UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"ANALISIS DE RESIDUALES EN DATOS CON DOBLE CLASIFICACION"

T E S I S

QUE PARA OBTENER

EL TITULO DE :

A C T U A R I O

P R E S E N T A

VICTOR ALFREDO BUSTOS Y DE LA TIJERA

DICIEMBRE 1974

A MIS PADRES

NATALIA

ANTONIO

CON INFINITO AGRADECIMIENTO

A MI NOVIA EVANGELINA
CON TODO MI AMOR

A MIS HERMANOS

JOSE ANTONIO

MARIA EUGENIA

LETICIA

LOURDES

SERGIO

Agradezco al Dr. Ignacio Méndez Ramírez la dirección de este trabajo así como la proposición del tema. Asimismo agradezco al CIMAS la ayuda prestada en el tiempo en que este trabajo se llevó a cabo y las facilidades brindadas con respecto al uso de tiempo de computadora. También a los Sres. Act. Miguel Cervera Flores, Dr. Federico O'Reilly Togno, Act. Santiago Rincón Gallardo, Act. Lucio Pérez Rodríguez la revisión de este trabajo.

I N D I C E

CAPITULO 1

1.- OBJETIVO

El objetivo del presente trabajo es investigar las propiedades distribucionales del error en un modelo de diseño de experimentos con dos criterios de clasificación y una sola observación por celda, cuando los efectos de interacción existen y son estimados mediante el método propuesto por el Dr. John Mandel, así el modelo propuesto es:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^{K} \theta_r u_{ki} v_{kj} + \epsilon_{ij}$$

Donde i se refiere al nivel del primer criterio en el que fué obtenida la información; j al nivel del segundo criterio,

$$\sum_{k=1}^{K} \theta_k u_{ki} v_{kj}$$

es la representación de la interacción; K es el número de términos de la forma $\theta_k u_{ki} v_{ki}$, este número no se fija de antemano sino que ha de determinarse de alguna forma; ϵ_{ij} es un error aleatorio, el cual se supone debe cumplir con las siguientes propiedades:

a) $E(\epsilon_{ij}) = 0$

b) $E(\epsilon_{ij}^2) = \sigma^2$

c) $E(\epsilon_{ij}\epsilon_{i'j'}) = 0$

El valor de K se determina mediante el examen de las propiedades de los ϵ_{ij}. Así se pretende encontrar el valor de K y en consecuencia la forma real de la relación que cumplen las observaciones con los criterios de acuerdo a los cuales se han clasificado las mismas. Es decir, una vez que las propiedades de los errores se cumplan satisfactoriamente diremos entonces que el modelo propuesto explica satisfactoriamente los datos. El criterio usado por el Dr. Mandel para fijar el valor de K consiste en ordenar los elementos $\hat{\theta}_k$ en forma decreciente, e incluir en el modelo solamente aquellos términos de la forma $\theta\, u_{ki}\, v_{kj}$ cuyo valor del cuadrado medio correspondiente es significativamente mayor que aquellos que le suceden en órden de magnitud.

Sin embargo no hace el Dr. Mandel referencia alguna a las hipótesis acerca de la distribución de los errores las cuales permitirán extraer información extra acerca de las observaciones además de las inferencias que acerca de los parámetros sea posible hacer.

Se hace a continuación una breve exposición del modelo propuesto por el Dr. Mandel y se trabaja un criterio para determinar la forma real del modelo a ser ajustado a los datos. Se discuten los métodos usados y se presenta la discusión de los resultados obtenidos al aplicar todos los métodos a varios grupos de datos.

CAPITULO 2

2.- INTRODUCCION

MODELOS

A lo largo de la historia, el hombre siempre ha intentado relacionar los fenómenos que observa con las causas de estos; así pués, ha propuesto relaciones ó modelos entre los fenómenos y sus posibles explicaciones.

Acerca de lo anterior se transcribe lo que al respecto señalan los autores Hillier y Lieberman (4):

"Los modelos son parte integral de la vida común y corriente. Ejemplos simples de modelos, pueden ser representaciones tales como aviones a escala, retratos, globos terráqueos, etc. Por otra parte, los modelos juegan un papel muy importante en la ciencia o en los negocios; así pués, el hombre ha creado también modelos del átomo, modelos de la estructura genética en los seres vivos, modelos matemáticos para describir leyes físicas de movimiento o reacciones químicas, gráficas y sistemas contables industriales. Tales modelos son de inmenso valor para abstraer la esencia del objeto por cuyas características nos preguntamos, mostrando interrelaciones y facilitando el análisis.

MODELOS MATEMATICOS

"Los modelos matemáticos son también representaciones idealizadas, pero son expresados en términos de símbolos matemáticos y expresiones. Leyes físicas tales como:

$$F = m a$$

o

$$E = m c^2$$

Son ejemplos con los que estamos familiarizados".

Más adelante esos mismos autores escriben:

Los modelos matemáticos tienen muchas ventajas sobre una descrinción verbal del problema. Una ventaja obvia es que el modelo matemático describe el problema en una forma mucho más concisa. Esto tiende a hacer la estructura total del problema comprensible y ayuda a encontrar importantes relaciones causa-efecto. De esta manera indica más claramente cuáles datos adicionales resultan relevantes para el análisis".

Una observación crucial para el desarrollo de la presente tesis es hecha por los autores antes mencionados, en el siguiente párrafo:

"Un modelo matemático es necesariamente una idealización abstracta del problema, y generalmente se requiere de suposiciones y aproximaciones que simplifican el modelo haciéndolo manejable. De ahí que se debe tener mucho cuidado de asegurarse que el modelo sigue siendo una representación válida del fenómeno. El criterio adecuado para juzgar la validez de un modelo es determinar si predice o no los efectos relativos de los cursos alternativos de acción con precisión suficiente para dar lugar a una decisión adecuada. De ahí que no es necesario incluir

detalles irrelevantes o factores que tienen aproximadamente el mismo efecto para todos los cursos alternativos de acción".

En su libro "Introducción a la metodología estadística" el Dr. Ignacio Méndez Ramírez (10) hace las siguientes observaciones acerca de los modelos matemáticos:

"Los modelos matemáticos, como los puramente simbólicos, deberán ser probados en la práctica para que puedan considerarse útiles; este proceso de pruebas es también, como parte del método científico, un proceso helicoidal que consta de los siguientes pasos:

"1.- Problemas y situaciones del mundo real se trasladan al modelo cómo problemas y situaciones matemáticas. Esto se hace mediante una abstracción; así de todas las características del mundo real, se representan en el modelo aquellas que se consideran relevantes para el fenómeno en estudio. Así en las leyes que ligan la masa de los cuerpos con otras variables, usualmente no interesa la forma, color, etc., de los cuerpos."

"2.- El modelo matemático debe ser funcional matemáticamente, en el sentido de que un problema del mundo real pueda ser trasladado a un problema matemático y que este último pueda ser resuelto manejando la lógica y postulados matemáticos inherentes al modelo."

"3.- Las soluciones matemáticas obtenidas por el modelo se interpretan términos de soluciones plausibles de los problemas rea-

les."

En la misma referencia se presenta el siguiente diagrama de funcionamiento del modelo:

MUNDO REAL		MUNDO SIMBOLICO
DATOS	- - - - - - - - - >	MODELO
PROBLEMAS	- - - - - - - - - >	PROBLEMAS AL MODELO
SOLUCION A PROBLEMAS	< - -	SOLUCION MATEMATICA

TAXONOMIA DE MODELOS MATEMATICOS

Dentro de los modelos matemáticos podemos establecer la siguiente taxonomía:

1.- Modelos determínisticos

2.- Modelos aleatorios

Los primeros pretenden, y a veces lo consiguen, establecer relaciones causa-efecto entre ciertas características del fenómeno en estudio y los resultados observados, así por ejemplo, en el modelo:

$$E = m c^2$$

basta con conocer la masa de un cuerpo (m) para determinar la energía (E) que es posible liberar de ese cuerpo. Obsérvese que modelos como el anterior dependen en la mayor parte de los casos, de constantes o parámetros que las más de las veces es necesario determinar: así el ejemplo arriba citado depende de el parámetro c^2 (la velocidad de la luz, en centímetros/segundo al cuadrado). La determinación de estos parámetros puede ser hecha

en base a experimentación, a repetición del experimento variando las condiciones en que este es realizado y obteniendo distintas mediciones de los efectos; y así a partir del modelo determinar los parámetros del mismo.

Los modelos aleatorios (o estocásticos o probabilísticos) por su parte no pretenden establecer relaciones causa-efecto entre los resultados de un experimento y las condiciones en que este fué realizado. Este tipo de modelos incluye un efecto o error aleatorio debido a las condiciones no controladas en las que fué realizado el experimento, proporcionando a su vez una idea de la distribución espacial de este error. Se puede decir que un modelo aleatorio es un modelo determinístico aumentado con una explicación de las fluctuaciones o variaciones observadas al realizar el experimento en condiciones más o menos homogéneas.

Es decir, supóngase que las condiciones A, B, C, D, E, ..., Z tienen una mayor o menor influencia con respecto a los resultados de un cierto fenómeno. Entonces el modelo aleatorio puede ser planteado como sigue:

$$Y_{ij...s} = f(A_i, B_j, \ldots, Z_r) + \psi_s$$

Donde A_i significa: el criterio A en su nivel i.

B_j significa: el criterio B en su nivel j, etc.

Y ψ_{ij} es el error o la explicación aleatoria de las fluctuaciones observadas al realizar repetidas veces el experimento monte-

niendo A, B, ..., Z en los mismos niveles; Ψ_{ij} también puede ser un error de medición.

Ahora bien puede suponerse que los criterios G, H, I, ..., Z tienen pequeña influencia en los resultados del experimento, lo que nos permitiría reescribir el modelo como:

$$Y_{ijk...s} = f_1 (A_i, B_j, ..., F_k) + \varepsilon_s$$

Donde ε_s incluye tanto al error aleatorio Ψ_s cómo la "pequeña" influencia de los criterios G,H, ..., Z .

En este caso es también de observarse que la forma analítica de la función f_1 depende también de un cierto número de parámetros, mismos que a su vez han de ser determinados de alguna manera (estimados). Sin embargo en este tipo de modelos, con el objeto de explicar las desviaciones aleatorias, se debe contar también con un modelo acerca de la distribución de las variables aleatorias ε_s, la forma de la distribución de las variables aleatorias nos indicará a su vez cuantos parámetros es preciso determinar.

La consideración de un elemento aleatorio en el modelo da a su vez un carácter aleatorio a las observaciones, y en consecuencia es preciso determinar la forma de su distribución y los parámetros de los que esta depende.

Con el objeto de aclarar lo antes mencionado se incluye el siguiente ejemplo:

El rendimiento de trigo en una parcela no puede ser predicho con precisión. Muchos de los factores que afectan este rendimiento

son conocidos, pero la ecuación que relaciona estas cantidades es obscura. Se sabe que la temperatura X_1, lluvia X_2, cantidad de luz solar X_3 y muchos otros factores afectan el rendimiento Y. Aunque todos los factores que afectan este rendimiento no se conoce y aunque la relación funcional tampoco se conoce, es útil sin embargo suponer que existe un número finito de factores $X_1, X_2, \ldots, X_n$ y una función g tal que el rendimiento Y puede ser exactamente determinado mediante:

$$Y = g \quad (X_1, \ldots, X_n)$$

Sin embargo si consederamos que ciertos factores tales como la distancia entre el sol y la tierra en el momento de la siembra, la cantidad de rayos gamma absorbidos por las plantas durante su crecimiento, etc., tienen poca influencia sobre el resultado final, las variables $X_6, \ldots, X_n$ podrían ser ignoradas y su influencia podría ser considerada como aleatoria, así el modelo podría ser expresado como:

$$Y = g_1 \quad (X_1, \ldots, X_5) \quad + \varepsilon$$

Donde ε sigue una distribución normal con media cero y dispersión σ común. Así el modelo completo quedaría expresado como:

$$Y = g\,(X_1, \ldots, X_5) \quad + \varepsilon$$
$$\varepsilon \sim N\,(0, \sigma^2)$$

MODELO LINEAL

Una parte importante de los modelos estocásticos la componen los modelos llamados lineales; una definición de estos modelos esta dada en el libro de Graybill (12).

"Definición 5.2- Por un modelo lineal entenderemos una ecuación que incluye variables aleatorias, variables matemáticas y parámetros, y que es lineal en los parámetros y en las variables aleatorias".

En el ejemplo anterior podemos suponer:

$$Y=g_1(X_1,\ldots,X_5)+\varepsilon = \alpha_0+\alpha_1X_1+\alpha_2X_1^2+\alpha_3X_2+\alpha_4X_3X_4+\alpha_5\log X_5+\varepsilon$$

En este caso podemos decir que nuestro modelo es lineal en el sentido de la definición arriba mencionada.

Una clasificación de estos modelos está dada en el libro de Graybill; aparece reproducida en seguida:

"Modelo 1.- Suponganos que $X_1, X_2, \ldots, X_n$ son variables matemáticonocidas, Y una variable aleatoria, $\alpha_0, \alpha_1, \ldots, \alpha_n$ son parámetros desconocidos y ε una variable aleatoria no observable con media 0. Bajo estas condiciones el modelo:

$$Y= \alpha_0 + \alpha_1X_1+ \ldots+\alpha_n X_n +\varepsilon$$

se define como el modelo 1.

"Modelo 2.- Modelos funcionalmente relacionados con variables sujetas a errores de medida. Supónese que las variables matemáticas $Y, X_1, X_2, \ldots, X_n$ son no observables (medidas con error), pero que $y, x_1, x_2, \ldots, x_n$ pueden ser observados, don-

de $y=Y+\varepsilon$, $x_i = X_i + \varepsilon_i$

Suponga además que existe una relación funcional entre las variables matemáticas, dada por:

$$Y = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_n X_n$$

Substituyendo tenemos:

$$y = \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_n x_n + \varepsilon - \alpha_1 \varepsilon_1 - \ldots - \alpha_n \varepsilon_n$$

Cuando las especificaciones anteriores se cumplen, llamemos al modelo Modelo 2.

"Modelo 3.- Modelos de Regresión. Suponga $y, x_1, \ldots x_n$ son conjunto de variables aleatorias conjuntamente distribuidas tales que el valor esperado en la distribución condicional de y dados $x_i = X_i$ $(i= 1,2,\ldots, n)$ está dado por $\alpha_0 + \alpha_1 X_1 + \ldots + \alpha_n X_n$.

Podemos entonces escribir:

$$y_x = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_n X_n + \varepsilon$$

Cuando las condiciones arriba apuntadas se cumplen, definiremos el modelo como el Modelo 3.

"Modelo 4.- Modelos de Diseños Experimentales. Sea y una variable aleatoria, y sean $X_1, \ldots, X_n$ variables que toman valores cero ó uno. Si $\alpha_1, \alpha_2, \ldots, \alpha_n$ son parámetros desconocidos, entonces:

$$y = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n + \varepsilon$$

Lo llamaremos Modelo 4. Este modelo es un caso especial del Modelo 1, pero dada su importancia, será discutido por separado.

Este modelo será escrito como sigue:

$$y_{ij\cdots k} = \mu + \alpha_i + \beta_j + \cdots + \varepsilon_{ij\cdots k}$$

Donde α_i, β_j,..., son parámetros desconocidos y $\varepsilon_{ij\cdots k}$ es una variable aleatoria.

"Modelo 5.- Modelos de componentes de varianza. Sean $a_1, \ldots, a_p$ variables aleatorias no observables de una distribución con medio 0 y varianza σ_a^2; sean $b_{11}, b_{12}\cdots, b_{pq}$ variables aleatorias no observables de una distribución con media 0 y varianza σ_b^2. Sea Y_{ij} una variable aleatoria tal que:

$$y_{ij} = \mu + a_i + b_{ij}$$

Donde μ es un parámetro desconocido. Modelos como este serán llamados Modelo 5."

Modelo de Diseños Experimentales

En el presente trabajo estaremos interesados en estudiar casos especiales del Modelo 4 de Diseños Experimentales.

Un ejemplo de este modelo es el siguiente (Graybill, (12,p.223)). "Suponga que un fabricante de focos, buscando incrementar la duración de sus productos, ha desarrollado dos recubrimientos químicos para los filamentos de los focos. Si el filamento no es tratado con cubrimiento alguno, los focos durarán un promedio de μ (desconocido) hrs. El fabricante supone que, si trata un filamento con el cubrimiento 1, este aumentará la vida del filamento en τ_1 (desconocido) hrs., y que si usa el cubrimiento 2 este

aumentará la duración de τ_2 (desconocido) hrs.

Ahora bien, a él le gustaría encontrar τ_1 y τ_2 el número de horas que los cubrimientos 1 y 2 respectivamente prolongan la vida de los focos (τ_1 y τ_2 pueden ser cero o negativos), y $\tau_1 - \tau_2$, la diferencia entre ambos tratamientos. Para evaluar estas constantes el fabricante podría producir un foco con cubrimiento 1 y probarlo para observar su duración, podría entonces hacer un foco con cubrimiento 2 y observar su duración. El modelo puede ser escrito:

$$Y_1 = \mu + \tau_1 + \varepsilon_1$$

$$Y_2 = \mu + \tau_2 + \varepsilon_2$$

Donde Y_1 es el número de horas que dura el foco con cubrimiento 1, y ε_1 es un error aleatorio debido a todos los errores no controlados, tales como voltaje variable, imperfecciones en la manufactura de los distintos componentes, temperatura, humedad, etc.

"Ahora bien, el fabricante probablemente no estará contento en sacar conclusiones basadas en la observación de solamente un foco por recubrimiento. Supongamos que decide fabricar 6 focos y cubrir 3 con cubrimiento 1 y 3 con cubrimiento 2. El modelo puede ser escrito como:

$$Y_{11} = \mu + \tau_1 + \varepsilon_{11}$$

$$Y_{12} = \mu + \tau_1 + \varepsilon_{12}$$

$$Y_{13} = \mu + \tau_1 + \varepsilon_{13}$$
$$Y_{21} = \mu + \tau_2 + \varepsilon_{21}$$
$$Y_{22} = \mu + \tau_2 + \varepsilon_{22}$$
$$Y_{23} = \mu + \tau_3 + \varepsilon_{23}$$

Que expresado en forma matricial quedaría:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{bmatrix}$$

o $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ "

Es de hacerse notar que para este tipo de modelos el rango de la matriz X es siempre menor que el número de columnas. Entonces la matriz X'X que aparece frecuentemente no tiene inversa.

A el modelo anterior se le conoce como modelo de un diseño experimental con un criterio de clasificación (cubrimiento).

Ahora bien, en el presente trabajo estaremos interesados en los modelos de diseño de experimentos con dos clasificaciones y una sola observación por celda. Para este tipo de modelos nuestras observaciones pueden ser conjuntadas en una tabla como sigue:

CRITERIO 1

CRITERIO 2	1	2	3	...	n
1	Y_{11}	Y_{12}	Y_{13}	...	Y_{1n}
2	Y_{21}	Y_{22}	Y_{23}	...	Y_{2n}
3	Y_{31}	Y_{32}	Y_{33}	...	Y_{3n}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
m	Y_{m1}	Y_{m2}	Y_{m3}		Y_{mn}

Donde Y_{ij} es el resultado obtenido al realizar el experimento con el criterio 1 en su nivel J, y el criterio 2 en su nivel i.

n es el total de niveles considerados para el criterio 1.

m es el total de niveles considerados para el criterio 2.

Si lo que nos interesa medir es, por ejemplo, la absorción de rayos ultravioleta, los que hemos llamado criterios pueden ser, intensidad de emisión y distancia entre emisor y receptor.

Si en cambio nuestras observaciones se refieren a la densidad de soluciones acuosas del alcohol etílico, los criterios pueden ser temperatura (distintos niveles) y concentración.

Para este tipo de datos el modelo lineal general que se ajusta es el siguiente:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \eta_{ij} + \varepsilon_{ij}$$

$$i = 1, \ldots, m \qquad j = 1, \ldots, n \qquad (1)$$

$$E(\varepsilon_{ij}) = 0 \quad ; \quad E(\varepsilon_{ij}^2) =$$

$$E(\varepsilon_{ij}\varepsilon_{i'j'}) = 0$$

Es decir, se supone que las observaciones son explicadas por:

μ un efecto común a todas las observaciones cuyo valor es desconocido.

α_i el efecto de que el criterio 2 se encuentre en su nivel i. Es desconocido también, pero es función de i solamente.

β_j el efecto de tener el criterio 1 en su nivel j. Este efecto es también conocido; es función de j solemente.

η_{ij} el efecto de interacción de los niveles i y j de los criterios 2 y 1 respectivamente. Este efecto es deconocido y es función de i y de j simultáneamente. Su forma no es aditiva ya que todo lo que podía considerarse aditivo ha sido extraido por α_i y β_j.

ε_i error aleatorio debido a las características no controladas en el experimiento y aquellas variaciones naturales en el fenómeno que se estudia.

La interacción queda definida, según Graybill como:

"Definición.- f (X, Z) será definida como una función sin interacción si y sólo si existen funciones $g(X)$ y $h(Z)$ tales que:

$f(X, Z) = g(X) + h(Z)$"

Cabe hacerse notar que la estructura para Y_{ij} dada por el modelo (1), se simplifica enormemente si el efecto de interacción η_{ij} es igual a cero; en este caso el modelo (1) se simplifica a:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

$$i = 1, \ldots, n \qquad (2)$$

$$j = 1, \ldots, m$$

$$E(\varepsilon_{ij}) = 0, \quad E(\varepsilon_{ij}^2) = \sigma^2, \quad E(\varepsilon_{ij}\,\varepsilon_{i'j'}) = 0$$

Este modelo sin interacción se conoce también como el modelo estrictamente aditivo y tiene la ventaja de descomponer la función de dos variables $E(Y_{ij})$ en funciones de una sola variable (α_i, β_j).

La confusión de ambos modelos acarrea graves problemas a las inferencias y pruebas de hipótesis que se hacen con respecto a los parámetros del modelo. Bancroft (1) menciona algunos de los problemas a que lleva el uso del modelo equivocado: "Si efectos de interacción están presentes en la población, pero el modelo escogido fué un modelo aditivo estrictamente (como el (2)) entonces cualquier suma de cuadrados de los efectos principales es solamente un estimador aproximado de la magnitud poblacional de este efecto. Si por otro lado, no hay interacción entre los criterios en la población, pero se supuso que un modelo con interacción era el adecuado entonces ambas sumas de cuadrados de los efectos principles serán insesgadas pero ineficientemente estimadas".

En vista de lo anterior fué necesario desarrollar criterios para determinar la existencia o no de efectos de interacción en la población y en base a ellos elegir el modelo adecuado a ser ajusta-

do a los datos. Así Tukey (19) desarrolló un método para probar la hipótesis $H_o: \eta_{ij} = 0$, i=1,..., m, j=1,...,n, bajo la suposición de una forma especial de interacción.

Este método es explicado en Graybill (12), Rao (13). El método consiste brevemente en:

1.- Suma de cuadrados debida a no aditividad:

$$N_{ss} = \frac{\sum_{ij} Y_{ij} (\bar{Y}_{i.} - \bar{Y}..) (\bar{Y}_{.j} - \bar{Y}..)}{\sum_{i} (\bar{Y}_{i.} - \bar{Y}..)^2 \sum_{j} (\bar{Y}_{.j} - \bar{Y}..)^2}^2$$

2.- Suma de cuadrados debida a balance (residuo):

$$R_{ss} = \sum_{ij} (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}..)^2 - N_{ss}$$

Si $\eta_{ij} = 0$ para todos los valores de i y de j entonces N_{ss}/σ^2 se distribuye como χ^2 (1). R_{ss}/σ^2 se distribuye como $\chi^2_{((m-1)(n-1)-1)}$ y es independiente de N_{ss}/σ^2. Entonces bajo la hipótesis $H_o: \eta_{ij} = 0$ para todas i y j, la cantidad:

$$F_c = (N_{ss}/R_{ss}) \ ((m-1)(n-1)-1)$$

Se distribuye como F (1, (m-1) -1). Si $F_c > F_\alpha$ la hipótesis $H_o: \eta_{ij} = 0$ puede ser rechazada al nivel α.

La derivación de la prueba anterior puede ser resumida como sigue:

Si se supone que $\eta_{ij} = \lambda \alpha_i \beta_j$

La prueba H_0: $\eta_{ij} = 0$, para todo valor de i y de j puede ser escrita como:

Además, sea λ igual a cero o no se tiene:

$$E\ (\bar{Y}_{i.} - \bar{Y}_{..}) = \alpha_i$$

$$E\ (\bar{Y}_{.j} - \bar{Y}_{..}) = \beta_j$$

$$E\ (\bar{Y}_{..}) = \mu$$

De manera que estimadores insesgados de α_i, β_j, μ están dados por:

$$\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..}$$

$$\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}_{..}$$

$$\hat{\mu} = \bar{Y}_{..}$$

Además:

$$E(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}) = \lambda \alpha_i \beta_j$$

y si se supone que α_i y β_j son conocidas, se puede dar un estimador de mínimos cuadrados para λ, se obtiene:

$$\hat{\lambda} = \frac{\sum_i \sum_j \alpha_i \beta_j (Y_{ij} - \alpha_i - \beta_j - \bar{Y}_{..})}{\sum_i \sum_j \alpha_i^2 \beta_j^2}$$

Se sustituyen los estimadores $\hat{\alpha}_i$ y $\hat{\beta}_j$ por α_i y β_j respectivamente, en la expresión para $\hat{\lambda}$, se obtiene:

$$\hat{\lambda} = \frac{\sum_i \sum_j (Y_{ij})\ (\bar{Y}_{i.} - \bar{Y}..)\ (\bar{Y}_{.j} - \bar{Y}..)}{\sum_i (\bar{Y}_{i.} - \bar{Y}..)^2 \sum_j (\bar{Y}_{.j} - \bar{Y}..)^2}$$

con una varianza, para $\hat{\alpha}_i$ y $\hat{\beta}_j$ dados, igual a:

$$\frac{m^2 n^2 \sigma^2 \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}..)^2 \cdot (\bar{Y}_{.j} - \bar{Y}..)^2}{\left[\frac{1}{m}\sum_i \bar{Y}^2_{i.} - \frac{1}{mn}\bar{Y}^2..\right]^2 \left[\frac{1}{n}\sum_j \bar{Y}^2_{.j} - \frac{1}{mn}\bar{Y}^2..\right]^2}$$

o equivalentemente:

$$\frac{m\ n\ \sigma^2}{\left[\frac{1}{m}\sum_i \bar{Y}^2_{i.} - \frac{1}{mn}\bar{Y}^2..\right]\left[\frac{1}{n}\sum_j \bar{Y}^2_{.j} - \frac{1}{mn}\bar{Y}^2..\right]}$$

con la estadísitca:

$$N_{ss} = \frac{m\ n\ \left[\sum_i \sum_j Y_{ij}\ (\bar{Y}_{i.} - \bar{Y}..)\ (\bar{Y}_{.j} - \bar{Y}..)\right]^2}{\left[\frac{1}{m}\sum_i \bar{Y}_{i.} - \frac{\bar{Y}^2..}{m\ n}\right]\left[\frac{1}{n}\sum_j \bar{Y}_{.j} - \frac{\bar{Y}^2..}{m\ n}\right]^2} \sim \chi^2_{(1)}$$

pero:

$$\frac{\sum_i \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} - \bar{Y}..)^2}{\sigma^2} \sim \chi^2_{((m-1)(n-1)-1)}$$

Siguiéndose de aquí la construcción de R_{ss} y F_c además de la prueba.

Lo anterior es un resumen de los teoremas dados en Graybill (12, pp.324-331).

Debe hacerse énfasis en el hecho de que la prueba de Tukey supone una forma particular para el efecto de interacción η_{ij}. Como se verá en la siguiente sección esta forma para el efecto de interacción queda incluida como un caso muy particular de la forma propuesta por Mandel (8,9,), para expresar el efecto de interacción (un sólo termino multiplicativo con $\theta = \lambda$, $u_i = \alpha_i$ para toda i, $v_j = \beta_j$ para toda j). 1/

La necesidad de suponer una forma especial para el efecto de interacción se debe a que los estimadores, obtenidos mediante el método de mínimos cuadrados, sin hacer ninguna suposición extra acerca de los parámetros del modelo, ajustan perfectamente a nuestras observaciones, lo cual da como resultado que la suma de cuadrados debida al error sea cero.

Esta suma de cuadrados dividida entre sus grados de libertad es usada como el denominador de la estadística que se usa para las pruebas de hipótesis y como estimador de la varianza del error σ^2. Es obvio que si se toma el valor cero ninguna prueba de hipótesis puede ser llevada a cabo.

1/ Si en cada celda se tiene más de una observación, no es necesario hacer suposición alguna acerca de la forma que debe tener el efecto de interacción, es decir, la prueba de la hipótesis H_0: $\eta_{ij} = 0$ para todos los valores de i y de j, puede ser llevada a cabo sin suponer ninguna forma particular para η_{ij}.

En el caso en que no se tiene evidencia alguna, por experiencia anterior o por una muestra independiente de la actual, será preciso aplicar una prueba de este tipo para determinar el modelo que se ajusta mejor a la evidencia disponible (la muestra). Sin embargo la aplicación previa al análisis de los datos de una prueba tal, alterará los niveles de significancia de las pruebas que se realicen con esta misma muestra y con respecto a los parámetros del modelo usado. En efecto, Bancroft (1,p11) propone que la prueba de noaditividad sea realizada a un nivel de significancia del 25% si las pruebas siguientes han de realizarse al 5% para mantenerse en un nivel próximo al 5%.

Sin embargo, asegura, esto disminuirá la potencia de las últimas pruebas.

Es de esperarse que la anterior presentación dé un punto de vista suficientemente amplio para la mejor comprensión de lo que sigue.

CAPITULO 3

3.- EL MODELO GENERAL PROPUESTO POR MANDEL (8,9)

Antes de plantear la derivación y estimación del modelo, se incluyen algunas consideraciones que resultan ser interesantes.

A diferencia de otros métodos de estimación de la interacción entre renglones y columnas, el presente método se propone atacar el problèma de la interacción de una manera completamente sistemática, así se parte este término del modelo en tantos términos individuales como sean requeridos por los datos. No se hace ninguna suposición previa sobre linealidad. En buena medida son los mismos datos lo que generan el número de términos para representar la interacción en el modelo.

Este método se deriva del método de componentes principales usado para el análisis de factores en análisis estadístico multivariado. El estudio de esta técnica se debe principalmente a Harold Hotelling (14).

Por otro lado, mientras que el método presentado aquí se apoya en resultados matemáticos conocidos, el punto de vista es esencialmente novedoso.

En particular, mientras que practicamente todos los discusiones acerca de este método asignan funciones asimétricas a los renglones y a las columnas de la tabla, el presente enfoque trata a los renglones y a las columnas de la misma manera, quedando el papel de renglón o columna sujeto a una elección subjetiva de

los criterios que se han mencionado con anterioridad.

Además el presente método puede ser usado para probar aditividad, de efectos de los criterios; para ver esto es preciso hacer ciertas suposiciones acerca de la forma particular en que se expresa el efecto de interacción.

Supóngase que el efecto de interacción (η_{ij} en el modelo) puede ser partido y expresado como sigue:

$$\eta_{ij} = \theta_1 u_{1i} v_{1j} + \theta_2 u_{2i} v_{2j} + \cdots + \theta_k u_{ki} v_{kj} + \varepsilon_{ij} \qquad (3.1)$$

con $E(\varepsilon_{ij}) = 0$, $E(\varepsilon_{ij}^2) = \sigma^2$, $E(\varepsilon_{ij}\ \varepsilon_{i'j'}) = 0$

De tal modo que la función de 2 variables η_{ij} se expresa como una combinación de funciones de una sola variable (u_{ri}, v_{rj}) nuevamente, lo cual es realmente una simplificación.

Sin pérdida de generalidad se imponen las siguientes restricciones:

$$\sum_i u_{i1} = \sum_i u_{i2} = \cdots = \sum_j v_{j1} = \sum_j v_{j2} = \cdots = 0 \qquad (3.2)$$

$$\sum_i u_{i1}^2 = \sum_i u_{i2}^2 = \cdots = \sum_j v_{j1}^2 = \sum_j v_{j2}^2 = \cdots = 1$$

El problema esencial consiste en encontrar estimadores de los nuevos parámetros θ_1, θ_2, ..., u_{i1}, u_{i2}, ..., v_{j1}, v_{j2},... y para σ, la desviación estandard del error aleatorio.

La estimación de los parámetros se lleva a cabo haciendo uso del método de mínimos cuadrados. Se sabe que un estimador de mínimos cuadrados de el efecto de interacción, lo constituye el resi-

dual d_{ij} que esta dado por la siguiente relación:

$$d_{ij} = Y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$$

donde

$$\hat{\mu} = \bar{Y}_{..} = \frac{1}{mn} \sum_i \sum_j Y_{ij}$$

$$\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..} = \frac{1}{n} \sum_j Y_{ij} - \bar{Y}_{..}$$

$$\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}_{..} = \frac{1}{m} \sum_i Y_{ij} - \bar{Y}_{..}$$

Con el objeto de (estimar) ajustar la interacción supóngase primero que:

$$\eta_{ij} = \theta u_i v_j + \varepsilon_{ij}$$

Donde u_i y v_j están sujetos a las restricciones (3.2) los estimadores de mínimos cuadrados se obtienen minimizando la siguiente expresión:

$$G = \sum_i \sum_j (d_{ij} - \theta u_i v_j)^2 - \lambda_1 (\sum_i u_i^2 - 1) - \lambda_2 (\sum_j v_j^2 - 1)$$

$$- 2\mu_1 (\sum_i u_i) - 2\mu_2 (\sum_j v_j)$$

Donde λ_1, λ_2, μ_1, μ_2, son multiplicadores de Lagrange. Después de llevar a cabo los cálculos respectivos se obtienen las siguientes relaciones:

$$u_t = \theta^{-1} \sum_j d_{tj} v_j \qquad (3.4)$$

$$v_1 = \theta^{-1} \sum_i d_{i1} u_i \qquad (3.5)$$

sustituyendo (3.5) en (3.4) y definiendo

$$S_{it} = \sum_j d_{ij} d_{tj}$$

obtenemos

$$\theta^2 u_t = \sum_i u_i S_{it}$$

En notación matricial (3.6) y (3.7) se pueden expresar como:

$$S = D\,D' \quad , \quad S \in M_{mXm} \quad , \quad D \in M_{mXn} \qquad (3.6)'$$

donde D es la matriz de residuales y apostrofe representa la transpuesta. Entonces (3.7) se convierte en:

$$\theta^2 \underline{u} = S \underline{u} \quad , \quad \underline{u} \in R^m \qquad (3.7)'$$

De la anterior relación se desprende que θ^2 es un eigenvalor de la matriz S y $\underline{u}$, el eigenvector asociado. De manera similar puede demostrarse que:

$$\theta^2 \underline{v} = (D'D) \underline{v} \quad , \quad \underline{V} \in R^n$$

y así la minimización de G se logra tomando como θ^2 el eigenvalor más grande de la matriz S. Más aún, puede probarse el hecho de que, habiendo escogido como θ^2 el máximo del conjunto de eigenvalores, si se desea minimizar con respecto a θ_2, u_{2i}, v_{2j} la siguiente expresión.

$$\sum_i \sum_j ((d_{ij} - \theta_1 u_{1i} v_{1j}) - \theta_2 u_{2i} v_{2j})^2$$

con u_{2i}, v_{2j} sujetos a restricciones similares a (3.2) se tendrá que el valor mínimo de esa expresión se logra tomando a θ_2 como el segundo más grande eigenvalor de la matriz S original con $\underline{u}_2$, el eigenvector asociado, y con $\underline{v}_2$ en situación similar. Así pues, basta con obtener el conjunto total de eigenvalores de la matriz S= DD', el conjunto $\underline{u}$, $\underline{u}_2$, ... de eigenvectores asocia-

dos, para tener los estimadores de mínimos cuadrados para todos los parámetros de (3.1).

Por otra parte, ya que los residuales se obtienen a partir de los estimadores obtenidos por el método de mínimos cuadrados, se tiene que la matriz $DD' = S$ tiene a lo más, rango igual a min $(m,n) -1$; es decir, habrá a lo más min $(m,n) -1$ eigenvalores distintos de cero y en consecuencia se tendrá a lo más min $(m,n) -1$ términos del tipo. $\theta\, u_i\, v_j$.

El método para determinar el número de términos de la forma $\theta\, u_i\, v_j$ que se han de incluir en el modelo, y que fué desarrollado por el Dr. Mandel (8,9) es interesante e ilustrativo. Es en cierta forma una generalización del criterio usado en el método de componentes principales para determinar cuales son las combinaciones lineales de las componentes del vector de observaciones que explican la mayor varianza de las mismas observaciones. Este método es explicado por Mandel de la siguiente manera:

"Supóngase que calculamos todos los términos del tipo $\theta\, u_i\, v_j$, de modo que ε_{ij} es hecho igual a cero.

"Entonces obtenemos lo siguiente:

$$\sum_i \sum_j d_{ij}^2 = \theta_1^2 + \theta_2^2 + \theta_3^2 + \ldots \qquad (3.8)$$

"Esto se obtiene a partir de la condición de ortogonalidad de los eigenvectores, de donde todos los productos cruzados son cero, y de la relación (3.1).

"La ecuación (3.8) corresponde a una partición de la suma de cuadrados de la interacción. Entonces se ve uno tentado a intentar un tratamiento del tipo de análisis de varianza. Sin embargo, las θ^2 no son formas cuadráticas de las observaciones originales Y_{ij} 1/ . A pesar de lo anterior, es posible formular el problema en un lenguaje de análisis de varianza en virtud de las siguientes consideraciones.

Supóngase primero que las Y_{ij} son una muestra aleatoria de una población N (0,1). Entonces las cantidades θ_1^2 , θ_2^2 , ... formarán cada una poblaciones estadísticas particulares. Sean:

$$M_1 = E_N(\theta_1^2) , \qquad M_2 = E_N(\theta_2^2), \ ...$$

"Donde E es el operador esperanza y N está por una distribución normal estander (N (0,1)).

"Si la población normal tuviera una varianza igual a σ^2 en vez de igual a 1, entonces las cantidades M_1, M_2, ..., sería simplemente multiplicadas por σ^2.

"Entonces, los cocientes de los valores θ_i^2 obtenidos a partir de una población N $(0, \sigma^2)$ entre los valores M correspondientes, son todos estimadores de σ^2.

"Si lo que tenemos dado ahora es una matriz de observaciones Y_{ij} donde los términos de interacción no son sino errores normales

---

1/ Si el modelo es estrictamente aditivo, la suma de cuadrados de los residuales estima insesgadamente a un múltiplo de la varianza de ε_{ij}.

(con distribución de probabilidad normal, entonces los cocientes de los valores θ^2 obtenidos a partir de estos datos, entre los correspondientes valores de M obtenidos de una matriz de observaciones N (0,1) de las mismas dimensiones son simplemente estimadores de σ^2.

"Yendo un poco más adelante con este argumento, y haciendo un razonamiento de tipo heurístico, podemos esperar que si el modelo real contiene K términos del tipo $\theta u_i v_j$, entonces los correspondientes K valores de θ^2 serán inflados por los efectos sistemáticos de estos términos, mientras que los términos restantes θ^2_{k+1}, θ^2_{k+2}... sólo serán estimadores de:

$$M_{k+1}\,\sigma^2, \quad M_{k+2}\,\sigma^2, \ldots.$$

"Luego

$$\frac{\theta^2_{k+1}}{M_{k+1}}, \quad \frac{\theta^2_{k+2}}{M_{k+2}}, \quad \frac{\theta^2_{k+3}}{M_{k+3}}, \ldots$$

serán estimadores de σ^2.

"Es de esta manera en que uno se puede hacer un juicio sobre el número k de términos del tipo $\theta u_i v_j$ que han de ser considerados en el modelo".

Haciendo uso de una simulación Monte-Carlo fueron determinados los valores de M_1, M_2, M_3, ... para distintos valores de m y n. Valiéndose de estos valores se planteó una analogía de las tablas de análisis de varianza.

Esta se representa a continuación:

Fuente	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios
Total	mn	$\sum_i \sum_j Y^2_{ij}$	
α_i	$m-1$	$\sum_i \frac{Y^2_{i.}}{n} - \frac{Y^2_{..}}{mn}$	
β_j	$n-1$	$\sum_j \frac{Y^2_{.j}}{m} - \frac{Y^2_{..}}{mn}$	
η_{ij}	$(m-1)(n-1)$	$\sum_i \sum_j d^2_{ij}$	$d^2_{ij}/(m-1)(n-1)$
$\theta_1 u_{1i} v_{1j}$	M_1	$\hat{\theta}^2_1$	$\hat{\theta}^2_1/M_1$
$\theta_2 u_{2i} v_{2j}$	M_2	$\hat{\theta}^2_2$	$\hat{\theta}^2_2/M_2$
$\theta_3 u_{3i} v_{3j}$	M_3	$\hat{\theta}^2_3$	$\hat{\theta}^2_3/M_3$
.	.	.	.
.	.	.	.
.	.	.	.

Ahora se examinan los cuadrados medios correspondientes a la partición de η_{ij} en la suma de términos $\theta u_i v_j$. En este momento no se tiene una distribución teórica para estos cuadrados medios, y las pruebas exactas de significancia no se pueden realizar. Sin embargo, aún la interpretación intuitiva d los cuadrados medios nos llevará a determinar fácilmente el número de términos multiplicativos que han de ser considerados en el modelo.

Así, si ninguno de los cuadrados medios relacionado con términos $\theta u_i v_j$ resulta ser drásticamente mayor que los demás, por lo expuesto antes, ningún término multiplicativo será considerado en el modelo que se ha de ajustar a los datos que se estudien. Es decir el modelo es un modelo estrictamente aditivo:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Este puede ser una alternativa a la prueba de aditividad de Tukey; sin embargo como se apunta arriba se desconoce la distribución de $\hat{\theta}^2$ y en consecuencia el nivel real de significancia de la prueba.

Si en cambio, resulta ser que solamente el primer cuadrado medio, es decir, el correspondiente a $\theta u_i v_j$ es significativamente mayor que los correspondientes a los términos semejantes posteriores. Así el modelo sería

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta u_i v_j + \varepsilon_{ij}$$

que incluye como caso particular la suposición de Tukey para su prueba de no aditividad

$$Y_{ij} = \mu + \alpha_i + \beta_j + \lambda \alpha_i \beta_j + \varepsilon_{ij}$$

Así podría continuarse usando los valores de la tabla de análisis de varianza.

CAPITULO 4

4.- BREVE DESCRIPCION DE LAS PRUEBAS DE HIPOTESIS USADAS

En el presente trabajo se hace uso de un criterio esencialmente distinto, para determinar la forma del modelo que ha de ser ajustado a los datos, a aquel presentado en la sección anterior y debido a esto se expone a continuación:

Se planteó en la sección 2 la forma general de un modelo estadístico, que era:

$$Y_{AB...Z} = f(A,B,...,Z) + \varepsilon_{AB...Z}$$

Donde E es una variable aleatoria cuya distribución cumple con las siguientes propiedades:

1) $E(\varepsilon_{AB...Z}) = 0$

2) $E(\varepsilon^2_{AB...Z}) = \sigma^2$ (constante e independiente de A,B, ...Z)

3) $\varepsilon_{AB...Z}$ se distribuye normalmente

4) $E(\varepsilon_{AB...Z}\ \varepsilon_{A'B'...Z'}) = 0$

Es decir, tenemos las suposiciones de:

2') Homogeneidad de Varianzas

3') Normalidad

4') Independencia de los errores

Lo que para este modelo en particular puede ser expresado como:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{r=1}^{k} \theta_r u_{ri} v_{rj} + \varepsilon_{ij}$$

con

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$E(\varepsilon_{ij}\ \varepsilon_{i'j'}) = 0$$

Esta última definición de el modelo, en particular la parte que se refiere a las propiedades que deben satisfacer los errores aleatorios asociados con cada observación, permiten atacar el problema de la determinación del modelo más adecuado (número de términos de la forma $\theta u_i v_j$ que expresan la interacción), para representar el fenómeno que generó los datos. Es decir, el criterio que se usará en adelante será:

Aumentar a un modelo del tipo $Y_{ij} = \mu + \alpha_i + \beta_j$ tantos términos de la forma $\theta u_i v_j$ como sea necesario hasta lograr que las propiedades que deben cumplir los errores sean satisfechas.

Aquí cabe hacer un paréntesis para preguntarse por el objeto de llevar a cabo tarea semejante. La razón por la que se hace lo anterior es lo siguiente:

a) La hipótesis de distribución normal de los errores lleva implícitas varias hipótesis más, como son:

1) La distribución es simétrica con respecto a su media.

2) El coeficiente de Kurtosis ($E(X - \mu)^4/\sigma^4 - 3$) es cero, es decir, la gráfica de la función de densidad no es ni achatada ni picuda.

3) La media es independiente de la varianza, es decir, una

translación lineal del origen no afecta la forma de la gráfica de la distribución.

4) Los estimadores de media y varianza ($\bar{X}$, S^2) resumen toda la información obtenible a partir de la muestra (son suficientes minimales). Además son estocásticamente independientes. El hecho de que esta hipótesis no se cumpla, en general, afectará fuertemente las inferencias que se hagan con respecto a la varianza y en menor grado las que se hagan con respecto a la media. Los efectos más graves de dicho incumplimiento son el que tampoco se cumpla alguna de las hipótesis que se enumeran arriba (Sesgo, Kurtosis, etc.), se obtendrán demasiados resultados significativos en algunas pruebas de hipótesis hechas con respecto a los parámetros del modelo y se afectará la potencia de alguna de estas pruebas.

Para probar la hipótesis de normalidad se han desarrollado distintos criterios, ampliamente difundidos en la literatura. En el presente trabajo se hace uso de cuatro criterios para probar algunas de las hipótesis, tanto de normalidad como de simetría y kurtosis. Dos de estas pruebas fueron usadas para probar normalidad y estas son:

I) Prueba de Kolmogorov-Smirnov.

II) Prueba de Ji-cuadrada asintótica.

Además se hizo uso de una prueba para simetría y otra para kurtosis.

La primera de las pruebas para normalidad fué usada cuando el número de celdas en las tablas en que se resumió la información era menor o igual a sesenta. La razón de hacer esto fué que en el programa de computadora que se hizo para llevar a cabo las pruebas de hipótesis, la rutina o procedimiento para obtener integrales numéricas, y en consecuencia la que realiza pruebas de Kolmogorov-Smirnov, lleva mucho tiempo de proceso.

Además para muestras grandes, la prueba de Ji-cuadrada asintótica es suficientemente buena.

Se intenta a continuación explicar el funcionamiento de estas pruebas.

La prueba de Ji-cuadrada asintótica.- En esta prueba los datos se agrupan en clases para formar una distribución de frecuencias, además se calculan la $\overline{X}$, la media, y S, la desviación estandard: a partir de estas cantidades se ajusta una distribución normal considerando $=\overline{X}$ y $\sigma = S$.

Para cada clase se encuentra la siguiente cantidad:

$$\frac{(f_i - F_i)^2}{F_i} = \frac{(\text{OBSERVADOS} - \text{ESPERADOS})^2}{\text{ESPERADOS}}$$

Siendo el criterio de prueba en que a continuación se presenta

$$D^2 = \sum_i (f_i - F_i)^2 / F_i$$

Donde la suma corre sobre todas las clases. Si todos los da-

tos realmente pertenecen a una distribución normal, entonces D^2 se distribuye aproximadamente como una Ji-cuadrada con (R-?) grados de libertad con R, el número de clases usadas para el cálculo de D^2. Si en cambio los datos pertenecen a alguna otra distribución, la diferencia entre f_i, la frecuencia observada, y F_i, la frecuencia esperada, será grande; consecuentemente el valor de D^2 sera grande. Lo que causa que si se obtienen valores grandes de D^2, la hipótesis de normalidad sea rechazada.

El teorema en que se propone el uso de esta prueba (Cramér (2)), requiere que el número esperado F_i no sea demasiado pequeño. Estos pueden ocurrir en las clases de las colas. Como regla práctica para trabajar esos casos extremos se aconseja construir las clases de los extremos de manera tal que los valores esperados en ellas sean al menos uno, siempre que casi todas las restantes tengan al menos un valor esperado de 5.

La prueba de Ji-cuadrada no propone ninguna hipótesis alternativa para la distribución de los errores. Las tablas usadas en esta prueba se encuentran reproducidas en la tabla 1 del apéndice II.

Kolmogorov-Smirnov.- La distribución muestral

$$F_n(x) = \frac{j}{n} \quad \text{para} \quad X_{(j)} < x \leq X_{(j+1)}, \quad j=0,1,\ldots,n$$

con $X_{(0)} = -\infty$ y $X_{(n+1)} = \infty$, en general resultará ser diferente de la función de distribución poblacional. Pero si esa diferencia resulta ser grande podemos usarla como una base para rechazar la hipótesis de que F (x) es realmente la distribución de la población de la cual fué extraída la muestra. Es decir, la diferencia existente entre la distribución empírica y la establecida en la hipótesis debe servir como una estadística muy adecuada para determinar si la distribución que se supuso es efectivamente la poblacional.

De hecho es la diferencia $\left| F_n(x) - F(x) \right|$ la que se usa para esta prueba. Sin embargo determinar esta diferencia para todos los valores de X, y decidir cual de ellas usar para la prueba puede resultar difícil; por lo que la estadística de Kolmogorov-Smirnov que llamaremos D_n, porque depende del tamaño de muestra, se toma como:

$$D_n = \sup_{-\infty < x < \infty} \left| F_n(x) - F(x) \right|$$

Esta estadística, bajo H_0, tiene una distribución independiente de la distribución F (x), que es la que se propone como hipótesis nula.

Para decidir si la hipótesis H_0 se rechaza o no, el valor de D_n se compara con valores que se encuentran tabulados (V.ej. Lindgren (16), P.456). Esta tabla, por que la distribución de D_n es independiente de F (x), es válida para cualquier que sea

nuestra elección de F(x).

Por otra parte, la hipótesis alternativa, como en el caso de la prueba de Ji-cuadrada asintótica, no se especifica. Esto impide conocer con exactitud la potencia de la prueba.

Debe notarse que hasta el momento se ha hablado de "La distribución F(x)" o "La hipótesis H_0", es decir, se ha hablado de establecer una hipótesis simple; sin embargo, para el presente trabajo, la hipótesis nula es: "La distribución es normal con media cero" por lo que las tablas usuales no son útiles en este caso ya que la distribución no se ha especificado completamente. Por lo anterior se hará uso de las tablas obtenidas por Lilliefors (7).

Estas tablas resultan ser más restrictivas que las tablas usuales, en el sentido de proporcionar valores críticos más pequeños para un mismo tamaño de muestra y un mismo nivel de significancia. Además estas tablas fueron construidas para hipótesis del tipo: "La distribución es normal", con media y varianza no especificadas. Las tablas se encuentran reproducidas en la tabla 2 del Apéndice número II, al final del volumen.

Prueba de Simetría.- Medida de la simetría de una población dada por el valor promedio de $(X-\mu)^3$ $(E(X-\mu)^3/\sigma^3)$, tomado sobre toda la población. Si los valores de X menores que μ se encuentran agrupados cerca de la media, pero los valores de X ma-

vores que μ se encuentran muy alejados de este valor, entonces la medida de simetría tomará un valor positivo ya que la contribución de los valores positivos de $(X-\mu)^3$ predominará sobre la de los valores de $(X-\mu)^3$ negativos. También se puede dar el caso de poblaciones con medida de simetría negativa.

Este coeficiente involucra las unidades (al cubo), en que fueron medidas las observaciones; para evitar esta propiedad no deseable, esto es, hacer el coeficiente independiente de las unidades, se divide este coeficiente entre σ^3. De lo anterior resulta el llamado coeficiente de asimetría el cual es en ocasiones denotado por γ_1.

El estimador muestral de este coeficiente, llamado g_i, se calcula como sigue:. Sean:

$$m_3 = \sum (X_i - \bar{X})^3 / n$$

$$m_2 = \sum (X_i - \bar{X})^2 / n$$

entonces

$$g_1 = \frac{m_3}{m_2^{3/2}}$$

Obsérvese que en el cálculo de m_2, la varianza muestral, el denominador es n y no (n-1), lo que facilita los cálculos posteriores.

Si la muestra se toma de una población normal entonces g_1 se distribuye aproximadamente normal con media cero y desviación

estandar $(6/n)^{1/2}$.

La suposición de que g_1 es normal es suficientemente buena para un tamaño de muestra mayor que 50 (para tamaños de muestra entre 25 y 200, los límites de significancia para pruebas de una cola, están dados para .05 y .01 en la tabla 3 del apéndice II).

PRUEBA DE KURTOSIS

Otro tipo de medida que es útil para caracterizar una población normal es la de Kurtosis (picudez). En general la medida de Kurtosis se define como el promedio de $(X-\mu)^4$ dividido entre σ^4. Para una población normal este cociente toma el valor 3. Si el cociente excede el valor 3 eso indica que hay una gran concentración de los valores de X alrededor de la media; consecuencia de lo anterior es que los valores alejados de la media tienen poca o ninguna importancia. Es decir la gráfica de la curva tiene una forma "picuda" o "puntiaguda". Por otra parte, en caso de que el coeficiente resulte tomar un valor menos que 3, la gráfica de la curva tiene una forma "achatada" o "más achatada" que la curva normal. Este coeficiente se denota usualmente por γ_2.

El estimador muestral del coeficiente de Kurtosis está dado por

$$g_2 = \frac{m_4}{m_2^2}$$

donde $$m_4 = \frac{\sum_i (X_i - \bar{X})^4}{n}$$

es el cuarto valor muestral con respecto a la media.

Desgraciadamente la distribución de g_2 está lejos de parecerse a la normal para tamaños de muestra menores que 1000. Para tamaños de muestra entre 50 y 1000, la tabla 4 del apéndice II nos da valores de los límites de significancia para niveles de significancia del 5% y del 1%.

b) La hipótesis de independencia de los errores, en caso de no satisfacerse puede traer problemas en, por ejemplo, inferencias hechas acerca de las medias poblacionales. Se sabe con una correlación positiva de los errores vicia las pruebas de t y de F produciendo muchos resultados significativos.

Esta hipótesis plantea la no existencia de la relación entre diferentes observaciones, la cual no se toma en cuenta por los términos del modelo que representan los efectos. Si los individuos u observaciones se toman al azar y se miden por separado, la suposición es muy razonable; sin embargo pueden aparecer casos en los que esta suposición no se cumple.

Se han desarrollado varios métodos para probar tal hipótesis como son la de Theil y Nagar, la de Durbin-Watson, la de Von Neumann; en el presente trabajo, ya que en general las hipóte-

sis que se pide que cumplan las observaciones no se satisfacen (como en Theil y Nagar) o no son conluyentes en algunos casos (Durbin Watson), se hará uso de la prueba Von Neumen para la cual se han tabulado los extremos de significancia. La tabla aparece reproducida en el apéndice II como la tabla 5. Brevemente esta prueba se lleva a cabo como sigue:

El criterio que se usa para probar la hipótesis de independencia, es en este caso el cociente entre el cuadrado medio de las diferencias sucesivas y varianza, es decir:

$$V = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2/(n-1)}{\sum_{t=1}^{n} (e_t - \bar{e})^2/n} = \frac{\delta^2}{S^2}$$

Para tamaños de muestra (n) grandes, δ^2/S^2 se distribuye aproximadamente como normal con

$$E(V) = \frac{2n}{n-1} \quad ; \qquad Var\ (V) = \frac{4n^2\ (n-2)}{(n+1)\ (n-1)^3}$$

entonces, para n grande, una prueba para autocorrelación puede llevarse a cabo mediante la comparación de los valores de la razón de Von-Neumann con los límites de la región crítica elegida para una normal con media y varianza dadas arriba, sin embargo es importante enfatizar que las fórmulas para media y varianza son ciertas solamente si los errores se distribuyen independientemente, y esto no se cumple para los residuales obtenidos

por mínimos cuadrados, aún cuando los errores reales si se distribuyen independientemente.

Como se dijo anteriormente, para tamaños de muestra pequeños ($4 \leq n \leq 60$), se usan los límites de significancia para niveles de .01 y .05 dados en la tabla 5 del apéndice II.

c) La hipótesis de igualdad de varianzas de los errores o como usualmente se le conoce, la homoscedasticidad, es realmente una suposición fuerte. Es más difícil de ser satisfecha que, por ejemplo, la independencia de los errores; sin embargo el análisis de varianza formal está basado en esta suposición. De hecho, una característica escencial de varias técnicas es la estimación de esta, supuestamente común, varianza.

Muchas circunstancias pueden viciar esta suposición. Es frecuente que distintos métodos de medición produzcan variaciones de diferentes magnitudes; y valores esperados mayores están comunmente asociados con desviaciones estandar mayores. Así pues es necesario considerar posibles variaciones en las desviaciones estandar antes aún de aplicar el análisis de varianza a cualquier conjunto de datos.

El incumplimiento de esta suposición tiende a dar, mediante la prueba de F, demasiados resultados significativos. Cuando en un modelo de diseño de experimentos se supone la igualdad de varianzas se combina la evidencia muestral con el fin de obtener un estimador ponderado de la varianza, obtenido a partir de

las observaciones muestrales; este estimador tiene la propiedad de dar intervalos de confianza más estrechos para la media y diferencia entre distintas medias, y en consecuencia dar pruebas de significancia más potentes. Si la igualdad de varianza no se satisface, el estimador ponderado de la varianza no es adecuado como estimador de la varianza de las observaciones en cada grupo.

A la desigualdad de las varianzas se le ha dado el nombre de heteroscedasticidad, acerca de la cual se ha concluido que si se supone un modelo con heteroscedasticidad, cuando esta realmente existe, es posible obtener un mejor estimador del vector de coeficientes en el modelo. Dicho estimador es un mejor estimador lineal e insesgado para el vector de coeficientes, y es obtenido mediante los llamados mínimos cuadrados generalizados.

El criterio para probar la violación de esta hipótesis, utilizado en este trabajo, fué propuesto por Cochran y hace uso de la siguiente estadística

$$C = \frac{\max_t \sum_{i=1}^{n_t} (X_{ti} - \bar{X}_{t.})^2}{\sum_{t=1}^{k} \sum_{i=1}^{nt} (X_{ti} - \bar{X}_{t.})^2} = \frac{\max_t S_t}{\sum_t S_t}$$

Para esta prueba, el libro de Johnson y Leone (5) proporciona una tabla de los límites de significancia para el 1% de significancia. Esta tabla es la número 6 del apéndice II.

Esta prueba se puede aplicar solamente cuando el cálculo de todo S_t está basado en el mismo número de observaciones.

El criterio para rechazar la hipótesis de homoscedasticidad es el siguiente:

Rechace si C > valor tabulado.

De otra manera no rechace.

CAPITULO 5

5.- EL PROGRAMA DE COMPUTADORA

Para el cálculo de los residuales, de los estimadores y la utilización de las pruebas de hipótesis a las que se ha hecho mención a lo largo de este trabajo, se desarrollo un programa de computadora en lenguaje ALGOL (un listado de dicho programa aparece en el apéndice I).

Para el desarrollo de este programa se procuró que las rutinas (procedures) que realizan tal o cual prueba fuesen lo más independientes posible de las restantes, con el objeto de facilitar su utilización en otros programas; aunque esto haya provocado una reducción fuerte en la eficiencia del programa total.

En la presente sección se hace un breve resumen de dicho programa con el fin de que la gente interesada pueda hacer uso del mismo.

Debido a la estructura que debe guardar un programa en ALGOL, se hará referencia únicamente de las declaraciones llamadas (procedures) procedimientos.

Procedimientos LEEMATRIZ E IMPRIMEMATRIZ.- Son rutinas de entradas y salidas de datos. LEEMATRIZ, como su nombre lo indica, lee un arreglo de datos dados en forma matricial. IMPRIMEMATRIZ a su vez imprime matrices; este último procedimiento fué usado para la impresión de las matrices de observaciones, de residuos, de eigenvectores y de diferencias en el procedimiento Kolmogorov-

Smirnov.

Procedimiento ESTIMAB.- Este procedimiento calcula los estimadores de mínimos cuadrados de los parámetros que representan los efectos principales (μ , α_i, β_j) si se considera que el modelo es del tipo.

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

los estimadores se encuentran en las salidas del programa bajo "residuos 1".

Procedimiento RESIDUO1.- Este procedimiento hace el cálculo de los residuos $d_{ij} = Y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$, donde, como es usual ($\hat{}$) representa estimador. Estos residuos son expresados en forma de matriz y se encuentran impresos en cada salida del programa.

Procedimiento ORDENA.- Este procedimiento arregla, según su magnitud, los elementos de una matriz. Fué usado en el procedimiento KOLMOGSMIR para obtener la estadística de orden de los residuos.

Procedimiento INTEGRA.- La función de este procedimiento es calcular integrales numéricas. Tiene como una de sus declaraciones al procedimiento INTEGRAL el cual es el que realmente hace el cálculo de las integrales. La función que ha de ser integrada le es proporcionada en el "Define" F (x,s), donde x es el argumento y s es la varianza de la función; la función que se usó en la

de densidad normal con media cero y varianza s (la media es una de las hipótesis del modelo). Es además un procedimiento (integra) iterativo con el fin de lograr que el valor de la integral obtenida sea "bueno" (se permitió un error hasta 10^{-5}).

Procedimiento KOLMOGSMIR.- Este procedimiento lleva a cabo la prueba de la hipótesis de normalidad (con valor esperado igual a cero). En virtud de que la función consiste en rechazar o no rechazar una hipótesis, le fué asignado un tipo (BOOLEAN); es decir, el procedimiento "toma" el valor (false) cero si la hipótesis es rechazada, y el valor uno (true) si no.

En este caso la prueba fué hecha a cinco distintos niveles de significancia (20%, 15%, 10%, 5%, 1%), siendo el tamaño de muestra el total de elementos de la matriz de residuos, y entendiendo por rechazar la hipótesis, el que esta sea rechazada a uno de los cinco niveles de significancia.

Un estimador de la varianza lo da el procedimiento SUMCUAD el cual se explica a continuación.

El procedimiento KOLMOGSMIR trabaja como se indica a continuación:

Dada una matriz de residuos, estos son ordenados de menor a mayor mediante el procedimiento ORDENA. Después se calcula la función de distribución mediante el procedimiento INTEGRA; esta función es comparada con la función de distribución empírica pero sólo en los valores de los residuos, ya que las mayores diferen-

Este estimador es

$$S_1^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij}^2}{(m-1)(n-1)}$$

Para los modelos que incluíen k términos de la forma $\theta u_i v_j$, el estimador de la varianza se obtuvo a partir de la siguiente formula:

$$S_k^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij}^2}{(m-1)(m-1)-M_1 \ldots -M_k}$$

Los valores de M_j fueron obtenidos de las tablas reportadas por Mandel en los artículos en los que expone su método de estimación de los efectos de interacción; estos valores como se anunto anteriormente fueron obtenidos mediante una simulación Monte-Carlo.

Procedimiento HOMOSCED.- Este es otro procedimiento con tipo (Booleano) ya que fué hecho para probar la hipótesis de igualdad de varianzas (homoscedasticidad). La prueba usada en este procedimiento es la propuesta por Cochran. Como se explicó antes dicha prueba consiste en calcular el cociente del segundo momento muestral con respecto a su media para una de las subpoblaciones entre la suma de los segundos momentos muestrales con respecto a la media de cada una de las subpoblaciones. Tanto los renglones como las columnas de la matriz fueron tomados co-

mo las subpoblaciones que se usan en la prueba y en cada caso fué realizada la prueba de homoscedasticidad para el nivel de significancia del 1% por ser este el único nivel para el cual se tabularon los límites de significancia. Se consideró no rechazada la hipótesis sólo en el caso en que simultáneamente no se rechazaba la hipótesis por renglones y por columnas; en este caso el procedimiento tomaba el valor true, de otro modo era false el valor.

Procedimiento JICUAD.- Este procedimiento realiza la prueba de la hipótesis de normalidad con media cero, por lo que también fué asignado el carácter Booleano. Como se dijo antes, este procedimiento fué usado cuando el número de elementos de la matriz de residuo resultó ser mayor o igual que 60. Para llevar a cabo la prueba de la hipótesis se hizo uso del estimador de la desviación estandar de las desviaciones, que se calculó con el procedimiento SUMCUAD. Las observaciones (elementos de la matriz) fueron divididos en 6 clases equiprobables (es decir con probabilidad 1/6 cada una de ellas) y por lo tanto con el mismo valor esperado, lo que motivó que las longitudes de los intervalos de clase fueran distintos, aunque resulten ser simétricos. Ya que la hipótesis que se desea probar es la de normalidad con valor esperado igual a cero, se usaron los límites de significancia obtenidos en tablas de la distribución Ji-cuadrada con mn-2 grados de libertad.

Procedimiento TRANSPON.- Es bien simple la función de este procedimiento, la cual consiste en encontrar la matriz transpuesta de otra matriz que le es dada como entrada.

Procedimiento MULTIPLICAMATRIZ.- Como su nombre claramente lo indica, este procedimiento encuentra la matriz producto de dos matrices que son dadas como entrada para el procedimiento.

El uso dado a los procedimientos anteriores (TRANSPON, MULTIPLICAMATRIZ), fué el encontrar la matriz DD' = S, y fueron usados dentro del procedimiento ESTIMAU.

Procedimiento SIMETRIA.- Este es otro procedimiento del tipo Booleano ya que fué usado para probar la hipótesis de que la distribución de los residuos era simétrica. Ya que los límites de significancia tabulados están dados para pruebas de una cola para los niveles de 5% y 1%, la prueba de hipótesis de simetría se llevo a cabo para niveles de significancia del 10% y del 2% (la distribución del estimador es simétrica). Como en los casos anteriores se consideró no rechazada la hipótesis sólo en el caso en que la hipótesis fuera no rechazada a ambos niveles, lo cual, por el acortamiento de la longitud del intervalo de no rechazo, implica que la hipótesis no fué rechazada con un nivel de significancia del 10%.

Procedimiento KURTOSIS.- Como se indicó al explicar las pruebas de hipótesis usadas, la prueba KURTOSIS fué usada como una prueba paralela para probar la hipótesis de normalidad. Por lo ante-

rior, a este procedimiento también le fué asignado el tipo Booleno.

Los niveles de significancia usados fueron del 10% y del 2%.

La hipótesis se considera no rechazada si el valor de la estadística calculada se encontraba entre los límites de significancia dados en las tablas para el nivel de significancia más grande; en ese caso el procedimiento toma el valor (true) uno y consecuentemente (false) cero si la hipótesis se rechazó a ese nivel.

Procedimiento NONSYMEIGENVALUESANDEIGENVECTORS.- Este es un procedimiento que fué tomado de la cinta Bibliomat de la biblioteca de la máquina. Fué usado para encontrar los eigenvalores y los eigenvectores de la matriz DD'=S, los cuales se recordará, son básicos para el método de estimación de los sumandos en que es dividida el efecto de interacción. Este fué usado también dentro del procedimiento ESTIMAU.

Procedimiento ESTIMAU.- En este procedimiento, haciendo uso de los procedimientos TRANSPON, MULTIPLICAMATRIZ, NONSYMEIGENVALUES-ANDEIGENVECTORS y ORDENA, se calcula a cada paso la matriz D', después la matriz S y después se encuentran los eigenvalores y los eigenvectores de la matriz S; de entre los eigenvectores de la matriz S se escoge el mayor y el eigenvector asociado a dicho eigenvalor. A este procedimiento se llega solamente si alguno de

los hipótesis fué rechazada.

Procedimiento ESTIMAV.- Haciendo uso del eigenvector encontrado en el procedimiento ESTIMAU y el valor propio asociado a dicho vector, se calcula el vector $\underline{v}$ cuyas componentes son los valores v_j que se usan en el cálculo de los términos de la forma $\theta u_i v_j$; este cálculo se lleva a cabo haciendo uso de la relación

$$v_j = \theta^{-1} \sum_i d_{ij} u_i$$

que fué introducida al comentar el método de Mandel.

Procedimiento RESIDUOS2.- En este procedimiento se calculan los nuevos residuos resultantes de suponer un modelo de la forma

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta_1 u_{1i} v_{1j} + \ldots + \theta_r u_{ri} v_{rj} + \varepsilon_{ij}$$

donde el número de términos (r) que explican el efecto de interacción se incrementa en uno cada vez que se llega a este procedimiento, es decir, cada vez que estadísticamente alguna de las suposiciones hechas acerca de la distribución del error aleatorio no se cumple.

El programa, propiamente dicho, trabaja de la siguiente manera: Una matriz de observaciones es leída mediante LEEMATRIZ. Información adicional como son las dimensiones de la matriz, también se proporcionan al programa, además de límites de significancia para las distintas pruebas, etc.

Para esta matriz se calcula la suma de cuadrados de sus componentes (SUMCUAD), se estima la media general y los efectos princips-

les por renglón y por columna (ESTIMAB) y se calculan los residuos (RESIDUO1) que se obtendrían si se supone el siguiente modelo:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Para la matriz de residuos resultantes se calcula la suma de cuadrados de sus componentes y, si el número de elementos de la matriz es menor o igual que 60, se prueban sobre estos residuos las hipótesis de:

1) Homoscedasticidad (HOMOSCED)
2) Normalidad (KOLMOGSMIR)
3) Independencia de los residuos (AUTOCORR)
4) Simetría o Sesgo ((SIMETRIA)
5) Kurtosis (KURTOSIS)

Si el número de elementos de la matriz es mayor que 60 entonces se realizan las mismas pruebas, con la única diferencia de que la prueba, de normalidad es realizada mediante JICUAD.

Si ninguna de las hipótesis es rechazada entonces el programa termina y esto nos indica que el modelo

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

es suficientemente bueno para explicar el comportamiento de las observaciones.

Si en cambio, alguna de las hipótesis es rechazada, se procede a encontrar un término de la forma $\theta u_i v_j$ con el objeto de ajustar

a los datos un modelo como el siguiente:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta_1 u_{1i} v_{1j} + \varepsilon_{ij} \qquad (2)$$

(ESTIMAU, ESTIMAV), entonces se calcula la nueva matriz de residuos (RESIDUOS2). Sobre estos nuevos residuos se prueban una vez más todas las hipótesis.

Si ninguna hipótesis es rechazada se concluye que el modelo (2) es suficientemente bueno. Si alguna hipótesis es rechazada entonces se procede a encontrar un nuevo término de la forma $\theta_2 u_{2i} v_{2j}$ para lo cual se hace uso de la nueva matriz de residuos. Este procedimiento se espera termine al incluir 2 o 3 términos multiplicativos para explicar la interacción.

Para probar la efectividad del modelo y de la explicación del efecto de interacción como una suma de términos multiplicativos, se corrió el programa sobre varias matrices de datos. Los resultados se discuten en la siguiente sección.

CAPITULO 6

6.- RESULTADOS OBTENIDOS Y DISCUSION

El objeto de esta sección es probar las suposiciones hechas acerca de la distribución del error en dos tipos de grupos de datos. Primero sobre varios grupos de datos reales tomados de varias publicaciones. Después se realizan todas las pruebas sobre un grupo de 25 matrices de datos las cuales se generaron mediante modelos con interacción y para las cuales se simularon variables aleatorias normales con media cero y distintas desviaciones estandar. Los grupos de datos fueron generados con distintos efectos de interacción. Para ello se hizo uso del siguiente modelo:

$$Y_{ij} = 2\,(i-1) + 3j + (\,(i-1)\,(j-3)\,)^{k} + e_{ij}$$

Donde a k le fueron asignados valores 1, 3, 5, 7 y 9; además para cada valor de k se generaron 5 poblaciones normales con media cero y desviación estandar igual a 1/20, 6/20, 11/20, 16/20, 21/20, lo que dió como resultado 25 matrices de datos, todos ellos independientes entre sí. (El programa usado para generar estas poblaciones está en el apéndice I).

1) El primer grupo de datos a los que se aplicó todo el juego de pruebas fué tomado del artículo en el que el Dr. Mandel expone su método de estimación. La matriz de observaciones está constituida por mediciones de la densidad de soluciones acuosas del alcohol etílico a 7 niveles de temperatura y a 6 nive-

les de concentración. La matriz aparece representada en la tabla E 1.

TABLA E 1

DENSIDAD DE SOLUCIONES ACUOSAS DE ALCOHOL ETILICO

CONCENTRACION	10	15	20	25	30	35	40
30.086	.95965	.95672	.95369	.95052	.94725	.94387	.94039
39.988	.94241	.93885	.93521	.93150	.92772	.92387	.91994
49.961	.92170	.91784	.31392	.90993	.90588	.90178	.89758
59.976	.89932	.89528	.89120	.88704	.88284	.87857	.87423
70.012	.87598	.87184	.86764	.86337	.85906	.85468	.85024
80.036	.85188	.84764	.84336	.83903	.83464	.83020	.82569

Para estos datos los estimadores de la media general y de los efectos de renglones y de columna están dados por:

$$\hat{\mu} = 0.8968$$

$\hat{\alpha}_1 = 0.0535$ $\quad$ $\hat{\beta}_1 = 0.0117$

$\hat{\alpha}_2 = 0.0346$ $\quad$ $\hat{\beta}_2 = 0.0079$

$\hat{\alpha}_3 = 0.0130$ $\quad$ $\hat{\beta}_3 = 0.0041$

$\hat{\alpha}_4 = -0.0098$ $\quad$ $\hat{\beta}_4 = 0.0001$

$\hat{\alpha}_5 = -0.0335$ $\quad$ $\hat{\beta}_5 = -0.0039$

$\hat{\alpha}_6 = -0.0578$ $\quad$ $\hat{\beta}_6 = -0.0079$

$\hat{\beta}_7 = -0.0121$

Estos estimadores están sujetos a la restricción no estimable:

$$\sum \hat{\alpha}_i = \sum \hat{\beta}_j = 0$$

Para la primera matriz de residuos los resultados de las pruebas fueron los siguientes:

Homoscedasticidad.- Esta hipótesis fué rechazada cuando las subpoblaciones fueron formadas por los renglones de la matriz (se dirá en adelante, "rechazada (o no rechazada) por renglones (o por columnas)"). Pero la misma hipótesis fué aceptada por columnas (el primer rechazo era suficiente para llevar a cabo todas las pruebas una vez más).

Normalidad.- La prueba de Kolmogorov-Smirnov para normalidad no rechazó esta hipótesis a ninguno de los cinco niveles (0.2, 0.15, 0.10, 0.05, 0.01) a los que fué probada.

Autocorrelación.- La prueba de Von Neuman rechazó esta hipótesis a los niveles de 0.10 y 0.02.

Simetría.- La hipótesis de simetría de los residuos no fué rechazada ni al nivel de significancia de 0.10 ni a nivel de 0.02.

Kurtosis.- Esta hipótesis fué rechazada al nivel 0.10 pero en cambio no lo fué al nivel 0.02.

Como resultado del rechazo de varias de las hipótesis, en particular de la autocorrelación (para la cual la razón de Von Neuman dió un valor de 0.861023 siendo la cota inferior más pequeña (para 10%) igual a 1.5408), se hace necesaria la obtención de al menos un término para explicar la interacción.

Los estimadores del primer término del efecto de interacción están dados por:

$$\hat{\theta}_1^2 = 0.000026108$$

$\hat{u}_1 = 0.78475$	$\hat{v}_1 = -0.59028$
$\hat{u}_2 = 0.23149$	$\hat{v}_2 = -0.37743$
$\hat{u}_3 = -0.05214$	$\hat{v}_3 = -0.17315$
$\hat{u}_4 = -0.22048$	$\hat{v}_4 = 0.01832$
$\hat{u}_5 = -0.33470$	$\hat{v}_5 = 0.20187$
$\hat{u}_6 = -0.40893$	$\hat{v}_6 = 0.37602$
	$\hat{v}_7 = 0.54466$

Los vectores $\underline{u}$ y $\underline{v}$ han sido normalizados, es decir:

$$\sum_i \hat{u}_i^2 = \sum_j \hat{v}_j^2 = 1$$

Para los residuos resultantes al ajustar un modelo con un término de interacción, las hipótesis no rechazadas, con los respectivos niveles de significancia, fueron los siguientes:

Homoscedasticidad	(Por columnas al 0.1)
Normalidad	(0.20, 0.15, 0.10, 0.05, 0.01)
Autocorrelación	(0.02)
Simetría	(0.10, 0.02).
Kurtosis	(0.10, 0.02).

Mientras que las hipótesis que se rechazaron y los niveles de significancia para los que se rechazaron, fueron:

Homoscedasticidad	(Por renglones 0.01)
Autocorrelación	(0.10)

El rechazo de estas dos hipótesis hizo necesario la inclusión de

un término de la forma $\theta u_i v_j$ y los estimadores para estos valores están dados por:

$$\hat{\theta}_2^2 = 0.000000003$$

$\hat{u}_1 = -0.45933$	$\hat{v}_1 = 0.51188$
$\hat{u}_2 = 0.72448$	$\hat{v}_2 = 0.00852$
$\hat{u}_3 = 0.32933$	$\hat{v}_3 = -0.33764$
$\hat{u}_4 = -0.04954$	$\hat{v}_4 = -0.46830$
$\hat{u}_5 = -0.32126$	$\hat{v}_5 = -0.35555$
$\hat{u}_6 = -0.22368$	$\hat{v}_6 = 0.12990$
	$\hat{v}_7 = 0.51118$

con $\sum_i \hat{u}_i^2 = \sum_j \hat{v}_j^2 = 1.$

Para los nuevos residuos las hipótesis no rechazadas son las siguientes:

Homoscedasticidad	(Por renglones y por columnas, 0.01)
Normalidad	(0.10, 0.05, 0.01)
Simetría	(0.10, 0.02)
Kurtosis	(0.10, 0.02)

Siendo las hipótesis rechazadas, las siguientes:

Normalidad	(0.20, 0.15)
Autocorrelación	(0.10, 0.02)

De lo anterior se desprende el que un modelo que incluya sólo dos términos multiplicativos no cumple con dos hipótesis centrales

(aunque la normalidad se rechaza con niveles altos de α, probabilidad de error tipo I, por lo que no es tan grave este aspecto), hechas acerca de la distribución conjunta de los errores, sin embargo según el criterio del Dr. John Mandel, este modelo explica suficientemente bien los datos del ejemplo.

Si se aumenta un término multiplicativo a la explicación del efecto de interacción, siendo los estimadores de los parámetros los siguientes:

$$\hat{\theta}_3^2 = (1.47535854)\ 10^{-10}$$

$\hat{u}_1 = 0.00425$	$\hat{v}_1 = -0.23143$
$\hat{u}_2 = -0.39407$	$\hat{v}_2 = 0.36943$
$\hat{u}_3 = 0.79460$	$\hat{v}_3 = -0.13611$
$\hat{u}_4 = -0.26636$	$\hat{v}_4 = 0.27288$
$\hat{u}_5 = 0.18845$	$\hat{v}_5 = -0.62633$
$\hat{u}_6 = -0.32686$	$\hat{v}_6 = 0.53833$
	$\hat{v}_7 = -0.18677$

Entonces las hipótesis no rechazadas son:

Homoscedasticidad	(Por renglones, por columnas, 0.01)
Normalidad	(0.20, 0.15, 0.10, 0.05, 0.01)
Autocorrelación	(0.02)
Simetría	(0.10, 0.02)
Kurtosis	(0.02)

Y las hipótesis rechazadas

Autocorrelación (0.10)

Kurtosis (0.10)

Es decir, dos de las hipótesis centrales se rechazaron pero sólo a uno de los niveles a los que fueron probados recuérdese que ocurrió lo mismo con un sólo término para la interacción, con lo cual un modelo que incluya uno o tres términos multiplicativos mejora el ajuste de los datos a las hipótesis planteadas acerca de la distribución del error. Entonces los modelos que mejor se ajustan a los datos según los criterios establecidos en el presente trabajo son los siguientes:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta_1 u_{1i} v_{1j} + \varepsilon_{ij}$$

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^{3} \theta_k u_{ki} v_{kj} + \varepsilon_{ij}$$

Para discriminar entre ambos puede usarse el criterio de Mandel, a partir del cual, el modelo que incluye tres términos es mejor. Aquí terminaría el ajuste del modelo, sin embargo si por fines de estudio teórico o por curiosidad se ocurriese el aumentar un término multiplicativo a la explicación del efecto de interacción, entonces se tendría que las hipótesis rechazadas serían las siguientes:

Homocedasticidad (Por renglones 0.01)

Normalidad (0.20, 0.15, 0.10).

Kurtosis (0.10, 0.02)

Y en consecuencia, las no rechazadas serían:

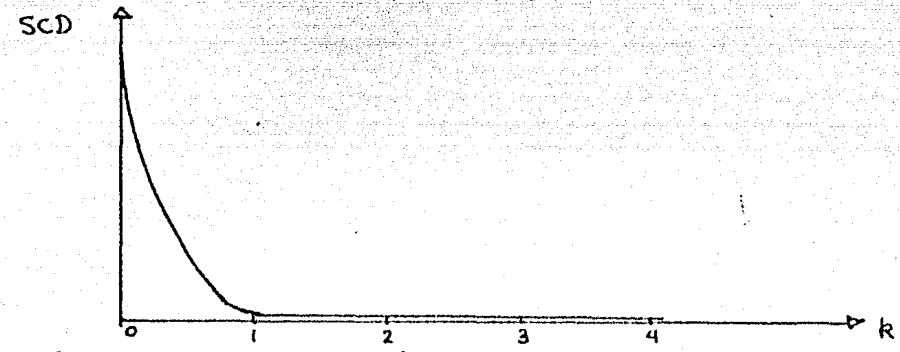
Homoscedasticidad (Por columnas 0.01)

Normalidad (0.05, 0.01)

Autocorrelación (0.10, 0.02)

Simetría (0.10, 0.02)

Además es interesante considerar lo ocurrido con la suma de cuadrados de los residuos; para ello simplemente se plotea la suma de cuadrados contra el número de términos multiplicativos incluidos en el modelo, en la gráfica G 1.



Gráfica G 1.- Reducción en suma de cuadrados contra número de sumandos del efecto de interacción.

2) El siguiente grupo de datos fué tomado de el artículo publicado por D.H. Levcock (Tron. Agriculture, Trin, Vol 32 No. 2, Abril 1955), en el cual se describe el efecto de las formas de las parcelas en la reducción del error en experimentos de té. Para este artículo se formaron 54 parcelas en un territorio rectangular con curvas de fertilidad definidas, las cuales se presentan en la tabla E 2 con los rendimientos de cada parcela incluidos.

TABLA N 2

RENDIMIENTOS DE TE CON CURVAS DE FERTILIDAD DEL CAMPO EXPERIMENTAL

28.5	27.1	30.4	27.6	25.1	23.2
32.7	20.5	22.3	24.8	22.1	22.8
31.7	23.0	23.9	29.4	22.5	29.0
26.7	21.9	23.0	25.3	26.9	29.4
25.8	21.4	20.0	23.7	31.1	37.3
25.1	20.2	21.2	22.2	27.4	34.7
28.1	25.3	21.4	25.1	28.5	35.7
30.6	37.0	52.7	41.9	43.9	45.1
41.4	49.7	49.2	47.4	47.8	43.7

La matriz de observaciones dió como resultado los siguientes estimadores de los parámetros:

$$\hat{\mu} = 30.2667$$

$\hat{\alpha}_1 = -3.2833$

$\hat{\alpha}_2 = -6.0667$

$\hat{\alpha}_3 = -3.6833$

$\hat{\alpha}_4 = -4.7333$

$\hat{\alpha}_5 = -3.7167$

$\hat{\alpha}_6 = -3.4667$

$\hat{\alpha}_7 = -2.0167$

$\hat{\alpha}_8 = 11.6000$

$\hat{\alpha}_9 = 16.2667$

$\hat{\beta}_1 = -0.2000$

$\hat{\beta}_2 = -2.9222$

$\hat{\beta}_3 = -0.9222$

$\hat{\beta}_4 = 0.5556$

$\hat{\beta}_5 = 0.3222$

$\hat{\beta}_6 = 3.1667$

A la matríz de residuos que de como resultado el uso de estos estimadores se aplicaron las pruebas de hipótesis, rechazando las

siguientes:

Normalidad (0.20, 0.15, aunque es a niveles altos de α).

Y no rechazadas las restantes hipótesis. De lo anterior es posible suponer que un modelo que incluya unicamente los términos aditivos es, para todos los fines prácticos, suficientemente bueno. Para este modelo, la suma de cuadrados del error es igual a 901.03555.

Si se aumenta un término multiplicativo para explicar los efectos de la interacción, si alguno de estos efectos existe, los estimadores de los parámetros involucrados en este término son:

$$\hat{\theta}_1^2 = 453.1528$$

$\hat{u}_1 = 0.26127$ $\hat{v}_1 = -0.52572$

$\hat{u}_2 = -0.16968$ $\hat{v}_2 = 0.17493$

$\hat{u}_3 = -0.18439$ $\hat{v}_3 = 0.72779$

$\hat{u}_4 = -0.10770$ $\hat{v}_4 = 0.01233$

$\hat{u}_5 = -0.34007$ $\hat{v}_5 = 0.01437$

$\hat{u}_6 = -0.23988$ $\hat{v}_6 = -0.40371$

$\hat{u}_7 = -0.28756$

$\hat{u}_8 = 0.65282$

$\hat{u}_9 = 0.40918$

Un modelo con un término multiplicativo para explicación de los efectos de interacción ajustado a los datos originales dió como resultado que las hipótesis rechazadas fueron las siguientes:

Normalidad (0.20)

Es decir ahora los errores son "más normales" que los obtenidos con un modelo estrictamente aditivo. Por otro lado la suma de cuadrados del error alcanzó un valor de sólo 447.88275, que es menor que el 50% de la suma de cuadrados obtenida con un modelo estrictamente aditivo, luego la inclusión de ese término de interacción mejora el ajuste del modelo.

Para estos datos la inclusión de un nuevo término de interacción sólo empeora las propiedades que tenían los errores con el modelo anterior, como lo demuestra el rechazo de las hipótesis de:

Normalidad	(0.20, 0.15, 0.10, 0.05)
Autocorrelación	(0.10, 0.02)
Simetría	(0.10, 0.02)
Kurtosis	(0.10, 0.02)

Los estimadores de este nuevo término de la explicación de la interacción son:

$$\hat{\theta}_2^2 = 318.9956$$

$\hat{u}_1 = 0.40551$	$\hat{v}_1 = 0.63069$
$\hat{u}_2 = 0.50841$	$\hat{v}_2 = 0.21988$
$\hat{u}_3 = 0.29973$	$\hat{v}_3 = 0.04704$
$\hat{u}_4 = -0.11220$	$\hat{v}_4 = 0.09778$
$\hat{u}_5 = -0.43815$	$\hat{v}_5 = -0.34488$
$\hat{u}_6 = -0.2619$	$\hat{v}_6 = -0.65051$
$\hat{u}_7 = -0.18909$	
$\hat{u}_8 = -0.42210$	
$\hat{u}_9 = 0.11220$	

Aunque se ha ganado en suma de cuadrados del error logrando reducir el valor hasta 128.8871, se han perdido casi todas las propiedades deseables para los errores.

Si para el mismo grupo de datos se construye la tabla de análisis de la varianza propuesta por el Dr. Mandel, esta tendría el aspecto siguiente:

FUENTE	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	CUADRADOS MEDIOS
$\theta_1 u_{1i} v_{1j}$	18.6	453.1528	48.4428
$\theta_2 u_{2i} v_{2j}$	10.89	318.9956	41.1269
$\theta_3 u_{3i} v_{3j}$	6.24	128.8871	20.6549

De aquí es fácil ver que, de acuerdo al criterio propuesto por Mandel, se tienen dos alternativas:

A) No incluir ningún término para explicación de la interacción ya que los cuadrados medios tienen el mismo orden de magnitud. Con esto se tendría una suma de cuadrados igual a 901.03555.

B) Incluir al menos tres términos para la explicación de la interacción. Esta alternativa daría como resultado errores poco apropiados para realizar inferencias y pruebas de hipótesis, por contradecir casi todas las hipótesis hechas acerca de la distribución del error.

De acuerdo a las propiedades de los errores y la suma de cuadrados del error el "mejor" modelo es el que incluye unicamente θ_1, u_{1i}, v_{1j}, lo que no coincide en ninguno de los casos a los que conduce el criterio de Mandel.

3) El programa también se aplicó a una serie de matrices que representan los rendimientos promedio de árboles frutales en libras por parcela para los años 1921, 1923, 1925 y 1927. Estos rendimientos fueron reportados por Parker-Batchelor en Hilgardia (Variation in yields of Fruit Trees, Hilgardia, October 1932 Vol. 7, No. 32).

Las matrices usadas representan sólo una parte del experimento total y son las siguientes:

1921

18	34	27	28	15
22	35	19	22	40
29	36	35	24	40
14	36	34	24	31
23	31	20	14	16
30	16	25	28	12
12	23	32	32	13
19	30	27	21	14
15	33	27	32	20
16	22	24	28	15
15	31	19	28	28

1923

50	76	60	72	63
66	58	72	63	62
74	65	52	56	79
52	60	42	66	68
65	73	65	73	55
71	59	72	79	55
46	75	85	86	51
57	51	83	74	75
47	61	64	75	60
62	62	72	81	68
66	107	80	91	86

1925

121	138	125	128	145
116	135	136	125	151
129	132	136	114	171
95	126	121	140	164
115	120	128	124	145
113	134	129	145	141
107	158	149	149	131
119	143	143	134	123
105	133	130	122	135
106	132	136	155	135
109	141	148	144	153

1927

108	104	107	103	128
128	152	169	129	144
136	153	156	114	183
107	153	146	129	157
145	162	174	135	168
134	168	179	134	132
159	186	186	145	146
154	167	170	144	140
159	180	145	152	164
152	177	187	164	167
135	180	185	186	163

Para la matriz que representa los rendimientos en el año de 1921 la aplicación del programa dió los siguientes resultados:

Los estimadores de los parámetros fueron:

$$\hat{\mu} = 24.6364$$

$\hat{\alpha}_1 = -0.2634$ $\qquad$ $\hat{\beta}_1 = -5.2727$

$\hat{\alpha}_2 = 2.9636$ $\qquad$ $\hat{\beta}_2 = 5.0909$

$\hat{\alpha}_3 = 8.1636$ $\qquad$ $\hat{\beta}_3 = 1.7273$

$\hat{\alpha}_4 = 3.3636$ $\qquad$ $\hat{\beta}_4 = 0.9091$

$\hat{\alpha}_5 = -3.86634$ $\qquad$ $\hat{\beta}_5 = -2.4545$

$\hat{\alpha}_6 = -2.4363$

$\hat{\alpha}_7 = -2.2364$

$\hat{\alpha}_8 = -2.4364$

$\hat{\alpha}_9 = 0.7636$

$\hat{\alpha}_{10} = -3.6364$

$\hat{\alpha}_{11} = -0.4364$

A los residuos resultantes de substituir los valores anteriores de los estimadores, se les aplicaron todas las pruebas de las hipótesis, lo que dió como resultado el que ninguna de las hipótesis se rechazara; en otras palabras, se supone que un modelo estrictamente aditivo es adecuado para ser ajustado a los datos. El valor de la suma de cuadrados de los residuos fué igual a 185.70909, con un estimador de la varianza igual a 43.3939272.

Para la matriz que representa los rendimientos en el año de 1923 los resultados variaron únicamente en la estimación de los parámetros del modelo aditivo y en el valor de la suma de cuadrados de los residuos . Los estimadores obtenidos se presentan a continuación.

$\hat{\mu} = 67.0545$

$\hat{\alpha}_1 = -2.8545$

$\hat{\alpha}_2 = -2.8545$

$\hat{\alpha}_3 = -1.8545$

$\hat{\beta}_1 = -7.4182$

$\hat{\beta}_2 = 0.8545$

$\hat{\beta}_3 = 0.8545$

$\hat{\alpha}_4 = -9.4545$ $\quad\quad$ $\hat{\beta}_4 = 7.1273$

$\hat{\alpha}_5 = -0.8545$ $\quad\quad$ $\hat{\beta}_5 = -1.4182$

$\hat{\alpha}_6 = 0.1455$

$\hat{\alpha}_7 = 1.5455$

$\hat{\alpha}_8 = 0.9455$

$\hat{\alpha}_9 = -5.6545$

$\hat{\alpha}_{10} = 1.9455$

$\hat{\alpha}_{11} = 18.9455$

Y la suma de cuadrados de los residuos aumentó hasta el valor 4635.309090, dando un estimador de la varianza igual a 115.88272; sin embargo ninguna de las hipótesis distribucionales del error fué rechazada, ajustándose en consecuencia solamente un modelo aditivo.

Para la matriz que representa los rendimientos para el año 1925 tampoco fué rechazada, ninguna de las hipótesis hechas al modelo, por lo que el modelo estrictamente aditivo se ajustó a los datos. Lo anterior trajo como consecuencia que la suma de cuadrados del error aumentara aún más hasta alcanzar el valor 5351.30909, dando un estimador de la varianza de los errores igual a 133.782727.

Los estimadores obtenidos fueron los siguientes:

$$\hat{\mu} = 132.4000$$

$\hat{\alpha}_1 = -1.0$ $\quad$ $\hat{\beta}_1 = -20.1273$

$\hat{\alpha}_2 = 0.2$ $\quad$ $\hat{\beta}_2 = 3.2364$

$\hat{\alpha}_3 = 4.0$ $\quad$ $\hat{\beta}_3 = 2.2364$

$\hat{\alpha}_4 = -3.2$ $\quad$ $\hat{\beta}_4 = 2.1455$

$\hat{\alpha}_5 = -6.0$ $\quad$ $\hat{\beta}_5 = 12.5091$

$\hat{\alpha}_6 = 0.0$

$\hat{\alpha}_7 = 6.4$

$\hat{\alpha}_8 = 9.3132(10)^{-10}$

$\hat{\alpha}_9 = -7.4$

$\hat{\alpha}_{10} = 0.4$

$\hat{\alpha}_{11} = 6.6$

Sin embargo para la matriz que presenta los rendimientos frutales para el año de 1927, los resultados cambiaron notablemente. Los estimadores de los parámetros para el modelo lineal aditivo resultaron ser los siguientes:

$$\hat{\mu} = 151.4545$$

$\hat{\alpha}_1 = -41.4545$ $\quad$ $\hat{\beta}_1 = -13.5455$

$\hat{\alpha}_2 = -7.0545$ $\quad$ $\hat{\beta}_2 = 10.5455$

$\hat{\alpha}_3 = -3.0545$ $\quad$ $\hat{\beta}_3 = 12.5455$

$\hat{\alpha}_4 = -13.0545$ $\quad$ $\hat{\beta}_4 = -11.9091$

$\hat{\alpha}_5 = 5.3455$ $\quad$ $\hat{\beta}_5 = 2.3636$

$\hat{\alpha}_6 = -2.0545$

$\hat{\alpha}_7 = 12.9455$

$\hat{\alpha}_8 = 3.5455$

$\hat{\alpha}_9 = 8.5455$

$\hat{\alpha}_{10} = 17.9455$

$\hat{\alpha}_{11} = 18.3455$

Para los residuos obtenidos al sustituir los estimadores arriba anotados la suma de cuadrados fué igual a 8022.43636. Al aplicar las pruebas a estos residuos, la hipótesis de independencia de los mismos fué rechazada únicamente a el nivel de significancia del 0.10 (el valor de la razón de Von Neuman fué igual a 2.510922 que es mayor que el límite superior del intervalo de no rechazo el cual tiene el valor de 2.4819). Lo anterior trajo como consecuencia que se aumentara un término multiplicativo para intentar explicar el efecto de interacción, los estimadores para los parámetros de dicho término estuvieron dados por:

$\hat{\theta}_1^2 = 3899.55294$

$\hat{u}_1 = 0.37193$ $\qquad$ $\hat{v}_1 = 0.09339$

$\hat{u}_2 = -0.09818$ $\qquad$ $\hat{v}_2 = -0.2308$

$\hat{u}_3 = 0.57935$ $\qquad$ $\hat{v}_3 = -0.42886$

$\hat{u}_4 = 0.19704$ $\qquad$ $\hat{v}_4 = -0.26174$

$\hat{u}_5 = 0.14902$ $\qquad$ $\hat{v}_5 = 0.828$

$\hat{u}_6 = -0.39709$

$\hat{u}_7 = -0.33477$

$\hat{u}_8 = -0.23751$

$\hat{u}_9 = 0.1783$

$\hat{u}_{10} = -0.12004$

$\hat{u}_{11} = -0.28806$

La suma de cuadrados para los nuevos residuos fué reducida hasta 4122.88342 al ajustar el nuevo modelo, con lo que se obtuvo un estimador de la varianza común de los errores igual a 103.072085. Para estos residuos ninguna de las hipótesis fué rechazada con lo que según el criterio desarrollado aquí el modelo que satisfactoriamente explica los datos originales es el dado por:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta u_i v_j + \varepsilon_{ij}$$

Es decir, es notable la pérdida en eficiencia del modelo lineal aditivo para explicar los rendimientos de los árboles frutales al aumentar la edad de los mismos. Así al corregir la falta de independencia se logra reducir la suma de cuadrados del error aproximadamente a la mitad.

4) El siguiente grupo de datos al que se le aplicó el método de estimación de Mandel fué publicado por R.J. Garber, T.C. McIlvane, y M.M. Hoover (18 pp. 286-298). El experimento realizado consistió en sembrar maíz en 1927, avena en 1928 y trigo en 1929, en el mismo campo experimental. Los rendimientos, para su comparación, fueron expresados como desviaciones con respecto a la producción promedio de todo el campo, para el mismo tipo de cul-

tivo. Esto es lo que se presenta en la tabla E 4, aunque sólo se presenta una parte del total del campo utilizado en el experimento.

TABLA E 4.- DESVIACIONES PORCIENTO CON RESPECTO A LA MEDIA

- 0.7	6.7	2.0	- 8.1	5.4	- 8.1	0.1	2.4	-11.4
-11.0	16.6	4.2	-11.6	-13.9	-21.6	-10.9	3.4	9.6
- 8.5	10.4	5.3	5.4	7.9	- 3.7	6.6	-13.8	- 3.2
- 8.5	2.9	-13.8	-13.7	-12.2	- 0.4	14.6	-20.2	-2.1
- 3.3	- 8.3	- 6.7	-15.2	- 6.3	10.6	5.1	-11.7	- 0.1
- 9.3	- 2.1	-12.7	-17.7	-34.0	-17.7	-26.5	-12.5	-10.1
-27.6	-32.1	-17.7	-17.0	-29.8	-18.7	-11.4	2.9	-15.6
-26.4	-13.1	-22.2	-20.9	-28.8	-26.1	-19.1	-16.6	- 2.1
-10.9	0.7	- 0.3	- 6.0	-18.8	-31.4	-31.7	-33.9	-21.2
13.9	-10.5	- 9.5	3.8	-20.1	-12.6	-41.1	-18.5	-15.1
- 1.3	18.4	11.0	2.4	- 2.1	- 1.3	-13.8	- 8.1	7.9

MAIZ

-25.9	39.3	32.2	25.3	- 0.9	12.8	24.1	3.0	22.0
-38.7	23.2	24.7	21.0	17.4	- 7.6	- 3.0	18.6	32.3
-35.7	38.7	7.9	41.8	11.0	2.7	13.1	-0 .6	15.2
-42.7	-23.8	2.4	12.8	7.3	24.1	18.0	-19.2	-13.7
-31.7	- 7.0	- 2.1	-16.2	8.2	17.4	- 7.0	-26.2	-15.9
-31.4	-23.8	8.8	8.8	-29.0	-26.3	-49.4	-33.8	-39.0
-58.8	-30.5	- 9.1	5.5	-14.6	-20.1	- 7.9	- 7.9	-29.6
-60.4	-28.7	-11.3	- 0.9	-25.0	-33.5	-16.8	-15.5	- 4.6
-22.3	-22.9	- 1.8	- 1.8	-16.2	-40.9	-34.8	-48.5	-25.0
-18.3	-28.7	-15.5	3.4	-18.9	-21.0	-24.1	-18.0	-16.2
-15.9	- 7.3	-17.7	6.1	- 9.1	- 5.2	-26.2	- 4.6	- 0.9

AVENA

10.8	68.6	61.9	29.4	31.4	34.0	51.0	28.9	14.9
- 0.5	25.1	49.0	30.4	12.9	12.4	17.0	21.6	44.3
- 3.6	45.4	33.5	34.5	35.6	12.9	18.6	26.3	25.8
2.6	2.1	18.6	26.3	20.1	16.0	17.0	-12.4	8.8
- 9.8	3.1	9.8	32.0	40.2	2.6	-20.6	-37.1	-12.4
4.1	10.8	11.9	23.2	-15.5	-33.5	-31.4	-32.0	-3.1
-25.8	-22.2	- 3.1	9.8	7.7	-21.6	7.7	-11.9	-26.3
-42.3	-16.5	-1.0	16.0	- 1.5	-18.6	- 1.5	- 9.3	- 5.2
2.6	- 5.7	-12.9	-16.5	-17.0	-30.9	-18.6	-22.7	-13.9
-10.8	-24.7	-17.5	-17.0	-24.2	-23.7	-10.3	-14.4	- 5.7
4.6	2.6	14.9	-10.8	- 6.7	- 8.2	-14.9	3.6	22.7

TRIGO

La primera matriz esta formada por las desviaciones, en porcentaje, de la media total del rendimiento de maíz obtenida.

Los restantes representan lo mismo pero para avena y trigo respectivamente.

Para el caso del maíz se obtuvo que los parámetros del modelo estrictamente aditivo eran los siguientes:

$$\hat{\mu} = -8.7212$$

$\hat{\alpha}_1 = 7.5768$	$\hat{\beta}_1 = 0.3394$
$\hat{\alpha}_2 = 4.8101$	$\hat{\beta}_2 = 7.7758$
$\hat{\alpha}_3 = 9.4323$	$\hat{\beta}_3 = 3.2303$
$\hat{\alpha}_4 = 2.7879$	$\hat{\beta}_4 = -0.2424$
$\hat{\alpha}_5 = 4.7323$	$\hat{\beta}_5 = -5.1606$
$\hat{\alpha}_6 = -7.1232$	$\hat{\beta}_6 = -3.3879$
$\hat{\alpha}_7 = -9.8343$	$\hat{\beta}_7 = -2.9242$
$\hat{\alpha}_8 = -10.7566$	$\hat{\beta}_8 = -2.7879$
$\hat{\alpha}_9 = -8.3343$	$\hat{\beta}_9 = 2.9576$

$\hat{\alpha}_{10} = -3.4677$

$\hat{\alpha}_{11} = 10.1768$

La suma de cuadrados de los residuos obtenidos, después de sustituir los estimadores arriba apuntados, resultó ser igual a 8498.5422; la aplicación de las pruebas de hipótesis a la matriz de residuos únicamente resultó en rechazo de la hipótesis de independencia de los residuos a un nivel de significancia del 10%. El que esta hipótesis haya sido rechazada motivó la inclusión de un término para explicar el efecto de interacción. Los estimadores de los parámetros que componen dicho término fueron:

$$\hat{\theta}_1^2 = 3497.86699$$

$u_1 = -0.06258$
$u_2 = 0.0382$
$u_3 = -0.06717$
$u_4 = -0.37186$
$u_5 = -0.33709$
$u_6 = 0.12015$
$u_7 = -0.37387$
$u_8 = -0.15874$
$u_9 = 0.45068$
$u_{10} = 0.59102$
$u_{11} = 0.21757$

$\hat{v}_1 = 0.38530$
$\hat{v}_2 = 0.27183$
$\hat{v}_3 = 0.25858$
$\hat{v}_4 = 0.33753$
$\hat{v}_5 = -0.00916$
$\hat{v}_6 = -0.22594$
$\hat{v}_7 = -0.70058$
$\hat{v}_8 = -0.20624$
$\hat{v}_9 = -0.11131$

La inclusión de este término trajo como resultado una reducción de la suma de cuadrados, la que alcanzó el valor de 5000.6763417;

además ninguna hipótesis fue rechazada con lo que se supone que el modelo que explica los datos de una manera satisfactoria, de acuerdo a los criterios establecidos, es un modelo que incluye un sólo término como explicación del efecto de interacción existente. Este hecho hizo suponer que existía cierta interacción entre los renglones y las columnas en que fueron arregladas las parcelas lo cual sería apoyado por los resultados obtenidos en los otros grupos de datos. Así pues se procedió a realizar un análisis semejante en los restantes grupos de datos. Para el rendimiento de avena los estimadores obtenidos para los parámetros del modelo estrictamente aditivo fueron:

$$\hat{\mu} = -8.5707$$

$\hat{\alpha}_1 = 23.2263$
$\hat{\alpha}_2 = 18.3374$
$\hat{\alpha}_3 = 19.0263$
$\hat{\alpha}_4 = 4.7040$
$\hat{\alpha}_5 = -0.2737$
$\hat{\alpha}_6 = -16.4404$
$\hat{\alpha}_7 = -10.6515$
$\hat{\alpha}_8 = -13.2848$
$\hat{\alpha}_9 = -15.2293$
$\hat{\alpha}_{10} = -8.9071$
$\hat{\alpha}_{11} = -0.4071$

$\hat{\beta}_1 = -26.1384$
$\hat{\beta}_2 = 2.0707$
$\hat{\beta}_3 = 10.2525$
$\hat{\beta}_4 = 18.1889$
$\hat{\beta}_5 = 2.2253$
$\hat{\beta}_6 = -1.2111$
$\hat{\beta}_7 = -1.7929$
$\hat{\beta}_8 = -5.3111$
$\beta_9 = 1.7162$

La suma de cuadrados de los residuos obtenidos de la aplicación de estos valores fué igual a 16561.013939. Las pruebas de las hipótesis resultaron todas en no rechazo de ninguna de las hi-

pótesis. De lo anterior se concluye que el comportamiento de los errores obtenidos es bueno si se supone un modelo estrictamente aditivo, sin embargo el tamaño de la suma de cuadrados de los residuos es aún muy grande lo que, como se verá más adelante, puede ser indicio de que existe un efecto de interacción pequeño.

La matriz de datos que reporta el rendimiento de trigo para el siguiente año tiene el inconveniente de haber sido tomada en un año en que los fenómenos meteorológicos tuvieron grandes cambios, lo que afectó los rendimientos; esto puede explicar la diferencia tan grande en el comportamiento de los datos. Mediante la aplicación del programa el modelo que, según los criterios usados, mejor se ajusta los datos es un modelo que contiene cuatro términos multiplicativos para explicar el efecto de interacción. La historia del proceso de este grupo de datos es la siguiente. Los estimadores de los parámetros del modelo estrictamente aditivo son:

$$\hat{\mu} = -14.9394$$

$\hat{\alpha}_1 = 51.7061$	$\hat{\beta}_1 = 7.3485$
$\hat{\alpha}_2 = 38.5172$	$\hat{\beta}_2 = 20.5121$
$\hat{\alpha}_3 = 40.2828$	$\hat{\beta}_3 = 27.0020$
$\hat{\alpha}_4 = 25.9505$	$\hat{\beta}_4 = 122.5061$
$\hat{\alpha}_5 = 15.3061$	$\hat{\beta}_5 = 23.0939$
$\hat{\alpha}_6 = 7.6616$	$\hat{\beta}_6 = 10.3576$
$\hat{\alpha}_7 = 5.4172$	$\hat{\beta}_7 = 17.5667$

$\hat{\alpha}_8 = 6.0616$ $\quad$ $\hat{\beta}_8 = 9.2121$

$\hat{\alpha}_9 = -0.1273$ $\quad$ $\hat{\beta}_9 = 17.4121$

$\hat{\alpha}_{10} = -189.8384$

$\hat{\alpha}_{11} = -1.5384$

La suma de cuadrados de los residuos es igual a 2628175.64077.

Las hipótesis rechazadas fueron:

Homoscedasticidad	(Por renglones y por columnas al 0.01)
Normalidad	(Al 0.10, 0.05, 0.01, usando Ji cuadrada)
Simetría	(0.10, 0.02)
Kurtosis	(0.10, 0.02).

Con la inclusión de un término multiplicativo para la interacción se tuvo que los estimadores fueron los siguientes:

$$\hat{\theta}_1^2 = 2615029.9695$$

$\hat{u}_{11}$	-0.082246	$\hat{v}_{11}$	0.12103
$\hat{u}_{12}$	-0.09146	$\hat{v}_{12}$	0.10367
$\hat{u}_{13}$	-0.02299	$\hat{v}_{13}$	0.10392
$\hat{u}_{14}$	-0.0968	$\hat{v}_{14}$	-0.94253
$\hat{u}_{15}$	-0.10751	$\hat{v}_{15}$	0.11736
$\hat{u}_{16}$	-0.10728	$\hat{v}_{16}$	0.1261
$\hat{u}_{17}$	-0.09939	$\hat{v}_{17}$	0.12173
$\hat{u}_{18}$	-0.10308	$\hat{v}_{18}$	0.12729
$\hat{u}_{19}$	-0.08593	$\hat{v}_{19}$	0.12152
$\hat{u}_{110}$	0.95312		
$\hat{u}_{111}$	-0.08622		

La suma de cuadrados de los residuos fue igual a 13145.6713.

Esto representa una reducción muy grande relativo al modelo adi-

tivo.

Las hipótesis rechazadas fueron:

Normalidad (0.10, 0.05)

Independencia (0.10, 0.02)

Con el segundo término multiplicativo para la interacción los resultados obtenidos fueron los siguientes. Los estimadores de los parámetros incluidos resultaron ser:

$$\hat{\theta}_2^2 = 4019.9959$$

$\hat{u}_{21} = -0.27793$ $\quad$ $\hat{v}_{21} = 0.73429$

$\hat{u}_{22} = 0.0162$ $\quad$ $\hat{v}_{22} = 0.16557$

$\hat{u}_{23} = -0.16643$ $\quad$ $\hat{v}_{23} = -0.0045$

$\hat{u}_{24} = -0.07734$ $\quad$ $\hat{v}_{24} = -0.00421$

$\hat{u}_{25} = 0.00575$ $\quad$ $\hat{v}_{25} = -0.26657$

$\hat{u}_{26} = 0.64464$ $\quad$ $\hat{v}_{26} = -0.22283$

$\hat{u}_{27} = -0.36355$ $\quad$ $\hat{v}_{27} = -0.44413$

$\hat{u}_{28} = -0.37$ $\quad$ $\hat{v}_{28} = -0.21808$

$\hat{u}_{29} = 0.41973$ $\quad$ $\hat{v}_{29} = 0.26046$

$\hat{u}_{210} = 0.00167$

$\hat{u}_{211} = 0.16725$

La suma de cuadrados de los residuos resultó ser 9125.67539.

Las hipótesis rechazadas fueron:

Homoscedasticidad (Por renglones)

Normalidad (0.10, 0.05)

Simetría (0.10)

Independencia (0.10)

Kurtosis (0.10, 0.02)

Para el tercer término multiplicativo los estimadores de los parámetros fueron:

$$\hat{\theta}_3^2 = 3700.86987$$

$\hat{u}_{31} = -0.14125$ $\hat{v}_{31} = 0.16714$

$\hat{u}_{32} = -0.3991$ $\hat{v}_{32} = -0.0407$

$\hat{u}_{33} = -0.09276$ $\hat{v}_{33} = -0.05772$

$\hat{u}_{34} = 0.21373$ $\hat{v}_{34} = -0.00293$

$\hat{u}_{35} = 0.81293$ $\hat{v}_{35} = 0.71261$

$\hat{u}_{36} = -0.00574$ $\hat{v}_{36} = 0.24791$

$\hat{u}_{37} = 0.07648$ $\hat{v}_{37} = -0.21491$

$\hat{u}_{38} = -0.16384$ $\hat{v}_{38} = 0.51162$

$\hat{u}_{39} = -0.05057$ $\hat{v}_{39} = -0.29978$

$\hat{u}_{310} = 0.01493$

$\hat{u}_{311} = -0.26481$

La suma de cuadrados de los residuos fué igual a 5424.80551.

Las hipótesis rechazadas fueron:

Homoscedasticidad (Por renglones y por columnas)

Kurtosis (0.10, 0.02)

Con el cuarto y último término multiplicativo los estimadores de los parámetros fueron:

$$\hat{\theta}_4^2 = 2814.99131$$

$\hat{u}_{41} = 0.6184$ $\quad$ $\hat{v}_{41} = -0.34011$

$\hat{u}_{42} = 0.08638$ $\quad$ $\hat{v}_{42} = 0.76162$

$\hat{u}_{43} = 0.29633$ $\quad$ $\hat{v}_{43} = 0.39283$

$\hat{u}_{44} = -0.25594$ $\quad$ $\hat{v}_{44} = -0.0199$

$\hat{u}_{45} = 0.08617$ $\quad$ $\hat{v}_{45} = -0.09921$

$\hat{u}_{46} = 0.27431$ $\quad$ $\hat{v}_{46} = -0.07093$

$\hat{u}_{47} = -0.31234$ $\quad$ $\hat{v}_{47} = -0.2449$

$\hat{u}_{48} = -0.15347$ $\quad$ $\hat{v}_{48} = -0.1528$

$\hat{u}_{49} = -0.16208$ $\quad$ $\hat{v}_{49} = -0.2266$

$\hat{u}_{410} = -0.00156$

$\hat{u}_{411} = -0.47619$

La suma de cuadrados de los residuos fué igual a 2609.4142.

Esta vez ninguna de las hipótesis fué rechazada.

Parecería que si el terreno tuviese alguna tendencia de fertilidad todos los modelos ajustados a los distintos grupos de datos deberían parecerse, sin embargo es notable la diferencia existente entre el tercero y los dos restantes; como se explicó antes aquel grupo de datos fué tomado en un año en que los cambios climatológicos resultaron particularmente serios, por lo que el tercer modelo resulta de poca aplicabilidad a los rendimientos de otros años. Además la configuración muy particular del experimento (1o. Maíz, 2o. Avena, 3o. Trigo), puede tener

influencia en los resultados obtenidos del mismo. Es decir parece ser necesario que el experimento se repita en las mismas condiciones para poder afirmar algo positivo o tomar otra parte de la superficie cultivada en el mismo experimento con el objeto de poder hacer alguna consideración al respecto. Esto es una indicación clara de que los rendimientos siguen un patrón de fertilidad del suelo pero este patrón cambia al considerar diferentes cultivos y diferentes condiciones ambientales.

5) Otro grupo de datos a los que se aplicó el análisis fué tomado del libro de V.G. Panse y P.V. Sukhatme (20), y consiste de un ensayo de uniformidad ("cosecha de un terreno con un cultivo particular con un tratamiento uniforme, dividiendo el terreno en pequeñas unidades, y cosechando y anotando separadamente la producción de cada una de estas unidades"), con algodón efectuado en el Institute of Plant Industry, de Indore. La matriz de datos que se presenta a continuación representa sólo una parte del experimento.

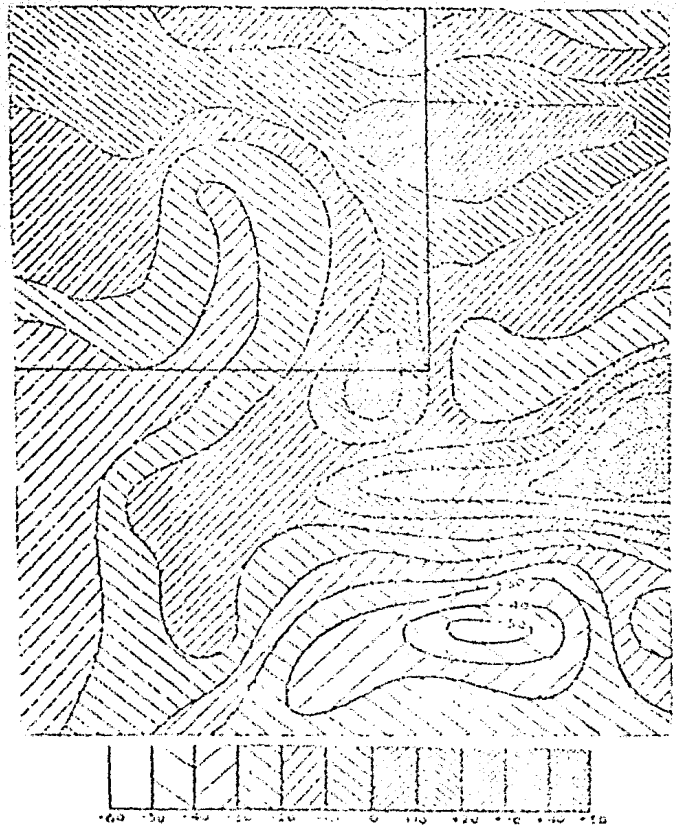
Esta matriz se reproduce en la tabla E 5.

TABLA E 5

ENSAYO DE UNIFORMIDAD CON ALGODON MALVI, 1933-34, INSTITUTE OF PLANT INDUSTRY, INDORE

93	95	111	116	97	82	80	131	80	102	97	95	87	136	118	89	57	106	118	97
49	63	50	115	83	78	69	93	85	84	64	69	51	44	115	61	54	54	36	88
89	90	95	129	97	133	72	115	97	76	91	106	40	147	119	66	122	63	72	34
85	91	115	107	99	77	76	63	70	71	59	78	58	85	61	104	84	104	59	56
63	96	68	69	101	120	96	100	70	110	88	75	101	61	106	113	93	62	92	109
86	82	113	74	117	28	83	81	109	91	53	108	83	74	129	89	79	89	75	114
62	86	119	80	79	92	73	103	89	77	67	88	97	50	101	111	100	96	66	113
94	76	119	75	125	96	95	74	90	63	68	66	96	66	85	74	160	148	84	110
45	96	115	130	99	77	104	116	64	77	32	88	67	76	72	87	59	123	67	59
51	79	111	79	84	84	67	63	66	40	56	80	58	68	52	76	87	72	91	81
81	80	72	102	104	81	42	76	85	80	56	74	93	88	60	111	128	124	117	118
74	51	66	86	71	73	98	87	61	111	81	61	73	61	63	85	115	80	82	76
85	89	112	88	105	76	76	78	72	92	59	81	90	52	82	68	88	94	68	106
84	64	101	90	69	101	87	70	63	75	63	88	74	81	68	87	82	80	64	105
85	55	88	87	73	119	60	76	60	73	77	44	83	82	74	52	97	65	59	75
82	83	102	80	88	57	81	55	75	69	70	84	59	73	77	81	98	74	95	110
46	54	83	60	57	75	73	50	66	31	56	66	43	70	59	35	90	62	89	96
71	68	39	80	73	86	64	89	101	88	63	83	74	78	74	50	126	98	104	109
58	52	85	99	72	96	67	86	37	71	62	48	63	64	106	90	91	86	83	104
65	45	77	78	93	60	94	82	72	81	84	64	75	97	114	109	111	74	107	86

En el mismo libro se presenta un "mapa de contorno de fertilidad", donde resulta obvio que tanto los renglones como las columnas interactúan, es decir los efectos principales por renglón no son constantes, ocurriendo lo mismo en las columnas. El mapa aparece reproducido a continuación



Mapa de contorno de fertilidad de los datos del ensayo de uniformidad de algodón.

En él se marcan los límites del terreno considerado en el grupo de datos al que fué aplicado el programa.

Los estimadores de los parámetros que representan los efectos principales son los siguientes:

$$\hat{\mu} = 83.7275$$

$\hat{\alpha}_1 = 15.6225$ $\quad$ $\hat{\beta}_1 = -11.3275$

$\hat{\alpha}_2 = -13.4775$ $\quad$ $\hat{\beta}_2 = -8.9775$

$\hat{\alpha}_3 = 8.4225$ $\quad$ $\hat{\beta}_3 = 8.3225$

$\hat{\alpha}_4 = -3.6275$ $\quad$ $\hat{\beta}_4 = 7.4725$

$\hat{\alpha}_5 = 5.9225$ $\quad$ $\hat{\beta}_5 = 5.5725$

$\hat{\alpha}_6 = 4.1225$ $\quad$ $\hat{\beta}_6 = 0.3225$

$\hat{\alpha}_7 = 3.6725$ $\quad$ $\hat{\beta}_7 = -5.9275$

$\hat{\alpha}_8 = 9.4725$ $\quad$ $\hat{\beta}_8 = 0.6725$

$\hat{\alpha}_9 = -1.0775$ $\quad$ $\hat{\beta}_9 = -8.1275$

$\hat{\alpha}_{10} = -11.4775$ $\quad$ $\hat{\beta}_{10} = -5.6725$

$\hat{\alpha}_{11} = 4.8725$ $\quad$ $\hat{\beta}_{11} = -16.4275$

$\hat{\alpha}_{12} = -5.9775$ $\quad$ $\hat{\beta}_{12} = -6.4275$

$\hat{\alpha}_{13} = -0.6775$ $\quad$ $\hat{\beta}_{13} = 22.8225$

$\hat{\alpha}_{14} = 29.3725$ $\quad$ $\hat{\beta}_{14} = -6.0775$

$\hat{\alpha}_{15} = -9.5275$ $\quad$ $\hat{\beta}_{15} = 3.0225$

$\hat{\alpha}_{16} = -4.0775$ $\quad$ $\hat{\beta}_{16} = -1.8275$

$\hat{\alpha}_{17} = -20.6775$ $\quad$ $\hat{\beta}_{17} = 12.3225$

$\hat{\alpha}_{18} = -2.8275$ $\quad$ $\hat{\beta}_{18} = 3.9725$

$\hat{\alpha}_{19} = -7.7275$ $\quad$ $\hat{\beta}_{19} = -2.3275$

$\hat{\alpha}_{20} = -0.3275$ $\quad$ $\hat{\beta}_{20} = 8.5725$

La suma de cuadrados de los residuos resultó ser igual a 536191.4025.

Las hipótesis rechazadas fueron:

Normalidad	(Ji-Cuadrada, 0.10, 0.05, 0.01)
Simetría	(0.10, 0.02)
Kurtosis	(0.10, 0.02)

El rechazo de estas hipótesis motivó la inclusión de un término del tipo $\theta\, u_i v_j$; los estimadores de los parámetros fueron los siguientes:

$$\hat{\theta}_1^2 = 411689.2818222$$

$\hat{u}_{11} = -0.05857$ $\quad$ $\hat{v}_{11} = -0.02827$

$\hat{u}_{12} = -0.06914$ $\quad$ $\hat{v}_{12} = -0.06505$

$\hat{u}_{13} = -0.12441$ $\quad$ $\hat{v}_{13} = -0.033$

$\hat{u}_{14} = -0.07018$ $\quad$ $\hat{v}_{14} = -0.05683$

$\hat{u}_{15} = -0.01744$ $\quad$ $\hat{v}_{15} = -0.07832$

$\hat{u}_{16} = -0.04653$ $\quad$ $\hat{v}_{16} = -0.01996$

$\hat{u}_{17} = -0.01918$ $\quad$ $\hat{v}_{17} = -0.03262$

$\hat{u}_{18} = -0.03224$ $\quad$ $\hat{v}_{18} = -0.07201$

$\hat{u}_{19} = -0.06247$ $\quad$ $\hat{v}_{19} = -0.06870$

$\hat{u}_{110} = -0.05765$ $\quad$ $\hat{v}_{110} = -0.04824$

$\hat{u}_{111} = -0.03053$ $\quad$ $\hat{v}_{111} = -0.05495$

$\hat{u}_{112} = -0.04341$ $\quad$ $\hat{v}_{112} = -0.0348$

$\hat{u}_{113} = -0.02386$ $\quad$ $\hat{v}_{113} = 0.97114$

$\hat{u}_{114} = 0.96903$ $\quad$ $\hat{v}_{114} = -0.05267$

$\hat{u}_{115} = -0.02016$ $\quad$ $\hat{v}_{115} = -0.07982$

$\hat{u}_{116} = 0.06785$ $\quad$ $\hat{v}_{116} = -0.0347$

$\hat{u}_{117} = -0.06674$ $\quad$ $\hat{v}_{117} = -0.0669$

$\hat{u}_{118} = -0.04982$ $\quad$ $\hat{v}_{118} = -0.05473$

$\hat{u}_{119} = -0.05600$ $\quad$ $\hat{v}_{119} = -0.07341$

$\hat{u}_{120} = -0.05287$ $\quad$ $\hat{v}_{120} = -0.01617$

Con una suma de cuadrados igual a 124502.120655.

La única hipótesis rechazada fué la siguiente:

Kurtosis $\quad$ (0.10, 0.02)

Lo que propició la inclusión de otro término multiplicativo para la explicación de la interacción; los estimadores de los parámetros considerados en este término fueron:

$$\hat{\theta}_2^2 = 27708.3075459$$

$\hat{u}_{21} = 0.33403$ $\quad$ $\hat{v}_{21} = -0.06925$

$\hat{u}_{22} = 0.3212$ $\quad$ $\hat{v}_{22} = 0.06354$

$\hat{u}_{23} = 0.506$ $\quad$ $\hat{v}_{23} = -0.02916$

$\hat{u}_{24} = 0.06267$ $\quad$ $\hat{v}_{24} = 0.34258$

$\hat{u}_{25}$= 0.04373 $\hat{v}_{25}$= -0.08131

$\hat{u}_{26}$= -0.0429 $\hat{v}_{26}$= 0.12726

$\hat{u}_{27}$= -0.06595 $\hat{v}_{27}$= 0.01489

$\hat{u}_{28}$= -0.46116 $\hat{v}_{28}$= 0.32595

$\hat{u}_{29}$= 0.2585 $\hat{v}_{29}$= -0.05737

$\hat{u}_{210}$= -0.09003 $\hat{v}_{210}$= 0.10497

$\hat{u}_{211}$= -0.33146 $\hat{v}_{211}$= 0.08738

$\hat{u}_{212}$= -0.05363 $\hat{v}_{212}$= 0.10279

$\hat{u}_{213}$= -0.08419 $\hat{v}_{213}$= 0.01272

$\hat{u}_{214}$= 0.05844 $\hat{v}_{214}$= 0.29172

$\hat{u}_{215}$= 0.098 $\hat{v}_{215}$= 0.26634

$\hat{u}_{216}$= -0.17952 $\hat{v}_{216}$= -0.05868

$\hat{u}_{217}$= -0.16371 $\hat{v}_{217}$= -0.45641

$\hat{u}_{218}$= -0.19248 $\hat{v}_{218}$= -0.30781

$\hat{u}_{219}$= 0.00732 $\hat{v}_{219}$= -0.24742

$\hat{u}_{220}$= -0.02487 $\hat{v}_{220}$= -0.43273

La suma de cuadrados de los residuos alcanzó el valor de 97793.8131.

Con hipótesis rechazadas, las siguientes:

Normalidad (0.10, 0.05, 0.01)

Kurtosis (0.10)

Los estimadores del tercer término multiplicativo fueron los siguientes:

$$\hat{\theta}_3^2 = 21307.0868892$$

$\hat{u}_{31} = 0.05993$	$\hat{v}_{31} = -0.5498$
$\hat{u}_{32} = -0.01911$	$\hat{v}_{32} = 0.24252$
$\hat{u}_{33} = -0.29119$	$\hat{v}_{33} = 0.48614$
$\hat{u}_{34} = 0.29149$	$\hat{v}_{34} = 0.08163$
$\hat{u}_{35} = -0.20065$	$\hat{v}_{35} = 0.20885$
$\hat{u}_{36} = 0.37377$	$\hat{v}_{36} = -0.37978$
$\hat{u}_{37} = 0.17063$	$\hat{v}_{37} = 0.08479$
$\hat{u}_{38} = 0.05414$	$\hat{v}_{38} = -0.03151$
$\hat{u}_{39} = 0.48417$	$\hat{v}_{39} = 0.00633$
$\hat{u}_{310} = 0.1156$	$\hat{v}_{310} = -0.09113$
$\hat{u}_{311} = -0.06157$	$\hat{v}_{311} = -0.31913$
$\hat{u}_{312} = -0.21736$	$\hat{v}_{312} = 0.18168$
$\hat{u}_{313} = 0.21659$	$\hat{v}_{313} = -0.00604$
$\hat{u}_{314} = -0.00588$	$\hat{v}_{314} = -0.20317$
$\hat{u}_{315} = -0.28581$	$\hat{v}_{315} = -0.05452$
$\hat{u}_{316} = 0.07974$	$\hat{v}_{316} = 0.16103$
$\hat{u}_{317} = -0.12516$	$\hat{v}_{317} = -0.39549$
$\hat{u}_{318} = -0.24209$	$\hat{v}_{318} = 0.29578$
$\hat{u}_{319} = -0.10623$	$\hat{v}_{319} = -0.2043$
$\hat{u}_{320} = -0.19101$	$\hat{v}_{320} = -0.0087$

La suma de cuadrados de los residuos obtenidos al sustituir estos estimadores descendió hasta 76486.766222.

La hipótesis rechazada fué:

Normalidad (0.10, 0.05)

Entonces se incluyó un cuarto término, con lo que se logró que ninguna de las hipótesis fuese rechazada; los estimadores de los parámetros que componen dicho término fueron los siguientes:

$$\hat{\theta}_4^2 = 18400.91374743$$

$\hat{u}_{41} = -0.03637$	$\hat{v}_{41} = 0.12066$
$\hat{u}_{42} = -0.37919$	$\hat{v}_{42} = 0.06741$
$\hat{u}_{43} = 0.40973$	$\hat{v}_{43} = 0.27358$
$\hat{u}_{44} = 0.30708$	$\hat{v}_{44} = 0.18766$
$\hat{u}_{45} = -0.38116$	$\hat{v}_{45} = 0.0039$
$\hat{u}_{46} = -0.36171$	$\hat{v}_{46} = 0.24898$
$\hat{u}_{47} = -0.16988$	$\hat{v}_{47} = -0.09789$
$\hat{u}_{48} = 0.26277$	$\hat{v}_{48} = -0.11825$
$\hat{u}_{49} = 0.18194$	$\hat{v}_{49} = -0.06469$
$\hat{u}_{410} = 0.22148$	$\hat{v}_{410} = -0.34737$
$\hat{u}_{411} = 0.12148$	$\hat{v}_{411} = -0.07114$
$\hat{u}_{412} = -0.09167$	$\hat{v}_{412} = -0.00633$
$\hat{u}_{413} = -0.06233$	$\hat{v}_{413} = -0.00305$
$\hat{u}_{414} = 0.04045$	$\hat{v}_{414} = 0.36386$
$\hat{u}_{415} = 0.20218$	$\hat{v}_{415} = -0.42572$
$\hat{u}_{416} = -0.04222$	$\hat{v}_{416} = -0.2021$
$\hat{u}_{417} = 0.12217$	$\hat{v}_{417} = 0.26236$

$\hat{u}_{418} = -0.04135$ $\quad\quad$ $\hat{v}_{418} = 0.235$

$\hat{u}_{419} = -0.11613$ $\quad\quad$ $\hat{v}_{419} = -0.00646$

$\hat{u}_{420} = -0.18726$ $\quad\quad$ $\hat{v}_{420} = -0.42039$

Con una suma de cuadrados, para los residuos, igual a 58085.812476.

Comparando este resultado con el que se hubiéra obtenido por el método de Mandel que ve que el número de términos incluído en el modelo es cuatro veces mayor; el estimador de la varianza para el modelo que incluye cuatro términos sería de 868.89, mientras que el obtenido con el modelo que incluye un solo término sería de 1863.39. Los valores requeridos para la aplicación del criterio de Mandel aparecen ordenados en la tabla de análisis de la varianza incluida en seguida.

TABLA 5.1

ANALISIS DE LA VARIANZA PARA LOS TERMINOS DE LA INTERACCION

FUENTE	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADOS MEDIOS
$\theta_1 u_{1i} v_{1j}$	411689.28	74.25	5558.13
$\theta_2 u_{2i} v_{2j}$	26708.31	62.24	429.11
$\theta_3 u_{3i} v_{3j}$	21307.09	53.50	398.26
$\theta_4 u_{4i} v_{4j}$	18400.91	48.13	382.31
errores	58085.81	66.85	868.89

6) Los siguientes grupos de datos son los obtenidos mediante una simulación en la computadora. El mecanismo generador de números aleatorios de la computadora produce números o variables aleatorias independientes que tienen una distribución uniforme en el intervalo (0,1). Haciendo uso de un teorema que dice "sean U y V variables aleatorias independientes y uniformemente distribuidas en el invervalo (0,1), entonces las variables aleatorias:

$$X_1 = (-2 \log_e U)^{1/2} \operatorname{sen} (2\pi V)$$
$$X_2 = (-2 \log_e U)^{1/2} \cos (2\pi V)$$

Son independientes y tienen una distribución normal con media cero y varianza uno", y de las variables aleatorias generadas en la computadora, se generaron variables aleatorias con distribución normal con media cero y desviación estandar σ . Estos errores aleatorios fueron sumados a distintos modelos (distintos en el efecto de interacción), y a esos datos simulados se les aplicó el programa de computadora.

Como se indicó anteriormente, el modelo usado en la simulación fué el siguiente:

$$Y_{ij} = 2(i-1) + 3j + \left((i-3)(j-1) \right)^r + \varepsilon_{ij}$$

$$i = 1, \ldots, 5$$

$$j = 1, \ldots, 4$$

con $\varepsilon_{ij} \sim N(0, \sigma^2)$

A r se le asignaron los valores 1,3,5,7,9 y σ tomó los valores 1/20, 6/20, 11/20, 16/20, 21/20; es decir, se generaron 25 matrices de datos.

Los resultados pueden resumirse en una tabla donde se indica cuáles de las hipótesis hechas acerca de los errores se rechazaron habiendo incluido k términos para explicación del efecto de interacción.

Además se incluyen los valores de la suma de cuadrados de los residuos y el valor del estimador de σ , para el mismo número de términos.

σ	r	NO. DE TERMINOS	SUMA DE CUADRADOS	$\hat{\theta}^2$	$\hat{\sigma}$	HIPOTESIS RECHAZADAS
.05	1	0	50.62431	2.057	-	Ninguna
.30	1	0	48.1089	2.0022	-	Ninguna
.55	1	0	52.5376	2.092	-	Ninguna
.80	1	0	79.90938	2.942	-	Ninguna
1.05	1	0	23.95009	1.8291	-	Ninguna
.05	3	0	61147.2634	71.383	-	N5,K2
		1	0.0204	.074	61143.243	I1
		2	0.00486	.073	.0155411	HR,I2,K1
		3	$(1.457)\times10^{-17}$	-	.0048584	HC,N5,S2

σ	r	NO. DE TERMINOS	SUMA DE CUADRADOS	$\hat{\theta}^2$	$\hat{\sigma}$	HIPOTESIS RECHAZADAS
0.30	3	0	60920.57	71.25	-	N5,K2
		1	0.407855	.333	60920.1628	I2
		2	.04155284	.25	0.3663	Ninguna
.55	3	0	60885.3151	71.23	-	N5,K2
		1	1.55388	.65	60883.1612	Ninguna
.80	3	0	61371.52956	71.51	-	N5,K2
		1	7.67	1.442	61363.8595	I 2
		2	0.466178	0.84	7.203824	N4,HR
		3	1.1778×10^{-16}	-	0.466178	HC,N5,S2
1.05	3	0	61192.4799	71.40	-	N5,K2
		1	1.94122	.72	61190.5387	Ninguna
.05	5	0	33703930.187	1675.90	-	HR,HC,N5, S2,K2
		1	4.587266	1.1178	33703925.599	HR,N3,K2
		2	0.00968326	0.1211	4.5776	N5,I2,K1
		3	$(3.3615) \times 10^{-15}$	-	0.00968	N5
.30	5	0	84112214.767	2647.51	-	HC,N5,K2

q	r	NO. DE TERMINOS	SUMA DE CUADRADOS	$\hat{\theta}^2$	$\hat{q}$	HIPOTESIS RECHAZADAS
		1	0.142344261	.1969	84112214.625	I2
		2	0.027256766	.2032	0.115087	Ninguna
.55	5	0	84132634.0647	2647.75	-	HC,N5,K2
		1	0.993415382	.52	84132633.072	N3
		2	0.2100978	.564	0.783317	N4,I2
		3	$(4.3831)X10^{-14}$	-	0.2100978	HC,N5,S2
.80	5	0	84120425.8584	2647.56	-	HC,N5,K2
		1	1.404441	.618	84120424.454	N3,S1,K2
		2	.2044	.556	1.20003735	I2
		3	$(7,0726)X10^{-14}$	-	0.204404	HC,N5,S2,K2
1.05	5	0	84132634.0647	2647.83	-	HC,N5,K2
		1	0.993415	.520	84132633.072	N3,S1
		2	.2100978	.564	0.7833176	N4,I2
		3	$(4.3831)X10^{-14}$	-	0.21009779	HC,N5,S2
.05	7	0	$(1.1328)X10^{11}$	$0(10^5)$	-	HC,N5,K2
		1	10.746	1.71	$0(10^{11})$	-

σ	r	NO. DE TERMINOS	SUMA DE CUADRADOS	$\hat{\theta}^2$	$\hat{\sigma}$	HIPOTESIS RECHAZADAS
.30	7	0	$(1.1328)X10^{11}$	$0(10^5)$	-	HC,N5,K2
		1	13.19405	1.89	$0(10^{11})$	Ninguna
.55	7	0	$(1.13331)X10^{11}$	$0(10^5)$	-	HC,N5,K2
		1	2.145226	.764	$0(10^{11})$	N4
		2	.008469	.1132	2.13675787	N1,I2
		3	$(2.8896)X10^{-11}$	-	0.0084689	N5,S2
.80	7	0	$(1.1333)X10^{11}$	$0(10^5)$	-	HC,N5,K2
		1	2.544475	.832	$0(10^{11})$	Ninguna
1.05	7	0	$(1.13332)X10^{11}$	$0(10^5)$	-	HC,N5,K2
		1	2.6159744	.844	$0(10^{11})$	N5,I2,
		2	.57813344	.935	2.037841	HR,N5,S2,K2
		3	$(2.4488)X10^{-10}$	-	0.57813344	N5,S2
.05	9	0	$(1.497964)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	0.016686518	0.0674	$0(10^{14})$	N5,S1,K1
		2	0.00232039	0.0592	.014366129	N2,I2,K1
		3	0.000000038	-	.00232035	N5,S2

σ	r	NO. DE TERMINOS	SUMA DE CUADRADOS	$\hat{\theta}^2$	$\hat{\sigma}$	HIPOTESIS RECHAZADAS
.30	9	0	$(1.4979648)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	0.507979325	.37203	$0(10^{13})$	N5
		2	0.125681986	.4363	0.3822997	N3,I2
		3	0.0000003	-	0.125682	HC,N5,S2
.55	9	0	$(1.49496487)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	18.639114	2.2536	$0(10^{13})$	HR,N2,S2,K2
		2	0.450466909	.8261	18.188647	N3,I2
		3	0.000000042	-	0.4504668	HC,N5,S2
.80	9	0	$(1.497964456)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	10.123865964	1.6609	$0(10^{13})$	Ninguno
1.05	9	0	$(1.49796476)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	8.6271067	1.533	$0(10^{13})$	N4,K1
		2	0.310009369	.6853	8.3170972	I2
		3	0.00000003	-	0.3100093	HC,N5,S2,K2

Donde por ejemplo KR quiere decir Kurtosis en sus R niveles de significancia mayores. Por otra parte $0(10^r)$ representa un número cuyo orden de magnitud era de 10^r.

Cómo puede observarse de la tabla, cuando los efectos de interacción son pequeños, por ejemplo para r=1 no se detectó ningún efecto de interacción (puede haber ocurrido en los rendimientos de avena del ejemplo 4), en contraste con los efectos de interacción grandes, por ejemplo para r=9.

Con el objeto de aclarar un poco más los puntos de interés se desarrollan en seguida los resultados obtenidos por las simulaciones.

Como se indica en el cuadro que resume los resultados, en el caso en el que el modelo era el siguiente:

$$Y_{ij} = 2(i-1) + 3j + (i-3)(j-1) + \varepsilon_{ij}$$

$$i=1,\ldots,5$$

$$j=1,\ldots,4$$

Con $\varepsilon_{ij} \sim N(0,1/400)$

El modelo ajustado fué un modelo estrictamente aditivo.

Dando por resultado una suma de cuadrados del error igual a 50.62431 con un estimador de la desviación estander dado por 2.057 cuando la desviación estander real era de 0.05= 1/20; es decir la varianza real fué sobrestimada y el modelo real no fué descubierto ya que no detectó el efecto de interacción, y es esto lo que produce sobre estimaciones de σ^2_1.

Algo semejante ocurrió con los modelos para los cuales las desviaciones estander de los errores eran 0.30, 0.55, 0.80, 1.05

σ	r	NO. DE TERMINOS	SUMA DE CUADRADOS	$\hat{\theta}^2$	$\hat{\sigma}$	HIPOTESIS RECHAZADAS
.30	9	0	$(1.4979648)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	0.507979325	.37203	$0(10^{13})$	N5
		2	0.125681986	.4363	0.3822997	N3,I2
		3	0.0000003	-	0.125682	HC,N5,S2
.55	9	0	$(1.49496487)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	18.639114	2.2536	$0(10^{13})$	HR,N2,S2,K2
		2	0.450466909	.8261	18.188647	N3,I2
		3	0.000000042	-	0.4504668	HC,N5,S2
.80	9	0	$(1.497964456)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	10.123865964	1.6609	$0(10^{13})$	Ninguna
1.05	9	0	$(1.49796476)X10^{14}$	$0(10^6)$	-	HC,N5,K2
		1	8.6271067	1.533	$0(10^{13})$	N4,K1
		2	0.310009369	.6853	8.3170972	I2
		3	0.00000003	-	0.3100093	HC,N5,S2,K2

Donde por ejemplo HR quiere decir Kurtosis en sus R niveles de significancia mayores. Por otra parte $0(10^r)$ representa un número cuyo orden de magnitud era de 10^r.

con los estimadores respectivos dados por 2.0022, 2.092, 2.942, 1.8291; las sumas de cuadrados para los modelos fueron 48.1089, 52.5376, 79.90938, 23,95009 respectivamente. Así pues, mientras que las desviaciones estandar eran sobreestimadas, los efectos de interacción no fueron siquiera detectados. Cabe aquí hacer notar que las dimensiones de las matrices de datos son pequeñas con sólo 20 elementos por cada matriz. Además puede verse la forma del modelo, que este es lineal en i y también lineal en j. Esto es, se tiene una interacción de tipo lineal, que no es detectada en el modelo.

Algo diferente ocurrió cuando el modelo real era:

$$Y_{ij} = 2(i-1) + 3j + ((i-3) (j-1))^3 + \varepsilon_{ij}$$

Con $\varepsilon_{ij} \sim N(0, \sigma^2)$

Donde a σ se le asignaron los valores de .05, .30, .55, .80, 1.05 y en cada caso los resultados obtenidos, aparte de los estimadores, fueron los que en seguida se reportan.

Para el modelo con $\sigma = .05$ el ajuste de un modelo estrictamente aditivo dió como resultado una suma de cuadrados del error de 61147.2634, con un estimador de la desviación estandar $\hat{\sigma} = 71.3$. Las hipótesis rechazadas con los residuos resultantes fueron la de normalidad a los cinco niveles de significancia y la de kurtosis en sus dos niveles.

Al aumentar un término para el efecto de interacción, el estima-

dor de θ^2 alcanzó un valor de 61147.243 con lo que la suma de cuadrados se redujo hasta 0.0204, dado un estimador de la desviación estandar 1/ $\hat{\sigma}$ = 0.074, el cuál se encuentra apenas 0.024 arriba del valor real de σ . La única hipótesis rechazada fué la de independencia de los errores al 10% de significancia. Esto indica que con 5% de significancia, no se hubiéra rechazado y se tendría un modelo bastante aceptable.

El aumento de un nuevo término para el efecto de interacción redujo aún más la suma de cuadrados siendo el valor de este igual a 0.00486 pero el estimador de la desviación estandar fué de 0.073; es decir la ganancia al introducir el nuevo término fué pequeña. Las hipótesis rechazadas fueron las de homoscedasticidad por renglones, independencia de los residuos en sus dos niveles y kurtosis al 10%.

El estimador de θ^2 para el tercer término del efecto de interacción fué de 0.00485?4 y para su estimación se hizo uso de 0.66 grados de libertad, con lo que se agotaron los grados de liber-

---

1/ Los estimadores de σ se calcularon haciendo uso de los valores dados para los grados de libertad en la tabla 7 del apéndice II; esto se hizo de la siguiente manera:

$$\hat{\sigma}_k = \left\{ \frac{SCD_k}{\nu_{k+1}+\ldots+\nu_3} \right\}^{1/2}$$

Donde SCD_k es la suma de cuadrados de los residuos obtenidos después de aumentar k términos y $\nu_1 = 8.23$, $\nu_2 = 3.01$ nótese que

$$\nu_1 + \nu_2 + \nu_3 = 12$$

tad y en consecuencia no fué posible dar un estimador de σ para este modelo; sin embargo la suma de cuadrados de los residuos fué de tan sólo 1.457×10^{-17}. Las hipótesis rechazadas fueron las de homoscedasticidad por columnas, normalidad en sus cinco niveles y simetría en sus dos niveles.

Para el caso en el cual la desviación estandar era igual a 0.30, un modelo aditivo dió como resultado una suma de cuadrados de 60920.57 y las hipótesis rechazadas fueron las de normalidad en sus cinco niveles y kurtosis en sus dos niveles.

El estimador de θ^2 para el primer término de la interacción fué de 60920.1628 y los correspondientes grados de libertad de 8.33. La suma de cuadrados de los residuos se redujo hasta el valor 0.407855 dando un estimador de la desviación estandar de 0.333. La hipótesis rechazada fué unicamente la de independencia de los residuos en sus dos niveles de significancia.

En vista del tamaño de $\hat{\theta}^2$ y el de la suma de cuadrados la aplicación del criterio de Mandel habría incluído únicamente un término para la explicación del efecto de interacción, sin embargo la inclusión de otro término para la interacción en el modelo, con un estimador de θ^2 igual a 0.3663 y con 3.01 grados de libertad, trajo como resultado una reducción en el valor del estimador de la desviación estandar hasta el valor de 0.25 y el no rechazo de ninguna de las hipótesis, mejorando notablemente el ajuste

de los datos.

Para el valor de $\sigma = 0.55$, la aplicación de ambos criterios para la selección del modelo habrían coincidido ya que la inclusión de un término para el efecto de interacción, con $\hat{\theta}^2 = 60833.7612$ y 8.33 grados de libertad, trajo como consecuencia una suma de cuadrados de 1.55388, un estimador de la desviación estandar igual a 0.65, estimado con 3.67 grados de libertad, y el no rechazo de ninguna de las hipótesis. Es preciso hacer notar que las hipótesis rechazadas para los residuos obtenidos después de ajustar un modelo aditivo fueron normalidad en sus cinco niveles y kurtosis en sus dos niveles.

Para el mismo modelo pero con $\sigma = 0.80$, el ajuste de un modelo estrictamente aditivo dió como resultado una suma de cuadrados de 61371.5295 y el rechazo de las hipótesis de normalidad en sus cinco niveles y kurtosis en sus dos niveles.

Al incluir un término para la explicación del efecto de interacción, para el cual $\hat{\theta}^2 = 61363.8595$, el estimador de σ, la desviación estandar, fué de 1.442, es decir, 0.642 arriba del valor real del parámetro. La única hipótesis rechazada en este caso fué la de independencia en sus dos niveles. La inclusión de un segundo término para el efecto de interacción produjo la reducción de la suma de cuadrados hasta el valor 0.466178, la desviación estandar fué estimada por 0.84 y las hipótesis rechazadas

fueron las de homoscedasticidad por renglones y normalidad en sus cuatro niveles de significancia mayores; para este término el estimador de θ^2 fué igual a 7.203824. Para un tercer término para el efecto de interacción se tuvo que el estimador de θ^2 fué 0.466178.

Para este término no restan grados de libertad para la estimación de σ . El programa reporta como suma de cuadrados el valor 1.1778×10^{-16}, aunque este valor puede deberse a errores de aproximación propios del sistema de cómputo. Además se tuvo que las hipótesis rechazadas para los residuos obtenidos fueron las de homoscedasticidad por columnas, normalidad en sus cinco niveles y simetría en sus dos niveles.

Cuando el valor de la desviación estandar cambió a 1.05, los resultados obtenidos al ajustar un modelo aditivo fueron para la suma de cuadrados un valor de 61192.4799, rechazando las hipótesis de normalidad en sus cinco niveles y kurtosis en sus dos niveles. Cuando se incrementó un término para el efecto de interacción, con $\hat{\theta}^2 = 61190.53871$, se obtuvo una suma de cuadrados de 1.94122 de donde el estimador de la desviación estandar alcanzó el valor de 0.72; es decir, el estimador estuvo 0.33 abajo del valor real de el parámetro. En este caso no se rechazó ninguna de las hipótesis, de donde el modelo ajustado fué uno que incluyese un término multiplicativo para el efecto de interac-

ción.

Cuando el modelo que generó los datos fué cambiado a:

$$Y_{ij} = 2\,(i-1) + 3j + (\,(1-3)\,(j-1)\,)^5 + \varepsilon_{ij}$$

Con $\varepsilon_{ij} \sim N(0,\sigma^2)$, $\sigma = 0.05, 0.30, 0.55, 0.80, 1.05$; se tuvo que para los valores de σ iguales a 0.05, 0.80, y 1.05 al menos una de las hipótesis fué rechazada a cada paso, es decir, ninguno de los distintos modelos ajustados satisfizo completamente las hipótesis hechas acerca de la distribución conjunta de los errores.

Los resultados obtenidos para este modelo con $\sigma = 0.05$ se pueden resumir de la siguiente manera. Cuando se ajustó un modelo estrictamente aditivo se obtuvo una suma de cuadrados igual a 33703930.187 y las hipótesis rechazadas fueron homoscedasticidad por renglones y por columnas, normalidad en sus cinco niveles, simetría en sus dos niveles y kurtosis también su sus dos niveles. Cuando se incluyó un término para el efecto de interacción con $\hat{\theta}^2 = 33703925.5996$, la suma de cuadrados de los residuos se redujo hasta alcanzar el valor de 4.587266 dando como resultado $\hat{\sigma} = 1.1178$, nótese que este valor es de casi 22 veces el del valor real de la desviación estandar. Las hipótesis rechazadas fueron la homoscedasticidad por renglones, la de normalidad en sus tres niveles de significancia mayores y la de kurtosis a los dos niveles de significancia usados. El aumento de un término multiplicativo para el efecto de interacción, para el

cual se tuvo que $\hat{\theta}^2 = 4.5776$ dió como resultado que la suma de cuadrados fuese igual a 0.00968326 y el estimador de la desviación estandar tuviese un valor de 0.1211, el cual es casi 2.5 veces el valor real de la desviación estandar. En términos generales las propiedades distribucionales del error no fueron mucho mejores ya que las hipótesis rechazadas fueron las de normalidad en sus cinco niveles; independencia en los dos niveles y kurtosis en un sólo de los niveles de significancia. Al obtenerse el estimador del tercer término para explicación del efecto de interacción se hizo uso de los grados de libertad restantes con lo cual no es posible tener un estimador de la desviación estandar de los errores, sin embargo la suma de cuadrados, incluye errores de redondeo por la máquina, se redujo hasta 3.3615×10^{-15}.

La única hipótesis rechazada en este caso fué la de normalidad en los cinco niveles usados.

El uso de el criterio de Mandel, dada la magnitud del primer estimador ($\hat{\theta}^2 = 33703925.5996$), hubiése aceptado un sólo término para la explicación del efecto de interacción lo que habría dado como resultado un estimado de la varianza de 440 veces el valor real y residuos con propiedades muy pobres.

Cuando se cambió el valor de σ a 0.30 los resultados fueron algo poco común. De el ajuste de un modelo aditivo se obtuvo un va-

lor de la suma de cuadrados de los residuos igual a 84112214.767 y el rechazo de las hipótesis de homoscedasticidad por columnas, normalidad a los cinco niveles usados y kurtosis a los dos niveles utilizados. Al incluir un término multiplicativo al modelo, el ajuste de este modelo trajo como consecuencia que la suma de cuadrados fuera igual a 0.142344261, el estimador de la desviación estandar tomó el valor de 0.1969 y la única hipótesis rechazada fuera la de independencia a los niveles de 0.10 y 0.02. Ahora bien al incluir un nuevo término al efecto de interacción, con $\hat{\theta}^2$ = 0.115087, la suma de cuadrados se redujo hasta el valor 0.027256766, pero el estimador de la desviación estandar "creció" y tuvo el valor de 0.2032 y con este valor para $\hat{\sigma}$ ninguna de las hipótesis fué rechazada. Así pues, el modelo ajustado a este grupo de datos fué uno que incluyó dos términos multiplicativos para explicación del efecto de interacción. La aplicación del criterio de Mandel habría resultado en el ajuste de un modelo con un sólo término para el efecto de interacción.

Para el mismo modelo, pero con ∇ = .55 el ajuste de un modelo aditivo dió una suma de cuadrados igual a 84126634.0647 y se tuvo que las hipótesis rechazadas fueron las de homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en los dos niveles usados. La inclusión de un término para la explicación del efecto de interacción, con $\hat{\theta}^2$ = 84126633.072 con 8.33

grados de libertad, se tuvo que la suma de cuadrados se redujo hasta 0.993415382 lo que propició un estimador de la desviación estandar de 0.52 y la única hipótesis rechazada fué la de normalidad en sus tres niveles.

El rechazo de esta hipótesis motivó la inclusión de un nuevo término, para el cual $\hat{\theta}^2 = 0.783317$ con 3.01 grados de libertad, el valor de $\hat{\theta}^2$ hizo que la suma de cuadrados se redujese hasta 0.2100978 y en consecuencia el valor $\hat{\sigma} = 0.564$; esta vez el estimador de la desviación estandar también creció, como en el ejemplo anterior, sin embargo rebasó el valor real del parámetro.

Las hipótesis rechazadas fueron las de normalidad en cuatro de sus cinco niveles y la de indenendencia en los dos niveles utilizados. La consideración de un nuevo término para la explicación del efecto de interacción, para el cual $\hat{\theta}^2 = 0.21009779$, dió una suma de cuadrados igual a $(4.3831)X10^{-14}$ y se tuvo el rechazo de las hipótesis de homoscedasticidad por columnas, normalidad en sus cinco niveles y simetría en los dos niveles. La aplicación del criterio de Mendel hubiera ajustado un modelo con un sólo término explicativo para el efecto de interacción, lo que en este caso particular resulta ser lo más adecuado.

Al cambiar el valor de σ a 0.30, el ajuste de un modelo aditivo dió como resultado una suma de cuadrados igual a 84120425.8584 y las hipótesis rechazadas fueron las de homoscedasticidad por

columnas, normalidad en sus cinco niveles y kurtosis en los dos niveles usados. La inclusión de un término multiplicativo para la explicación del efecto de interacción, para el cual $\hat{\theta}^2 = 84120424.4541$, trajo como consecuencia una suma de cuadrados igual a 1.404441, con el estimador de la desviación estandar igual a 0.618; las hipótesis rechazadas fueron normalidad en los tres niveles mayores de significancia, simetría al 10% y kurtosis en los dos niveles de significancia usados, es decir, los estimadores de los errores resultan poseer muchas propiedades no deseables. Cuando se incluyó un nuevo término para el efecto de interacción, para el cual $\hat{\theta}^2 = 1.20003735$, la suma de cuadrados se redujo hasta el valor 0.2044 y en consecuencia $\hat{\sigma} = 0.556$; la única hipótesis rechazada fué la de independencia a ambos niveles de significancia usados. Si se observa que el estimador de la desviación estandar es más pequeño que el valor real de dicho parámetro, se puede pensar que hemos quitado "demasiado" a los errores de modo que estos ya no son independientes. La consideración de un tercer término para explicación del efecto de interacción, con $\hat{\theta}^2 = 0.2044$, se obtuvo una suma de cuadrados igual a $(7.0726) \times 10^{-14}$, que puede deberse a los ya mencionados errores de redondeo; los errores obtenidos no cumplieron con las hipótesis de homocedasticidad por columnas, normalidad en los cinco niveles, simetría y kurtosis en los dos niveles de significa-

cancia utilizados. En este caso, a través del criterio presentado en este trabajo no es posible encontrar o suponer un modelo óptimo, en todo caso, el modelo "menos malo" sería aquel que incluyese dos términos para explicar el efecto de interacción conclusión que difiere de aquella a la que se llega por el criterio de Mandel.

El comportamiento del modelo para el cual $\tau = 1.05$ fué aún peor ya que las estimaciones de la desviación estandar fueron muy malas. El ajuste de el modelo aditivo dió como resultado una suma de cuadrados igual a 84132634.0647 resultando rechazadas las hipótesis de homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en sus dos niveles. La inclusión de un término multiplicativo, con $\hat{\theta}^2 = 84132633.0720$, redujo la suma de cuadrados hasta el valor 0.993415 y dió un estimador de la desviación estandar igual a 0.52, el cual es menor que la mitad del valor real del parámetro; las hipótesis rechazadas fueron las de normalidad en sus tres niveles mayores y simetría al 0.02. Cuando se aumentó un término el efecto de interacción, con $\hat{\theta}^2 = 0.7833176$, la suma de cuadrados se redujo hasta alcanzar el valor 0.2100978 con un estimador de la desviación estandar igual a 0.464 que sigue siendo significativamente menor que el valor real del parámetro. Las hipótesis que resultaron rechazadas fueron las de normalidad en los cuatro mayores niveles e independencia en los dos niveles de significancia utilizados. Las

propiedades distribucionales de los errores son aún peores al incluir un nuevo término, para el cual $\hat{\theta}^2 = 0.21009779$, ya que las hipótesis rechazadas fueron las de homoscedasticidad por columnas, normalidad en sus cinco niveles y simetría en sus dos niveles de significancia. Para este término no se cuenta ya con un estimador de la desviación estandar.

Debe hacerse notar en este punto que para los últimos cinco modelos y para los siguientes el efecto de interacción es un polinomio de grado mayor o igual que cinco, en los valores de los renglones y las columnas y no se le ha relacionado con los niveles de los criterios considerados en el desarrollo del experimento, lo que por supuesto es imposible en los ejemplos que se presentan adelante. Lo anterior puede ser la causa del mal comportamiento de los errores y el consecuente mal ajuste de los modelos. En el caso de experimentos reales se espera que los efectos de interacción, si los hay sean de un orden de magnitud pequeño.

El siguiente modelo considerado fué el siguiente:

$$Y_{ij} = 2(i-1) + 3j + ((i-3)(j-1))^7 + \epsilon_{ij}$$

Con los cinco valores, considerados en los modelos anteriores, para la desviación estandar de los errores ϵ_{ij}.

Los resultados obtenidos para el valor de α igual a 0.05 son sorprendentes. Para el primer modelo ajustado, es decir un modelo estrictamente aditivo, dió un valor para la suma de cuadrados igual a $(1.1328)X10^{11}$ y las hipótesis rechazadas son las de

homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en sus dos niveles. Ahora bien, a pesar del grado del polinomio en i y en j, la inclusión de el primer término para explicación del efecto de interacción trajo como consecuencia el no rechazo de ninguna de las hipótesis, es decir, se acentó un modelo con un sólo término para la explicación del efecto de interacción. En este caso la suma de cuadrados fué igual a 10.746 dando, para $\hat{\sigma}$, el valor de 1.75. La aplicación del criterio de Mandel habría dado precisamente el mismo modelo dado que la magnitud del cuadrado medio para el efecto de interacción es del orden de 10^{11}, mientras que los restantes cuadrados medios sería todos del orden de 10.

Para el caso en que $\sigma = 0.30$ el ajuste de un modelo aditivo dió una suma de cuadrados de los residuos igual a $(1.1328)X10^{11}$ con el rechazo de las hipótesis de homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en sus dos niveles. Al incluir un término para la explicación del efecto de interacción resultó, nuevamente, que ninguna de las hipótesis fué rechazada, y como consecuencia que el ajuste de un modelo con un sólo término para la interacción con θ^2 del orden de 10^{10}. La suma de cuadrados alcanzó el valor de 13.19405 y $\hat{\sigma}$ =1.89, valor que no es tan exageradamente mayor como el del ejemplo anterior. Así mismo el modelo que se hubiéra ajustado haciendo

uso del criterio del Dr. Mandel, sería precisamente el mismo. En cambio, cuando $\sigma = 0.55$, la decisión acerca de cuál modelo debe ajustarse a los datos no resultó obvia. El ajuste del modelo aditivo dió una suma de cuadrados de $(1.13331)X10^{11}$ resultando rechazadas las hipótesis de homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en los dos niveles considerados. La inclusión del primer término multiplicativo para la explicación del efecto de interacción, para el cual $\hat{\theta}^2$ fué de un orden de magnitud de 10^{11}, dió una suma de cuadrados del error igual a 2.145226, valor que junto con los 3.67 grados de libertad restantes dió un estimador de la desviación estandar de tan solo 0.764; la hipótesis rechazada fué la de normalidad en cuatro de sus niveles únicamente. Al considerar un nuevo término para la interacción, con $\hat{\theta}^2 = 2.13675787$, la suma de cuadrados del error fue igual a 0.008469 lo que dió $\hat{\sigma} = .1132$, valor que es ya cuatro veces menor que el valor real del parámetro; las hipótesis que resultaron rechazadas fueron las de normalidad al 20% de significancia e independencia a los dos niveles considerados. La inclusión de un tercer término multiplicativo para el efecto de interacción, para el cual $\hat{\theta}^2 = 0.0084689$, dió una suma de cuadrados de $2.8896\ X10^{-11}$. En este caso ya no es posible obtener un estimador de la desviación estandar del error. Se rechazaron las hipótesis de normalidad y simetría, ambas a

todos los niveles usados. Obviamente el criterio de Mandel habría considerado el ajuste de un modelo con un sólo término para el efecto de interacción como el adecuado.

En contraste con lo anterior, el caso en que $\sigma = 0.80$ tuvo un comportamiento mejor ya que aunque el ajuste de un modelo aditivo dió una suma de cuadrados de $(1.1333)X10^{11}$ y las hipótesis que se rechazaron fueron las de homoscedasticidad por columnas, normalidad y kurtosis, ambas en todos sus niveles, el ajuste de un modelo que incluye un término para la explicación del efecto de interacción, con $\hat{\theta}^2$ del orden de 10^4, dió como resultado el que ninguna de las hipótesis fuese rechazada, una suma de cuadrados de tan sólo 2.544475 y un estimador de la desviación estandar de 0.832.

Para los datos con $\sigma = 1.05$ ocurre que la hipótesis de normalidad se rechaza siempre y a todos los niveles para los tres términos aumentados al modelo aditivo. Los estimadores de la desviación estandar fueron 0.844 y 0.935 para los modelos que incluían 1 y 2 términos respectivamente. Otras hipótesis rechazadas fueron las de homoscedasticidad por columnas y kurtosis en sus dos niveles para el modelo aditivo; la de independencia en los dos niveles para el modelo con un término de la interacción; homoscedasticidad por renglones, simetría y kurtosis para los dos niveles en ambas pruebas, cuando el modelo incluía dos terminos

para explicación del efecto de interacción; finalmente, además de la hipótesis de normalidad, para un modelo con tres términos multiplicativos se rechazó la hipótesis de simetría al 10% y al 2%. En este caso, de acuerdo a los resultados obtenidos, tampoco es posible hablar de un modelo mejor para representar los datos.

Las últimas cinco matrices de datos simulados correspondieron al siguiente modelo:

$$Y_{ij} = 2\,(i-1) + 3j + \left((1-3)\,(j-1) \right)^{9} + \varepsilon_{ij}$$

En cada caso, se hizo variar a σ como en todos los datos simulados anteriormente.

Brevemente los resultados obtenidos para cada grupo de datos son:

Para $\sigma = 0.05$ el ajuste de un modelo aditivo, para el cual la suma de cuadrados fué igual a 1.497964×10^{14}, dió como resultado que se rechazaran las hipótesis de homocedasticidad por columnas, normalidad y simetría, todas en todos los niveles usados.

Para el primer término de la interacción se tuvo una suma de cuadrados de 0.507979325 y un estimador de σ igual a 0.06740; se rechazaron las hipótesis de normalidad a los cinco niveles de significancia, simetría y kurtosis ambas al 10%. Para el segundo término multiplicativo la reducción en suma de cuadrados fué de 0.232297329 ($=\hat{\theta}^2$) y el valor de $\hat{\sigma}$ fué 0.0502; se rechazaron

las hipótesis de normalidad en los niveles 20% y 15%, independencia en los dos niveles y kurtosis al 10% de significancia, es decir, la hipótesis cuyo rechazo es más grave en este caso es la independencia para el tercer término la suma de cuadrados fué de 0.000000038; no fué posible ya dar un estimador de la desviación estandar y se rechazaron las hipótesis de normalidad y simetría en todos sus niveles. Para σ=0.30 cuando se ajustó un modelo lineal se obtuvo una suma de cuadrados de 1.497964×10^{14} y se rechazaron las hipótesis de homoscedasticidad por columnas, normalidad y kurtosis en todos los niveles. Con el primer término de la interacción la suma de cuadrados fué de 0.50797932 y $\hat{\sigma}$ = 0.37203; la única hipótesis que se rechazó fué la de normalidad a todos los níveles. Para el segundo término de la interacción se tuvo una suma de cuadrados de 0.12566198 y la estimación para σ fué 0.4363; se rechazaron las hipótesis de normalidad al 20%, 15% y 10% e independencia al 10% y al 2%. Para el mismo modelo, pero con σ =0.55, al ajustar un modelo estrictamente aditivo se obtuvo una suma de cuadrados de los residuos igual a $(1.49496487) \times 10^{14}$ y las hipótesis rechazadas con los mismos residuos fueron las de homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en los dos niveles usados. La inclusión de un término multiplicativo para la explicación del efecto de interacción, para el cual $\hat{\theta}^{\circ}$ resultó

un valor del orden de 10^{13}, dió como resultado el que la suma de cuadrados de los nuevos residuos fuese igual a 18.639114 lo que daba como resultado un valor para el estimador de la desviación estandar de 2.2536; como hipótesis rechazadas resultaron ser las de homoscedasticidad por renglones, normalidad al 20% y al 15%, simetría y kurtosis, ambas a 10% y 2%. Con la inclusión de un nuevo término para la explicación del efecto de interacción la suma de cuadrados se redujo a 0.450466909 lo que dió como resultado que $\hat{\sigma} = 0.8261$; se rechazaron las hipótesis de normalidad a los tres niveles mayores y la de independencia a los niveles 10% y 2%. Para el tercer término explicativo, la suma de cuadrados se redujo hasta el valor 0.00000004?; para este término no se tiene ya un estimador de la desviación estandar; las hipótesis rechazadas fueron las de homoscedasticidad por columnas, normalidad y kurtosis ambas a todos los niveles usados.

Para $\sigma = 0.80$ el ajuste de un modelo estrictamente aditivo dió unos residuos cuya suma de cuadrados fué igual a $(1.497964456) \times 10^{14}$ y con los cuales se rechazaron las hipótesis de homoscedasticidad por columnas, normalidad en sus cinco niveles y kurtosis en los dos niveles usados. En este caso la inclusión de un término para la interacción redujo la suma de cuadrados hasta

10.123865964 con un estimador de σ igual a 1.6609; en este caso ninguna de las hipótesis fué rechazada con lo que se concluyó que un modelo que incluya un término multiplicativo para explicación del efecto de interacción explica satisfactoriamente el comportamiento de los datos.

En contraste con lo anterior para el caso en que $\sigma = 1.05$ los residuos obtenidos después del ajuste de un modelo aditivo tuvieron una suma de cuadrados igual a $(1.49796476)X10^{14}$ y con ellos se rechazaron las hipótesis de homoscedasticidad por columnas, normalidad en los cinco niveles utilizados y kurtosis al 10% y 2%. Con la inclusión de un término para el efecto de interacción la suma de cuadrados de los residuos fué igual a 8.6271067 y el estimador de la desviación estandar fué igual a 1.533; se rechazaron las hipótesis de normalidad en cuatro de sus niveles y kurtosis al 10%. Para el segundo término, para explicación del efecto de interacción, para el cual $\hat{\theta}^2$=8.31709725, el valor de la suma de cuadrados descendió hasta 0.310009369 con lo que $\hat{\sigma}$ = 0.6853; resultó en rechazo la prueba de independencia usada a los niveles de 10% y 2%. Con la inclusión del tercer término para el efecto de interacción, con $\hat{\theta}^2$=0.31000[illegible]; la suma de cuadrados de los residuos resultó igual a 0.00000003, no se tuvo un estimador para la desviación estandar y se rechazaron las hipótesis de homoscedasticidad por columnas, normalidad,

kurtosis y simetría, estas últimas tres lo fueron a todos los niveles a los que se probaron.

Los resultados obtenidos en las últimas matrices de datos simulados pueden deberse a la magnitud de los efectos de interacción, los cuales representaban una gran influencia en los valores originales obtenidos, lo que opacaba los efectos aleatorios (en general las desviaciones estandar resultaban pequeñas) lo que trajo como consecuencia que los criterios usados para las pruebas de las hipótesis resultaran "viciados".

En general es posible observar que en experimentos reales los efectos de interacción no alcanzan valores tan exagerados como los incluidos en las matrices para las cuales $k \geq 5$. El asignar valores tan grandes a los efectos de interacción se hizo con el fin de calibrar las bondades del método para casos extremos.

CAPITULO 7

7.- RESUMEN Y CONCLUSIONES

Se ha propuesto como modelo mas general y más útil para representar algunos fenómenos que son observados en el mundo real, ya sea mediante experimientación u observación directa de la naturaleza al siguiente

$$Y_i = \mu + \varepsilon_i$$

con $$\varepsilon_i \sim N(0, \sigma^2)$$

La suposición de normalidad de los errores o fluctuaciones aleatorias se justifica en algunos casos por la no consideración explícita de aquellos factores o condiciones cuya influencia es nula o pequeña y que individualmente, son causa de una variabilidad pequeña en los resultados observados. El equivalente teórico a esta justificación está dado por el teorema central de límite que plantea más o menos el hecho de que el resultado de sumar variables aleatorias independientes con varianzas finitas, habiendo estandarizado adecuadamente estas variables, y con un número grande de sumandos, es una variable que sigue una distribución normal con media cero y varianza 1.

Como parte importante de estos modelos, se cuentan los modelos lineales de la forma:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \ldots + \alpha_p X_{pi} + \varepsilon_i$$

con $$\varepsilon_i \sim N(0, \sigma^2)$$

Casos particulares de estos modelos, en los cuales se ha concen-

trado toda la atención en este trabajo, son los modelos de diseños de experimentos para los cuales X_{ji} toma únicamente los valores cero o uno. En este tipo de modelos se considera que los efectos de los criterios con los cuales se han clasificado las observaciones pueden descomponerse en una parte aditiva, con cada uno de los sumandos compuestos por funciones de una sola de las variables o criterios, y otra parte posiblemente aditiva pero donde los sumandos son funciones de dos o más de las variables. Cuando se consideran grupos de datos con doble criterio de clasificación y una sola observación por cada pareja de niveles de los criterios, la estimación por mínimos cuadrados de los parámetros del modelo trae como consecuencia un ajuste perfecto de los datos por el modelo; razón por la cual se hace imposible el realizar pruebas de hipótesis o el construir intervalos de confianza para los parámetros que forman parte del modelo.

Por lo anterior se han propuesto distintos métodos de estimación de los efectos de interacción, en uno de los cuales, propuesto por el Dr. John Mandel, se base el desarrollo del presente trabajo. Este método propone que los efectos de interacción pueden descomponerse como una suma de términos de la forma $\theta u_i v_j$ donde los valores de θ son estimados por los eigenvalores de la matriz $G = DD'$, siendo la matriz D la matriz de residuos d_{ij}; los valores de u_i son estimados por las componentes de los vectores propios

de la misma matriz asociados con los estimadores del parámetro θ; asimismo los parámetros v_j son estimados por los eigenvectores de la matriz D'D. Es decir, el modelo que se propone para ser ajustado a las observaciones es el siguiente:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{r=1}^{k} \theta_r u_{ri} v_{rj} + \varepsilon_{ij}$$

El número de términos multiplicativos que forman la descomposición del efecto de interacción ha de ser determinado mediante criterios adecuados. El criterio propuesto por el Dr. Mandel hace uso de un enfoque parecido al análisis de varianza comparando magnitudes de los cuadrados medios asociados a los términos del modelo. Por otra parte se espera que una vez que los efectos de interacción sean incluidos en el modelo, aquellos factores no controlados producirán variaciones o errores con distribución de probabilidad normal con media cero y desviación estandar común, además de ser independientes.

De acuerdo a este segundo punto de vista se buscó de una manera recursiva eliminar de los residuos tantos términos de la forma $\theta u_i v_j$ como fuera necesario para obtener errores independientes y normalmente distribuidos con media cero y varianza común. Con este punto de vista se ajustaron los modelos que produjeron los errores con las propiedades que se mencionan arriba.

A partir de los resultados obtenidos en los ejemplos de la sección anterior se puede afirmar que el método de descomposición de los

efectos de interacción como una suma de términos, todos ellos múltiplos del producto de funciones de una sola variable, produce errores suficientemente apegados a las hipótesis hechas acerca de la distribución de los errores reales asociados a cada una de las observaciones.

Esto se logra, en ocasiones, haciendo uso de un número distinto de terminos a los que se incluirían haciendo uso del enfoque de análisis de la varianza propuesto por el Dr. John Mandel. Lo anterior se observa principalmente en el conjunto de ejemplos reales de la sección anterior, en los cuales las decisiones hechas acerca del número de términos multiplicativos mediante el enfoque de análisis de la varianza produjeron errores no normales (o menos normales), y/o correlacionados y/o heteroscedásticos y/o sobreestimaciones de la varianza real de los errores; algunos de estas desventajas eran eliminadas, parcial o totalmente, con la inclusión de uno o dos términos más aún cuando los cuadrados medios relacionados con estos términos no fueran "significativos" bajo el enfoque de análisis de la varianza; es decir, ir un poco mas adelante en la estimación de los efectos de interacción produjo, en ocasiones, notables mejorías en cuanto a los errores obtenidos y a sus propiedades.

Sin embargo para los datos construidos artificialmente, en los casos en que el efecto de interacción era pequeño en magnitud, con

respecto a los valores de los términos que representaban los efectos principales, dichos efectos no fueron detectados ya que los errores obtenidos al ajustar un modelo estrictamente aditivo cumplían satisfactoriamente con las hipótesis planteadas; este hecho propiciaba en la mayoría de los casos que la varianza fuese sobreestimada, en ocasiones era demasiado grande esta diferencia como para ser pasada por alto.

En casos como este el uso previo de una prueba de no aditividad, tal vez la de Tukey, podría proporcionar una mayor confianza en la decisión tomada acerca del modelo que se debe ajustar a los datos, desgraciadamente esta prueba no toma en cuenta las propiedades distribucionales de los errores obtenidos al rechazar la hipótesis de aditividad. Cabe aquí recordar que algo parecido ocurrió con conjuntos de datos reales (ejemplo 3), en los que se sospechaba que ocurría algo parecido, en vista de la magnitud de las sumas de cuadrados por lo que este caso no debe ser pasado por alto.

Por otro lado, también en los datos construidos artificialmente, es notable la poca eficiencia del método para producir errores satisfactorios en los casos en que el efecto de interacción era tan grande como un polinomio de grado mayor o igual a cinco. Esto es sin embargo una desventaja de tipo teórico unicamente, ya que como se observó en los ejemplos reales, estos casos difícilmente ocurren en la práctica.

Para este tipo de datos sería deseable el contar con información acerca de la distribución de los estimadores de los parámetros incluidos en los términos multiplicativos en los cuales se descompone el efecto de interacción ya que en ese caso sería posible el realizar pruebas estadísticas para niveles de significancia prefijados. Sin embargo, aún cuando se contase con esta información, los errores producidos después de realizar la prueba de significancia adecuada serían poco satisfactorios en cuanto a las propiedades distribucionales de los mismos. La magnitud de los valores originales y las sumas de cuadrados de los residuos, que decrecen rapidamente, pueden servir de indicios de que el efecto de interacción es muy grande.

En todos los casos anteriores es posible observar que los errores obtenidos parecen alcanzar un nivel óptimo para un cierto número de términos del efecto de interacción, a partir del cual los errores que se obtienen aumentando un término empeoran en cuanto a las condiciones que satisfacen. Más aún hay ocasiones en que los errores obtenidos no satisfacen completamente las hipótesis establecidas para ningún número de términos en el efecto de interacción, aunque para un cierto número de ellos las propiedades satisfechas parecen ser mejores que las que se obtuvieron con un término menos y que las que se obtendrían aumentando otro término multiplicativo; si en este caso se quiere pensar que la función "hipótesis no rechazadas" o "propiedades deseables", fun-

ción del número de términos incluidos en el modelo o de los valores de los estimadores de los parámetros, fuese continua, una "interpolación" entre los valores de los estimadores de los parámetros y "pesada" por aquellas hipótesis no satisfechas podría producir un modelo cuyo ajuste diese errores "óptimos" en el sentido que todas las hipótesis fuesen satisfechas.

Es decir, el modelo podría aumentarse con un pequeño conjunto de parámetros (a lo más tres), cuyos valores se determinarían tomando en cuenta las hipótesis rechazadas y los niveles para los cuales tuvo lugar ese rechazo. El modelo aumentado tomaría una forma parecida a la siguiente:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{r=1}^{k} \lambda_r \theta_r u_{ri} v_{rj} + \varepsilon_{ij}$$

Donde los perámetros λ_r se han aumentado al modelo; los valores posibles para dichos parámetros serían los comprendidos en el intervalo cerrado (0,1).

Para aquellos casos en los que el ajuste de un modelo aditivo diese como resultado el no rechazo de ninguna de las hipótesis pero se sospechase que la varianza ha sido sobreestimada o la hipótesis de aditividad hubiese sido rechazada podría aumentarse un término de la forma $\lambda_1 \theta_1 u_{1i} v_{1j}$, determinando el tamaño de λ_1 de manera tal que las propiedades de los errores siguiesen siendo óptimas. De este modo se obtendría una reducción en suma

de cuadrados y en consecuencia en el valor del estimador de la varianza, logrando además un mejor ajuste del modelo a los datos. Para el caso en que el modelo "óptimo" incluye t términos, aún cuando alguna hipótesis haya sido rechazadas, los valores de λ_k serían uno para aquellos términos de con $k \leq t-2$, fijándose λ_{t-1}, λ_t, λ_{t+1} de acuerdo a las hipótesis rechazadas y pudiendo alguno de estos parámetros alcanzar el valor cero.

Finalmente, es recomendable como tema de futura investigación el trabajo en la distribución exacta de los parámetros incluidos en el modelo con el objeto de obtener un criterio preciso que nos lleve a determinar aquellos cuadrados medios que resulten realmente significativos. Acerca de esto (la distribución de los eigenvalores de la matriz de varianza y covarianzas), Mandel ha publicado ya un interesante artículo que aparece con el número (15) en la lista de referencias.

APENDICE I

U-N-I-V-E-R-S-I-D-A-D N-A-C-I-O-N-A-L A-U-T-O-N-O-M-A D-E M-E-X-I-C-O

C E N T R O D E S E R V I C I O S D E C O M P U T O

COMPILADOR ALGOL, NIVEL 2.5.009, BURROUGHS B6700

MARTES 22 DE OCTUBRE DE 1974 06:10 PM.

PROG/PRUEBA

```
BEGIN                                                                    000:0000:0
                                                                  B.0000 IS SEGMENT 00003
FILE ENT(KIND=READER), SAL(KIND=PRINTER);                             1  003:0000:0
                                                                      DATA IS 0005 LONG
                                                                      DATA IS 0005 LONG
ARRAY  ALFA[0: 1,1:5],PSIME,SIME[1:2],KURTO,VON[1:4],COCH[1:2];          003:0000:0
INTEGER M,N,     PTE;ARRAY JITA[1:3],PNOR[1:2];                          003:0007:2
REAL PSC,SIGMA,MED,SCDA,X;                                               003:0007:2
LABEL  LEE,FIN;                                                          003:0007:2
     DEFINE ANORMA(A,S)= FOR I:=1 STEP 1 UNTIL S DO                      003:0007:2
       SC:= -A[CONT,I]**2;                                               003:0007:2
       SC:=SC**(1/2);                                                    003:0007:2
      FOR I:=1 STEP 1 UNTIL S DO                                         003:0007:2
       A[CONT,I]:=-/SC #;                                                003:0007:2
PROCEDURE  IMPRIMEMATRIZ(SAL,A,M,N);                                     003:0007:2
   VALUE  M,N;                                                           003:0007:2
   FILE  SAL;                                                            003:0007:2
   ARRAY  A[*,*];                                                        003:0007:2
   INTEGER  N,M;                                                         003:0007:2
  BEGIN                                                                  003:0007:2
     INTEGER  I,J;                                                       003:0007:2
                                                          IMPRIMEMATRIZ IS SEGMENT 00004
     FOR I:=1 STEP 1 UNTIL M DO                             2            004:0000:0
     BEGIN                                                               004:0004:3
     WRITE (SAL,<X10,"*",/("----------------*")>,MIN(N,10));             004:0004:3
   WRITE(SAL,<X10,"*",6(R15.4,X1,"*")>,FOR J:=1 STEP 1 UNTIL N DO        004:0014:1
     A[I,J]);                                                            004:0021:0
     WRITE (SAL,<X10,"*",/("----------------*")>,MIN(N,10));             004:0029:1
     END;  END;                                                          004:0039:1
                                                          IMPRIMEMATRIZ(004) IS 0041 LONG
BOOLEAN PROCEDURE  LEEMATRIZ(FE,FS,A,M,N,IMPRIME);                       2  003:0007:2
   VALUE  M,N,IMPRIME;                                                   003:0007:2
   FILE  FE,FS;                                                          003:0007:2
   ARRAY  A[*,*];                                                        003:0007:2
   INTEGER  M,N;                                                         003:0007:2
   BOOLEAN  IMPRIME;                                                     003:0007:2
     BEGIN                                                               003:0007:2
     INTEGER  I,J;                                                       003:0007:2
                                                          LEEMATRIZ IS SEGMENT 00007
     LABEL  FINDATOS,FORMATO,FIN;                           2            007:0000:0
     FOR I:=1 STEP 1 UNTIL M DO                                          007:0000:0
```

```
                GO TO FIN;                                                                  007:001E:0
              FINDATOS:                                                                     007:001E:3
                WRITE(FS,<"ERROR, SE TERMINARON LAS TARJETAS EN LA HILERA ",I3>,           007:001E:3
                 I); LEEMATRIZ:=TRUE;  GO TO FIN;                                          007:0021:2
              FORMATO:                                                                      007:002D:5
                WRITE(FS,<"ERROR, FORMATO INCOMPATIBLE EN LA HILERA ",I3>,I);             007:002D:5
                LEEMATRIZ:=TRUE;                                                          007:003C:1
              FIN:                                                                        007:003C:5
                IF IMPRIME THEN IMPRIMEMATRIZ(FS,A,M,N);                                  007:003C:5
            END;                                                                          007:0041:0
                                                                          LEEMATRIZ(007) IS 0049 LONG
            PROCEDURE ESTIMAB(M,N,Y,B);                                        2          003:0007:2
            VALUE M,N;  INTEGER M,N; ARRAY Y[*,*],B[*];                                   003:0007:2
              BEGIN                                                                       003:0007:2
                INTEGER J,K;                                                              003:0007:2
                                                                            ESTIMAB IS SEGMENT 00008
                REAL YPP;                                                          2      008:0000:0
                WRITE(SAL,</"ESTIMA B:"/>);                                               008:0000:0
                FOR J := 1 STEP 1 UNTIL M DO                                              008:0005:3
                FOR K:=1 STEP 1 UNTIL N DO BEGIN                                          008:000A:0
                YPP:=* + Y[J,K];                                           3              008:000E:3
                B[J+1] := * + Y[J,K]/N;                                                   008:0011:3
                B[M+K+1] := * + Y[J,K]/M;                                                 008:0016:2
              END;                                                                        008:001B:4
                B[1] := YPP/(M*N);                                         3              008:001C:4
                FOR J := 1 STEP 1 UNTIL M DO                                              008:001F:2
                B[J+1]:= * - B[1];                                                        008:0023:5
                FOR K := 1 STEP 1 UNTIL N DO                                              008:0027:4
                B[M+K+1] := * - B[1];                                                     008:002C:1
               WRITE(SAL,<X10,"LA MEDIA GENERAL ES: ",R9,4,/X10,"Y LOS EFECTOS POR        008:0030:3
              "RENGLON Y COLUMNA ESTAN DADOS POR LOS SIGUIENTES VECTORES:"/>,B[1]          008:0032:2
                );                                                                        008:003A:1
              WRITE(SAL,<X10,*(R14,4,"*")>,M,FOR J:=2 STEP 1 UNTIL M+1 DO B[J]);          008:003F:1
              WRITE(SAL,<X10,*(R14,4,"*")>,N,FOR J:= M+2 STEP 1 UNTIL M+N+1 DO            008:0055:4
                B[J]);                                                                    008:0065:2
              END;                                                                        008:006C:4
                                                                          ESTIMAB(008) IS 0074 LONG
            PROCEDURE RESIDUO1(M,N,Y,D,B,FS);                                  2          003:0007:2
            VALUE M,N;      INTEGER M,N;ARRAY Y,D[*,*],B[*]; FILE FS;                     003:0007:2
              BEGIN                                                                       003:0007:2
                INTEGER I,J;                                                              003:0007:2
                                                                            RESIDUO1 IS SEGMENT 00009
                WRITE(FS,</"RESIDUO 1:",/>);                                       2      009:0000:0
                FOR I:=1 STEP 1 UNTIL M DO                                                009:0005:4
                FOR J:=1 STEP  1 UNTIL N DO                                               009:000A:1
                D[I,J]:=Y[I,J]-B[1]-B[I+1]-B[J+M+1];                                      009:000E:4
                WRITE(FS,</X10,"LA MATRIZ DE RESIDUOSPRIMERA ES: "/>);                    009:0019:1
            END RESIDUOS;                                                                 009:001E:5
                                                                          RESIDUO1(009) IS 0020 LONG
            PROCEDURE ORDENA(A,M,N,AOR,PTE);                                   2          003:0007:2
            VALUE M,N;   INTEGER M,N; ARRAY A,AOR[*,*];INTEGER PTE;                       003:0007:2
              BEGIN                                                                       003:0007:2
                INTEGER I,J;                                                              003:0007:2
                                                                             ORDENA IS SEGMENT 0000A
                LABEL  L;                                                          2      00A:0000:0
                ARRAY B,C[0:M*N];                                                         00A:0000:0
                     B[0]:=10**5;                                                         00A:0004:0
                FOR I:=1 STEP 1 UNTIL M DO                                                00A:0006:1
```

```
                C[(I-1)*N+J]:= JJ END;                                                    00A:0013:3
            FOR I:=1 STEP 1 UNTIL M*N-1 DO BEGIN                              3           00A:0017:1
            IF B[I] GTR  B[I+1] THEN GO TO L ELSE                             3           00A:001C:3
             BEGIN      B[I]:= READLOCK(B[I],B[I+1]);                                    00A:001E:5
                        C[I]:=READLOCK(C[I],C[I+1]);                          4           00A:0022:0
             FOR J:=I STEP -1  UNTIL 1 DO                                               00A:0025:1
             IF B[J] LSS B[J-1] THEN GO TO L ELSE BEGIN                                 00A:0028:5
              B[J-1]:= READLOCK(B[J-1],B[J]);                                 5         00A:002B:1
              C[J-1]:= READLOCK(C[J-1],C[J]);  END; END;                              00A:002E:4
            L:  END;                                                          4         00A:0032:4
            FOR I:=1 STEP  1 UNTIL M DO                                       3         00A:0033:1
            FOR J:=1 STEP 1 UNTIL N DO                                                  00A:0037:4
                ADR[I,J]:=B[(I-1)*N+J];                                                 00A:003C:1
                PTE:=C[1];                                                              00A:0041:4
        END ORDENA;                                                                     00A:0043:0
                                                                                ORDENA(00A) IS 0047 LONG
        PROCEDURE INTEGRA(P,A,B,INT,SCD);                                     2         003:0007:2
        VALUE P,A,B; INTEGER P; REAL A,B,INT,SCD;                                       003:0007:2
        BEGIN                                                                           003:0007:2
            REAL X,CIF,INT1; INTEGER L;REAL  VAR;                                       003:0007:2
                                                                              INTEGRA IS SEGMENT 0000B
            LABEL ITERA;                                                      2         00B:0000:0
            DEFINE F(X,S)=(2*3,1416*S)**(-1/2)*EXP((-X**2)/(2*S))#;                     00B:0000:0
            PROCEDURE INTEGRAL(P,  A,X,B,INT);                                          00B:0000:0
            VALUE   A,B; REAL   A,B,INT,X; INTEGER P;                                   00B:0000:0
             BEGIN                                                                      00B:0000:0
              INTEGER I,J; ARRAY UES[0:P+1];                                            00B:0000:0
                                                                              INTEGRAL IS SEGMENT 0000C
              I:=-1; INT:=0;                                                  3         00C:0003:2
              FOR X:=A STEP (B-A)/P UNTIL  B DO  BEGIN                                  00C:0005:0
              UES[I:=I+1]:=F(X,VAR);                                          4         00C:000E:4
              END;                                                                      00C:0018:3
              FOR J:=0 STEP 1 UNTIL P DO BEGIN                                4         00C:0019:0
              IF J EQL 0 OR J EQL P THEN                                      4         00C:001D:3
              INT:=INT + UES[J]                                                         00C:001F:1
              ELSE IF J MOD(2) EQL 0 THEN                                               00C:0020:5
                 INT:= * + 2* UES[J]                                                    00C:0023:1
              ELSE INT:= * + 4*UES[J]; END;                                             00C:0025:1
              INT:=INT*(B-A)/(P*3);                                           4         00C:0029:2
            END;                                                                        00C:002C:2
                                                                              INTEGRAL(00C) IS 0031 LONG
                VAR:=SCD/(M*N-1);                                             3         00B:0000:0
                CIF:=10**(-5);                                                          00B:0002:1
        ITERA: INT1:=INT;                                                               00B:0004:3
            L:=L+1; P:=2*P; IF A EQL B THEN  INT:=0 ELSE                                00B:0005:2
            INTEGRAL(P,A,X,B,INT);                                                      00B:000A:2
            IF ABS(INT-INT1) GTR CIF THEN GO TO ITERA;                                  00B:000D:4
        END;                                                                            00B:0010:2
                                                                              INTEGRA(00B) IS 0014 LONG
        BOOLEAN PROCEDURE KOLMOGSMIR (M,N,D,SCD,FS,ARR,X,ORD);                2         003:0007:2
        VALUE M,N,SCD;   INTEGER M,N; ARRAY D[*,*],ARR[*,*];REAL   SCD;FILE FS;         003:0007:2
            REAL  X; ARRAY ORD[*,*];                                                    003:0007:2
        BEGIN                                                                           003:0007:2
            INTEGER  I,J,K,PTE;                                                         003:0007:2
                                                                              KOLMOGSMIR IS SEGMENT 0000F
            ARRAY DIF,ODIF[1:M,1:N];                                          2         00F:0000:0
            REAL Q,QP,COTA,INT;                                                         00F:0003:5
            DEFINE  COMPARA(X)=                                                         00F:0003:5
```

```
                "CATIVA AL ",F6.3,"%."//X25,"EL VALOR DE LA ESTADISTICA ES"                    00F:0003:5
                ,X5,R10.8/>, ARR[0,I],DDIF[1,1])                                               00F:0003:5
            ELSE  BEGIN    Q:=Q+1;                                                             00F:0003:5
            WRITE(FS,</"KOLMOGOROV-SMIRNOV:"/X20,"LA NORMALIDAD SE ACEPTA CON",                00F:0003:5
            R6.3,"% SIENDO EL VALOR DE LA ESTADISTICA: ",X5,R10.8/>, ARR[0,I],                 00F:0003:5
            DDIF[1,1]);  END;                                                                  00F:0003:5
            INTERPOLA(A,B,C,D,E,F): F:=C+((D-C)/(B-A))*(D-A);                                  00F:0003:5
            ASINT(X): ARR[1,X]/((M*N)**(1/2));                                                 00F:0003:5
            ORDENA(D,M,N,ORD,PTE);                                                             00F:0003:5
             K:=0;                                                                             00F:0008:2
            FOR I:=M STEP -1 UNTIL 1 DO                                                        00F:0009:0
            FOR J:=N STEP -1 UNTIL 1 DO BEGIN                                                  00F:000C:5
            IF J EQL N AND  I EQL M THEN                                          3            00F:0010:5
            INTEGRA( 6,ORD[M,N]**4*SCD,ORD[M,N],INT,SCD) ELSE IF J EQL N THEN                  00F:0012:A
            IF  I EQL M THEN ELSE INTEGRA( 6,ORD[I+1,1],ORD[I,J],INT,SCD)                      00F:001C:3
            ELSE INTEGRA(6,ORD[I,J+1],ORD[I,J],INT,SCD);                                       00F:0024:5
            QP:=QP+INT;                                                                        00F:002D:2
             K:=K+1;                                                                           00F:002E:4
            DIF[I,J]:=MAX(ABS(K/(M*N)-QP),ABS((K-1)/(M*N)-QP));                                00F:0030:0
            END;                                                                               00F:0038:3
            ORDENA(DIF,M,N,DDIF,PTE);                                             3            00F:0039:3
            WRITE(FS,</X10,"LA MATRIZ DE DIFERENCIAS ORDENADAS ES:"/>);                        00F:003E:0
            IMPRIMEMATRIZ(FS,DDIF,M,N);                                                        00F:0043:A
            FOR I:=1 STEP 1 UNTIL 5 DO BEGIN                                                   00F:0047:0
            COTA:= ASINT(I);                                                      3            00F:0047:A
            COMPARA(COTA);END;                                                                 00F:004C:4
        IF Q EQL 5  THEN KOLMOGSMIR:=TRUE ELSE KOLMOGSMIR:=FALSE;                 3           00F:0077:0
        END;                                                                                   00F:007A:1
                                                                                  KOLMOGSMIR(00F) IS 0084 LONG
        PROCEDURE SUMCUAD(M,N,D,CONT,FCDA,SAL);                                    2           003:0007:2
        VALUE M,N,CONT;   INTEGER M,N,CONT; ARRAY D[*,*];REAL FCDA;FILE SAL;               003:0007:2
           BEGIN                                                                           003:0007:2
            INTEGER I,J;                                                                   003:0007:2
                                                                                  SUMCUAD IS SEGMENT 0010
            REAL SCD,DIF;                                                          2       010:0000:0
            FOR I:=1 STEP 1 UNTIL N DO                                                     010:0000:0
            FOR J:=1 STEP 1 UNTIL N DO                                                     010:0004:3
                SCD:=* + D[I,J]**2;                                                        010:0009:0
            WRITE(SAL,</"SUMA DE CUADRADOS:"/X15,"SCD= ",R19.9/>,SCD);                     010:000D:2
            IF CONT GTR 0 THEN   BEGIN                                                     010:0018:4
            SCD:=*/PSC;  FCDA:=*/PSC;                                              3       010:001C:5
            IF ABS(SCD-FCDA) LEQ 0.1 THEN BEGIN                                            010:001F:4
            WRITE (SAL,<X15,"LA SUMA DE CUADRADOS DE LOS RESIDUOS CONVERGEPARA"            4       010:0023:4
            I3," )"/X15,"Y CONVERGE ALREDEDOR DEL VALOR: ",R19.9>,CONT,SCD);               010:0025:4
             END                                                                           010:0033:4
            ELSE BEGIN                                                                     4       010:0033:4
            DIF:= SCD-FCDA;                                                                4       010:0034:1
            WRITE(SAL,<X15,"LA SUMA DE LOS CUADRADOS DE LOS RESIDUOS ACTUALES"                     010:0035:3
            " DIFIERE DE LA DE LOS ANTERIORES EN :",R10.5," )",/X15,"SIENDO LA"                    010:0037:3
            " ACTUAL SUMA DE CUADRADOS IGUAL A: ",R19.9," )">,DIF,SCD);END;END;                    010:0037:3
            IF CONT GTR 0 THEN FCDA:=SCD+PSC ELSE                                  3       010:0045:4
            FCDA:=SCD;                                                                     010:0048:2
        END;                                                                               010:0049:5
                                                                                  SUMCUAD(010) IS 0051 LONG
        BOOLEAN PROCEDURE HOMOSCED(D,M,N,SAL,COCH);                                2       003:0007:2
        VALUE M,N;INTEGER M,N;ARRAY D[*,*],COCH[*];   FILE SAL;                            003:0007:2
        BEGIN                                                                              003:0007:2
           REAL SUM,DIV;                                                                   003:0007:2
```

```
ARRAY RENG,MEDR[1:M],COL,MEDC[1:N]);                                                    011:0000:0
DEFINE MAXIMO(A,S)= FOR I:=1 STEP 1 UNTIL S-1 DO                                        011:0006:2
  IF A[I] GTR  A[I+1] THEN A[I]:=READLOCK(A[I],A[I+1]) #,                                011:0006:2
  SUMA(A,S,K)  =  SUM:=0;                                                               011:0006:2
  FOR I:=1 STEP 1 UNTIL S DO SUM:=SUM+A[I];                                             011:0006:2
       DIV:= A[S]/SUM;                                                                  011:0006:2
       IF DIV LEQ COCH[K]  THEN BEGIN                                                   011:0006:2
  WRITE(SAL,<X10,"LA HOMOSCEDASTICIDAD SE ACEPTA AL 1%, SIENDO",                        011:0006:2
  "LA COTA IGUAL A ",R7.4,", Y EL VALOR DE LA ESTADISTICA IGUAL A",                     011:0006:2
  X1,R7.4/>,COCH[K],DIV); Q:=Q+1;                                                       011:0006:2
  END ELSE BEGIN                                                                        011:0006:2
  WRITE(SAL,<X10,"LA HOMOSCEDASTICIDAD SE RECHAZA AL 1%,SIENDO LA ",                    011:0006:2
  "COTA IGUAL A"R10.6" EL VALOR DE LA ESTADISTICA ES"R11.6/>>,                          011:0006:2
  COCH[K],DIV);                                                                         011:0006:2
  Q:=0; END #;                                                                          011:0006:2
  FOR I:=1 STEP 1 UNTIL M DO                                                            011:0006:2
  FOR J:=1 STEP 1 UNTIL N DO BEGIN                                                      011:000A:5
  MEDR[I]:=++D[I,J]/N;                                                    3             011:000F:2
  MEDC[J]:=++D[I,J]/M;                                                                  011:0013:4
  END;                                                                                  011:0018:0
       FOR I:=1 STEP 1 UNTIL M DO                                         3             011:0019:0
       FOR J:=1 STEP 1 UNTIL N DO BEGIN                                                 011:001D:3
  RENG[I]:=++(D[I,J]-MEDR[I])**2;                                         3             011:0022:0
  COL[J]:=++(D[I,J]-MEDC[J])**2; END;                                                   011:0027:2
  MAXIMO(RENG,M);  MAXIMO(COL,N);                                         3             011:002D:4
  SUMA(RENG,M,1); SUMA(COL,N,2);                                                        011:0046:4
                                                                                  DATA IS 00F3 LONG
  IF Q EQL 2 THEN HOMOSCED :=TRUE  ELSE HOMOSCED:=FALSE;                  3             011:00A5:2
END;                                                                                    011:00A8:3
                                                                          HOMOSCED(011) IS 0086 LONG
BOOLEAN PROCEDURE JICUAD(D,M,N,SCD,JITA,PNOR,SAL,ORD);                    2             003:0007:2
VALUE M,N,SCD; INTEGER M,N;REAL SCD; ARRAY D[*,*],PNOR[*],JITA[*];                      003:0007:2
FILE SAL; ARRAY ORD[*,*];                                                               003:0007:2
  BEGIN                                                                                 003:0007:2
  REAL EXPE,JISQ;                                                                       003:0007:2
                                                                                JICUAD IS SEGMENT 00013
  INTEGER I,J,Q;                                                          2             013:0000:0
  ARRAY OBS[1:6];                                                                       013:0000:0
  DEFINE COMPOR(COI,VAR,COS,S)=                                                         013:0000:0
  IF VAR GTR COI*(SCD/(M*N-1))**(1/2) AND VAR LEQ COS*(SCD/(M*N-1))**                   013:0000:0
 (1/2) THEN OBS[S]:=++1 ELSE #;                                                         013:0000:0
  ORDEN(D,M,N,ORD,PTE);                                                                 013:0000:0
  EXPE:=M*N/6.1666;                                                                     013:0004:3
  FOR I:=1 STEP 1 UNTIL  M DO                                                           013:0007:4
  FOR J:=1 STEP 1 UNTIL N DO                                                            013:000C:1
  IF ORD[I,J] LEQ -PNOR[1]*(SCD/(M*N-1))**(1/2) THEN OBS[1]:=++1 ELSE                   013:0010:4
  COMPOR(-PNOR[1],ORD[I,J],-PNOR[2],2)                                                  013:0019:0
  COMPOR(-PNOR[2],ORD[I,J],0,3)                                                         013:0019:3
  COMPOR(0,ORD[I,J],PNOR[2],4)                                                          013:002A:0
  COMPOR(PNOR[2],ORD[I,J],PNOR[1],5)                                                    013:0039:4
  IF ORD[I,J] GTR PNOR[1]*(SCD/(M*N-1))**(1/2) THEN OBS[6]:=++1;                        013:0049:1
  FOR I:=1 STEP 1 UNTIL 6 DO                                                            013:0064:1
  JISQ:=++((OBS[I]-EXPE)**2)/EXPE;                                                      013:0064:5
  FOR J:=1 STEP 1 UNTIL 3 DO                                                            013:006A:3
  IF JISQ LSS JITA[J] THEN BEGIN                                                        013:006B:1
  Q:=Q+1;                                                                 3             013:006D:2
  WRITE(SAL,<"BONDAD DE AJUSTE:"/,X10,"LA NORMALIDAD SE ACEPTA AL ",                    013:006E:4
  I2,"%, Y EL VALOR DE LA ESTADISTICA ES ",R7.4/>,ALFA[0,2+J],JISQ);                    013:0070:4
```

```
                WRITE(SAL,<"SONDAD DE AJUSTE!"/,X10,"LA NORMALIDAD NO ES ACEPTADA "      013:0081:2
                ,I2,"%)"/>,ALFA[0,2*J])) END;                                          013:0083:2
                IF Q EQL 3 THEN JICUAD:=TRUE ELSE JICUAD:=FALSE;          3            013:0093:3
            END JICUAD;                                                                013:0095:4
                                                                        JICUAD(013) IS 00A1 LONG
            PROCEDURE TRANSPON(D,M,N,DP);                                 2            003:0007:2
            VALUE M,N;  INTEGER M,N; ARRAY D[*,*],DP[*,*];                             003:0007:2
               BEGIN                                                                   003:0007:2
                 INTEGER I,J;                                                          003:0007:2
                                                                        TRANSPON IS SEGMENT 00014
                 FOR I :=1 STEP 1 UNTIL M DO                              2            014:0000:0
                 FOR J := 1 STEP  1 UNTIL N DO                                        014:0004:3
                     DP[J,I] :=  D[I,J];                                              014:0009:0
               END;                                                                   014:000E:3
                                                                        TRANSPON(014) IS 0010 LONG
            PROCEDURE MULTIPLICAMATRIZ(A,B,C,M,N,P,FS,FE);                2            003:0007:2
               VALUE  M,N,P;                                                        003:0007:2
               FILE  FS,FE;                                                         003:0007:2
               INTEGER  M,N,P;                                                      003:0007:2
               ARRAY   A[*,*],  B[*,*],  C[*,*];                                    003:0007:2
                 BEGIN                                                              003:0007:2
                 INTEGER   I,J,K;                                                   003:0007:2
                                                                        MULTIPLICAMATRIZ IS SEGMENT 00015
                 LABEL   FIN;                                             2            015:0000:0
                 FOR I:=1 STEP 1 UNTIL M DO                                           015:0000:0
                 FOR J:=1 STEP 1 UNTIL P DO                                           015:0004:3
                 FOR K:=1 STEP 1 UNTIL N DO                                           015:0009:0
                     C[I,J]:= *+A[I,K]*B[K,J]                                         015:000D:3
            END MULTIPLICAMATRIZ;                                                     015:0016:1
                                                                        MULTIPLICAMATRIZ(015) IS 0016 LONG
            BOOLEAN PROCEDURE AUTOCORR(D,M,N,NOR,FS,ARR);                 2            003:0007:2
            VALUE M,N; INTEGER M,N; ARRAY D[*,*],NOR[*]; FILE FS;ARRAY ARR[*,*];       003:0007:2
            BEGIN                                                                      003:0007:2
            INTEGER I,J,L;                                                             003:0007:2
                                                                        AUTOCORR IS SEGMENT 00016
            ARRAY B[1:M*N];                                               2            016:0000:0
            REAL  Z,Q,DEN,NUM,DB;                                                      016:0003:1
                 WRITE(FS,<"AUTOCORRELACION!"/>);                                      016:0003:1
                 FOR I:=1 STEP 1 UNTIL M DO                                            016:0003:5
                 FOR J:=1 STEP 1 UNTIL N DO BEGIN                                      016:0008:2
                 B[(I-1)*N+J]:=D[I,J];  DB:=*+D[I,J]; END;                   3         016:0011:5
                      DB:=/(M*N);                                            3         016:001A:3
                 FOR I:=2 STEP 1 UNTIL M*N DO BEGIN                                    016:001C:2
                 NUM:=*+(B[I]-B[I-1])**2;                                    3         016:0021:3
                 DEN:=*+(B[I]-DB)**2;  END;                                            016:0025:2
                 DEN:=*+(B[I]-DB)**2;                                        3         016:0025:4
                 Q:= M*M*NUM/((M*M-1)*DEN);                                            016:0028:0
                 WRITE(FS,<X15,"EL VALOR DE LA RAZON DE VON NEWMAN ES! ",R10.6/>,Q);   016:002E:4
                  FOR I:=1 STEP 1 UNTIL 2 DO                                           016:003C:4
                 IF Q GEQ NOR[2*I-1] AND Q LEQ NOR[2*I] THEN BEGIN                     016:0030:2
                 L:=L+1;                                                     3         016:0042:4
                 WRITE(FS,<X15,"LA INDEPENDENCIA DE LOS RESIDUOS SE ACEPTA AL ",       016:0044:0
                 X1,I2,"%"/,X15,"SIENDO LOS EXTREMOS DEL INTERVALO IGUALES A ",        016:0046:0
                 2(R10.7,X1,","),/>,PSIMB[I],NOR[2*I-1],NOR[I*2]);END                  016:0046:0
                 ELSE  BEGIN L:=0;                                           3         016:005A:1
                 WRITE(FS,<X15,"LA INDEPENDENCIA DE LOS RESIDUOS NO ES ACEPTADA AL",   016:005B:2
                 X1,I2,"%"/,X15,"SIENDO LOS EXTREMOS DEL INTERVALO IGUALES A ",        016:0050:2
                 2(R10.7,X1,","),/>,PSIMB[I],NOR[2*I-1],NOR[I*2]);END;                 016:0050:2
```

```
END;                                                                                    AUTOCORR(016) IS 0083 LONG
   BOOLEAN PROCEDURE SIMETRIA(D,M,N,K2,CONT,SAL,Z,ARR);                          2      003:0007:2
VALUE   M,N,CONT; INTEGER M,N,CONT; ARRAY D[*,*];REAL Z,K2; FILE SAL;                   003:0007:2
ARRAY ARR[*];                                                                           003:0007:2
BEGIN                                                                                   003:0007:2
   INTEGER I,J,C;                                                                       003:0007:2
                                                                                    SIMETRIA IS SEGMENT 00017
   REAL K1,SKEW;                                                                2      017:0000:0
     WRITE(SAL,<//"SIMETRIA: ",/>);                                                     017:0000:0
     K2:=0;                                                                             017:0005:4
     FOR I:=1 STEP 1 UNTIL M DO                                                         017:0006:3
     FOR J:=1 STEP 1 UNTIL N DO                                                         017:0008:0
     Z:=+D[I,J];                                                                        017:000F:3
     Z:=Z/(N*M);                                                                        017:0013:4
     FOR I:=1 STEP 1 UNTIL M DO                                                         017:0015:4
     FOR J:=1 STEP 1 UNTIL N DO BEGIN                                                   017:001A:1
     K1:=K1 +(D[I,J]-Z)**3;                                                     3       017:001E:4
     K2:= K2 + (D[I,J]-Z)**2;  END;                                                     017:0022:5
     SKEW:=(K1*(N*M)**(3/2))/(N*M*K2**(3/2));                                   3       017:0027:5
     WRITE(SAL,<X15,"EL COEFICIENTE DE SIMETRIA PARA LA MATRIZ DE "                     017:0030:0
     "RESIDUOS EN LA ITERACION ",I3," ES: "R6.4/>,CONT,SKEW);                           017:0032:0
     FOR I:=1 STEP 1 UNTIL 2 DO                                                         017:0040:1
     IF ABS(SKEW) LSS  SIME[I] THEN BEGIN C:=C+1;                                       017:0040:5
     WRITE(SAL,<X15,"DE DONDE, LOS DATOS TIENEN UNA SIMETRIA MARCADA AL"        3       017:0044:3
     ,I2,"%"/X15,"EL VALOR DE LA COTA ES IGUAL A ",R10.6/>,PSIME[I],                    017:0046:3
     SIME[I]); END                                                                      017:004F:3
     ELSE  BEGIN   C:=0;                                                        3       017:0056:1
     WRITE(SAL,<X15," DE DONDE LOS RESIDUOS NO SON SIMETRICOS AUN.",/           3       017:0057:2
     X15,"EL VALOR DE LA COTA ES IGUAL A ",R10.6/>,SIME[I]);                            017:0059:2
END; IF C EQL 2 THEN  SIMETRIA :=TRUE ELSE SIMETRIA := FALSE; END;                      017:0066:1
                                                                                    SIMETRIA(017) IS 0074 LONG
BOOLEAN PROCEDURE KURTOSIS(D,M,N,CONT,SAL,Z,K2,ARR);                            2      003:0007:2
VALUE   M,N,CONT,Z,K2; INTEGER M,N,CONT;REAL K2,Z; ARRAY D[*,*],ARR[*];                 003:0007:2
FILE SAL;                                                                               003:0007:2
BEGIN                                                                                   003:0007:2
   INTEGER I,J; INTEGER C;                                                              003:0007:2
                                                                                    KURTOSIS IS SEGMENT 00018
   REAL R1,KURT;                                                                2      018:0000:0
     WRITE(SAL,<//"KURTOSIS:"/>);                                                       018:0000:0
     FOR I:=1 STEP 1 UNTIL M DO                                                         018:0005:4
     FOR J:=1 STEP 1 UNTIL N DO                                                         018:000A:1
     R1:=R1 + (D[I,J]-Z)**4;                                                            018:000E:4
     KURT:=R1*M*N/(K2**2);                                                              018:0013:5
     WRITE(SAL,<X15,"EL COEFICIENTE DE KURTOSIS EN LA ITERACION ",I3,                   018:0016:3
     " ES IGUAL A ",R10.6/>,CONT,KURT);                                                 018:001B:3
     FOR J:=1 STEP 1 UNTIL 2 DO                                                         018:0026:4
     IF ABS(KURT) GEQ KURTO[2*J-1] AND ABS(KURT) LEQ KURTO[2*J] THEN                    018:0027:2
     BEGIN C:=C+1;                                                                      018:002C:3
     WRITE(SAL,<X15,"POR LO QUE LOS DATOS TIENEN UNA KURTOSIS NORMAL.",/,       3       018:002E:2
     X15,"LOS EXTREMOS DEL INTERVALO SON ",2(R10.6,X2),/>,KURTO[2*J-1],                 018:0030:2
     KURTO[2*J]);                                                                       018:003A:1
     END                                                                                018:0041:4
     ELSE BEGIN C:=0;                                                           3       018:0041:4
     WRITE(SAL,<X15,"POR LO QUE LOS DATOS AUN SE ENCUENTRAN LEJOS DE ",         3       018:0042:5
     "TENER UNA KURTOSIS NORMAL"/,X15,"LOS EXTREMOS DEL INTERVALO SON"                  018:0044:5
     ,X2,2(R10.6,X1)/>,KURTO[2*J-1],KURTO[2*J]); END;                                   018:0044:5
     IF C EQL 2 THEN  KURTOSIS := TRUE ELSE KURTOSIS := FALSE;                  3       018:0058:3
END;                                                                                    018:005B:4
```

```
          $INCLUDE "BIBLIO/EIGENVECTOR."                                                    2          003:0007:2
*****WARNING-1-AFTER-THIS-RELEASE-ONLY-DOLLAR-CARDS-WITH-$-IN-COL-2-OR->-WILL-BE-WRITTEN-ON-THE-NEWTAPE*****
                                                                  NONSYMEIGENVALUESANDEIGENVECTORS IS SEGMENT 00019
                                                                                    DATA IS 00DC LONG
                                                                                    DATA IS 0006 LONG
                                                                          INNERPRODUCT IS SEGMENT 00018
                                                                         INNERPRODUCT(018) IS 000A LONG
                                                                            HESSENBERG IS SEGMENT 0001D
                                                                           HESSENBERG(01D) IS 0076 LONG
                                                                              TRANSFORM IS SEGMENT 0001E
                                                                             TRANSFORM(01E) IS 0022 LONG
                                                                              NORMREAL IS SEGMENT 0001F
                                                                             NORMREAL(01F) IS 0023 LONG
                                                                           NORMCOMPLEX IS SEGMENT 00020
                                                                          NORMCOMPLEX(020) IS 004E LONG
                                                                                GAUSSM IS SEGMENT 00021
                                                                               GAUSSM(021) IS 0039 LONG
                                                                                SOLVE IS SEGMENT 00022
                                                                               SOLVE(022) IS 0014 LONG
                                                                                  MOR IS SEGMENT 00023
                                                                                 MOR(023) IS 0065 LONG
                                                                          EIGENVALUES IS SEGMENT 00024
                                                                                B.0001 IS SEGMENT 00025
                                                                               B.0001(025) IS 0043 LONG
                                                                          EIGENVALUES(024) IS 0008 LONG
                                                                          EIGENVECTORS IS SEGMENT 00026
                                                                         EIGENVECTORS(026) IS 0110 LONG
                                                        NONSYMEIGENVALUESANDEIGENVECTORS(019) IS 00AD LONG
PROCEDURE ESTIMAU(U,TETA,M,D,SAL,CONT,PTE,ENT);                                     2          003:0007:2
VALUE M;INTEGER M,CONT,PTE;ARRAY U[*,*],TETA[*],D[*,*];FILE SAL,ENT;                           003:0007:2
BEGIN                                                                                          003:0007:2
   REAL  SC;                                                                                   003:0007:2
                                                                                    ESTIMAU IS SEGMENT 00029
   INTEGER I,J;                                                                     2          029:0000:0
   ARRAY  DP[1:N,1:M],UP,T[0:M,0:M],TETAP,TETAI[0:N],TOR[1:M,1:M];                             029:0000:0
   TETAPR[1:1,1:N];                                                                            029:000E:4
   DEFINE  NORMA(A,X)=FOR I:=1 STEP 1 UNTIL N DO                                               029:0011:4
                      FOR J:=1 STEP 1 UNTIL N DO                                               029:0011:4
                      A[I,J]:=*/X#;                                                            029:0011:4
     WRITE(SAL,<//"ESTIMA U:"/>);                                                            029:0011:4
     TRANSPON(D,N,N,DP);                                                                     029:0017:2
     MULTIPLICAMATRIZ(D,DP,T,M,N,M,SAL,ENT);                                                 029:0018:2
     ORDENA(T,M,M,TOR,PTE);                                                                  029:0021:4
     IF MAX(ABS(TOR[1,1]),ABS(TOR[M,M])) GTR 1 THEN                                          029:0025:5
     IF ABS(TOR[1,1]) LEQ ABS(TOR[M,M]) THEN                                                 029:002A:5
       NORMA( T ,ABS(TOR[M,M]))                                                              029:002E:3
     ELSE NORMA( T ,ABS(TOR[1,1]));                                                          029:002E:3
     NONSYMEIGENVALUESANDEIGENVECTORS(M,T,TETAP,TETAI,UP);                                   029:004A:3
     FOR J:=1 STEP 1 UNTIL M DO                                                              029:0049:4
       TETAPR[1,J]:=TETAP[J];                                                                029:0052:1
     ORDENA(TETAPR,1,M,TETAPR,PTE);                                                          029:0055:0
     IF MAX(ABS(TOR[1,1]),ABS(TOR[M,M])) GTR 1 THEN                                          029:0059:1
     TETA[CONT]:=TETAPR[1,1]*MAX(ABS(TOR[1,1]),ABS(TOR[M,M])) ELSE                           029:005E:1
     TETA[CONT]:=TETAPR[1,1];                                                                029:0065:4
     FOR J:=1 STEP 1 UNTIL M DO                                                              029:0068:2
     U[CONT,J]:= UP [J,PTE];                                                                 029:006C:5
     ANDA(U,M);                                                                              029:0071:1
     IMPRIMEMATRIZ(SAL,U,MIN(M,N),M);                                                        029:0085:0
     WRITE(SAL,<X15,"EN LA ITERACION ",I3," EL EIGENVALOR ELEGIDO RESULTO                    029:008A:2
```

```
"    ,R19.9,/X15,"Y EL EIGENVECTOR CORRESPONDIENTE FUE: ",/>,CONT,
        TETA[CONT]);                                                          029:0094:4
      WRITE(SAL,<X10, *(R10.5,X1,"*")/>,H,FOR J:=1 STEP 1 UNTIL H DO          029:009B:4
         U[CONT,J]);                                                          029:00AA:3
   END;                                                                       029:00B2:4
                                                      ESTIMAU(029) IS 00BC LONG
   PROCEDURE ESTIMAV(CONT,TE,U,V,H,N,D,SAL);                  2               003:0007:2
   VALUE TE,H,N;INTEGER H,N;ARRAY U[*,*],V[*,*],D[*,*];REAL TE;INTEGER CONT   003:0007:2
   ;FILE  SAL;                                                                003:0007:2
      BEGIN                                                                   003:0007:2
      INTEGER I,J; REAL SCV;                                                  003:0007:2
                                                      ESTIMAV IS SEGMENT 002A
      REAL  SC;                                               2               02A:0000:0
        WRITE(SAL,</"ESTIMA V:"/>);                                           02A:0000:0
        FOR J:=1 STEP 1 UNTIL N DO BEGIN                                      02A:0005:4
        FOR I:=1 STEP 1 UNTIL H DO                            3               02A:000A:1
             V[CONT,J]:=*+D[I,J]*U[CONT,I];                                   02A:000E:4
             V[CONT,J]:=*/(TE**(1/2));  END;                                  02A:0016:4
        ANORMA(V,N);                                          3               02A:001C:1
        WRITE(SAL,<X15,"EL VECTOR V ES:"/>);                                  02A:0030:0
        WRITE(SAL,<X10, *(R10.5,X1,"*")/>,N,FOR J:=1 STEP 1 UNTIL N DO        02A:0035:4
           V[CONT,J]);                                                        02A:0044:3
      END;                                                                    02A:004C:4
                                                      ESTIMAV(02A) IS 0050 LONG
   PROCEDURE RESIDUOS2(H,N,D,U,V,CONT,TETA,SAL);              2               003:0007:2
   VALUE H,N,CONT;INTEGER H,N,CONT;FILE SAL; ARRAY D[*,*],U[*,*],V[*,*],      003:0007:2
      TETA[*];                                                                003:0007:2
      BEGIN                                                                   003:0007:2
        INTEGER I,J,K;                                                        003:0007:2
                                                      RESIDUOS2 IS SEGMENT 002B
        WRITE(SAL,</"RESIDUOS 2:"/>);                         2               02B:0000:0
        FOR I:=1 STEP 1 UNTIL H DO                                            02B:0005:4
   FOR J:=1 STEP 1 UNTIL N DO                                                 02B:000A:1
   D[I,J]:=*-TETA[CONT]**(1/2)*U[CONT,I]*V[CONT,J];                           02B:000E:4
        WRITE(SAL,<X10,"LA MATRIZ DE RESIDUOS, PARA LA ITERACION ",I3,        02B:0019:5
        "RESULTA SER: "/>,CONT);                                              02B:001B:5
        IMPRIMEMATRIZ(SAL,D,H,N);                                             02B:0028:1
   END RESIDUOS2;                                                             02B:0029:3
                                                      RESIDUOS2(02B) IS 002F LONG
      READ(ENT,/,ALFA[0,*]);                                  2               003:0007:2
      READ(ENT,/,ALFA[1,*]);                                                  003:0019:2
      READ(ENT,/,PSIHE[*],JITA[*],PNOR[*]);                                   003:002B:2
      READ(ENT,/,SIHE[*],KURTO[*],VON[*],COCH[*]);                            003:004A:2
   LEE: READ(ENT,/,H,N)[FIN];                                                 003:006F:5
        BEGIN                                                                 003:0081:1
           REAL CONT;                                         2               003:0081:1
                                                        B.0002 IS SEGMENT 002D
        LABEL REG,YA,AUNNO;                                                   02D:000C:0
        ARRAY D[0:H,0:N],TETA[1:MIN(H,N)],U[1:MIN(H,N),1:N], ORD[1:M,1:N],    02D:0000:0
        V[1:MIN(H,N),1:N], BETA[1:H+N+1],DP[1:N,1:M],T[0:M,0:M];              02D:0011:2
          LEEMATRIZ(ENT,SAL,D,H,N,TRUE);                                      02D:0021:3
              SUMCUAD(H,N,D,CONT, PSC,SAL);                                   02D:0025:2
          ESTIMAB(H,N,D,BETA);                                                02D:0029:2
          RESIDUO1(H,N,D,D,BETA,SAL);                                         02D:002C:4
       IMPRIMEMATRIZ(SAL,D,H,N);                                              02D:0031:2
   REG:   SUMCUAD(H,N,D,CONT,SCDA,SAL);                                       02D:0034:2
        IF H*N LEQ 60 THEN                                                    02D:0038:2
             IF HOMOSCED(D,H,N,SAL,COCH) AND                                  02D:0039:4
          KOLMOGSMIR(H,N,D,SCDA,SAL,ALFA,X,ORD)  AND                          02D:003E:0
          AUTOCORR(D,H,N, VON,SAL,ALFA) AND                                   02D:0044:1
```

```
                SIMETRIA(D,M,N,SIGMA,CONT,SAL,MED,SIME) AND                                  02D:004E:4
                KURTOSIS(D,M,N,CONT,SAL,MED,SIGMA,KURTO) THEN                                02D:0053:5
                GO TO YA ELSE GO TO AUNNO                                                    02D:0054:5
              ELSE  IF HOMOSCED(D,M,N,SAL,COCH) AND                                          02D:0059:1
                JICUAD(D,M,N,SCDA,JITA,PNOR,SAL,ORD) AND                                     02D:005F:3
                SIMETRIA(D,M,N,SIGMA,CONT,SAL,MED,SIME) AND                                  02D:0065:0
                AUTOCORR(D,M,N, VON,SAL,ALFA) AND                                            02D:006A:0
                KURTOSIS(D,N,N,CONT,SAL,MED,SIGMA,KURTO) THEN                                02D:006F:1
                GO TO YA ELSE GO TO AUNNO  ;
             YA: WRITE(SAL,<X15,"LOS RESIDUOS TIENEN YA UNA DISTRIBUCION MUY  ",             02D:0070:1
                "PROXIMA ALA NORMAL"/>)) ; GO TO LEE;                                        02D:0072:0
            AUNNO: CONT:=CONT+1;                                                             02D:0076:5
                ESTIMAU(U,TETA,N,D,SAL,CONT,PTE,ENT));                                       02D:0078:0
                ESTIMAV(CONT,TETA[CONT],U,V,M,N,D,SAL));                                     02D:007E:0
                RESIDUOS2(M,N,D,U,V,CONT,TETA,SAL));                                         02D:0084:5
                GO TO REG;                                                                   02D:0088:3
              END;                                                                           02D:008C:0
                                                                    B,0002(02D) IS 0092 LONG
            FIN: END.                                           2               003:0081:5
                                                                    B,0000(003) IS 00C4 LONG
                                                                          DATA  IS 0048 LONG
```

RESUMEN DE LA COMPILACION EN ALGOL

COMPILACION CORRECTA.

NOMBRE DEL PROGRAMA = PROG/PRUEBA.

TIEMPO EN LA MEZCLA = 51 SEG.
TIEMPO DE COMPILACION = 15 SEG.
TIEMPO DE ENTRADA/SALIDA = 11 SEG.

MEMORIA ESTIMADA = 7630 PALABRAS.
TORRE DE APILAMIENTO = 123 PALABRAS.

NUMERO DE SEGMENTOS = 43.
LONGITUD DE SEGMENTOS = 3684 PALABRAS.
NUMERO DE TARJETAS = 1061.

NUMERO DE ELEMENTOS SINTACTICOS = 9301.
NUMERO DE SEGMENTOS DE DISCO = 277.

```
100 BEGIN
102 FILE SAL(KIND=PRINTER);
104 INTEGER K,L;
106 PROCEDURE IMPRIMEMATRIZ(SAL,A,M,N);
108 VALUE M,N;INTEGER M,N;ARRAY A[*,*];FILE SAL;
110 BEGIN INTEGER I,J;
111  WRITE(SAL,<"ARREGLOS DE DATOS NO. ",2I5,/>,K,L);
112 FOR I:=1 STEP 1 UNTIL M DO
114 BEGIN
118 WRITE(SAL,<"*",*("-------------*")>,N);
120 WRITE(SAL,<"*",5(R12.8,X1,"*")>,FOR J:=1 STEP 1 UNTIL N DO
122 A[I,J]);
126 END;END;
128 PROCEDURE NORMAL(DE,EX);
130 VALUE DE,EX;REAL DE,EX;
132 BEGIN
134 ARRAY NOR,UNI[1:4,1:5];
136 REAL BASE;
138 INTEGER I,J;
140 BASE:=(944154*K)*L;
142 FOR I:=1 STEP 1 UNTIL 4 DO
144 FOR J:=1 STEP 1 UNTIL 5 DO
146 BEGIN
148  UNI[I,J]:=5*I+2*(J-1)+((I-1)*(J-3))**EX*RANDOM(BASE);
150 NOR[I,J]:=SQRT(-2*LN(RANDOM(BASE)))*SIN(6.2832*RANDOM(BASE))*
152 (DE/20)+5*I+2*(J-1)+((I-1)*(J-3))**EX;
154 END;
156 IMPRIMEMATRIZ(SAL,UNI,4,5);
158 IMPRIMEMATRIZ(SAL,NOR,4,5);
160 END NORMAL;
162 FOR K:=1 STEP 5 UNTIL 81 DO
164 FOR L:=1 STEP 2 UNTIL 9 DO
166 NORMAL(K,L);
168 END.
```

APENDICE II

TABLA 1

VALORES CRITICOS PARA LA PRUEBA DE COCHRAN AL 1%

	4	5	6	7	8	9	10
2	0.864	0.788	0.722	0.664	0.615	0.573	0.536
3	0.781	0.696	0.626	0.568	0.521	0.481	0.447
4	0.721	0.633	0.564	0.508	0.463	0.425	0.393
6	0.641	0.533	0.487	0.435	0.393	0.359	0.331
8	0.590	0.504	0.440	0.391	0.352	0.321	0.294
10	0.554	0.470	0.408	0.362	0.325	0.295	0.270

Tomada de Johnson & Leone, Tabla 13.17, p.54

TABLA 2

VALORES CRITICOS PARA LA PRUEBA DE KOLMOGOROV-SMIRNOV

n	.20	.15	.10	.05	.01
4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.149	.153	.165	.180	.203
30	.131	.136	.144	.161	.187
	$\frac{.736}{\sqrt{n}}$	$\frac{.768}{\sqrt{n}}$	$\frac{.805}{\sqrt{n}}$	$\frac{.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

TABLA No.3

VALORES PARA LA PRUEBA DEL SESGO

PUNTOS PORCENTUALES DE UNA COLA PARA LA DISTRIBUCION DE q,*

TAMAÑO DE MUESTRA n	PUNTOS PORCENTUALES 5%	1%	DESVIACION ESTANDAR
25	0.711	1.061	0.4354
30	0.662	0.986	.4052
35	0.621	0.923	.3804
40	0.587	0.870	.3596
45	0.558	0.825	.3418
50	0.534	0.787	.3264
60	0.492	0.723	.3009
70	0.459	0.673	.2806
80	0.432	0.631	.2638
90	0.409	0.596	.2498
100	0.389	0.567	.2377
125	0.350	0.508	.2139
150	0.321	0.464	.1961
175	0.298	0.430	.1820
200	0.280	0.403	.1706
250	0.251	0.360	.1531
300	0.230	0.329	.1400
350	0.213	0.305	.1298
400	0.200	0.285	.1216
450	0.188	0.269	.1147
500	0.179	0.255	.1089

* Como la distribución de q, es simétrica respecto al cero, los puntos por centuales representan valores del 10% y 2% en pruebas de dos colas.

Tomada de: Snedecor & Cochran (34), Tabla A6 i), del Apéndice.

TABLA 4

VALORES PARA LA PRUEBA DE CURTOSIS (g_2)

TAMAÑO DE MUESTRA n	PUNTOS PORCENTUALES			
	SUPERIOR		INFERIOR	
	1%	5%	1%	5%
50	4.88	3.99	1.95	2.15
75	4.59	3.87	2.08	2.27
100	4.39	3.77	2.18	2.35
125	4.24	3.71	2.24	2.40
150	4.13	3.65	2.29	2.45
200	3.98	3.57	2.37	2.51
250	3.87	3.52	2.42	2.55
300	3.79	3.47	2.46	2.59
350	3.72	3.44	2.50	2.62
400	3.67	3.41	2.52	2.64
450	3.63	3.39	2.55	2.66
500	3.60	3.37	2.57	2.67
550	3.57	3.35	2.58	2.69
600	3.54	3.34	2.60	2.70
650	3.52	3.33	2.61	2.71
700	3.50	3.31	2.62	2.72
750	3.48	3.30	2.64	2.73
800	3.46	3.29	2.65	2.74
850	3.45	3.28	2.66	2.74
900	3.43	3.28	2.66	2.75
950	3.42	3.27	2.67	2.76
1000	3.41	3.26	2.68	2.76
1200	3.37	3.24	2.71	2.78
1400	3.34	3.22	2.72	2.80
1600	3.32	3.21	2.74	2.81
1800	3.30	3.20	2.76	2.82
2000	3.28	3.18	2.77	2.83

Tomada de: SNEDECOR & COCHRAN (34), Tabla A6 ii), del Apéndice.

TABLA 5

VALORES DE V* (RAZON DE VON NEUMANN)
PARA LOS NIVELES DE SIGNIFICACION DEL 1% Y 5%

n	Valores de K P=.01	Valores de K P=.05	Valores de K' P=.95	Valores de K' P=.99
4	0.8341	1.0406	4.2927	4.4992
5	.6724	1.0255	3.9745	4.3276
6	.6738	1.0682	3.7318	4.1262
7	.7163	1.0919	3.5748	3.9504
8	.7575	1.1228	3.4486	3.8139
9	.7974	1.1524	3.3476	3.7025
10	.8353	1.1803	3.2642	3.6091
11	.8706	1.2062	3.1938	3.5294
12	.9033	1.2301	3.1335	3.4603
13	.9336	1.2521	3.0812	3.3996
14	.9618	1.2725	3.0352	3.3458
15	.9880	1.2914	2.9943	3.2977
16	1.0124	1.3090	2.9577	3.2543
17	1.0352	1.3253	2.9247	3.2148
18	1.0566	1.3405	2.8948	3.1787
19	1.0766	1.3547	2.8675	3.1456
20	1.0954	1.3680	2.8425	3.1151
21	1.1131	1.3805	2.8195	3.0869
22	1.1298	1.3923	2.7982	3.0607
23	1.1456	1.4035	2.7784	3.0362
24	1.1606	1.4141	2.7599	3.0133
25	1.1748	1.4241	2.7426	2.9919
26	1.1888	1.4336	2.7264	2.9718
27	1.2012	1.4426	2.7112	2.9528
28	1.2135	1.4512	2.6969	2.9348
29	1.2252	1.4594	2.6834	2.9177
30	1.2363	1.4672	2.6707	2.9016
31	1.2469	1.4746	2.6587	2.8864
32	1.2670	1.4817	2.6473	2.8720
33	1.2667	1.4885	2.6365	2.8583
34	1.2761	1.4951	2.6262	2.8451
35	1.2852	1.5014	2.6163	2.8324
36	1.2940	1.5075	2.6068	2.8202
37	1.3025	1.5135	2.5977	2.8085
38	1.3108	1.5193	2.5889	2.7973
39	1.3188	1.5249	2.5804	2.7865
40	1.3266	1.5304	2.5722	2.7760
41	1.3342	1.5357	2.5643	2.7658
42	1.3415	1.5408	2.5567	2.7560
43	1.3486	1.5458	2.5494	2.7466
44	1.3554	1.5506	2.5424	2.7376
45	1.3620	1.5552	2.5357	2.7289
46	1.3684	1.5596	2.5293	2.7205
47	1.3745	1.5638	2.5232	2.7125
48	1.3802	1.5678	2.5173	2.7049
49	1.3856	1.5716	2.5117	2.6977
50	1.3907	1.5752	2.5064	2.6908
51	1.3957	1.5787	2.5013	2.6842
52	1.4007	1.5822	2.4963	2.6777
53	1.4057	1.5856	2.4914	2.6712
54	1.4107	1.5890	2.4866	2.6648
55	1.4156	1.5923	2.4819	2.6585
56	1.4203	1.5955	2.4773	2.6524
57	1.4249	1.5987	2.4728	2.6465
58	1.4294	1.7019	2.4684	2.6407
59	1.4389	1.6051	2.4640	2.6350
60	1.4384	1.6082	2.4596	2.6294

*Si $V \leq K$ se concluye que existe autocorrelación positiva, y si $V \geq K'$ se concluye que existe autocorrelación negativa

Tomada de: Yamane T. (39). Tabla 11 de Apéndice.

BIBLIOGRAFIA

(1) Bancroft, T.A., "Topics in Intermediate Statistical Methods" Ames, Iowa State University, 1968.

(2) Cramér, H., "Metodos Matemáticos de la Estadística" Ed. Aguilar, 1953.

(3) Hernández, F. y Guerrero, V.M., "El uso de transformaciones en los modelos lineales", Tesis Profesional de Actuario, 1974.

(4) Hillier, F.S. y Lieberman, G.J.; "Introduction to Operations Research", E. Holden Day, 1972.

(5) Johnson, N.L. y Leone, F.C., "Statistics and Experimental Design in Engineering on the Physical Sciences", V.II, Ed. John Wiley & Sons, Inc. Segunda Impresión, 1968.

(6) Johnston, J. "Econometric Methods", New York, McGraw Hill, 1963.

(7) Tillicfors, Journal of the American Statistical Association, Vol.62, 1967.

(8) Mandel, J., "The Partitioning of Interaction in Analysis of Variance", Journal of Research, National Bureau of Standards, Section B, Math. Sciences, No. 4, 1973.

(9) Mandel, J., "A new Analysis of Variance Model for Non-Additive Data", Technometrics Vol. 13, No.1, February 1971.

(10) Mendez, R., I., "Introduccion a la Metodologia Estadística", Ed. Prtena, 1971.

(11) Yamane, T., "Statistics, an Introductory Analysis", 2 Ed., New York Harper & Row, 1967.

(12) Graybill, F.A., "Introduction to linear Statistical Models", V.I., McGraw Hill, 1961.

(13) Rao, P., "Linear Statistical Inference, and its applications", John Wiley & Sons, Inc., 1965.

(14) Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components", Journal of Educational Psychology 24, 417-441; 498-521 (1933).

(15) Mandel, J.,"The Distribution of Eigenvalues of Covariance Matrices of Residuals in Analysis of Variance", Journal of Research, National Bureau of Standards-B. Math. Sciences Vol 74 B, No.3, 1970.

(16) Lindgren, B.W., "Statistical Theory" Collier MacMillan Ltd. London, 1968.

(17) Anderson, T.W., "Introduction to Multivariate Statistical Analysis", John Willey & Sons Ltd.

(18) Garber, R.J., McIlvane, T.C., and Hoover, M.M."A Method of Laying ourt Experiment Plats", Journal of the American Society of Agronomy, 1931.

(19) Tukey, J.W., "One Degree of Freedom for Non-Additivity" Biometrics 5, 232-242 (1949)

(20) Panse. V.G. y Sukhatme, P.V., "Metodos Estadísticos para Investigadores Agrícolas", Fondo de Cultura Económica, 2a. Ed., 1963.