
0	260	270	0	270
180	230	290	290	240
0	260	250	270	260
0	250	230	0	230

El interés es el de conocer a cual de los antibióticos presentan mayor susceptibilidad, en base al mecanismo de acción de éstos:

T1-CTX: (CEFALOSPORINA) INHIBE SÍNTESIS DE PARED CELULAR

T2-PEF: (FLOUROQUINOLONA) INHIBE SINTESIS DE ÁCIDO NUCLEICOS

T3-TE: (TETRACICLINA) INHIBE SINTESIS DE PROTEÍNAS

T4- GE: (AMINOGLUCOSIDOS) INHIBE SINTESIS DE PROTEÍNAS

T5: CAZ: (CEFALOSPORINA) INHIBE SINTESIS DE PARED CELULAR

Solo se pueden formular 4 preguntas debido a la restricción de t-1 contrastes. Formule y evalúe, resuelva y conteste todas las preguntas que se pueden hacer. Sin olvidar los análisis primarios.

Referencias

Arnau Gras, J.. Diseños Experimentales Multivariados.



Alianza Editorial, Madrid. 1990.

Box, G. E. y Draper, N., "Empirical Model Building and Response Surfaces". Ed. Jhon Wiley. 1986.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Cox, D.R. Planning of Experiments. Wiley, New York. 1992.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C.: "Análisis Multivariante", Ed. Prentice Hall, 5ª edición. 1999,

Hinkelmann, K. Design and analysis of experiments. John Wiley & Sons, N.Y. (2005).

Jobson, J.D. Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design. Springer-Verlag, New York. 1991.

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lara, A.M. Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS. Proyecto Sur de Ediciones, Granada. 2001.

Lindman, H.R. Analysis of Variance in Experimental Design. Springer-Verlag, N.Y. 1992.

Montgomery, D.C. Diseño y Análisis de Experimentos. Limusa, México. 2003.

Peña Sánchez de Rivera, D. Estadística, Modelos y Métodos. Volumen II: Modelos Lineales y Series Temporales. Alianza Universidad, Madrid. 2000.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.

Wiley. N.Y. Toutenburg, H. Statistical Analysis of Designed Experiments. 2a edición. Springer-Verlag, New York. 2002.



Parte III: Arreglos Factoriales



Capítulo IX: Principios Básicos de los Experimentos Factoriales

Lo visto hasta ahora corresponde al efecto de un factor sobre las unidades experimentales, sin embargo en muchas ocasiones, se examina el efecto simultáneo de varios factores sobre una característica en estudio, ya no pudiéndose ocupar los diseños simples, por lo que se tiene que abordar el problema mediante la experimentación factorial, en el cual se tiene que considerar el diseño de los tratamientos.

Para poder ejecutar una investigación de este tipo, existen varias opciones como es, el de ensayar los factores, haciendo variar uno de ellos, manteniendo fijo los otros; o también, haciendo variar cada factor, ensayando varios niveles de cada uno de ellos.

Con lo anterior se pueden deducir las posibles relaciones entre los factores en estudio y estimar su efecto, ya que, antes de iniciar un experimento se desconoce que factores son importantes y si ejercen su acción en forma independiente o combinada.

Por lo que el ensayo simultáneo de varios factores tiene la ventaja sobre la técnica de investigar un factor cada vez que se experimenta, el de medir la interacción, sobre el carácter en estudio de los factores considerados.

La interacción es un efecto en ocasiones positivo y otras veces negativo, sobre la suma de los efectos que aisladamente ejercen los factores experimentales. Sin embargo, en el sentido práctico se considera que una interacción es un efecto de menor importancia que los efectos principales.



Experimentos Factoriales

Se conocen como experimentos factoriales a aquellos en los que se estudian conjuntamente dos o más factores en un mismo tratamiento diferenciándose de los experimentos simples en los que solamente se estudia un factor por tratamiento.

Antes que nada es necesario saber que un factor es una clase de tratamientos, y en experimentación factorial, todo factor proporcionará varios tratamientos, a través de las combinaciones de los niveles de cada factor, por lo que se considera que: “Un experimento factorial es aquel en el que el conjunto de tratamientos consiste en todas las combinaciones posibles de los niveles de varios factores”.

El término factorial implica el concepto de un diseño de tratamientos, es decir, la selección de los factores a estudiar, sus niveles y la combinación de ellos, siendo independiente del Diseño Experimental.

En virtud de que en estos experimentos los tratamientos provienen de combinaciones de varios niveles de diferentes factores y se evalúan con un diseño simple, también se dicen que son arreglos experimentales, y no un diseño propiamente dicho.

Puede considerarse dos tipos fundamentales de experimentos factoriales:

- 1) El factorial incompleto: Ensaya solo algunos de las posibles combinaciones de tratamientos que pueden generarse.
- 2) El factorial completo: Ensaya todas las posibles combinaciones de tratamiento que se generan con los distintos niveles de los factores de estudio.

Dentro de los factoriales completos pueden distinguirse:

- Factoriales Simétricos: Comprenden el mismo número de niveles de cada factor,
- Factoriales Asimétricos; Ensayan números distintos de niveles de los factores en estudio.



Ventajas

La razón principal de estudiar dos o más factores en forma conjunta es ver como cambia la respuesta ante los niveles de un factor en presencia de los niveles del otro factor, es decir, ver si los factores interactúan entre si.

Otras ventajas, en el uso de este tipo de arreglos son:

- 1) Economía en el material experimental, al obtener información sobre varios factores que aumentan el tamaño del experimento.
- 2) Se amplía el rango de validez del experimento, es decir se amplía el rango de la inferencia en relación a un factor ya que se estudia en las diferentes condiciones representadas por los niveles de los otros factores.

Desventajas

Debido a que los experimentos factoriales requieren de un gran número de tratamientos, sobre todo cuando se estudian muchos factores con muchos niveles, se tienen las siguientes desventajas:

- 1) Al usar bloques, es difícil encontrar grupos de unidades experimentales homogéneas para asignar los tratamientos. Esto se elimina usando el concepto de confusión, que es reduciendo el tamaño del bloque pasando por alto la precisión en ciertas interacciones de orden superior.
- 2) Se aumenta el costo del experimento al tener muchas unidades experimentales. Se minimiza usando los experimentos factoriales fraccionados, donde se prueba sólo una parte de los tratamientos posibles.
- 3) Es difícil interpretar las interacciones de orden superior.

Los experimentos factoriales son útiles en investigaciones exploratorias, en los que poco se sabe acerca de niveles óptimos de los factores o cuáles son importantes

Por ejemplo: Para la siembra de un microorganismo, es conveniente estudiar de manera simultanea temperatura, pH, tiempo de crecimiento, nutrientes del medio de cultivo, etc.; en situaciones similares es conveniente incluir testigos en el experimento.



El Modelo

En los arreglos factoriales se combinan el diseño de tratamientos y el diseño simple del experimento, correspondiéndole un modelo matemático; por ejemplo: si el Diseño Experimental es un Diseño de Bloques al Azar, el modelo es:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \text{con} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

En el experimento factorial, los tratamientos se forman combinando niveles de los factores en estudio, de manera que el efecto de tratamientos τ_i se considera a su vez compuesto de los efectos de los factores y sus interacciones. Por lo que el modelo de los tratamientos, por ejemplo, para dos factores en estudio será:

$$\tau_i = A_k + B_l + (AB)_{kl}$$

en donde: $A_k =$ Efecto del nivel k del factor A.
 $B_l =$ Efecto del nivel l del Factor B.
 $(AB)_{kl} =$ Efecto de la interacción de los niveles k y l de los factores A y B, respectivamente.

Entonces el modelo para el arreglo factorial bajo el diseño de bloques al azar es:

$$Y_{klj} = \mu + A_k + B_l + (AB)_{kl} + \beta_j + \varepsilon_{klj}$$

con: $k = 1, 2, \dots, n_1$
 $l = 1, 2, \dots, n_2$
 $j = 1, 2, \dots, r$

También se puede escribir como:

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \beta_k + \varepsilon_{ijk}$$

con: $i = 1, 2, \dots, n_1$
 $j = 1, 2, \dots, n_2$
 $k = 1, 2, \dots, r$

Este es el modelo para el caso más simple de los arreglos factoriales, modelos más complicados se verán al estudiar casos en donde se tengan más de dos factores.

Es importante asociar a cada experimento factorial un modelo, ya que a partir del modelo es posible obtener los grados de libertad, las sumas de cuadrados y las esperanzas de los cuadrados medios, mediante reglas muy simples que se estudiarán posteriormente.



Notación

La mayoría de los autores coinciden en hacer la notación de los factores por letras mayúsculas; por ejemplo: los factores A, B, C, los niveles de los factores se denotan con letras minúsculas con subíndices; por ejemplo: el factor A tiene los niveles $a_0, a_1, a_2 \dots$, siendo común iniciar la numeración de los niveles en cero.

Los tratamientos se pueden denotar considerando la combinación de los niveles de los factores A y B con niveles, por ejemplo: a_0, a_1 y b_0, b_1 respectivamente, los tratamientos se denotan por $a_0b_0, a_0b_1, a_1b_0, a_1b_1$. Otra forma de denotar los tratamientos es considerar solo el subíndice del nivel; por ejemplo, los tratamientos anteriores se pueden denotar de la forma: 00, 01, 10, 11; en cada par, el primer número se refiere al factor A y el segundo al factor B y así sucesivamente.

Para denotar el tipo de experimento factorial en cuanto al número de factores y niveles de cada factor, se considera la siguiente notación:

La notación en el caso de los factoriales asimétricos, se hace indicando el producto de los niveles de cada factor, es decir si se tienen 3 factores: A con “p” niveles, B con “q” niveles y C con “r” niveles, entonces se dice que es un factorial $p \times q \times r$.

La notación en el caso de los factoriales simétricos, se hace mediante una expresión exponencial en el cual la base indica el número de niveles y el exponente el número de factores, es decir, si se tienen n factores a dos niveles cada uno, se dice que es un factorial 2^n , si se tienen 3 niveles de cada factor, es un factorial 3^n .

En el caso del factorial 2^n se tiene:

- n efectos principales.
- $N(n-1)/2!$ Interacciones de 2 factores (primer orden).
- $N(n-1)(n-2)/3!$ Interacciones de 3 factores (segundo orden).
- $\{N(n-1)\dots(n-r+1)\}!$ interacciones de r factores (r-1 orden)

Cuando abordemos el tema del factorial 2^2 se introducirá la notación complementaria para el estudio de los experimentos factoriales.



Representacion Grafica de los Diseños Factoriales

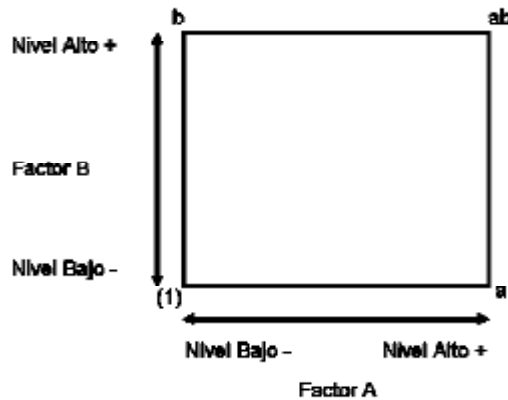


Figura 5.- Representación Gráfica del Diseño Factorial 2^2

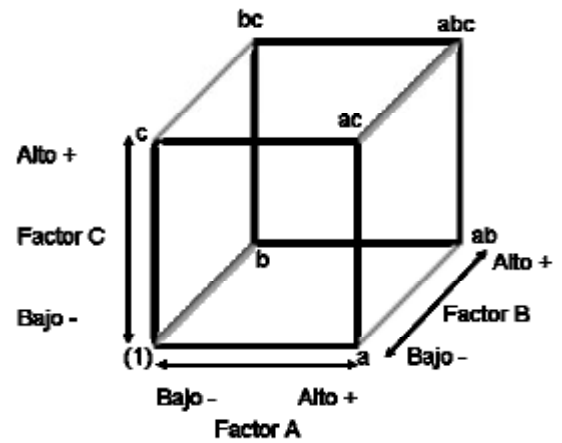


Figura 6.- Representación Gráfica del Diseño Factorial 2^3

Análisis Estadístico

El ANVA para un arreglo factorial se obtiene calculando las sumas de cuadrados de los efectos principales y de las interacciones, las que se pueden obtener por medio de dos procedimientos:

- A) Por contrastes ortogonales.
- B) Por medio del modelo estadístico correspondiente.

Análisis de las Interacciones

Si la interacción no es significativa (N.S.), entonces los factores son independientes.

Si hay interacción un factor tiene un determinado comportamiento frente a otro factor.

Por lo que, si hay interacción tenemos 2 formas de analizarlas:

1. Si la interacción es significativa, entonces existe evidencia experimental de que hay comportamientos diferentes de un factor en presencia de otro, por lo tanto se debe hacer un análisis de comportamiento de medias de los niveles de un factor dentro de los niveles del otro factor.



Quien tiene la Fc menor se calcula dentro de los niveles del que tiene la Fc mayor.

Consideremos un caso en el que los efectos promedios de los medios de cultivo (M.C.) son diferentes en función del microorganismo M.O.) a aislar. Se construye una tabla de medias de la siguiente forma:

		M.O.	
		0	1
M.C.	0	17.2	19.2
	1	17.2	25.4

Se calcula el estimador:

$$DMS = t_{(\alpha, gle)} \sqrt{\frac{2 CME}{r}} = t_{(0.01, 12)} \sqrt{\frac{2 CME}{r}} = (3.428)(0.3439) = 1.1791$$

Se procede a hacer las comparaciones, encontrándose en este caso que en presencia del nivel de M.O. alto la mayor respuesta está en el 2º M.C.

2. Cuando no hay diferencia significativa (N.S.) en la interacción, pero si en los efectos principales; se comparan las medias de los factores principales:

Se comparan las medias del factor (A)	Se comparan las medias del factor (B)
$EE = \sqrt{\frac{2CMErr}{ra}}$	$EE = \sqrt{\frac{2CMErr}{rb}}$

Donde “a” y “b” son el número de niveles de cada factor y EE el Error Estándar

La figura 7, muestra la manera en como se dan los efectos principales y las interacciones de los diseños factoriales con tres dimensiones

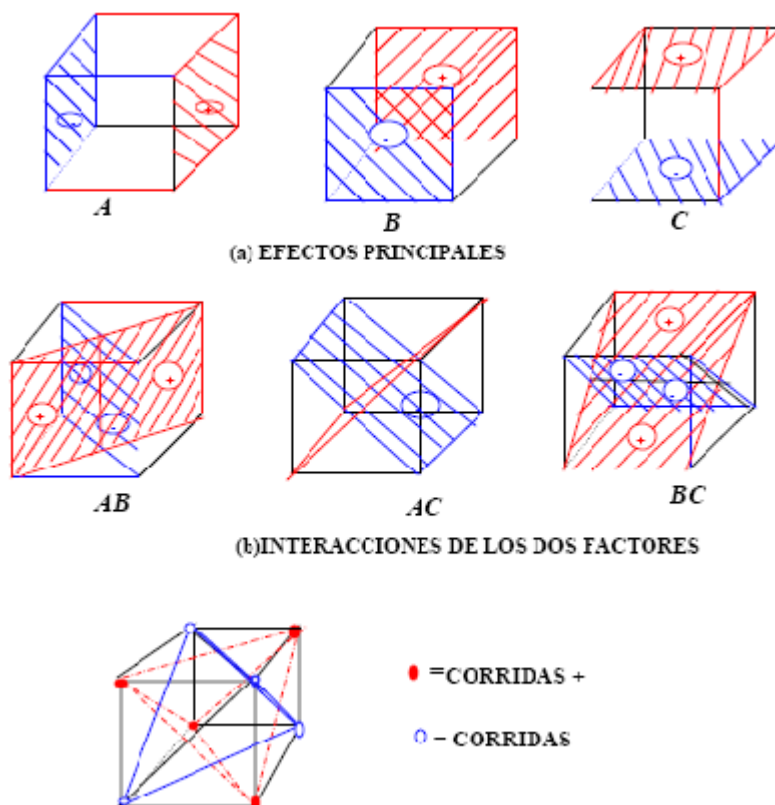


Figura 7.- Representación Gráfica de los efectos principales e interacciones en modelos factoriales de tres dimensiones

Referencias

Arnau Gras, J.. Diseños Experimentales Multivariables. Alianza Editorial, Madrid. 1990.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Cox, D.R. Planning of Experiments. Wiley, New York. 1992.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C.: "Análisis Multivariante", Ed. Prentice Hall, 5ª edición. 1999,



Hinkelmann, K. Design and analysis of experiments. John Wiley & Sons, N.Y. (2005).

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lindman, H.R. Analysis of Variance in Experimental Design. Springer-Verlag, N.Y. 1992.

Mason, R.L., Gunst, R.F. and Hess, J.L. Statistical Design and Analysis of Experiments. Wiley, N.Y. 2003.

Montgomery, D.C. Diseño y Análisis de Experimentos. Limusa, México. 2003.

Rencher, A.C. Multivariate Statistical Inference and Applications. 1998.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.



Parte III: Arreglos Factoriales



Capítulo X: Análisis de un Arreglo Factorial por Contrastes Ortogonales.

Todo arreglo factorial es evaluado bajo un diseño simple (DCA, DBA, DCL, etc.); por lo tanto, primeramente se obtiene el ANVA para el diseño simple utilizado, a través del Modelo Estadístico correspondiente, probando el juego de hipótesis para tratamientos.

Ho: El efecto de los tratamientos VS Ha: Al menos uno es diferente
es el mismos

Posteriormente, se aplica la metodología de análisis por contrastes ortogonales para conocer los efectos de los factores principales y de las interacciones que resulten; con el ANVA correspondiente se prueban las hipótesis para los contrastes planeados.

Arreglo Factorial 2 x 2 (2^2).

Consideremos un experimento factorial que incluya dos factores (A y B) con dos niveles (0 y 1), desarrollado bajo un Diseño de Bloques al Azar, este arreglo factorial es conocido como 2^2 , entonces las combinaciones y los tratamientos son:

Factores: A y B

niveles: 0 y 1

Diseño de los tratamientos

Factor A	Factor B	Tratamiento	
a0	b0	a0b0	T1
a0	b1	a0b1	T2
a1	b0	a1b0	T3
a1	b1	a1b1	T4

Tratamientos:

$$T_1 = a_0b_0$$

$$T_2 = a_0b_1$$

$$T_3 = a_1b_0$$

$$T_4 = a_1b_1$$



Las hipótesis a probar para el diseño simple son:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 \quad \text{vs} \quad H_a : \text{al menos una } \tau_j \neq \tau_k$$

En el caso de este ejemplo, las hipótesis se prueban con el ANVA para el Diseño de Bloques al Azar, cuyo modelo estadístico, las sumas de cuadrados para bloques, tratamientos, error y el total son exactamente los mismos del DBA.

En un experimento factorial el interés es conocer si los factores principales a los niveles evaluados son diferentes o independientes

Por lo que las hipótesis a probar bajo estas condiciones son:

$$H_{01} : \text{no hay efecto del factor A} \quad \text{vs} \quad H_{a1} : \text{hay efecto del factor A}$$

$$H_{02} : \text{no hay efecto del factor B} \quad \text{vs} \quad H_{a2} : \text{hay efecto del factor B}$$

$$H_{03} : \text{no hay interacción AxB} \quad \text{vs} \quad H_{a3} : \text{hay interacción AxB}$$

(Se comparan los niveles bajos contra los niveles altos de cada factor)

Las hipótesis equivalentes a probar por contrastes ortogonales son:

$$H_{01} : \tau_1 + \tau_2 = \tau_3 + \tau_4 \quad \text{vs} \quad H_{a1} : \tau_1 + \tau_2 \neq \tau_3 + \tau_4$$

$$H_{02} : \tau_1 + \tau_3 = \tau_2 + \tau_4 \quad \text{vs} \quad H_{a2} : \tau_1 + \tau_3 \neq \tau_2 + \tau_4$$

$$H_{03} : \tau_1 + \tau_4 = \tau_2 + \tau_3 \quad \text{vs} \quad H_{a3} : \tau_1 + \tau_4 \neq \tau_2 + \tau_3$$

Estas hipótesis representadas como contrastes de análisis y la tabla de coeficientes respectiva, es:

CONTRASTES

$$C_1 : \tau_1 + \tau_2 - \tau_3 - \tau_4$$

$$C_2 : \tau_1 - \tau_2 + \tau_3 - \tau_4$$

$$C_3 : \tau_1 - \tau_2 - \tau_3 + \tau_4$$

TABLA DE COEFICIENTES

	τ_1	τ_2	τ_3	τ_4
C_1	1	1	-1	-1
C_2	1	-1	1	-1
C_3	1	-1	-1	1



Las sumas de cuadrados para los efectos principales y la interacción se calculan con:

$$SC_{(A)} = \frac{C_1^2}{r \sum_{i=1}^t c_{i1}^2} \quad conC_1 = \sum_{i=1}^t c_{i1} T_i \quad T_i = \text{Total del trat.}$$

$$SC_{(B)} = \frac{C_2^2}{r \sum_{i=1}^t c_{i2}^2} \quad conC_2 = \sum_{i=1}^t c_{i2} T_i \quad T_i = \text{Total del trat.}$$

$$SC_{(A \times B)} = \frac{C_3^2}{r \sum_{i=1}^t c_{i3}^2} \quad conC_3 = \sum_{i=1}^t c_{i3} T_i \quad T_i = \text{Total del trat.}$$

Cuadro 19. Elementos del Análisis de la Varianza para un Diseño Factorial 2^2 por DBA, mediante contrastes ortogonales

F V	G L	SC
Bloques	r-1	SC (β)
Tratamientos	t-1 = 3	SC (Trat)
A	1*	SC (A)
B	1*	SC (B)
AxB	1*	SC (AxB)
Error	(t-1)(r-1)	por diferencia
Total	rt-1	SC (total)

*El grado de libertad se saca del número de niveles menos uno; ejemplo: 2^2 tiene 2 factores con 2 niveles, por lo que tiene G.L.=2-1=1.

Otro ejemplo sería en el caso de un 3^2

F.V.	G.L.
Tratamientos	t-1=(9-1)=8
A	3-1=2
B	3-1=2
(A*B)	(3-1)(3-1)=4



Ejemplo 2²: Se desarrolla un experimento para evaluar en que medio de cultivo y que especie de *Brucella* es la que crece mejor en la zona. Se tienen 2 medios de cultivo (agar soya tripticasa y Ruiz Castañeda) y las brucellas que se desean aislar son *Brucella abortus* y *Brucella mellitensis*. En el experimento se bloqueo por 4 diferentes portadores de las brucellas. La variable de respuesta fueron el número de colonias que crecieron en los medios de cultivo.

		2	←Factores
Diseño:	2		←Niveles

Factores	A:	Brucellas	a ₀	Br. abortus
			a ₁	Br. mellitensis
	B:	Medios de cultivo	b ₀	Agar soya tripticasa
			b ₁	Ruiz Castañeda

Diseño del experimento: D.B.A.

Bloques	T1	T2	T3	T4	
	a ₀ b ₀	a ₀ b ₁	a ₁ b ₀	a ₁ b ₁	
1	20	15	19	12	66
2	19	14	18	17	68
3	14	12	18	14	58
4	25	18	20	14	77
	78	59	75	57	269

Antes que nada se resuelve el diseño simple para el juego de hipótesis:

Ho: El efecto de los tratamientos VS Ha: Al menos uno es diferente es el mismos

F.V.	G.L.	S.C.	C.M.	F	Pr.
Tratamiento	3	87.1875	29.0625	5.28	0.0225
Bloque	3	45.6875	15.2291	2.764	0.103
Error	9	49.5625	5.5069		
Total	15	182.4375			

Se observa que existen diferencias altamente significativas en los tratamientos.



Una vez resuelto el diseño simple se procede a verificar las hipótesis de los factores en los tratamientos:

HIPOTESIS:

$$H_0(A): T1+T2 = T3 + T4$$

$$H_0(B): T1+T3 = T2 + T4$$

$$H_0(AB): T1+T4 = T2 + T3$$

CONTRASTES:

$$C1: T1+T2-T3-T4=0$$

$$C2: T1-T2+T3-T4=0$$

$$C3: T1-T2-T3+T4=0$$

Tabla de coeficientes

	T1	T2	T3	T4
C1	1	1	-1	-1
C2	1	-1	1	-1
C3	1	-1	-1	1

Evaluación de ortogonalidad

$$C1 \text{ y } C2 = (1)(1) + (1)(-1) + (-1)(1) + (-1)(-1) = 0$$

$$C1 \text{ y } C3 = (1)(1) + (1)(-1) + (-1)(-1) + (-1)(1) = 0$$

$$C2 \text{ y } C3 = (1)(1) + (-1)(-1) + (1)(-1) + (-1)(1) = 0$$

$$SC(A) = \frac{C_1^2}{r \sum c_{i1}^2} = \frac{(78 + 59 - 75 - 57)^2}{4(1^2 + 1^2 + 1^2 + 1^2)} = \frac{25}{16} = 1.5625$$

$$SC(B) = \frac{C_2^2}{r \sum c_{i2}^2} = \frac{(78 - 59 + 75 - 57)^2}{4(1^2 + 1^2 + 1^2 + 1^2)} = \frac{1369}{16} = 85.5625$$

$$SC(AB) = \frac{C_3^2}{r \sum c_{i3}^2} = \frac{(78 - 59 - 75 + 57)^2}{4(1^2 + 1^2 + 1^2 + 1^2)} = \frac{1}{16} = 0.0625$$

Como se puede observar en la presente tabla, los tratamientos son descompuestos en los efectos principales de los factores que lo integran y las interacciones de estos.



F.V.	G.L.	S.C.	C.M.	F	Pr.
A	1	1.5625	1.5625	0.2837 ^{n.s.}	0.612
B	1	85.5625	85.5625	15.5372 ^{**}	0.004
AB	1	0.0625	0.0625	0.0113 ^{n.s.}	0.914
Bloque	3	45.6875	15.2291	2.764 ^{n.s.}	0.103
Error	9	49.5625	5.5069		
Total	15	182.4375			

La lectura de la tabla queda de la siguiente forma : Los tratamientos del factor A (brucellas), tienen el mismo efecto, es decir que las dos crecen de igual forma. Por otra parte el factor B (medios de cultivo) muestra diferencias altamente significativas, lo que indica que uno de los medios funciona mejor que el otro. No hay significancia en las interacciones, lo que indica que los factores actúan de manera INDEPENDIENTE.



Arreglo Factorial 2x2x2 (2^3).

Consideremos un experimento factorial que incluya 3 factores (A, B y C) con 2 niveles (0 y 1), desarrollado bajo un D.B.A., este arreglo factorial es conocido como 2^3 . Entonces las combinaciones y los tratamientos son:

Factores: A, B y C

Niveles: 0 y 1

Diseño de los tratamientos

Factor A	Factor B	Factor C	Tratamientos	
a0	b0	c0	a0b0c0	T1
a0	b0	c1	a0b0c1	T2
a0	b1	c0	a0b1c0	T3
a0	b1	c1	a0b1c1	T4
a1	b0	c0	a1b0c0	T5
a1	b0	c1	a1b0c1	T6
a1	b1	c0	a1b1c0	T7
a1	b1	c1	a1b1c1	T8

Tratamientos: $T_1=a_0b_0c_0$ $T_2=a_0b_0c_1$ $T_3=a_0b_1c_0$ $T_4=a_0b_1c_1$
 $T_5=a_1b_0c_0$ $T_6=a_1b_0c_1$ $T_7=a_1b_1c_0$ $T_8=a_1b_1c_1$

Las hipótesis a probar para el diseño simple son:

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = \tau_6 = \tau_7 = \tau_8$ vs H_a : al menos una $\tau_1 \neq \tau_2$

Estas hipótesis se prueban con el ANVA para el Diseño de Bloques al Azar, cuyo modelo estadístico, las sumas de cuadrados para bloques, tratamientos, error y el total son los del DBA.



El interés es conocer si los factores principales a los niveles evaluados son diferentes e independientes, por lo que las hipótesis a probar son:

- H_{01} : no hay efecto del factor A vs H_{a1} : hay efecto del factor A
 - H_{02} : no hay efecto del factor B vs H_{a2} : hay efecto del factor B
 - H_{03} : no hay efecto del factor C vs H_{a3} : hay efecto del factor C
 - H_{04} : no hay interacción AxB vs H_{a4} : hay interacción AxB
 - H_{05} : no hay interacción AxC vs H_{a5} : hay interacción AxC
 - H_{06} : no hay interacción BxC vs H_{a6} : hay interacción BxC
 - H_{07} : no hay interacción AxBxC vs H_{a7} : hay interacción AxBxC
- (Se comparan los niveles bajos contra los niveles altos de cada factor)

Las hipótesis equivalentes a probar por contrastes ortogonales son:

$$\begin{array}{ll}
 H_{01} : \tau_1 + \tau_2 + \tau_3 + \tau_4 = \tau_5 + \tau_6 + \tau_7 + \tau_8 & \text{vs} \quad H_{a1} : \tau_1 + \tau_2 + \tau_3 + \tau_4 \neq \tau_5 + \tau_6 + \tau_7 + \tau_8 \\
 \begin{array}{c} a_0 \swarrow \\ b_0 \searrow \end{array} & \begin{array}{c} a_1 \swarrow \\ b_1 \searrow \end{array} \\
 H_{02} : \tau_1 + \tau_5 + \tau_6 = \tau_3 + \tau_4 + \tau_7 + \tau_8 & \text{vs} \quad H_{a2} : \tau_1 + \tau_2 + \tau_5 + \tau_6 \neq \tau_3 + \tau_4 + \tau_7 + \tau_8 \\
 \begin{array}{c} c_0 \swarrow \\ c_1 \searrow \end{array} & \begin{array}{c} c_1 \swarrow \\ c_0 \searrow \end{array} \\
 H_{03} : \tau_1 + \tau_3 + \tau_5 + \tau_7 = \tau_2 + \tau_4 + \tau_6 + \tau_8 & \text{vs} \quad H_{a3} : \tau_1 + \tau_3 + \tau_5 + \tau_7 \neq \tau_2 + \tau_4 + \tau_6 + \tau_8 \\
 \begin{array}{c} a_0 b_0, a_1 b_1 \swarrow \\ a_0 b_1, a_1 b_0 \searrow \end{array} & \\
 H_{04} : \tau_1 + \tau_2 + \tau_7 + \tau_8 = \tau_3 + \tau_4 + \tau_5 + \tau_6 & \text{vs} \quad H_{a4} : \tau_1 + \tau_2 + \tau_7 + \tau_8 \neq \tau_3 + \tau_4 + \tau_5 + \tau_6 \\
 \begin{array}{c} a_0 c_0, a_1 c_1 \swarrow \\ a_0 c_1, a_1 c_0 \searrow \end{array} & \\
 H_{05} : \tau_1 + \tau_3 + \tau_6 + \tau_8 = \tau_2 + \tau_4 + \tau_5 + \tau_7 & \text{vs} \quad H_{a5} : \tau_1 + \tau_3 + \tau_6 + \tau_8 \neq \tau_2 + \tau_4 + \tau_5 + \tau_7 \\
 \begin{array}{c} b_0 c_0, b_1 c_1 \swarrow \\ b_0 c_1, b_1 c_0 \searrow \end{array} & \\
 H_{06} : \tau_1 + \tau_4 + \tau_5 + \tau_8 = \tau_2 + \tau_3 + \tau_6 + \tau_7 & \text{vs} \quad H_{a6} : \tau_1 + \tau_4 + \tau_5 + \tau_8 \neq \tau_2 + \tau_3 + \tau_6 + \tau_7 \\
 H_{07} : \tau_1 + \tau_4 + \tau_6 + \tau_7 = \tau_2 + \tau_3 + \tau_5 + \tau_8 & \text{vs} \quad H_{a7} : \tau_1 + \tau_4 + \tau_6 + \tau_7 \neq \tau_2 + \tau_3 + \tau_5 + \tau_8
 \end{array}$$

Estas hipótesis representadas como contrastes de Análisis y la Tabla de Coeficientes respectiva, es:

CONTRASTES:

$$\begin{array}{l}
 C_1 : \tau_1 + \tau_2 + \tau_3 + \tau_4 - \tau_5 - \tau_6 - \tau_7 - \tau_8 \\
 C_2 : \tau_1 + \tau_2 - \tau_3 - \tau_4 + \tau_5 + \tau_6 - \tau_7 - \tau_8 \\
 C_3 : \tau_1 - \tau_2 + \tau_3 - \tau_4 + \tau_5 - \tau_6 + \tau_7 - \tau_8 \\
 C_4 : \tau_1 + \tau_2 - \tau_3 - \tau_4 - \tau_5 - \tau_6 + \tau_7 + \tau_8 \\
 C_5 : \tau_1 - \tau_2 + \tau_3 - \tau_4 - \tau_5 + \tau_6 - \tau_7 + \tau_8 \\
 C_6 : \tau_1 - \tau_2 - \tau_3 + \tau_4 + \tau_5 - \tau_6 - \tau_7 + \tau_8 \\
 C_7 : \tau_1 - \tau_2 - \tau_3 + \tau_4 - \tau_5 + \tau_6 + \tau_7 - \tau_8
 \end{array}$$



Tabla de Coeficientes

T r a t a m i e n t o s

		T1	T2	T3	T4	T5	T5	T7	T8
Factor	Contraste	000	001	010	011	100	101	110	111
A	C ₁	1	1	1	1	-1	-1	-1	-1
B	C ₂	1	1	-1	-1	1	1	-1	-1
C	C ₃	1	-1	1	-1	1	-1	1	-1
AxB	C ₄	1	1	-1	-1	-1	-1	1	1
AxC	C ₅	1	-1	1	-1	-1	1	-1	1
BxC	C ₆	1	-1	-1	1	1	-1	-1	1
AxBxC	C ₇	1	-1	-1	1	-1	1	1	-1

Las sumas de cuadrados para los efectos principales y la interacción se calculan con:

$$SC_{(A)} = \frac{C_1^2}{r \sum_{i=1}^t c_{i1}^2} \quad conC_1 = \sum_{i=1}^t c_{i1} T_i$$

$$SC_{(B)} = \frac{C_2^2}{r \sum_{i=1}^t c_{i2}^2} \quad conC_2 = \sum_{i=1}^t c_{i2} T_i$$

$$SC_{(C)} = \frac{C_3^2}{r \sum_{i=1}^t c_{i3}^2} \quad conC_3 = \sum_{i=1}^t c_{i3} T_i$$

$$SC_{(AxB)} = \frac{C_4^2}{r \sum_{i=1}^t c_{i4}^2} \quad conC_4 = \sum_{i=1}^t c_{i4} T_i$$



$$SC_{(A \times C)} = \frac{C_5^2}{r \sum_{i=1}^t c_{i5}^2} \quad con C_5 = \sum_{i=1}^t c_{i5} T_i$$

$$SC_{(B \times C)} = \frac{C_6^2}{r \sum_{i=1}^t c_{i6}^2} \quad con C_6 = \sum_{i=1}^t c_{i6} T_i$$

$$SC_{(A \times B \times C)} = \frac{C_7^2}{r \sum_{i=1}^t c_{i7}^2} \quad con C_7 = \sum_{i=1}^t c_{i7} T_i$$

Cuadro 20. Elementos del Análisis de la Varianza para un Diseño Factorial 2³ por DBA

F V	G L	SC
Bloques	r-1	SC(β)
Tratamientos	t-1 = 7	SC(trat)
A	1	SC(A)
B	1	SC(B)
C	1	SC(C)
AxB	1	SC(AxB)
AxC	1	SC(AxC)
BxC	1	SC(BxC)
AxBxC	1	SC(AxBxC)
Error	(t-1)(r-1)	Por diferencia
Total	rt-1	SC(total)



Ejemplo 2³.- Se analiza el agua de pozos de diferentes partes de la ciudad, tomándose muestras de cuatro zonas suburbanas. Se procedió a cultivar las muestras en dos diferentes medios de cultivo (M.C.), bajo 2 diferentes temperaturas (t°) y pH para el crecimiento de coliformes fecales, ejecutándose bajo un DBA, en donde los niveles de cada factor fueron los siguientes:

Factor		Nivel 0	Nivel 1
A	M.C.	Mackonkey	Sulfito de Bismuto
B	t°	35°C	37°C
C	pH	6.2	6.6

El número de colonias que crecieron como resultado de la combinación de los factores estudiados fueron los siguientes:

Tratamientos

	T1	T2	T3	T4	T5	T6	T7	T8	$\sum_{j=1}^r Y^2_{.j}$
Bloq.	000	00 1	010	011	100	101	11 0	11 1	
I	10	12	12	16	11	13	14	20	108
II	12	14	13	16	12	14	15	22	118
III	11	14	13	19	12	16	13	24	122
IV	14	16	13	22	11	18	14	24	132
$\sum_{i=1}^t Y^2_{i.}$	47	56	51	73	46	61	56	90	480
\bar{x}	11.75	14	12.7 5	18.2 5	11.5	15.2 5	14	22 5	

Las hipótesis a probar para el diseño DBA son:

$H_0: T_1 = T_2 = \dots T_8$ vs H_a : al menos un T_i es diferente

F.V.	G.L.	S.C.	C.M.	Fc	Pr
Bloques	$r-1 = 3$	37	12.3333	7.6178**	0.0012
Trat.	$t-1 = 7$	387	55.2857	34.1480**	0.000
Error	$(t-1)(r-1) = 21$	34	1.6190		
Total	$rt - 1 = 31$	487			

Se observa que hay diferencias altamente significativas en tratamientos y en bloques



Una vez verificada la hipótesis de tratamientos, se procede a verificar la hipótesis de los factores:

H_{01} : no hay efecto del factor A	vs	H_{a1} : hay efecto del factor A
H_{02} : no hay efecto del factor B	vs	H_{a2} : hay efecto del factor B
H_{03} : no hay efecto del factor C	vs	H_{a3} : hay efecto del factor C
H_{04} : no hay interacción AxB	vs	H_{a4} : hay interacción AxB
H_{05} : no hay interacción AxC	vs	H_{a5} : hay interacción AxC
H_{06} : no hay interacción BxC	vs	H_{a6} : hay interacción BxC
H_{07} : no hay interacción AxBxC	vs	H_{a7} : hay interacción AxBxC

F.V.	G.L.	S.C.	C.M.	Fc	Pr
Bloques	$r-1 = 3$	37	12.3333	7.6178**	0.0012
Trat.	$t-1 = 7$	387	55.2857	34.1480**	0.0000
A	1	21.125	21.125	7.14*	0.0133
B	1	112.5	112.5	38.03**	0.0000
C	1	200.0	200.0	67.61**	0.0000
AB	1	10.125	10.125	3.42*	0.0767
AC	1	10.125	10.125	3.42*	0.0767
BC	1	32.0	32.0	10.82**	0.0031
ABC	1	1.125	1.125	0.38 ^{n.s.}	0.5433
Error	$(t-1)(r-1) = 21$	34	1.6190		
Total	$rt - 1 = 31$	487			

Con respecto a los factores, los medios de cultivo (A) presentan diferencias significativas y las temperaturas (B) y pH (C) presentan diferencias altamente significativas, lo cual indica que los niveles trabajan de manera diferente uno de otro. Por otra parte las interacciones del factor medios de cultivo (A), con temperatura (B) y pH (C), muestran diferencias significativas, es decir existen interacción entre ellos; y la interacción de la temperatura (B) con pH (C) muestra diferencias altamente significativas (una mayor interacción). Sin embargo al estar presentes los tres factores, no se da interacción alguna.



Problemario de Arreglos Factoriales por Contrastes Ortogonales

36.- Se realiza un estudio en nuestro estado evaluando a 2 fármacos antipalúdicos: Pirimetamina a 2 diferentes concentraciones 2 y 5 mg y Cloroquina a dos diferentes concentraciones, 3 y 6 mg. en pacientes con paludismo. Análisis previos indican que los agentes causales son los parásitos Plasmodium falciparum, Plasmodium vivax y Plasmodium malariae. Se determino la eficiencia de los fármacos en el número de pacientes curados. Los resultados obtenidos fueron los siguientes:

Pirimetamina	Cloroquina	Parásitos	Pacientes Curados
2mg.	3mg.	P. falciparum	21
2mg.	3mg.	P. vivax	20
2mg.	3mg.	P. malariae	17
2mg.	6mg.	P. falciparum	23
2mg.	6mg.	P. vivax	24
2mg.	6mg.	P. malariae	22
5mg.	3mg.	P. falciparum	19
5mg.	3mg.	P. vivax	21
5mg.	3mg.	P. malariae	16
5mg.	6mg.	P. falciparum	22
5mg.	6mg.	P. vivax	24
5mg.	6mg.	P. malariae	14

37.- En el Hospital General de Zona 1 del IMSS se analizó el índice de infecciones vaginales en mujeres con diferente actividad sexual (activas y pasivas) encontrándose una mayor prevalencia de dos tipos de virus (VPH y VHS); tomando en cuenta que el estudio se realizó a mujeres de diferentes edades (de 15 - 25, 26 – 35, 36 – 45, 46 – 55) y diferente nivel social (baja, media baja, media alta y alta).

VPH: virus del papiloma humano VHS: virus del herpes simple

FACTORES: A: Actividad Sexual ao: Activas a1: Pasivas
B: virus bo: VPH b1: VHS

	Nivel Social			
Edades	Baja	Media baja	Media alta	Alta
15-25	T4- 17	T1- 15	T2- 16	T3- 14
26-35	T1- 16	T2- 14	T3- 10	T4- 15
36-45	T2- 15	T3- 16	T4- 13	T1- 11
46-55	T3- 9	T4- 13	T1- 15	T2- 10

38.- Se realizó un experimento para evaluar la eficacia de 2 insecticidas a 2 vectores de Leishmania, para ver cual de los vectores es más sensible a estos; este estudio se realizó en 4 zonas del estado de Chiapas (zona selva, zona altos, zona costa, zona soconusco), se tienen 2 insecticidas: DDT y MALATION, los vectores que se desean eliminar son: Lutzomyia verrucarum y Lutzomyia noguchi. En el experimento se utilizaron unidades experimentales con 20 vectores y se evaluó el número de muertes de un tiempo de exposición.

	T1	T2	T3	T4
	00	01	10	11
I (ZONA SELVA)	16	15	13	10
II (ZONA ALTOS)	13	14	12	11
III (ZONA COSTA)	14	15	16	16
IV (ZONA SOCONUSCO)	10	12	12	14

FACTORES	A: INSECTICIDA	a0 : DDT a1: MALATION
	B: VECTORES	b0 : <i>Lutzomyia verrucarum</i> b1 : <i>Lutzomyia noguchi</i>

39.- Una microempresa se dedica a la producción de alimentos derivados de la leche. Emplean leche de vaca y cabra, ambos animales tienen dos fuentes de agua: río y pozo. Se realiza un análisis bacteriológico para determinar el control de calidad en la producción de queso, determinándose mediante un recuento de colonias de la enterobacteria E. Coli aisladas del medio EMB de quesos escogidos al azar (5 quesos con 4 repeticiones cada uno). La enterobacteria se encuentra aceptable en un límite de 1 a 4 UFC/ ml. aisladas de una dilución del queso 1:10. Los resultados obtenidos son en UFC/ ml.

T1	T2	T3	T4
00	01	10	11
5	4	5	3
7	5	4	5
5	2	6	4
7	4	5	2
6	3	7	2

T1	T2	T3	T4
00	01	10	11
5	4	5	3
7	5	4	5
5	2	6	4
7	4	5	2
6	3	7	2

Factores:	A:	LECHES:	a0: vaca
			a1: cabra
	B:	AGUA:	b0: Río
			b1: potable

40.- Una compañía dedicada a la fabricación de medicamentos desea probar un nuevo medicamento el cuál consta de la combinación de dos fármacos una penicilina (ampicilina) y una sulfamida (sulfametoxazol) pero se cuenta con dos concentraciones para cada fármaco 200 y 300 mg para Ampicilina y 100 y 400 mg para Sulfametoxazol. Dicho laboratorio



realiza un experimento para determinar la eficacia de la combinación de los fármacos y sus concentraciones. Los fármacos son antibióticos y se probaron en medios de cultivo con cepas de:

- E. Coli,
- Proteus spp.,
- Klebsiella spp.,
- Pseudomona aeruginosa;

midiendo el halo de inhibición de crecimiento en milímetros (mm) alrededor de un multidisco conteniendo los sensidiscos de los nuevos medicamentos a probar.

Ampicilina	Sulfametoxasol	Bacterias	Inhibición (mm)
200	100	E. coli	3
200	100	Proteus spp	4
200	100	Klebsiella spp	3
200	100	Pseudomona aeruginosa	2
200	400	E. coli	2
200	400	Proteus spp	2
200	400	Klebsiella spp	4
200	400	Pseudomona aeruginosa	3
300	100	E. coli	1
300	100	Proteus spp	3
300	100	Klebsiella spp	2
300	100	Pseudomona aeruginosa	2
300	400	E. coli	2
300	400	Proteus spp	3
300	400	Klebsiella spp	3
300	400	Pseudomona aeruginosa	2

Referencias

Arnau Gras, J.. Diseños Experimentales Multivariados. Alianza Editorial, Madrid. 1990.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Cox, D.R. Planning of Experiments. Wiley, New York. 1992.



Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C.: "Análisis Multivariante", Ed. Prentice Hall, 5ª edición. 1999,

Hinkelmann, K. Design and analysis of experiments. John Wiley & Sons, N.Y. (2005).

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lindman, H.R. Analysis of Variance in Experimental Design. Springer-Verlag, N.Y. 1992.

Mason, R.L., Gunst, R.F. and Hess, J.L. Statistical Design and Analysis of Experiments. Wiley, N.Y. 2003.

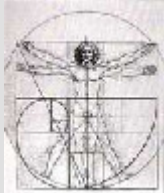
Montgomery, D.C. Diseño y Análisis de Experimentos. Limusa, México. 2003.

Rencher, A.C. Multivariate Statistical Inference and Applications. 1998.

Sixto Ríos. "Métodos Estadísticos", Ediciones del Castillo, Madrid. 1977.



Parte III: Arreglos Factoriales



Capítulo XI:

Análisis de un Arreglo Factorial por Medio del Modelo Factorial.

Al igual que en el capítulo anterior, el análisis de un arreglo factorial también puede realizarse construyendo el ANVA mediante el modelo factorial correspondiente.

Arreglo Factorial 2^2

En el caso de un arreglo factorial 2^2 con dos factores (A y B) y dos niveles (0 y 1) bajo un Diseño de Bloques al Azar, las combinaciones y los tratamientos son:

FACTORES: A y B

NIVELES: 0 y 1

Diseño de los tratamientos

Factor A	Factor B	Tratamiento	
a0	b0	a0b0	T1
a0	b1	a0b1	T2
a1	b0	a1b0	T3
a1	b1	a1b1	T4

TRATAMIENTOS:

$$T_1 = a_0b_0$$

$$T_2 = a_0b_1$$

$$T_3 = a_1b_0$$

$$T_4 = a_1b_1$$

El modelo matemático para este arreglo es descomponer el tratamiento en los diferentes factores e interacciones:

$$Y_{ijk} = \mu + \beta_j + A_j + B_k + (AB)_{jk} + \varepsilon_{ijk}$$

con: $y = 1, 2, \dots, r$

$j = 1, 2, \dots, a$

$k = 1, 2, \dots, b$

$a = 2$

$b = 2$



donde:

Y_{ijk} = Es la observación del nivel j del factor A, nivel k del factor B en el i-ésimo bloque.

μ = Es la media verdadera general.

β_i = Es el efecto verdadero del i-ésimo bloque.

A_j = Es el efecto verdadero del nivel j del factor A.

B_k = Es el efecto verdadero del nivel k del factor B.

$(AB)_{jk}$ = Es el efecto verdadero de la interacción de los niveles j y k de los factores A y B.

ϵ_{ijk} = El error experimental de la ijk-ésima observación.

Las hipótesis a probar son:

H_{01} : no hay efecto del factor A vs H_{a1} : hay efecto del factor A

H_{02} : no hay efecto del factor B vs H_{a2} : hay efecto del factor B

H_{03} : no hay interacción AxB vs H_{a3} : hay interacción AxB

(Se comparan los niveles bajos contra los niveles altos de cada factor)

La tabla de datos es de la siguiente forma:

BLOQUES	TRATAMIENTOS				Σy_i
	1 a_0b_0	2 a_0b_1	3 a_1b_0	4 a_1b_1	
I	y_{111}	y_{112}	y_{121}	y_{122}	$y_{1..}$
II	y_{211}	y_{212}	y_{221}	y_{222}	$y_{2..}$
III	y_{311}	y_{312}	y_{321}	y_{322}	$y_{3..}$
.
R	y_{r11}	y_{r12}	y_{r21}	y_{r22}	y_r
Σy_{jk}	$y_{..1}$	$y_{..2}$	$y_{.21}$	$y_{.22}$	$y_{..}$

Donde Y_{rab} = r:Bloque, a:Nivel del factor I, b:Nivel del Factor II.

El cuadro del análisis de la varianza se construye calculando los grados de libertad y las sumas de cuadrados mediante ecuaciones que se pueden obtener directamente del modelo. Son reglas válidas únicamente para arreglos completos balanceados (las posibles clasificaciones de estos tienen igual número de observaciones en las celdas, y para los cuales cada nivel de los factores se repite el mismo número de veces).

El estudio de los efectos del factorial 2^2 se hace en base a los totales de cada factor en los niveles dados y de la interacción, con la siguiente



notación: los efectos principales se denotan por letras mayúsculas, por ejemplo: A y B; la interacción por AB, entonces tendremos:

El total para el nivel bajo del factor A se denota por:

$$(A)_{j=0} = Ta_0b_0 + Ta_0b_1$$

El total para el nivel alto del factor A es:

$$(A)_{j=1} = Ta_1b_0 + Ta_1b_1$$

Donde: Ta_0b_0 , Ta_0b_1 , Ta_1b_0 , Ta_1b_1 denotan los totales de los tratamientos a_0b_0 , a_0b_1 , a_1b_0 , y a_1b_1 , respectivamente.

De igual forma se tienen los totales para el factor B:

$$(B)_{k=0} = Ta_0b_0 + Ta_1b_0$$

$$(B)_{k=1} = Ta_0b_1 + Ta_1b_1$$

Para la interacción: $(AB)_{j+k=0} = Ta_0b_0 + Ta_1b_1$

$$(AB)_{j+k=1} = Ta_0b_1 + Ta_1b_0$$

El total $(AB)_{j+k=0}$ se obtiene sumando los totales de los tratamientos a_0b_0 y a_1b_1 , observe que $j+k = 0$ indica la suma de los índices, o sea cuando $j+k=2$ equivale a $j+k=0$.

Los efectos se definen como:

$$A = \frac{1}{2r} \left[(A)_{j=1} - (A)_{j=0} \right]$$

$$B = \frac{1}{2r} \left[(B)_{k=1} - (B)_{k=0} \right]$$

$$AB = \frac{1}{2r} \left[(AB)_{j-k=0} - (AB)_{j-k=1} \right]$$

Para ver si estos efectos son significativos se prueban las hipótesis correspondientes en el cuadro del análisis de la varianza.



Las reglas para encontrar los G.L. y las S.C. son:

- 1) Se considera el modelo.
- 2) Para cada término del modelo habrá una hilera en el cuadro de análisis de varianza.
- 3) Para cada término del modelo se determina su “Producto Simbólico”, que es igual al índice menos uno o el producto de estos en el término; los grados de libertad se obtienen al sustituir el índice en el producto simbólico por el máximo valor que puede tomar; por ejemplo en arreglo 2^2 :

FACTOR	INDICE	P. SIMBOLICO	G.L.
Bloques	i	j-1	r-1
A	j	j-1	a-1
B	k	k-1	b-1
AxB	j, k	(j-1)(k-1)	(a-1)(b-1)

- 4) Para obtener las sumas de cuadrados se realiza:
 - a) Se suma sobre los índices que no están en el producto simbólico.
 - b) Se elevan al cuadrado los totales.
 - c) Se divide entre el máximo valor que pueden tomar los índices que forman el total.
 - d) Se suma sobre los índices que si están en el producto simbólico.

Consideremos el arreglo factorial antes descrito del modelo. Se tienen las fuentes de variación, los productos simbólicos y los grados de libertad siguientes (ver cuadro 21)

Cuadro 21.- Fuentes de variación, los productos simbólicos y los grados de libertad de un diseño factorial 2^2 bajo un DBA

F V	PS	G L
Media	1	1
Bloques	i-1	r-1
A	j-1	a-1
B	k-1	b-1
AB	(j-1)(k-1)	(a-1)(b-1)
Error	Por diferencia	Por diferencia
Total	ijk-1	rab-1



NOTA: El producto simbólico para la media es uno; el producto simbólico para el total es el producto simple de todos los índices; el producto simbólico para el error se obtiene por diferencia.

Las sumas de cuadrados, obtenidas con los productos simbólicos son:

Total: P.S. = $ijk - 1$ $SC_{total} = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^b y_{ijk}^2 - \frac{(y_{...})^2}{rab}$

Bloques: $i - 1$ $SC_{Bloques} = \sum_{i=1}^r \frac{y_{i..}^2}{ab} - \frac{(y_{...})^2}{rab}$

Factor A: $j - 1$ $SC(A) = \sum_{j=1}^a \frac{y_{.j.}^2}{rb} - \frac{(y_{...})^2}{rab}$

Factor B: $k - 1$ $SC(B) = \sum_{k=1}^b \frac{y_{..k}^2}{ra} - \frac{(y_{...})^2}{rab}$

Interacción AxB: $(j - 1)(k - 1) = (jk - 1) - (j - 1) - (k - 1)$

$$SC(AxB) = \sum_{i=1}^a \sum_{k=1}^b \frac{y_{.jk}^2}{r} - \frac{(y_{...})^2}{rab} - SC_{Fact}(A) - SC_{Fact}(B)$$

Error: $(ijk - 1) - (i - 1) - (j - 1) - (i - 1)(j - 1)$

$$SC \text{ Error} = SC \text{ Total} - SC \text{ Bloques} - SC(A) - SC(B) - SC(AxB)$$



Cuadro 22. Elementos del Análisis de la Varianza para un Diseño Factorial 2^2 por DBA mediante el modelo factorial

F V	G L	SC
Bloques	$r - 1$	SC (β)
Tratamientos	$t - 1 = 3$	SC (Trat)
A	$a - 1 = 1$	SC (A)
B	$b - 1 = 1$	SC (B)
AxB	$(a-1)(b-1) = 1$	SC (AxB)
Error	$(t-1)(r-1)$	Por diferencia
Total	$rt - 1$	SC (Total)

Ejemplo 2²: Se desarrolla un experimento para evaluar en que medio de cultivo y que especie de Brucella es la que crece mejor en la zona. Se tienen 2 medios de cultivo (agar soya tripticasa y Ruiz Castañeda) y las brucellas que se desean aislar son Brucella abortus y Brucella mellitensis. En el experimento se bloqueo por 4 diferentes portadores de las brucellas. La variable de respuesta fueron el número de colonias que crecieron en los medios de cultivo.

		2	←Factores
Diseño:	2		←Niveles

Factores	A:	Brucellas	a_0	Br. abortus
			a_1	Br. mellitensis
	B:	Medios de cultivo	b_0	Agar soya tripticasa
			b_1	Ruiz Castañeda

Diseño del experimento: D.B.A.

Bloques	T1	T2	T3	T4	
	a_0b_0	a_0b_1	a_1b_0	a_1b_1	
1	20	15	19	12	66
2	19	14	18	17	68
3	14	12	18	14	58
4	25	18	20	14	77
	78	59	75	57	269

$r = i, a = j, b = k$



Antes que nada se resuelve el diseño simple para el juego de hipótesis:

Ho: El efecto de los tratamientos VS Ha: Al menos uno es diferente es el mismos

F.V.	G.L.	S.C.	C.M.	F	Pr.
Tratamiento	3	87.1875	29.0625	5.28	0.0225
Bloque	3	45.6875	15.2291	2.764	0.103
Error	9	49.5625	5.5069		
Total	15	182.4375			

Se observa que existen diferencias altamente significativas en los tratamientos.

Una vez resuelto el diseño simple se procede a verificar las hipótesis de los factores en los tratamientos:

Sumatorias por los niveles de los factores:

a ₀	=137	b ₀	=153
a ₁	=132	b ₁	=116

$$SC_{.Bloq} = \sum_i^r \frac{y_{i..}^2}{ab} - \frac{y_{...}^2}{rab} = \frac{66^2 + 68^2 + 58^2 + 77^2}{(2)(2)} - \frac{269^2}{4(2)(2)} = 45.6875$$

$$SC_{.(A)} = \sum_i^r \frac{y_{.j.}^2}{rb} - \frac{y_{...}^2}{rab} = \frac{137^2 + 132^2}{4(2)} - \frac{269^2}{4(2)(2)} = 1.5625$$

$$SC_{.(B)} = \sum_i^r \frac{y_{.k}^2}{ra} - \frac{y_{...}^2}{rab} = \frac{153^2 + 116^2}{4(2)} - \frac{269^2}{4(2)(2)} = 85.5625$$

$$SC_{.(AB)} = \sum_i^r \frac{y_{.jk}^2}{r} - \frac{y_{...}^2}{rab} - SC_{(A)} - SC_{(B)} =$$

$$SC_{(AB)} = \frac{78^2 + 59^2 + 75^2 + 57^2}{4} - \frac{269^2}{4(2)(2)} - 1.5625 - 85.5625 = 0.0625$$



F.V.	G.L.	S.C.	C.M.	F	Pr.
A	1	1.5625	1.5625	0.2837	0.612
B	1	85.5625	85.5625	15.5372	0.004
AB	1	0.0625	0.0625	0.0113	0.914
Bloque	3	45.6875	15.2291	2.764	0.103
Error	9	49.5625	5.5069		
Total	15	182.4375			

La lectura de la tabla queda de la siguiente forma : Los tratamientos del factor A (brucellas), tienen el mismo efecto, es decir que las dos crecen de igual forma. Por otra parte el factor B (medios de cultivo) muestra diferencias altamente significativas, lo que indica que uno de los medios funciona mejor que el otro. No hay significancia en las interacciones, lo que indica que los factores actúan de manera INDEPENDIENTE.



Arreglo Factorial 2^3

Consideremos un arreglo factorial 2^3 con tres factores (A; B y C) y dos niveles (0 y 1), bajo un D.B.A., las combinaciones y los tratamientos son:

Factores: A, B y C

Niveles: 0 y 1

Diseño de los tratamientos

Factor A	Factor B	Factor C	Tratamientos	
a0	b0	c0	a0b0c0	T1
a0	b0	c1	a0b0c1	T2
a0	b1	c0	a0b1c0	T3
a0	b1	c1	a0b1c1	T4
a1	b0	c0	a1b0c0	T5
a1	b0	c1	a1b0c1	T6
a1	b1	c0	a1b1c0	T7
a1	b1	c1	a1b1c1	T8

Tratamientos: $T_1=a_0b_0c_0$ $T_2=a_0b_0c_1$ $T_3=a_0b_1c_0$ $T_4=a_0b_1c_1$
 $T_5=a_1b_0c_0$ $T_6=a_1b_0c_1$ $T_7=a_1b_1c_0$ $T_8=a_1b_1c_1$

Las hipótesis a probar en el diseño simple son:

$H_0: \tau_1 = \tau_2 = : \tau_3 = \tau_4 = \tau_5 = \tau_6 = : \tau_7 = \tau_8$ vs $H_a: \text{al menos una } \tau_i \neq \tau_j$

El modelo matemático para este arreglo es descomponer el tratamiento en los diferentes factores e interacciones:

$$Y_{ijkl} = \mu + \beta_i + A_j + B_k + C_l + (AB)_{jk} + (AC)_{jl} + (BC)_{kl} + (ABC)_{jkl} + \varepsilon_{ijkl}$$

con: $i = 1, 2, \dots, r$

$j = 1, 2, \dots, a$

$k = 1, 2, \dots, b$

$l = 1, 2, \dots, c$

$a = 2$

$b = 2$

$c = 2$



donde:

Y_{ijk} = Es la observación del nivel j del factor A, nivel k del factor B, nivel l en el i-ésimo bloque.

μ = Es la media verdadera general.

β_i = Es el efecto verdadero del i-ésimo bloque.

A_j = Es el efecto verdadero del nivel j del factor A.

B_k = Es el efecto verdadero del nivel k del factor B.

C_l = Es el efecto verdadero del nivel l del factor C.

$(AB)_{jk}$ = Es el efecto verdadero de la interacción de los niveles j y k de los factores A y B.

$(AC)_{jl}$ = Es el efecto verdadero de la interacción de los niveles j y l de los factores A y C.

$(BC)_{kl}$ = Es el efecto verdadero de la interacción de los niveles k y l de los factores B y C.

$(ABC)_{jkl}$ = Es el efecto verdadero de la interacción de los niveles k, k y l de los factores A, B y C.

ε_{ijkl} = El error experimental de la ijkl-ésima observación.

En el arreglo factorial el interés es conocer si los factores principales a los niveles evaluados son diferentes e independientes; las hipótesis a probar son:

H_{01} : no hay efecto del factor A	vs	H_{a1} : hay efecto del factor A
H_{02} : no hay efecto del factor B	vs	H_{a2} : hay efecto del factor B
H_{03} : no hay efecto del factor C	vs	H_{a3} : hay efecto del factor C
H_{04} : no hay interacción AxB	vs	H_{a4} : hay interacción AxB
H_{05} : no hay interacción AxC	vs	H_{a5} : hay interacción AxC
H_{06} : no hay interacción BxC	vs	H_{a6} : hay interacción BxC
H_{07} : no hay interacción AxBxC	vs	H_{a7} : hay interacción AxBxC

(Se comparan los niveles bajos contra los niveles altos de cada factor)



La Tabla de Datos es de la siguiente forma:

Bloques	T r a t a m i e n t o s								$\Sigma y_{i...}$
	1	2	3	4	5	6	7	8	
	$a_0b_0c_0$	$a_0b_0c_1$	$a_0b_1c_0$	$a_0b_1c_1$	$a_1b_0c_0$	$a_1b_0c_1$	$a_1b_1c_0$	$a_1b_1c_1$	
I	y_{1111}	y_{1112}	y_{1121}	y_{1122}	y_{1211}	y_{1212}	y_{1221}	y_{1222}	$y_{1...}$
II	y_{2111}	y_{2112}	y_{2121}	y_{2122}	y_{2211}	y_{2212}	y_{2221}	y_{2222}	$y_{2...}$
III	y_{3111}	y_{3112}	y_{3121}	y_{3122}	y_{3211}	y_{3212}	y_{3221}	y_{3222}	$y_{3...}$
.	
.	
R	y_{r111}	y_{r112}	y_{r121}	y_{r122}	y_{r211}	y_{r212}	y_{r221}	y_{r222}	$y_{r...}$
$\Sigma y_{.jkl}$	$y_{.111}$	$y_{.112}$	$y_{.121}$	$y_{.122}$	$y_{.211}$	$y_{.212}$	$y_{.221}$	$y_{.222}$	$y_{....}$

El cuadrado del análisis de la varianza se construye calculando los grados de libertad y las sumas de cuadrados mediante las reglas dadas en el tema anterior del factorial 2^2 .

El estudio de los efectos de los factores y de las interacciones se hace en base a los totales de cada factor en los niveles dados y de la interacción, con la misma notación; por lo que tenemos:

El total para los niveles del factor A son:

$$(A)_{j=0} = Ta_0b_0c_0 + Ta_0b_0c_1 + Ta_0b_1c_0 + Ta_0b_1c_1$$

$$(A)_{j=1} = Ta_1b_0c_0 + Ta_1b_0c_1 + Ta_1b_1c_0 + Ta_1b_1c_1$$

Los totales para el factor B son:

$$(B)_{k=0} = Ta_0b_0c_0 + Ta_0b_0c_1 + Ta_1b_0c_0 + Ta_1b_0c_1$$

$$(B)_{k=1} = Ta_0b_1c_0 + Ta_0b_1c_1 + Ta_1b_1c_0 + Ta_1b_1c_1$$

Los totales para el factor C son:

$$(C)_{l=0} = Ta_0b_0c_0 + Ta_0b_1c_0 + Ta_1b_0c_0 + Ta_1b_1c_0$$

$$(C)_{l=1} = Ta_0b_0c_1 + Ta_0b_1c_1 + Ta_1b_0c_1 + Ta_1b_1c_1$$



Para las interacciones:

$$\begin{aligned}(AB)_{j+k=0} &= Ta_0b_0c_0 + Ta_0b_0c_1 + Ta_1b_1c_0 + Ta_1b_1c_1 \\(AB)_{j+k=1} &= Ta_0b_1c_0 + Ta_0b_1c_1 + Ta_1b_0c_0 + Ta_1b_0c_1 \\(AC)_{j+l=0} &= Ta_0b_0c_0 + Ta_0b_1c_0 + Ta_1b_0c_1 + Ta_1b_1c_1 \\(AC)_{j+l=1} &= Ta_0b_0c_1 + Ta_0b_1c_1 + Ta_1b_0c_0 + Ta_1b_1c_0 \\(BC)_{k+l=0} &= Ta_0b_0c_0 + Ta_0b_1c_1 + Ta_1b_0c_0 + Ta_1b_1c_1 \\(BC)_{k+l=1} &= Ta_0b_0c_1 + Ta_0b_1c_0 + Ta_1b_0c_1 + Ta_1b_1c_0 \\(ABC)_{j+k+l=0} &= Ta_0b_0c_0 + Ta_0b_1c_1 + Ta_1b_0c_1 + Ta_1b_1c_0 \\(ABC)_{j+k+l=1} &= Ta_0b_0c_1 + Ta_0b_1c_0 + Ta_1b_0c_0 + Ta_1b_1c_1\end{aligned}$$

Los efectos se definen como:

$$A = \frac{1}{4r}[(A)_{j=1} - (A)_{j=0}]$$

$$B = \frac{1}{4r}[(B)_{k=1} - (B)_{k=0}]$$

$$C = \frac{1}{4r}[(C)_{l=1} - (C)_{l=0}]$$

$$AB = \frac{1}{4r}[(AB)_{j-k=0} - (AB)_{j-k=1}]$$

$$AC = \frac{1}{4r}[(AC)_{j-l=0} - (AC)_{j-l=1}]$$

$$BC = \frac{1}{4r}[(BC)_{k-l=0} - (BC)_{k-l=1}]$$

$$ABC = \frac{1}{4r}[(ABC)_{j-k-l=1} - (ABC)_{j-k-l=0}]$$

Para ver si estos efectos son significativos se prueban las hipótesis correspondientes en el cuadro del análisis de la varianza.



Consideremos el Arreglo Factorial antes descrito del modelo, se tienen las fuentes de variación los productos simbólicos y los grados de libertad siguientes:

Cuadro 23.- Fuentes de variación, los productos simbólicos y los grados de libertad de un diseño factorial 2^3 bajo un DBA

F V	PS	G L
Media	1	1
Bloques	i - 1	r - 1
A	j - 1	a - 1
B	k - 1	b - 1
C	l - 1	c - 1
AB	(j-1)(k-1)	(a-1)(b-1)
AC	(j-1)(l-1)	(a-1)(c-1)
BC	(k-1)(l-1)	(b-1)(c-1)
ABC	(j-1)(k-1)(l-1)	(a-1)(b-1)(c-1)
Error	Por diferencia	Por diferencia
Total	i j k l - 1	rabc - 1

Nota: El producto simbólico para la media es uno; el producto simbólico para el total es el producto simple de todos los índices; el producto simbólico para el error se obtiene por diferencia.

Las sumas de cuadrados, obtenidas con los productos simbólicos son:

$$\text{Total: P.S.} = ijkl - 1 \quad SC_{total} = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^b \sum_{l=1}^c y_{ijkl}^2 - \frac{(y....)^2}{rabc}$$

$$\text{Bloques: i - 1} \quad SC_{Bloques} = \sum_{i=1}^r \frac{y_{i...}^2}{abc} - \frac{(y....)^2}{rabc}$$

$$\text{Factor A: j - 1} \quad SC(A) = \sum_{j=1}^a \frac{y_{.j..}^2}{rbc} - \frac{(y....)^2}{rabc}$$

$$\text{Factor B: k - 1} \quad SC(B) = \sum_{k=1}^b \frac{y_{...k.}^2}{rac} - \frac{(y....)^2}{rabc}$$



Factor C: 1 - 1

$$SC(C) = \sum_{l=1}^c \frac{y_{..l}^2}{rab} - \frac{(y....)^2}{rabc}$$

Interacción AxB: (j - 1) (k - 1) = (jk - 1) - (j - 1) - (k - 1)

$$SC(AxB) = \sum_{i=1}^a \sum_{k=1}^b \frac{y_{.jk.}^2}{r} - \frac{(y....)^2}{rabc} - SCFact(A) - SCFact(B)$$

Interacción AxC: (j - 1) (l - 1) = (jl - 1) - (j - 1) - (l - 1)

$$SC(AxC) = \sum_{i=1}^a \sum_{l=1}^c \frac{y_{.jl.}^2}{r} - \frac{(y....)^2}{rabc} - SCFact(A) - SCFact(C)$$

Interacción BxC: (k - 1) (l - 1) = (kl - 1) - (k - 1) - (l - 1)

$$SC(BxC) = \sum_{k=1}^b \sum_{l=1}^c \frac{y_{..kl}^2}{r} - \frac{(y....)^2}{rabc} - SCFact(B) - SCFact(C)$$

Interacción AxBxC: (j-1)(k-1)(l-1)=(jkl-1)-(j-1)(k-1)-(l-1)-(jk-1)-(jl-1)(kl-1)

$$SC(AxBxC) = \sum_{j=1}^a \sum_{k=1}^b \sum_{l=1}^c \frac{y_{.jkl}^2}{r} - \frac{(y....)^2}{rabc} - SC(A) - SC(B) - SC(C) - SC(AxB) - SC(AxC) - SC(BxC)$$

Error: (ijk-1)-(i-1)-(j-1)-(k-1)-(j-1)(k-1)-(j-1)(l-1)-(j-1)(k-1)-(k-1)(l-1)-(j-1)(k-1)(l-1).

$$SC \text{ Error} = SC \text{ Total} - SC \text{ Bloques} - SC(A) - SC(B) - SC(C) - SC(AxB) - SC(AxC) - SC(BxC) - SC(AxBxC)$$



Cuadro 24. Elementos del Análisis de la Varianza para un Diseño Factorial 2^3 por DBA mediante el modelo factorial

F V	G L	SC
Bloques	$r - 1$	SC (β)
Tratamientos	$t - 1 = 7$	SC (Trat)
A	1	SC (A)
B	1	SC (B)
C	1	SC (C)
AxB	1	SC (AxB)
AxC	1	SC (AxC)
BxC	1	SC (BxC)
AxBxC	1	SC (AxBxC)
Error	$(t-1)(r-1)$	Por diferencia
Total	$rt - 1$	SC (Total)

Ejemplo 23.- Se analiza el agua de pozos de diferentes partes de la ciudad, tomándose muestras de cuatro zonas suburbanas. Se procedió a cultivar las muestras en dos diferentes medios de cultivo (M.C.), bajo 2 diferentes temperaturas (t°) y pH para el crecimiento de coliformes fecales, ejecutándose bajo un DBA, en donde los niveles de cada factor fueron los siguientes:

Factor		Nivel 0	Nivel 1
A	M.C.	Mackonkey	Sulfito de Bismuto
B	t°	35°C	37°C
C	pH	6.2	6.6

El número de colonias que crecieron como resultado de la combinación de los factores estudiados fueron los siguientes:

Tratamientos

	T1	T2	T3	T4	T5	T6	T7	T8	$\sum_{j=1}^r Y_j^2$
Bloq.	000	001	010	011	100	101	110	111	
I	10	12	12	16	11	13	14	20	108
II	12	14	13	16	12	14	15	22	118
III	11	14	13	19	12	16	13	24	122
IV	14	16	13	22	11	18	14	24	132
$\sum_{i=1}^t Y_i^2$	47	56	51	73	46	61	56	90	480
\bar{x}	11.75	14	12.75	18.25	11.5	15.25	14	22.5	



Las hipótesis a probar para el diseño DBA son:

$H_0: T_1 = T_2 = \dots T_8$ vs H_a : al menos un T_i es diferente

F.V.	G.L.	S.C.	C.M.	Fc	Pr
Bloques	$r-1 = 3$	37	12.3333	7.6178**	0.0012
Trat.	$t-1 = 7$	387	55.2857	34.1480**	0.000
Error	$(t-1)(r-1) = 21$	34	1.6190		
Total	$rt - 1 = 31$	487			

Se observa que hay diferencias altamente significativas en tratamientos y en bloques

Una vez verificada la hipótesis de tratamientos, se procede a verificar la hipótesis de los factores:

H_{01} : no hay efecto del factor A	vs	H_{a1} : hay efecto del factor A
H_{02} : no hay efecto del factor B	vs	H_{a2} : hay efecto del factor B
H_{03} : no hay efecto del factor C	vs	H_{a3} : hay efecto del factor C
H_{04} : no hay interacción AxB	vs	H_{a4} : hay interacción AxB
H_{05} : no hay interacción AxC	vs	H_{a5} : hay interacción AxC
H_{06} : no hay interacción BxC	vs	H_{a6} : hay interacción BxC
H_{07} : no hay interacción AxBxC	vs	H_{a7} : hay interacción AxBxC

F.V.	G.L.	S.C.	C.M.	Fc	Pr
Bloques	$r-1 = 3$	37	12.3333	7.6178**	0.0012
Trat.	$t-1 = 7$	387	55.2857	34.1480**	0.0000
A	1	21.125	21.125	7.14*	0.0133
B	1	112.5	112.5	38.03**	0.0000
C	1	200.0	200.0	67.61**	0.0000
AB	1	10.125	10.125	3.42*	0.0767
AC	1	10.125	10.125	3.42*	0.0767
BC	1	32.0	32.0	10.82**	0.0031
ABC	1	1.125	1.125	0.38 ^{n.s.}	0.5433
Error	$(t-1)(r-1) = 21$	34	1.6190		
Total	$rt - 1 = 31$	487			

Con respecto a los factores, el A presenta diferencias significativas y el B y C altamente significativas, lo cual indica que los niveles trabajan de



manera diferente uno de otro. Por otra parte las interacciones del factor A, con B y C, muestran diferencias significativas; y la interacción de B con C muestra diferencias altamente significativas. No hay interacción de los tres factores (ABC).

Problemario de Arreglos Factoriales por Modelos Factoriales

Resolver los ejercicios 36, 37, 38, 39 y 40

Referencias

Arnau Gras, J.. Diseños Experimentales Multivariados. Alianza Editorial, Madrid. 1990.

Cochran, W.G. y Cox, G.M. Diseños Experimentales. Trillas, México. 1995.

Cox, D.R. Planning of Experiments. Wiley, New York. 1992.

Dean, A. y Voss, D. Design and Analysis of Experiments. Springer-Verlag, New York. 1999.

Douglas C. Montgomery, "Diseño y Análisis de Experimentos". Grupo Editorial Iberoamerica, 1991.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C.: "Análisis Multivariante", Ed. Prentice Hall, 5ª edición. 1999,

Hinkelmann, K. Design and analysis of experiments. John Wiley & Sons, N.Y. (2005).

Kempthorne O. The Design and Analysis of Experiments. John Wiley and Sons. , New York, 1952,

Lindman, H.R. Analysis of Variance in Experimental Design. Springer-Verlag, N.Y. 1992.

Mason, R.L., Gunst, R.F. and Hess, J.L. Statistical Design and Analysis of Experiments. Wiley, N.Y. 2003.

Montgomery, D.C. Diseño y Análisis de Experimentos. Limusa, México. 2003.

Rencher, A.C. Multivariate Statistical Inference and Applications. 1998.



Sixto Ríos. "Métodos Estadísticos", Ediciones del
Castillo, Madrid. 1977.



Tablas



DISTRIBUCION "t"

Prueba de Dos Colas									
Probabilidad de valor más alto. Signo no tomado en cuenta									
G.L.E.	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001
1	1	1.376	3.078	6.314	12.706	25.452	63.657		
2	0.816	1.061	1.886	2.92	4.303	6.205	9.925	14.089	31.598
3	0.765	0.978	1.68	2.353	3.182	4.176	5.841	7.453	12.941
4	0.741	0.941	1.533	2.132	2.776	3.495	4.604	5.598	8.61
5	0.727	0.92	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	0.718	0.906	1.44	1.943	2.447	2.969	3.707	4.317	5.959
7	0.711	0.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	0.706	0.889	1.397	1.86	2.306	2.752	3.355	3.832	5.041
9	0.703	0.883	1.383	1.833	2.262	2.685	3.25	3.69	4.781
10	0.7	0.879	1.372	1.812	2.228	2.634	3.169	3.5581	4.587
11	0.697	0.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	0.695	0.873	1.356	1.782	2.179	2.56	3.055	3.428	4.318
13	0.694	0.87	1.35	1.771	2.16	2.533	3.012	3.372	4.221
14	0.692	0.868	1.345	1.761	2.145	2.51	2.977	3.326	4.14
15	0.691	0.866	1.341	1.753	2.131	2.49	2.947	3.286	4.073
16	0.69	0.865	1.337	1.746	2.12	2.473	2.921	3.252	4.015
17	0.689	0.863	1.333	1.74	2.11	2.458	2.898	3.222	3.965
18	0.688	0.862	1.33	1.734	2.101	2.445	2.878	3.197	3.922
19	0.688	0.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	0.687	0.86	1.325	1.725	2.086	2.423	2.845	3.153	3.85
21	0.686	0.859	1.323	1.721	2.08	2.414	2.831	3.135	3.819
22	0.686	0.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	0.685	0.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	0.685	0.857	1.318	1.711	2.064	2.391	2.797	3.09	3.745
25	0.684	0.856	1.316	1.708	2.06	2.385	2.787	3.078	3.725
26	0.684	0.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	0.684	0.855	1.314	1.703	2.052	2.373	2.771	3.056	3.69
28	0.683	0.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	0.683	0.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	0.683	0.854	1.31	1.697	2.042	2.36	2.75	3.03	3.646
35	0.682	0.852	1.306	1.69	2.03	2.342	2.724	2.996	3.591
40	0.681	0.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	0.68	0.85	1.301	1.68	2.014	2.319	2.69	2.952	3.52
50	0.68	0.849	1.299	1.676	2.008	2.31	2.678	2.937	3.496
55	0.679	0.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	0.679	0.848	1.296	1.671	2	2.299	2.66	2.915	3.46
70	0.678	0.847	1.294	1.667	1.994	2.29	2.648	2.899	3.435
80	0.678	0.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	0.678	0.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	0.677	0.846	1.29	1.661	1.982	2.276	2.625	2.871	3.39
120	0.677	0.845	1.289	1.658	1.98	2.27	2.617	2.86	3.373
□	0.6745	0.8416	1.2816	1.6448	1.96	2.2414	2.5758	2.807	3.2905



Tabla "F"

Valores "F" para $\alpha = 0.05$ y 0.01 para diferentes combinaciones de grados de libertad del numerador y del denominador																			
Gld		Grados de Libertad del númeroador																	
	a	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	30	120
1	0.05	161	200	216	225	230	234	237	239	241	242	243	244	244	245	246	248	250	253
	0.01	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6083	6106	6125	6142	6157	6209	6261	6339
2	0.05	18.50	19.00	19.20	19.20	19.30	19.30	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.50	19.50
	0.01	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.50	99.50
3	0.05	10.10	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.72	8.71	8.70	8.66	8.62	8.55
	0.01	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20	27.10	27.10	27.00	26.90	26.90	26.70	26.50	26.20
4	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.89	5.87	5.86	5.80	5.75	5.66
	0.01	21.20	18.00	16.70	16.00	15.50	15.20	15.00	14.80	14.70	14.50	14.50	14.40	14.30	14.30	14.20	14.00	13.80	13.60
5	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.65	4.63	4.62	4.56	4.50	4.40
	0.01	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10	9.96	9.89	9.82	9.77	9.72	9.55	9.38	9.11
6	0.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.02	4.00	3.97	3.95	3.94	3.87	3.81	3.70
	0.01	13.70	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56	7.40	7.23	6.97
7	0.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.52	3.51	3.44	3.38	3.27
	0.01	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.53	6.47	6.41	6.36	6.31	6.16	5.99	5.74
8	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.25	3.23	3.22	3.15	3.08	2.97
	0.01	11.30	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52	5.36	5.20	4.95
9	0.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.04	3.02	3.01	2.94	2.86	2.75
	0.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.25	5.18	5.11	5.05	5.00	4.96	4.81	4.65	4.40
10	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.88	2.86	2.85	2.77	2.70	2.58
	0.01	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56	4.41	4.25	4.00
11	0.05	0.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.81	2.79	2.76	2.73	2.72	2.65	2.57	2.45
	0.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25	4.10	3.94	3.69
12	0.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.71	2.69	2.65	2.63	2.62	2.54	2.47	2.34
	0.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01	3.86	3.70	3.45
13	0.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.57	2.55	2.53	2.46	2.38	2.25
	0.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.90	3.86	3.82	3.66	3.51	3.25
14	0.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.56	2.53	2.50	2.48	2.46	2.39	2.31	2.18
	0.01	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.74	3.70	3.66	3.51	3.35	3.09
15	0.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.50	2.48	2.44	2.42	2.40	2.33	2.25	2.11
	0.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52	3.37	3.21	2.96
16	0.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.39	2.37	2.35	2.28	2.19	2.06
	0.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.50	3.45	3.41	3.26	3.10	2.84
17	0.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.32	2.31	2.23	2.15	2.01
	0.01	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31	3.16	3.00	2.75
18	0.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.28	2.27	2.19	2.11	1.97
	0.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.31	3.27	3.23	3.08	2.92	2.66
19	0.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.33	2.31	2.27	2.25	2.23	2.16	2.07	1.93
	0.01	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15	3.00	2.84	2.58
20	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.30	2.28	2.24	2.22	2.20	2.12	2.04	1.90
	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09	2.94	2.78	2.52
30	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.12	2.09	2.06	2.03	2.01	1.93	1.79	1.68
	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.90	2.84	2.79	2.74	2.70	2.55	2.30	2.11
120	0.05	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.86	1.83	1.80	1.77	1.75	1.66	1.55	1.35
	0.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.39	2.34	2.28	2.23	2.19	2.03	1.86	1.53



DUNCAN

Amplitudes studentizadas significativas para 5% y 1% de la nueva prueba de amplitud múltiple															
GLE	P= Número de medias para la amplitud a probarse														
	a	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	0.05 0.01	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00	18.00 90.00
2	0.05 0.01	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00	6.09 14.00
3	0.05 0.01	4.50 8.26	4.50 8.50	4.50 8.60	4.50 8.70	4.50 8.80	4.50 8.90	4.50 8.90	4.50 9.00	4.50 9.00	4.50 9.00	4.50 9.10	4.50 9.20	4.50 9.30	4.50 9.30
4	0.05 0.01	3.93 6.51	4.01 6.80	4.02 6.90	4.02 7.00	4.02 7.10	4.02 7.10	4.02 7.20	4.02 7.20	4.02 7.30	4.02 7.30	4.02 7.40	4.02 7.40	4.02 7.50	4.02 7.50
5	0.05 0.01	3.64 5.70	3.74 5.96	3.79 6.11	3.83 6.18	3.83 6.26	3.83 6.33	3.83 6.40	3.83 6.44	3.83 6.50	3.83 6.60	3.83 6.60	3.83 6.70	3.83 6.70	3.83 6.80
6	0.05 0.01	3.46 5.24	3.58 5.51	3.64 5.65	3.68 5.73	3.68 5.81	3.68 5.88	3.68 5.95	3.68 6.00	3.68 6.00	3.68 6.10	3.68 6.20	3.68 6.20	3.68 6.30	3.68 6.30
7	0.05 0.01	3.35 4.95	3.47 5.22	3.54 5.37	3.58 5.45	3.60 5.53	3.61 5.61	3.61 5.69	3.61 5.73	3.61 5.80	3.61 5.80	3.61 5.90	3.61 5.90	3.61 6.00	3.61 6.00
8	0.05 0.01	3.26 4.74	3.39 5.00	3.47 5.14	3.52 5.23	3.55 5.32	3.56 5.40	3.56 5.47	3.56 5.51	3.56 5.50	3.56 5.60	3.56 5.70	3.56 5.70	3.56 5.80	3.56 5.80
9	0.05 0.01	3.20 4.60	3.34 4.86	3.41 4.99	3.47 5.08	3.50 5.17	3.52 5.25	3.52 5.32	3.52 5.36	3.52 5.40	3.52 5.50	3.52 5.50	3.52 5.60	3.52 5.70	3.52 5.70
10	0.05 0.01	3.15 4.48	3.30 4.73	3.37 4.88	3.43 4.96	3.46 5.06	3.47 5.13	3.47 5.20	3.47 5.24	3.47 5.28	3.47 5.36	3.47 5.42	3.47 5.48	3.47 5.54	3.48 5.55
11	0.05 0.01	3.11 4.39	3.27 4.63	3.35 4.77	3.39 4.86	3.43 4.94	3.44 5.01	3.45 5.06	3.46 5.12	3.46 5.15	3.46 5.24	3.46 5.28	3.46 5.34	3.47 5.38	3.48 5.39
12	0.05 0.01	3.08 4.32	3.23 4.50	3.33 4.68	3.36 4.76	3.40 4.84	3.42 4.92	3.44 4.96	3.44 5.02	3.46 5.07	3.46 5.13	3.46 5.17	3.46 5.22	3.47 5.24	3.48 5.26
13	0.05 0.01	3.06 4.26	3.21 4.48	3.30 4.62	3.35 4.69	3.38 4.74	3.41 4.84	3.42 4.88	3.44 4.94	3.45 4.98	3.45 5.04	3.46 5.08	3.46 5.13	3.47 5.14	3.47 5.15
14	0.05 0.01	3.03 4.21	3.18 4.42	3.27 4.55	3.33 4.63	3.37 4.70	3.39 4.78	3.41 4.83	3.42 4.87	3.44 4.91	3.45 4.96	3.46 5.00	3.46 5.04	3.47 5.06	3.47 5.07
15	0.05 0.01	3.01 4.17	3.16 4.37	3.25 4.50	3.31 4.58	3.36 4.64	3.38 4.72	3.40 4.77	3.42 4.81	3.43 4.84	3.44 4.90	3.44 4.94	3.45 4.97	3.46 4.99	3.47 5.00
16	0.05 0.01	3.00 4.13	3.15 4.34	3.23 4.45	3.30 4.54	3.34 4.60	3.37 4.67	3.39 4.72	3.41 4.76	3.43 4.79	3.44 4.84	3.45 4.88	3.46 4.91	3.47 4.93	3.47 4.94
17	0.05 0.01	2.98 4.10	3.13 4.30	3.22 4.41	3.28 4.50	3.33 4.56	3.36 4.63	3.38 4.68	3.40 4.72	3.42 4.75	3.44 4.80	3.45 4.83	3.46 4.86	3.47 4.88	3.47 4.89
18	0.05 0.01	2.97 4.07	3.12 4.27	3.21 4.38	3.27 4.46	3.32 4.53	3.35 4.59	3.37 4.64	3.39 4.68	3.41 4.71	3.43 4.76	3.45 4.79	3.46 4.82	3.47 4.84	3.47 4.85
19	0.05 0.01	2.96 4.05	3.11 4.24	3.19 4.35	3.26 4.43	3.31 4.50	3.35 4.56	3.37 4.61	3.39 4.64	3.41 4.67	3.43 4.72	3.44 4.76	3.46 4.79	3.47 4.81	3.47 4.82
20	0.05 0.01	2.95 4.02	3.10 4.22	3.18 4.33	3.25 4.40	3.30 4.47	3.34 4.53	3.36 4.58	3.38 4.61	3.40 4.65	3.43 4.69	3.44 4.73	3.46 4.76	3.46 4.78	3.47 4.79
22	0.05 0.01	2.93 3.99	3.08 4.17	3.17 4.28	3.24 4.36	3.29 4.42	3.32 4.48	3.34 4.53	3.37 4.57	3.39 4.60	3.42 4.65	3.44 4.68	3.45 4.71	3.46 4.74	3.47 4.75
24	0.05 0.01	2.92 3.96	3.07 4.14	3.15 4.24	3.22 4.33	3.28 4.39	3.31 4.44	3.34 4.49	3.37 4.53	3.38 4.57	3.41 4.62	3.44 4.64	3.45 4.67	3.46 4.70	3.47 4.72
26	0.05 0.01	2.91 3.93	3.06 4.11	3.14 4.21	3.21 4.30	3.27 4.36	3.30 4.41	3.34 4.46	3.36 4.50	3.38 4.53	3.41 4.58	3.43 4.62	3.45 4.65	3.46 4.67	3.47 4.69
28	0.05 0.01	2.90 3.91	3.04 4.08	3.13 4.18	3.20 4.28	3.26 4.34	3.30 4.39	3.33 4.43	3.35 4.47	3.37 4.51	3.40 4.56	3.43 4.60	3.45 4.62	3.46 4.65	3.47 4.67
30	0.05 0.01	2.89 3.89	3.04 4.06	3.12 4.16	3.20 4.22	3.25 4.32	3.29 4.36	3.32 4.41	3.35 4.45	3.37 4.48	3.40 4.54	3.43 4.58	3.44 4.61	3.46 4.63	3.47 4.65
40	0.05 0.01	2.86 3.82	3.01 3.99	3.10 4.10	3.17 4.17	3.22 4.24	3.27 4.30	3.30 4.34	3.33 4.37	3.35 4.41	3.39 4.46	3.42 4.51	3.44 4.54	3.46 4.57	3.47 4.59
60	0.05 0.01	2.83 3.76	2.98 3.92	3.08 4.03	3.14 4.12	3.20 4.17	3.24 4.23	3.28 4.27	3.31 4.31	3.33 4.34	3.37 4.39	3.40 4.44	3.43 4.47	3.45 4.50	3.47 4.53
100	0.05 0.01	2.80 3.71	2.95 3.86	3.05 3.98	3.12 4.06	3.18 4.11	3.22 4.17	3.26 4.21	3.29 4.25	3.32 4.29	3.36 4.35	3.40 4.38	3.42 4.42	3.45 4.45	3.47 4.48
□	0.05 0.01	2.77 3.64	2.92 3.80	3.02 3.90	3.09 3.98	3.15 4.04	3.19 4.09	3.23 4.14	3.26 4.17	3.29 4.20	3.34 4.26	3.38 4.31	3.41 4.34	3.44 4.38	3.47 4.41



TUKEY

Puntos porcentuales superiores de la amplitud studentizada [$q_{\alpha} = (\bar{y}_{\max} - \bar{y}_{\min}) / S_{\bar{y}}$]																				
GLE		P= Número de medias para la amplitud a probarse																		
	a	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
	0.01	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	0.05	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.49	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.09	7.17
	0.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
	0.01	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	0.05	3.20	3.96	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
	0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.32	8.41	8.49	8.57
10	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.20	6.27	6.34	6.40	6.47
	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.48	7.60	7.71	7.81	7.91	7.99	8.07	8.15	8.22
11	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.99	6.06	6.14	6.20	6.26	6.33
	0.01	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.62	5.71	5.80	5.88	5.95	6.03	6.09	6.15	6.21
	0.01	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.05	6.11
	0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.34	7.42	7.48	7.55
14	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.72	5.79	5.85	5.92	5.97	6.03
	0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.12	7.20	7.27	7.33	7.39
15	0.05	3.01	3.67	4.08	4.37	4.60	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.58	5.65	5.72	5.79	5.85	5.90	5.96
	0.01	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.72	5.79	5.84	5.90
	0.01	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	0.05	2.98	3.63	4.02	4.30	4.52	4.71	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
	0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.80	6.87	6.94	7.00	7.05
18	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.72	6.79	6.85	6.91	6.96
19	0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
	0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.29	6.37	6.45	6.52	6.59	6.65	6.71	6.76	6.82
24	0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.54	5.59
	0.01	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	0.05	2.89	3.49	3.84	4.10	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48
	0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.91	4.98	5.05	5.11	5.16	5.22	5.27	5.31	5.36
	0.01	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.16	5.20	5.24
	0.01	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.79	5.84	5.89	5.93	5.98	6.02
120	0.05	2.80	3.35	3.69	3.92	4.10	4.24	4.36	4.48	4.56	4.64	4.72	4.78	4.84	4.90	4.95	5.00	5.05	5.09	5.13
	0.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.51	5.56	5.61	5.66	5.71	5.75	5.79	5.83
□	0.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01
	0.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65



DUNNETT

Comparaciones entre p trat. y un control. Niveles de confianza de 0.95 y 0.99										
GLE		P= Número de medias para la amplitud a probarse								
Err	p	1	2	3	4	5	6	7	8	9
5	0.95	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
	0.99	4.03	4.63	5.09	5.44	5.73	5.97	6.18	6.36	6.53
6	0.95	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
	0.99	3.71	4.22	4.60	4.88	5.11	5.30	5.47	5.61	5.74
7	0.95	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3.86
	0.99	3.50	3.95	4.28	4.52	4.71	4.87	5.01	5.13	5.24
8	0.95	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
	0.99	3.36	3.77	4.06	4.27	4.44	4.58	4.70	4.81	4.90
9	0.95	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
	0.99	3.25	3.63	3.90	4.09	4.24	4.37	4.48	4.57	4.65
10	0.95	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
	0.99	3.17	3.53	3.78	3.95	4.10	4.21	4.31	4.40	4.47
11	0.95	2.20	2.53	2.76	2.92	3.05	3.15	3.24	3.31	3.38
	0.99	3.11	3.45	3.68	3.85	3.98	4.09	4.18	4.26	4.33
12	0.95	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
	0.99	3.05	3.39	3.61	3.76	3.89	3.99	4.08	4.15	4.22
13	0.95	2.16	2.48	2.69	2.84	2.96	3.06	3.14	3.21	3.27
	0.99	3.01	3.33	3.54	3.69	3.81	3.91	3.99	4.06	4.13
14	0.95	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
	0.99	2.98	3.29	3.49	3.64	3.75	3.84	3.92	3.99	4.05
15	0.95	2.13	2.44	2.64	2.79	2.90	2.99	3.07	3.13	3.19
	0.99	2.95	3.25	3.45	3.59	3.70	3.79	3.86	3.93	3.99
16	0.95	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
	0.99	2.92	3.22	3.41	3.55	3.65	3.74	3.82	3.88	3.93
17	0.95	2.11	2.41	2.61	2.75	2.85	2.94	3.01	3.06	3.13
	0.99	2.90	3.19	3.38	3.51	3.62	3.70	3.77	3.83	3.89
18	0.95	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
	0.99	2.88	3.17	3.35	3.48	3.58	3.67	3.74	3.80	3.85
19	0.95	2.09	2.39	2.58	2.72	2.82	2.90	2.97	3.04	3.09
	0.99	2.86	3.15	3.33	3.46	3.55	3.64	3.70	3.76	3.81
20	0.95	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
	0.99	2.85	3.13	3.31	3.43	3.53	3.61	3.67	3.73	3.78
24	0.95	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
	0.99	2.80	3.07	3.24	3.36	3.45	3.52	3.58	3.64	3.69
30	0.95	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
	0.99	2.75	3.01	3.17	3.28	3.37	3.44	3.50	3.55	3.59
40	0.95	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
	0.99	2.70	2.95	3.10	3.21	3.29	3.36	3.41	3.46	3.50
60	0.95	2.00	2.27	2.43	2.55	2.63	2.70	2.76	2.81	2.85
	0.99	2.66	2.90	3.04	3.14	3.22	3.28	3.33	3.38	3.42
120	0.95	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
	0.99	2.62	2.84	2.98	3.08	3.15	3.21	3.25	3.30	3.33
□	0.95	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75
	0.99	2.58	2.79	2.92	3.01	3.08	3.14	3.18	3.22	3.25