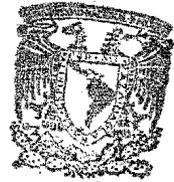


UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS



BIBLIOTECA
INSTITUTO DE ECOLOGIA
UNAM

DESARROLLO DE UN PROGRAMA COMPUTACIONAL PARA
TRATAR EL PROBLEMA DE REGRESION LINEAL MULTIPLE.

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 :

RODOLFO ANTONIO HILL GONZALEZ

México, D.F.

1978

A Rodolfo Antonio, mi padre

A Mélida, mi madre

A quienes todo debo porque

todo me lo dieron.

A Linett

e Ibeth del Carmen

mis hermanas.

A los maestros en ciencias

Gustavo Valencia y Francisco

Aranda, por su atinada direc

ción en la elaboración de --

esta tesis.

Quiero agradecer

Al Ing. Carlos Strassburguer,

Al Act. Mario Villalobos,

Al Dr. José Negrete y

Al Act. Héctor Mena-Brito,

la ayuda que me brindaron, sin

la cual este trabajo no hubiera

llegado a un feliz término.

Agradezco en todo lo que vale
la desinteresada ayuda que me
prestaron los Srs. Jorge Cár-
denas, Miguel Angel Romero, -
Alberto Córdoba y Enrique - -
Aguilar; todos ellos operado-
res nocturnos del C.S.C.

A la Lic. Dora Relux, muy especialmente,
por todos los trabajos que por mi se to-
ma; ayuda invaluable.

A Hugo Torrijos, por su apoyo,
valioso impulso para mi supera
ción personal.

A la familia Luermo-Solorio,
en especial a Pepe, y a la -
familia Díaz-López, por las-
atenciones recibidas durante
mi estancia en México.

A mis Amigos.

CONTENIDO

	Página
Introducción	1
Capítulo I. Conceptos del Análisis de Regresión utilizados en el programa SORA.	
I.1 El Modelo	6
I.2 Estimación Puntual	10
I.3 Estimación por Intervalos	12
I.4 Prueba de Hipótesis	25
I.5 Análisis de Residuales	31
Capítulo II. Generalidades del Programa SORA.	
II.1 Transformaciones	45
II.2 Leer y/o Simular	47
II.3 Capacidad, Mensajes, Resultados y Dis-- tribuciones	48
Capítulo III. Transformaciones de los Datos en- el SORA.	
III.1 Introducción	52
III.2 Modo de Uso	53
III.3 La Instrucción FOR	60
III.4 Ejemplos	63

Capítulo IV. Lectura de Datos	
IV.1 Introducción	73
IV.2 Ejemplos	75
Capítulo V. Simulación de Datos	84
Capítulo VI. Comandos del Programa	93
Capítulo VII. Diagramas de Flujo	105
Capítulo VIII. Ejemplo	118

ANEXOS

- I. Programa Auxiliar
- II. Funciones Utilizables en las Transformaciones
- III. Tabla de Tate y Klett
- IV. Tabla para Corridas
- V. Tabla de Von-Neumann y Durbin y Watson

Bibliografía

INTRODUCCION

Debido a la gran utilidad que representa el predecir fenómenos, como también el tener relacionado a un conjunto de variables que intervienen en un determinado proceso, la parte de la estadística que estudia la relación entre variables, ha cobrado gran importancia; en ocasiones con el fin de predecir el valor de una de estas variables y en otras con el fin solo de explicar los efectos de una (s) sobre la que se explica. A esta área estadística se le conoce como Análisis de Regresión. Los métodos involucrados en el Análisis de Regresión requieren de cálculos numéricos complicados o engorrosos; es por lo antes mencionado que surge la necesidad de tener un programa computacional capaz de resolver gran parte de los problemas a que esta área se dedica. Un intento para satisfacer esta necesidad ha sido realizado por el autor mediante la elaboración del programa "SORA".

Entre las motivaciones que se tuvieron para realizar este trabajo figura, la de proporcionar un auxiliar en los cursos de Análisis de Regresión y Estadística II que se imparten en esta facultad, conjuntar los mejores algoritmos para la solución de problemas que esta área involucra. Se consideró como buen algoritmo aquel que fuera rápido y exacto. Se le dió a estas propiedades mucha

importancia sobre todo en el algoritmo de inversión de matrices y en el de la descomposición de Cholesky. Dichas cualidades antes mencionadas no es posible encontrarlas en los programas de biblioteca (paquetes) de los Centros de Computo a que se tienen acceso, resultando tediosa la tarea de llegar a calcular todo lo que en este programa se obtiene, aún cuando, utilizando varios paquetes, se podrían obtener los mismos resultados que en este programa.

Para la realización del programa se utilizaron algunas subrutinas (procedures) del Collected ACM Algorithms (1960-1977), que fueron modificadas por el autor para satisfacer los objetivos y necesidades de este trabajo, - - siendo éstas: inversión de matrices, descomposición espectral, área de las distribuciones N (Normal), X^2 (Ji-Cuadrada), F (Snedecor), t (Student).

El programa está escrito en lenguaje ALGOL y fue diseñado para ser utilizado por personas con conocimientos del Análisis de Regresión, sin importar el área a la que se dediquen. Es necesario una terminal remota para la utilización del programa.

Este texto tiene como finalidad ser una guía para -- todas aquellas personas que desean utilizar el programa -

de manera mas eficiente.

CAPITULO I

**CONCEPTOS DEL ANALISIS DE REGRESION
UTILIZADOS EN EL PROGRAMA SORA**

I.1 El Modelo

En innumerables procesos se tiene que ciertas variables afectan (o parecen afectar) el valor de otra(s) variable(s). Por ejemplo al estudiar, en un proceso industrial, la "CALIDAD" del producto terminado, se considera que ésta se ve afectada por las respectivas "CALIDADES" de las materias primas utilizadas, así como por la temperatura de operación, por la presión dentro de las máquinas que intervienen en el proceso, etc.

Así pues, se considera que la respuesta podría ser aproximada a partir de una relación funcional, en la cual fuesen consideradas todas aquellas variables que tuviesen mas o menos influencia en el proceso. De una manera ideal, si se consideran todas las variables que influyen en el proceso, entonces los valores de la respuesta (Y) pueden ser obtenidos a partir de

$$Y = f(X_1, \dots, X_r)$$

mediante este tratamiento surgen dos problemas cuya solución, como se verá, puede ser difícil; éstos son:

- a) La forma analítica de f puede ser conocida y muy complicada, ó desconocida.

- b) El número r de variables en estudio puede ser tan grande que sea prácticamente imposible de manejar f en forma adecuada.

Entonces se ocurren dos alternativas las cuales son compatibles:

- i) Aproximar f mediante f^* , una función mas sencilla (posiblemente un polinomio).
- ii) Ignorar todas aquellas variables cuya influencia sea considerada "NO IMPORTANTE" reduciendo así el número de variables consideradas.

La consideración de (ii) traerá como resultado que las variables no consideradas causen fluctuaciones en la respuesta, mismas que pueden ser consideradas aleatorias, aún cuando se mantengan fijos los valores de las variables consideradas. Así pues, a partir de (i) y (b) es posible establecer la siguiente relación:

$$Y = f^* (x_1, \dots, x_m) + \epsilon$$

donde ϵ es una variable aleatoria tal, que estará determinada por aquellas variables cuya influencia fue considerada despreciable y por errores que se pudieron cometer en la medición de las variables involucradas.

En base a lo anterior, se antoja que el problema a cuya solución debe avocarse el esfuerzo es a la determina--ción de una forma adecuada para f^* .

Si se supone que dicha función es continua (al menos en el rango de valores de X_1, \dots, X_r que interesan), será posible aproximar sus valores mediante un polinomio (con- lo que un nuevo factor de error será incluido), entonces- es posible suponer que se cumple con la siguiente rela- -ción

$$Y = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p + \epsilon$$

donde Z_i ($i=1, 2, \dots, p$) es alguna función real de las variables X_1, X_2, \dots, X_m , con parámetros conocidos, y ϵ es una variable aleatoria.

El Análisis de Regresión está diseñado para aquellas situaciones donde se supone que una variable está rela- -cionada a una o más mediciones hechas, usualmente en el - mismo "objeto". El propósito del Análisis es usar los - datos para estimar la forma de estas relaciones. De es- tos modelos ya se habló y se propuso un modelo del siguien- te tipo

$$Y_i = \beta_1 z_{i1} + \beta_2 z_{i2} + \dots + \beta_p z_{ip} + \epsilon_i$$

el cual es llamado modelo lineal; lineal porque es expresado como una combinación lineal de los parámetros y de la variable aleatoria. Por supuesto podrían postularse innumerables modelos no lineales en los parámetros que ajustacen mejor a las observaciones, sin embargo el modelo lineal es el que ha recibido mayor atención tanto en teoría como en la práctica. Desde un punto de vista teórico, es matemáticamente manejable, y su función se ha visto adecuada en muy variadas aplicaciones prácticas. El modelo es conocido a menudo como modelo de regresión y, como su expresión es lineal, es más comunmente llamado modelo de Regresión Lineal. La variable denotada por Y es conocida como variable dependiente y las variables Z_1, Z_2, \dots, Z_p como variables independientes.

Supongase que se cuenta con n observaciones de las variables Y_i y las correspondientes $Z_{i1}, Z_{i2}, \dots, Z_{ip}$ ($n \geq p$), entonces el modelo puede ser expresado en forma matricial como sigue

$$Y = Z\beta + \epsilon$$

$$\text{Rango}(Z) = p$$

donde \underline{Y} es un vector columna n dimensional, es decir, \underline{Y} es de $n \times 1$, donde n es el número de renglones y 1 es el número de columnas, las componentes de dicho vector son las y_i , $\underline{\beta}$ es el vector p dimensional $p \times 1$ de parámetros desconocidos, \underline{Z} es una matriz $n \times p$ cuyo i -ésimo renglón es el vector $(z_{i1}, z_{i2}, \dots, z_{ip})$ y $\underline{\epsilon}$ es un vector aleatorio n dimensional cuyas componentes se consideran "errores".

1.2 Estimación Puntual

En esta sección se muestran los mejores estimadores de tipo puntual encontrados para la teoría del Análisis de Regresión.

El estimador de $\underline{\beta}$, por el método de mínimos cuadrados, está dado por:

$$\hat{\underline{\beta}} = (\underline{Z}'\underline{Z})^{-1} \underline{Z}'\underline{Y}$$

Si se supone que $E(\underline{\epsilon}) = \underline{0}$, entonces $\hat{\underline{\beta}}$ tiene la propiedad de insesgamiento. Si se supone también que los errores cumplen con

$$V(\epsilon_i) = E(\epsilon_i^2) = \sigma^2; \quad \text{Cov}(\epsilon_i, \epsilon_j) = 0; \quad \forall i \neq j \quad (i, j = \overline{1, n})$$

¹El signo $\hat{}$ se lee gorro y se usa para indicar estimada(o).

siendo $\hat{\beta}$ un estimador lineal insesgado de β y de acuerdo con el teorema de Gauss-Markoff se obtiene que $\hat{\beta}$ es de varianza mínima.

La matriz de varianzas y covarianzas de $\hat{\beta}$, está dada por

$$\text{Var}(\hat{\beta}) = \sigma^2 (Z'Z)^{-1}$$

Un estimador insesgado para σ^2 es

$$S^2 = \frac{Y'(I - Z(Z'Z)^{-1}Z')Y}{n-p}$$

Hasta aquí no se han hecho suposiciones acerca de la función de distribución del vector de errores. Si se supone que la distribución de $\underline{\epsilon}$ pertenece a la familia Normal, se obtiene lo siguiente (Graybill (1961)):

1) $\hat{\beta}$ y $\frac{n-p}{n} S^2$ son estimadores de máxima verosimilitud, por lo tanto tienen todas las propiedades de éstos (Mood y Graybill (1972)).

$$2) Y \sim N(Z\beta, \sigma^2 I)^1$$

¹ El signo \sim se lee distribuida y se utiliza para indicar distribuciones de variables aleatorias.



BIBLIOTECA
INSTITUTO DE ECOLOGIA
UNAM

$$3) \hat{\beta} \sim N(\beta, \sigma^2 (Z'Z)^{-1})$$

$$4) \frac{(n-p)}{\sigma^2} S^2 \sim \chi^2_{(n-p)}$$

5) $\hat{\beta}$ y $(n-p) S^2$ son independientes.

I.3 Estimación por Intervalos

Intervalo de confianza para la i -ésima componente de

β

Bajo la suposición de normalidad en los errores, se tiene que

$$\hat{\beta} \sim N(\beta, \sigma^2 (Z'Z)^{-1})$$

de aquí
$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{c_{ii} \sigma^2}} \sim N(0,1)$$

donde c_{ii} es la i -ésima componente de la diagonal de $(Z'Z)^{-1}$, $i = \overline{1, n}$. En vista de que $\hat{\beta}$ y S^2 se distribuyen independientemente, $\hat{\beta}_i$ y S^2 también lo cumplen. Se tiene:

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{c_{ii} \sigma^2}} \bigg/ \sqrt{\frac{(n-p) S^2}{(n-p) \sigma^2}} \sim t_{(n-p)}$$

debido también a que

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

de donde

$$\frac{\hat{\beta}_i - \beta_i}{S \sqrt{C_{ii}}} \sim t_{(n-p)}$$

por lo que se obtiene el siguiente intervalo

$$\hat{\beta}_i - t_{\frac{\alpha}{2}} S \sqrt{C_{ii}} \leq \beta_i \leq \hat{\beta}_i + t_{\frac{\alpha}{2}} S \sqrt{C_{ii}}$$

además cumple con

$$P\left(\hat{\beta}_i - t_{\frac{\alpha}{2}} S \sqrt{C_{ii}} \leq \beta_i \leq \hat{\beta}_i + t_{\frac{\alpha}{2}} S \sqrt{C_{ii}}\right) = 1 - \alpha$$

donde $(1 - \alpha) 100\%$ es la confiabilidad del intervalo.

Intervalo de confianza para combinaciones lineales de β

La forma de obtención del intervalo de confianza para combinaciones lineales de β va a ser encontrada por medio de un elipsoide de confianza.

Sean

$$\psi_i = \sum_{j=1}^p C_{ij} \beta_j \quad \text{para } i = \overline{1, q}$$

y

$$\hat{\psi}_i = \sum_{j=1}^p C_{ij} \hat{\beta}_j$$

en términos vectoriales se pueden escribir como

$$y \quad \begin{aligned} \underline{\psi} &= C \underline{\beta} \\ \underline{\hat{\psi}} &= C \underline{\hat{\beta}} \end{aligned}$$

donde C es una matriz de $q \times p$, con elemento (ij) -ésimo igual a C_{ij} ; sustituyendo en $\underline{\hat{\psi}}$ a $\underline{\hat{\beta}}$ se tiene:

$$\underline{\hat{\psi}} = C (Z'Z)^{-1} Z'Y = AY \quad \text{donde} \quad A = C (Z'Z)^{-1} Z'$$

Entre las propiedades de $\underline{\hat{\psi}}$ con las que se cuentan, están

$$E(\underline{\hat{\psi}}) = \underline{\psi}$$

$$V(\underline{\hat{\psi}}) = V(A\underline{Y}) = AV(\underline{Y})A' = A\sigma^2 I A' = \sigma^2 AA'$$

ya que \underline{Y} está distribuido normalmente, entonces

$$\underline{\hat{\psi}} \sim N(\underline{\psi}, \sigma^2 AA')$$

y si se define $B = AA'$, entonces $\underline{\hat{\psi}} \sim N(\underline{\psi}, \sigma^2 B)$. Se supone que $\text{Rango}(C) = q$, es decir, que los renglones de C son linealmente independientes. Si $\text{Rango}(C) \neq q$, sea $\text{Rango}(C) = r$ con $r < q$, entonces puede procederse de la siguiente forma: se reenumeran las ψ_i ($i = \overline{1, q}$), tal que, $\{\psi_{r+1}, \dots, \psi_q\}$ son combinaciones lineales con coefi

cientes conocidos de $\{\psi_1, \dots, \psi_r\}$. Entonces cualquier punto del conjunto $\psi_{r+1}, \dots, \psi_q$ es determinado por el conjunto ψ_1, \dots, ψ_r , es decir, si se tiene una elipsoide de confianza para ψ_1, \dots, ψ_r también se tiene para $\psi_{r+1}, \dots, \psi_q$, debido a que queda determinada inmediatamente (aunque sea de dimensión diferente). Por todo lo anterior solo se tratará el primer caso.

Como $\hat{\psi}$ se distribuye $N(\psi, \sigma^2 B)$ en forma independiente de $\sigma^2 \chi^2_{(n-p)}$ (ya que $\hat{\beta}$ y $(n-p)S^2$ son independientes) y por el teorema (Scheffé (1959), apéndice V), que dice:

Si $\underline{x} \sim N(\underline{\xi}, J^{-1})$, $\underline{x}_{n \times 1}$, $\underline{\xi}_{n \times 1}$, J real

y J es una matriz de varianzas-covarianzas

$$\leadsto J^{-1}(\underline{x} - \underline{\xi})' J^{-1}(\underline{x} - \underline{\xi}) \sim \chi^2_{(n)}$$

de donde al aplicar el teorema antes enunciado y el hecho de que B es no singular (Scheffé (1959)), se tiene que -

$$\frac{(\hat{\psi} - \psi)' B^{-1} (\hat{\psi} - \psi)}{\sigma^2} \sim \chi^2_{(q)}$$

y se sabe además que

$$\frac{(n-p) S^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

de aquí

$$\frac{(\hat{\underline{\psi}} - \underline{\psi})' B^{-1} (\hat{\underline{\psi}} - \underline{\psi})}{q \sigma^2} \Bigg/ \frac{(n-p) S^2}{(n-p) \sigma^2} \sim F_{q, n-p}$$

y simplificando se obtiene

$$\frac{(\hat{\underline{\psi}} - \underline{\psi})' B^{-1} (\hat{\underline{\psi}} - \underline{\psi})}{q S^2} \sim F_{q, n-p}$$

Esta es la forma general para una elipsoide de confianza. Este resultado será útil tanto aquí como en el desarrollo de intervalos simultáneos, sólo que con un desarrollo diferente como se mostrará en la siguiente sección.

El elipsoide de una confiabilidad de $1-\alpha$ resulta estar dado por

$$(\hat{\underline{\psi}} - \underline{\psi})' B^{-1} (\hat{\underline{\psi}} - \underline{\psi}) \leq q S^2 F_{\alpha; q, n-p}$$

dicha desigualdad determina un elipsoide en un espacio q dimensional determinado por $\underline{\psi}$, con centro en $\hat{\underline{\psi}}$. La probabilidad de que este elipsoide aleatorio cubra a $\underline{\psi}$ es $1-\alpha$.

Si lo que se desea obtener es información de una combinación lineal, resultando por consiguiente estar intere

sado en un intervalo de confianza para esta, entonces se llevan a cabo los siguientes pasos:

$$\psi = \underline{c}' \underline{\beta} \quad (\underline{c} \neq \underline{0}) \quad \text{donde} \quad \underline{c}'_{1 \times q}$$

El elipsoide unidimensional resultante es el intervalo

$$b^{-1} (\hat{\psi} - \psi)^2 \leq S^2 F_{\alpha; 1, n-p}$$

donde $\hat{\psi} = \underline{a}' \underline{y}$ tal que $\underline{a}'_{1 \times n}$

y además $b = \underline{a}' \underline{a}$,

$$V(\hat{\psi}) = \sigma^2 \underline{a}' \underline{a}$$

$$\hat{\sigma}_{\hat{\psi}}^2 = S^2 \underline{a}' \underline{a}$$

y se obtiene por lo tanto

$$\hat{\psi} - t_{\frac{\alpha}{2}; n-p} \hat{\sigma}_{\hat{\psi}} \leq \psi \leq \hat{\psi} + t_{\frac{\alpha}{2}; n-p} \hat{\sigma}_{\hat{\psi}}$$

Intervalo de confianza para combinaciones lineales simultáneas de $\underline{\beta}$

Se han construido intervalos de confianza para una combinación lineal. Sin embargo, si ahora se considera un conjunto de combinaciones lineales, y se quiere encontrar una región que contenga a todas las combinaciones lineales con probabilidad $1-\alpha$, no basta con encontrar cada

uno de los intervalos de confianza, ya que la región definida por ellos tendría probabilidad diferente de $1-\alpha$. (podría ser mayor, menor o la región podría contener información contradictoria). En éste caso se procede de la siguiente manera:

se continua con la forma de la elipsoide de confianza obtenida en la sección anterior; es decir, partiendo de

$$\frac{(\hat{\underline{\psi}} - \underline{\psi})' B^{-1} (\hat{\underline{\psi}} - \underline{\psi})}{q S^2} \sim F_{q, n-p}$$

donde

$$P\left(\frac{(\hat{\underline{\psi}} - \underline{\psi})' B^{-1} (\hat{\underline{\psi}} - \underline{\psi})}{q S^2} \leq F_{\alpha; q, n-p}\right) = 1-\alpha \quad (1)$$

Por la desigualdad de CAUCHY - SCHWARZ (Rao (1965), pag. 240)

$$D'W^{-1}D = \max_U \frac{(U'D)^2}{U'WU}$$

donde U y D son vectores columnas y W es una matriz definida positiva. Aplicando este resultado a lo obtenido, se tiene que

$$\frac{(\hat{\underline{\psi}} - \underline{\psi})' B^{-1} (\hat{\underline{\psi}} - \underline{\psi})}{q S^2} = \frac{1}{q S^2} \max_U \frac{(U' (\hat{\underline{\psi}} - \underline{\psi}))^2}{U' B U}$$

a partir de (1)

$$P \left\{ \max_U \frac{|U' (\hat{\underline{\psi}} - \underline{\psi})|}{\sqrt{U' B U}} \leq S \sqrt{q F_{\alpha}; q, n-p} \right\} = 1 - \alpha$$

esto es

$$P(|U' (\hat{\underline{\psi}} - \underline{\psi})| \leq S \sqrt{q F_{\alpha} U' B U} \quad \forall U) = 1 - \alpha$$

o equivalentemente

$$P(U' \hat{\underline{\psi}} - S \sqrt{q F_{\alpha} U' B U} \leq U' \underline{\psi} \leq U' \hat{\underline{\psi}} + S \sqrt{q F_{\alpha} U' B U} \quad \forall U) = 1 - \alpha$$

Esta ecuación provee intervalos de confianza simultáneos para todas las funciones de la forma $U' \underline{\psi} = U' C \underline{\beta}$. Como se observa, se obtienen intervalos de confianza para combinaciones lineales de $\underline{\beta}$ pero en términos de $U' C$ -- (se tendrá que estar probando la multiplicación del vec--

tor U' por la matriz C) y la única restricción que se tiene es que $\text{Rango}(C) \leq \min(q, p)$. Si $q=p$ y $C = I_{p \times p}$, se tiene entonces que esto da lugar a intervalos de confianza de la β como se querían.

Intervalo de confianza para $E(Y|X_0)$

Se tiene que para un vector X_0 el estimador de la media de Y_0 se obtiene manteniendo fijo X_0 , dicha observación está dada por

$$E(Y|X_0) = X_0' \beta \quad X_0 \text{ pxi}$$

que es precisamente el caso en que $C' = X_0'$ ($q=1, C = X_0$), de donde el intervalo de confianza para $E(Y|X_0)$ está dado por

$$X_0' \hat{\beta} - t_{\frac{\alpha}{2}} S \sqrt{X_0'(Z'Z)^{-1}X_0} \leq X_0' \beta \leq X_0' \hat{\beta} + t_{\frac{\alpha}{2}} S \sqrt{X_0'(Z'Z)^{-1}X_0}$$

Intervalo de confianza para $Y|X_0$

Se se desea el intervalo de confianza para el valor de la variable dependiente en X_0 , se dice que se desea encontrar el intervalo de predicción. Dicho intervalo será más grande que el calculado anteriormente debido a -

que no es la media de observaciones, sino un valor específico.

El intervalo de predicción para $y_0 = \beta' x_0 + \varepsilon_0$, se obtiene a partir de $\hat{y}_0 - y_0$, donde $\hat{y}_0 = \hat{\beta}' x_0$, es decir, $\hat{y}_0 - y_0 = x_0' \hat{\beta} - x_0' \beta - \varepsilon_0$ que tiene una distribución normal con media cero y varianza igual a:

$$\begin{aligned} V(\hat{y}_0 - y_0) &= V(x_0' \hat{\beta}) + V(x_0' \beta + \varepsilon_0) - 2 \text{Cov}(x_0' \hat{\beta}, x_0' \beta + \varepsilon_0) \\ &= x_0' V(\hat{\beta}) x_0 + V(\varepsilon_0) = \sigma^2 x_0' (Z'Z)^{-1} x_0 + \sigma^2 \\ &= \sigma^2 (x_0' (Z'Z)^{-1} x_0 + 1) \end{aligned}$$

donde el término $\text{Cov}(\hat{y}_0, x_0' \beta + \varepsilon_0)$ se anula en vista de que ε_0 es independiente de aquellos términos de error mediante los cuales se estima a $\hat{\beta}$. Entonces

$$\hat{y}_0 - y_0 \sim N(0, \sigma^2 (x_0' (Z'Z)^{-1} x_0 + 1))$$

y como

$$\frac{(n-p) S^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

entonces

$$\frac{\hat{y}_0 - y_0}{S \sqrt{\underline{x}_0' (Z'Z)^{-1} \underline{x}_0 + 1}} \sim t_{(n-p)}$$

de donde el intervalo queda dado por

$$\hat{y}_0 - t_{\frac{\alpha}{2}} S \sqrt{\underline{x}_0' (Z'Z)^{-1} \underline{x}_0 + 1} \leq y_0 \leq \hat{y}_0 + t_{\frac{\alpha}{2}} S \sqrt{\underline{x}_0' (Z'Z)^{-1} \underline{x}_0 + 1}$$

Intervalo de confianza para σ^2

Dado que $\frac{(n-p) S^2}{\sigma^2} \sim \chi^2_{(n-p)}$, es posible construir un intervalo de confianza para σ^2 . El problema se reduce a encontrar valores de A y B , tales que:

$$P\left(A \leq \frac{(n-p) S^2}{\sigma^2} \leq B\right) = 1 - \alpha$$

Estos valores pueden encontrarse en tablas de la distribución $\chi^2_{(n-p)}$. Ahora bien, lo ideal sería que estos valores diesen un intervalo lo más corto posible. El intervalo tiene la forma

$$\frac{(n-p) S^2}{B} \leq \sigma^2 \leq \frac{(n-p) S^2}{A}$$

entonces A y B deben de ser tales que si u es una variable aleatoria con distribución $\chi^2_{(n-p)}$, se tiene

$$\int_A^B f(u) du = 1 - \alpha$$

de modo que $\left(\frac{1}{A} - \frac{1}{B}\right)$ sea mínimo. Usualmente esto no es sencillo, por lo que prácticamente A y B se seleccionan de manera que se cumpla lo siguiente:

$$P[u \leq A] = P[B < u] = \alpha/2$$

Otra manera de seleccionar A y B se encuentra en los intervalos de confianza definidos por Tate y Klett.

Intervalos de Confianza para σ^2 según Tate y Klett.¹

Como se vió en intervalos de confianza, para la varianza se desea encontrar valores A y B de modo tal que dicho intervalo sea de longitud mínima. Tate y Klett (1959) desarrollaron una teoría concerniente a esto, encontraron valores que cumplen tal requisito y los tabularon. Dichos intervalos tienen la siguiente forma:

$$\left(\frac{1}{b_n} \sum_{i=1}^n (y_i - \bar{y})^2, \frac{1}{a_n} \sum_{i=1}^n (y_i - \bar{y})^2 \right)$$

donde a_n y b_n se encuentran en una tabla dada a conocer por los mismos autores (aparece en el anexo 3), en donde la n representa los grados de libertad.

¹ Las modificaciones a este método surgieron a partir de pláticas entre Valencia G., Aranda F y el Autor.

En vez de $\sum_i (y_i - \bar{y})$ se usó el numerador de los dos estimadores de la varianza en modelos de Análisis de Regresión, es decir, se hace uso del modelo. Dichos numeradores se escogieron dependiendo de si había o no carencia de ajuste, obteniéndose:

sin Carencia de Ajuste

$$\sum (y_i - \hat{y}_i)^2 \quad \text{con } n-p \text{ grados de libertad,}$$

con Carencia de Ajuste

sea $y_{11}, y_{12}, \dots, y_{1n_1}$ con n_1 repeticiones en X_1
 $y_{21}, y_{22}, \dots, y_{2n_2}$ con n_2 repeticiones en X_2
 $\vdots \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$
 $y_{k1}, y_{k2}, \dots, y_{kn_k}$ con n_k repeticiones en X_k

donde $\sum_{i=1}^k n_i = n$ y X_i ($i = \overline{1, k}$) son vectores observación de $1 \times p$. Se usa entonces

$$\sum_{i=1}^k \sum_{u=1}^{n_i} (y_{iu} - \bar{y}_i)^2$$

con $\sum_{i=1}^k n_i - k$ grados de libertad,

donde $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i$

I.4 Prueba de Hipótesis

Para probar hipótesis del tipo

$$H_0: C\beta = \underline{y} \quad \text{vs} \quad H_2: C\beta \neq \underline{y}$$

donde C es una matriz de $q \times p$ de rango q ($\leq p$), algunos ejemplos sobre el tipo de hipótesis que se pueden probar por medio de la hipótesis lineal general son:

i) Si β es un vector de 2×1 ,

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{y} \quad \underline{y} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ se estaría probando la hipótesis } H_0: \beta_1 = 0 \text{ y } \beta_2 = 0.$$

ii) Si β es un vector de $p \times 1$,

$$C = I_{p \times p} \quad \text{y} \quad \underline{y} = \underline{0}_{p \times 1}, \text{ la hipótesis a probar sería } H_0: \begin{matrix} \beta_1 = 0 \\ \vdots \\ \beta_p = 0 \end{matrix}$$

y iii) Siendo el modelo igual que ii),

$$C = \begin{pmatrix} 1 & -1 & 0 & \dots & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & 1 & -1 & 0 \\ 0 & \dots & \dots & 0 & 1 & -1 \end{pmatrix} \quad \text{y} \quad \underline{y}_{p \times 1} = \underline{0}_{p \times 1}$$

La prueba de hipótesis estaría dada por:

$$H_0: \begin{cases} \beta_1 - \beta_2 = 0 \\ \beta_2 - \beta_3 = 0 \end{cases}$$

\vdots

$$\beta_{p-1} - \beta_p = 0$$

es decir $H_0: \beta_1 = \beta_2 = \dots = \beta_p$

Se hará uso del cociente de verosimilitudes. El cociente de verosimilitudes está definido como:

$$\lambda = \frac{\max_{\omega} L(\beta, \sigma^2)}{\max_{\Omega} L(\beta, \sigma^2)}$$

donde ω y Ω son espacios definidos por $\beta_i \forall i$ y σ^2 , y $L(\beta, \sigma^2)$ se conoce como función de verosimilitud. Obteniendo el máximo de la función de verosimilitud en ω y Ω se obtiene que:

$$\lambda = \left[\frac{(\underline{y} - Z\hat{\beta})' (\underline{y} - Z\hat{\beta})}{(\underline{y} - Z\tilde{\beta})' (\underline{y} - Z\tilde{\beta})} \right]^{1/2}$$

donde $\tilde{\beta} = (Z'Z)^{-1} [Z'\underline{y} + C'\underline{\mu}]$

La región crítica queda determinada por

$$P_{H_0} [\lambda < k] = \alpha$$

donde P_{H_0} es la probabilidad debida a H_0 y K es una constante. Pedir que $\lambda < K$ es lo mismo que pedir que

$$\lambda^{2/n} = \frac{SC_{EMC}}{SC_{EMR}} = \frac{SC_{EMC}}{SC_{EMC} + SC_{H_0}} = \frac{1}{1 + \frac{SC_{H_0}}{SC_{EMC}}}$$

sea menor que K' . Donde

$$SC_{EMC} = (\underline{Y} - Z\hat{\beta})' (\underline{Y} - Z\hat{\beta})$$

$$SC_{EMR} = (\underline{Y} - Z\tilde{\beta})' (\underline{Y} - Z\tilde{\beta})$$

$$SC_{H_0} = (\underline{Y} - C\hat{\beta})' [C(Z'Z)^{-1}C'] (\underline{Y} - C\hat{\beta})$$

es decir

$$\lambda^{2/n} = \frac{1}{1 + \frac{SC_{H_0}}{SC_{EMC}}} < K' \quad \text{y esto equivale}$$

a que SC_{H_0} / SC_{EMC} sea grande, esto es, mayor que una K'' , por lo que H_0 se rechaza si:

$$\frac{SC_{H_0}}{SC_{EMC}} > K''$$

Se sabe

$$\frac{SC_{H_0}}{\sigma^2} = \frac{(C\hat{\beta} - \underline{Y})' [C(Z'Z)^{-1}C']^{-1} (C\hat{\beta} - \underline{Y})}{\sigma^2} \sim \chi^2(q)$$

y

$$\frac{SC_{EMC}}{\sigma^2} = \frac{(n-p) S^2}{\sigma^2} \sim \chi^2(n-p)$$

y como $C\hat{\beta}$ y S^2 son independientes ($\hat{\beta}$ y S^2 lo son)

$$\frac{\frac{SC_{Ho}}{q \sigma^2}}{\frac{SC_{EMC}}{(n-p) \sigma^2}} \sim F_{(q, n-p)}$$

quedando la región de rechazo de la siguiente manera:

$$\frac{(n-p) SC_{Ho}}{q SC_{EMC}} > F_{q, n-p}^{\alpha}$$

donde $F_{q, n-p}^{\alpha}$ es un valor tal que

$$P_{Ho} \left[\frac{(n-p) SC_{Ho}}{q SC_{EMC}} > F_{q, n-p}^{\alpha} \right] = \alpha$$

Todos los resultados importantes necesarios para --
realizar una prueba de hipótesis quedan resumidos en una --
tabla conocida con el nombre de "Tabla de Análisis de Va-
rianza" y es como la siguiente:

(La tabla de Análisis de Varianza aparece en la --
página 28-b.)

FV	GL	SC	CM	E(CM)	F
$H_0: C\beta = \delta$	q	SC_{H_0}	SC_{H_0} / q	$\sigma^2 + \frac{(C\beta - \delta)' [C(R^2Z)^{-1}C']^{-1} (C\beta - \delta)}{\sigma^2}$	CM_{H_0} / CM_{EMC}
ERROR	n-p	SC_{EMC}	$SC_{EMC} / n-p = CM_{EMC}$	σ^2	
TOTAL	n	$y'y$			



BIBLIOTECA
INSTITUTO DE ECOLOGIA
UNAM

donde FV es fuente de variación, GL grados de libertad, -
SC suma de cuadrados, CM cuadrados medios, E(CM) esperan-
za del cuadrado medio y F es la estadística de la prueba.

Es un resultado conocido que la suma de cuadrados --
del error representa la variación debida a la varianza --
del error y la variación debida al modelo. Debido a --
ésto y con objeto de conocer si la variación debida al --
modelo es grande y por lo tanto erróneo, es conveniente -
estimar la variación del error de alguna forma que no in-
volucre al modelo para separar estos efectos. Ahora --
bien, con tal motivo, es necesario que existan repeticio-
nes, pues de acuerdo a Draper y Smith (1966) se tiene:

$$SC_{\text{ERROR}} = SC_{\text{CARENCIA DE AJUSTE}} + SC_{\text{ERROR PURO}}$$

y

$$GL_{\text{ERROR}} = GL_{\text{CARENCIA DE AJUSTE}} + GL_{\text{ERROR PURO}}$$

donde la $SC_{\text{CARENCIA DE AJUSTE}}$ es la suma de cuadrados que representa la variación debida al modelo. Como esta variación depende del modelo, puede encontrarse otro conmenor $SC_{\text{CARENCIA DE AJUSTE}}$. Para saber si esta suma de cuadrados es satisfactoria se compara con $SC_{\text{ERROR PURO}}$. Esta última suma de cuadrados representa la variación involucrada exclusivamente por la varianza del error (σ^2).

Como se vió en intervalos de confianza, para carencia de ajuste, se tiene que:

$$y \quad SC_{\text{ERROR PURO}} = \sum_{i=1}^K \sum_{v=1}^{n_i} (y_{iv} - \bar{y}_i)^2$$
$$GL_{\text{ERROR PURO}} = \sum_{i=1}^K n_i - K$$

De aquí se obtiene que

$$CM_{EP} = SC_{EP} / GL_{EP}$$

donde EP es error puro y CM_{EP} es un estimador de la varianza del error. Se puede proceder ahora a calcular por diferencia SC_{CA} (CARENCIA DE AJUSTE), GL_{CA} y por lo tanto CM_{CA} , ya que se conoce SC_{ERROR} , SC_{EP} y GL_{ERROR} , GL_{EP} .

Ya habiendo obtenido GL_{CA} , SC_{CA} y CM_{CA} , se puede formar otra F calculada, la cual está dada por - -

(Searle (1971)).

$$F = \frac{CM_{CA}}{CM_{EP}}$$

y se utilizará para probar la hipótesis de que el modelo presenta carencia de ajuste; rechazándose dicha hipótesis si:

$$F < F^{\alpha}_{(GL_{CA}, GL_{EP})}$$

I.5 Análisis de Residuales

Para mejorar el análisis de datos habría 2 caminos, -tales son

- a) Mejorar las técnicas del Análisis de Regresión
- b) Usar mejor las técnicas de Análisis de Regresión y seguir más allá del sitio donde se detienen las técnicas convencionales.

Las dos alternativas de b son las que se tratan por medio del Análisis de Residuales, donde los residuos se definen como:

$$e_i = y_i (\text{valor observado}) - \hat{y}_i (\text{valor ajustado})$$

Para la primera alternativa se analizan los residuos con el objeto de aprender cualquier cosa de interés que -

ellos contengan. Para la segunda, se deben analizar los residuales obtenidos a partir de una primera aplicación de las técnicas de Regresión, con el objeto de determinar como alterar el modelo para una segunda aplicación de dichas técnicas, por medio de la cual obtener mejores resultados, lo cual resulta un proceso iterativo.

Estas mejoras, en una segunda aplicación, pueden lograrse mediante:

- i) Cambiando la forma del modelo o incluyendo nuevas variables (interacciones o términos de mayor orden, etc.)
- ii) Cambiando la forma de expresión de las observaciones y aplicando el análisis original a las nuevas variables (esto es, hacer uso de $F(\underline{Y})$ en vez de \underline{Y} en el análisis).
- iii) Discriminando entre las observaciones individuales, tal vez mediante la asignación de distintas ponderaciones, tal vez rechazando ciertas observaciones, ya porque se sospeche que se deban a un error (medición, etc.) o ya porque se desee analizarlas por separado; o tal vez sustituyendo algunos valores observados por valores modificados -- antes de aplicar un análisis de la forma original.

A partir de la definición de los residuales como diferencia entre lo observado y lo que se esperaría, puede verse entonces que los residuos son las cantidades que la ecuación de regresión no ha sido capaz de explicar. En consecuencia se puede considerar a los residuales como los errores estimados.

Para llevar a cabo el Análisis de Regresión se han hecho ciertas suposiciones acerca de la distribución de los errores, entonces si el modelo ajustado es correcto, los residuales obtenidos deben tener ciertas tendencias a satisfacer dichas suposiciones o al menos no deben contradecirlas. Esta es la idea central del Análisis de Residuales, a partir de la cual se puede concluir que:

- 1) Existe evidencia en contra de las suposiciones
- o
- 2) No existe evidencia en contra.

Nótese que 2 no afirma que las suposiciones son correctas, sino que en base a la información disponible no se puede afirmar que sean incorrectas.

El Análisis de Residuales es usualmente llevado a cabo de dos maneras: gráficamente o analíticamente.

Los métodos gráficos, muy fáciles de llevar a cabo, son los siguientes: Graficar e_i

- 1) Respecto al tiempo de ocurrencia.
- 2) Contra los valores ajustados \hat{y}_i .
- 3) Contra las variables independientes

$$x_{i1}, x_{i2}, \dots, x_{ir} \quad i = \overline{1, n}$$

- 4) Contra las variables Z 's (X 's transformadas)

$$z_{i1}, z_{i2}, \dots, z_{ip} \quad i = \overline{1, n}$$

Al realizar alguna(s) de éstas gráficas, se podrán notar distintas formas de violaciones a las hipótesis que pueden observarse. Si el modelo es adecuado se espera que los residuos se encuentren contenidos dentro de una banda horizontal. Sin embargo, puede ocurrir que las regiones que contienen a los residuos tengan otras formas, indicando por consiguiente alguno de los casos que a continuación se hacen mención:

- i) La varianza no es constante y crece, indicando -- que debería llevarse a cabo un ajuste por mínimos cuadrados ponderados, o debería de hacerse una -- transformación de las observaciones.
- ii) Un término lineal debió haberse incluido en el -- modelo.
- iii) Términos cuadráticos debieron haberse incluido.

- iv) Error en el Análisis. Puede ser causado por omitir erróneamente un término constante en el modelo.
- v) Modelo no adecuado. Se requieren términos extras en el modelo (cuadrados, productos-cruzados, etc.)
- vi) Etc.

Una discusión amplia sobre este punto se encuentra en Draper y Smith (1966).

Entre los métodos analíticos usados para el estudio sobre la suposición de la no covarianza de los errores, figuran, entre otros; el examen de las corridas, la estadística de Von-Neumann y la estadística de Durbin-Watson. Todas ellas basadas en los residuales y en el orden de ocurrencia respecto al tiempo.

Si los errores cumplen con la suposición de covarianza cero entonces tendrán autocorrelación cero (esto es, el coeficiente de autocorrelación entre el error al tiempo t y el error al tiempo $t-1$ será cero). Es por esto que en las pruebas de Durbin-Watson y Von-Neumann la prueba es sobre la autocorrelación y no sobre la covarianza.

El examen de las corridas es una técnica utilizada cuando se conoce el orden en el tiempo en que aparecieron

las observaciones de las cuales provienen los residuos. En ocasiones es notable que grupos de residuos positivos y negativos ocurran en lo que podría considerarse un patrón poco usual. Por ejemplo si de treinta residuos aparecen primero 10 positivos y enseguida 20 negativos podría sospecharse que una variable no considerada, pero con influencia en las observaciones, tuvo un fuerte cambio en sus niveles entre las observaciones décima y undécima. Se tratará entonces de asignar una causa a éste comportamiento. Este método es de gran ayuda con tal de efectuar decisiones basadas en la anormalidad de las corridas.

Supongase que se tiene la siguiente sucesión de signos de los residuos

$$(++)(-)(+)(----)(++)(-)(+++)$$

Entonces se tienen 14 signos en total, con 8 de ellos positivos y 6 negativos; y 7 corridas (paréntesis). Entonces cabe preguntarse si el arreglo (de signos) observado es extremo o no, es decir, aleatorio o no. Esta decisión se toma considerando la probabilidad de tener u corridas cuando se tienen n_1 valores positivos y n_2 negativos. Si ésta probabilidad es pequeña comparada con el

nivel de significancia, se considera que el arreglo de --
signos no ocurrió de manera aleatoria. En el anexo 4 --
aparecen tabulados los valores de dichas probabilidades --
para casos en que $n_1 \leq n_2 \leq 10$ y $\alpha = .05$. (cuando $n_1 > n_2$
intercambie n_1 y n_2).

Cuando $n_1 > 10$ y/o $n_2 > 10$ no se requieren --
los valores exactos de éstas probabilidades, ya que la --
distribución real puede ser aproximada con bastante pre-
cisión por una distribución normal. Sean

$$\mu = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$\sigma^2 = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}$$

Que son la media y la varianza reales de la distribución-
discreta de U . Entonces

$$Z = \frac{(u - \mu + 1/2)}{\sigma} \sim N(0,1)$$

donde $1/2$ es una corrección por continuidad.

La regla de decisión es la siguiente:

Si $n_1 \leq n_2 \leq 10$, se encuentra $P(u \text{ observado} \leq u)$ directamente con el uso de la tabla del anexo 4; y se rechaza la hipótesis sobre aleatoriedad de los residuales - si dicha probabilidad es menor o igual que α (.05 en este caso).

Si n_1 y/o $n_2 > 10$, se calcula Z ; habiendo calculado antes μ y r^2 , se procede entonces a calcular - - -
 $\int_{-\infty}^Z f(v) dv$ donde $f(v)$ es la función de densidad de una normal estandar. Si dicha probabilidad es mayor que α , no se rechaza la hipótesis de aleatoriedad de los residuales.

Para mayor detalle sobre el método de corridas ver Conover (Wold-Wolfowitz, pp-350, (1971)) y Draper y Smith (1966).

Las pruebas de hipótesis que se pueden probar con las estadísticas de Von-Neumann y Durbin-Watson son las siguientes:

- i) H_0 : autocorrelación cero.
 H_a : autocorrelación diferente de cero.
- ii) H_0 : autocorrelación cero.
 H_a : autocorrelación positiva.

iii) H_0 : autocorrelación cero.

H_a : autocorrelación negativa.

Se procederá primero a explicar la estadística de --
Von-Neumann.

Esta estadística está definida como el cociente de --
la diferencia sucesivas de las medias cuadráticas a la --
varianza, esto es:

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

donde e_t representa el t -ésimo residual y $\bar{e} = \sum_{t=1}^n e_t / n$.
Si $n \rightarrow \infty$, δ^2 / S^2 se distribuye normalmente con

$$E\left(\frac{\delta^2}{S^2}\right) = \frac{2n}{n-1}$$

y

$$V\left(\frac{\delta^2}{S^2}\right) = \frac{4n^2(n-2)}{(n+1)(n-1)^3}$$

La región crítica para cada una de las hipótesis --
está determinada por:

i) Si $\frac{\delta^2}{S^2} < K$ ó $\frac{\delta^2}{S^2} > K'$ se rechaza H_0 , --
donde si $4 \leq n \leq 60$ entonces K y K' son los --
valores tabulados en la tabla de Von-Neumann, la-
cual aparece en el anexo 5 (tabla 1). Como K
y K' son los cuantiles α_1 y $1-\alpha_2$ ($\alpha_1, \alpha_2 = .01,$
.05) de la distribución de $\frac{\delta^2}{S^2}$ (Hart (1942)),
el nivel de significancia de la prueba es $\alpha_1 + \alpha_2$ --
(= .02, .06, .1).

Si $n > 60$ se utiliza la aproximación normal y --
 K y K' son los cuantiles de orden $\alpha/2$ y --
 $1 - \alpha/2$ de la distribución normal, y la prueba --
tiene un nivel de significancia α .

ii) Si $\frac{\delta^2}{S^2} < K$ (donde K se determina de tal forma --
que el nivel de significancia sea α) se recha-
za H_0 y se concluye que existe autocorrelación-
positiva.

iii) Si $\frac{\delta^2}{S^2} > K'$ se acepta la hipótesis de autoco-
rrelación negativa.

A continuación se presenta la estadística de Durbin-
Watson la cual está dada por

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Esta estadística se relaciona con la de Von-Neumann de la siguiente manera:

$$\frac{S^2}{s^2} = \frac{n}{n-1} d$$

Durbin y Watson investigaron la distribución muestral de la estadística d . Es intuitivamente claro que para una serie de autocorrelación positiva, las primeras diferencias tenderán a ser pequeñas en valor absoluto comparadas con los valores absolutos de e , mientras que para una serie de autocorrelación negativa frecuentemente son más grande que el valor de e , de modo que d tendería a ser chico para series positivamente autocorrelacionadas, grande para series negativamente autocorrelacionadas y en algún sitio entre ellas para series aleatorias.

La distribución muestral de d depende de los valores de la matriz Z , de modo tal que solo fue posible para Durbin y Watson el establecer límite superior (d_u) e inferior (d_l) para los niveles de significancia de d . Estas son las pruebas de hipótesis de autocorrela

ción cero contra la hipótesis alternativa de autocorrelación positiva. Los valores d_U y d_L se presentan en la tabla 2, anexo 5. Es importante el enfatizar que esta prueba fue derivada para Z no estocástica.

Si $n \rightarrow \infty$ las estadísticas Durbin y Watson y Von-Neumann coinciden, y se utiliza la teoría expuesta anteriormente.

Si $15 \leq n \leq 100$ se utilizan las tablas obtenidas por Durbin-Watson.

Las regiones críticas para probar las hipótesis (i, ii, iii) establecidas anteriormente son respectivamente.

i) Rechazar H_0 si $d < d_L$ o $d > 4 - d_L$,

No rechazar H_0 si $d_U < d < 4 - d_U$,

La prueba es inconclusa si:

$$4 - d_U \leq d \leq 4 - d_L \quad \text{ó si}$$

$$d_L \leq d \leq d_U$$

ii) Rechazar H_0 si:

$$d < d_L$$

No rechazar H_0 si:

$$d > d_U$$

La prueba es inconclusa si:

$$d_L \leq d \leq d_U$$

iii) Para probar autocorrelación negativa se calcula $4-d$ y se procede como en ii).

Para mayor detalle de estas pruebas Von-Neumann y -- Durbin-Watson (op. cit.), así como para saber como proceder en el caso en que la prueba de Durbin-Watson sea inconclusa, ver Johnston (1972).

Las pruebas de Durbin-Watson y Von-Neumann sirven -- para probar autocorrelación, si $n < 15$ ó $n > 60$ entonces está clara la elección entre ellas; en el primer caso Von-Neumann y en el segundo i) $60 < n \leq 100$ Durbin-Watson y ii) Si $n > 100$ La aproximación normal.

Si $15 \leq n \leq 60$ las dos pruebas podrían usarse.

Se recomienda usar Durbin-Watson ya que resulta una prueba mas fina, en el sentido de que detecta una región en la que no se puede concluir, cosa que Von-Neumann no hace.

CAPITULO II

GENERALIDADES DEL PROGRAMA SORA.

En este capítulo se trata de dar brevemente una idea general de algunas partes del programa. Entre estas - - están: transformaciones, lectura y simulación de datos, y otras que son de interés general para el usuario. Además se mencionan algunos consejos útiles para el uso del programa.

II.1 Transformaciones

En innumerables casos se tiene que las variables observadas no son precisamente las variables con las cuales se desea plantear la regresión, sino que las segundas son transformación de las primeras; es decir, es posible - - transformar los datos observados de modo tal que se exprese Y en términos de variables que no necesariamente -- fueron observadas.

Si se define a X como la matriz de las variables -- observadas de $n \times r$, donde r no guarda relación de - orden con p , el caso mas simple se obtiene cuando $r=p$, y además la función de transformación general es la idéntica, es decir

$$Z_{ij} = X_{ij} \quad \forall i, j \quad i = \overline{1, n}, \quad j = \overline{1, p}$$

en términos matriciales se tiene

$$Z = X$$

Si se supone que $r > p$, se obtiene el caso en donde se observan mas variables que aquellas con las que se plantea la regresión; a manera de ejemplo se considera $r=5$ y $p=3$. En este caso se tiene que cada columna de Z está relacionada mediante una transformación con las columnas de X , es decir

$$Z_{ij} = T_j (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}) \quad i = \overline{1, n}, \quad j = \overline{1, 3}$$

donde T_j indica la j -ésima transformación. Esta T_j podría, o no, variar para cada caso, con la única restricción que la función debe de estar reconocida en lenguaje-ALGOL. Como un ejemplo para la primera transformación se podría tener

$$Z_{i1} = \text{Seno} (X_{i1} \cdot X_{i2})$$

Este tema se tratará con mas detalle en el capítulo siguiente.

II.2 Leer_y/o_Simular

El usuario cuenta con la posibilidad de leer, leer y simular, o nada más simular los datos con los cuales desea trabajar a lo largo de la ejecución del programa.

Cuando se simulan datos se hace uso del intrínseco "RANDOM", esta función da como resultado un número aleatorio con distribución uniforme en el intervalo $[0,1)$. Dicho número aleatorio se encuentra utilizando el método-congruencial mixto, el cual tiene la siguiente forma:

$$|N(t)| = (A * |N(t-1)| + 116177073375) \text{ MOD } 2^{39}$$

donde A es dado de la siguiente manera:

$$A = 152587890725 \quad \text{si } N \text{ es negativo}$$

ó

$$A = 277626315293 \quad \text{si } N \text{ es positivo.}$$

Al valor $N(0)$ se le conoce como semilla, y es dado por el usuario. Valores $N(0)$ que hacen que el método se comporte "mejor" deben de ser números nones cerca de 524288 ó 1048576 ó 2097152. Se habla de "mejor" en términos de $N(0)$ para aquellos valores que hacen: máxima-

la velocidad de computación, que el período sea suficientemente largo para finalidades practicas y que la sucesión obtenida satisfaga pruebas de independencia en términos estadísticos.

II.3 Capacidad, Mensajes, Resultados y Distribuciones

Capacidad

No existe restricción respecto a la cantidad de memoria utilizable, pero debido a la carga de trabajo en el día, los problemas que utilizan una gran cantidad de memoria (posiblemente por las dimensiones de Z , X) es mejor correrlos en la noche. Esto es sobre todo para no recargar mas a la máquina en términos de memoria, como también para no dilatarse en la solución del problema.

Mensajes

Cada vez que el programa haga una pregunta, necesite algún dato, dé principio o final a algún proceso, aparecerá un mensaje en la terminal; con tal de ser mas legible la historia que aparece en ésta, se les antepone una arroba (@).

Lectura

Todas las lecturas realizadas a lo largo de la eje-

cución del programa se hacen en formato libre, es decir, los valores deben de estar separados por comas. Existen filtros de información para la gran cantidad de lecturas, ya sean alfabéticas, alfanuméricas o numéricas, según sea el caso. Resultando estos de gran utilidad y flexibilidad.

Resultados

Debido a la utilidad que representa el tener un registro de lo que se hace por terminal durante la ejecución del programa, la mayoría de los resultados se obtienen tanto en la terminal como en impresora. En caso que ésto no suceda, se preguntará al usuario por donde desea los resultados, en este momento se puede contestar TERM o IMPRE para indicar si se desean por TERMINAL o IMPRESORA respectivamente.

Distribuciones

Para la teoría del Análisis de Regresión es de gran utilidad el uso de cuantiles de algunas funciones de distribución, con el objeto de obtener explícitamente los intervalos de confianza o determinar las regiones críticas para las pruebas de hipótesis que en este programa se cal

culan. De aquí la necesidad de tener dichos valores. --
Comunmente la forma de obtención es por tablas, esta op-
ción fue eliminada debido a la gran cantidad de memoria -
utilizable en caso de incluirlas como parte interna del -
programa; es por ésto que el programa tiene subrutinas en
cargadas de calcular dichos cuantiles, algunos con res- -
tricciones acerca de los valores y otros con mas flexibi-
lidad que las tablas. A continuación se pasa a mencio--
nar algunos detalles con respecto a ellos.

En cuanto a la Normal, X^2 y t no existe restric-
ción alguna, sin embargo con respecto a la F se tienen-
cotas inferiores y superiores, la restricción es que la -
confiabilidad esté entre .7 y .995 (inclusive). Si se -
tiene en cuenta que esta es la región de más uso, resulta
que no se pierde generalidad. Estas cotas se fijaron --
sobre todo porque la F se dispersa mucho dependiendo de
los grados de libertad que se tengan.

En el uso de las subrutinas antes mencionadas, se --
utiliza mucho menos memoria que si se hubiera optado por-
otra solución, además de ser de gran rapidez y por tanto-
un atractivo mas para los usuarios.

CAPITULO III

TRANSFORMACIONES DE LOS DATOS EN EL SORA

III.1 Introducción

Como se mencionó en II.1, las transformaciones son de gran utilidad debido a la posibilidad que dan de plantear regresiones con datos que han sido transformados de otros que son conocidos.

Entre sus principales usos figura la oportunidad de plantear regresiones no sólo como combinaciones lineales de las variables dependientes; es decir, se cuenta con la posibilidad de plantear regresiones como combinaciones no lineales en las variables dependientes. Las transformaciones también se utilizan en muy variados casos como cuando se quiere plantear una regresión en variables tales como x_1 , x_2 , $\text{SENO}(x_1+x_2)$; con su uso sólo es necesario dar al programa x_1 , x_2 y efectuar tres transformaciones. Se plantearía entonces una regresión con tres variables z_1 , z_2 , z_3 donde z_1 , z_2 serían iguales a x_1 y x_2 respectivamente y z_3 igual al $\text{SENO}(x_1+x_2)$. Como podrá observarse, con esta facilidad del programa, se evita calcular la(s) transformación(es) fuera del programa e introducirla(s) como una nueva variable, obteniendo una gran flexibilidad operativa.

III.2 Modo de Uso

Al iniciar la ejecución del programa este pregunta -- por el número de observaciones, columnas de la matriz X y número de betas o número de variables involucradas en el modelo. A la matriz con los datos originales se le conoce como X y a la matriz del modelo como Z (matriz de los datos transformados), la dimensión de cada -- una de ellas es $n \times r$ y $n \times p$ respectivamente. Entonces se tendrá que dar como datos en esta lectura los -- valores de n , r y p , dichos valores no guardan relación de orden alguna ($n \geq p$).

En caso de ser r y p iguales se preguntará al -- usuario si las matrices X y Z son iguales; si la -- respuesta es afirmativa no se pedirán las transformaciones de X . Claro está, no se excluye la posibilidad -- de dar todas las transformaciones, que en este caso serían la identidad. Se efectúa la igualdad como parte -- interna del programa con el fin de optimizar tiempo de -- ejecución; de no ser así tendrían que hacerse otras consideraciones como se verá mas adelante.

Se supone entonces que Z y X no son iguales, -- por lo que se necesitan p transformaciones, las cuales

van a relacionar cada columna de Z con las columnas de X . A estas p transformaciones se les reservará un espacio estipulado por el usuario en el momento en que el programa pregunte el número de tarjetas necesarias para llevar a cabo la definición de las transformaciones; este número debe ser mayor o igual a 1 y menor o igual que p ; en caso de no ser así el programa regresará a hacer la pregunta. Se entiende por tarjeta a un renglón de 72 posiciones en la terminal, se considera por lo tanto que el máximo de posiciones para definir una transformación es de 72, resultando éste un espacio suficientemente grande. Lo más conveniente es asignar un renglón por transformación, pero se pueden definir varias en uno solo, también se puede hacer uso de instrucciones del lenguaje ALGOL; la de mas utilidad (FOR) se explicará posteriormente.

Antes de continuar se deben hacer algunas consideraciones. En esta parte del programa se usa que las matrices Z y X esten definidas a partir de 0,0, es decir, si se supone que i representan renglones y j columnas, en vez de rotular los renglones $i = \overline{1, n}$ se rotularán $i = \overline{0, (n-1)}$, cosa similar sucede con las columnas. La razón de lo antes mencionado se debe a que se trabaja

con la parte interna del programa. Se tiene entonces lo siguiente:

$$Z_{ij} = T_j (X_{i0}, X_{i1}, \dots, X_{i(r-1)}) \quad \begin{matrix} i = \overline{0, (n-1)}, \\ j = \overline{0, (p-1)} \end{matrix}$$

donde Z_{ij} es el elemento del renglón i , columna j de Z y X_{ik} es el elemento del renglón i , columna k de X , el rango de k es $\overline{0, (r-1)}$; T_j indica la j -ésima transformación y puede ser cualquier función -- definida en lenguaje ALGOL, además de poder ser cualquiera de las básicas. En el anexo 2 se dan todas las funciones utilizables para T_j . El modo de expresar las transformaciones se hace utilizando ALGOL, es decir las propiedades de asignación, matrices, etc., se hacen en este lenguaje. A manera de ejemplo se considera el siguiente modelo

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \cdot X_2) + \epsilon$$

este modelo puede ser expresado en la siguiente forma:

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \epsilon$$

donde

$$Z_1 = X_1$$

$$Z_2 = X_2$$

$$Z_3 = X_1 \cdot X_2$$

en este caso se tiene $r=2$, $p=3$ y se supone que el número de tarjetas es 3, las transformaciones que se tienen son las siguientes:

$$Z_{i0} = t_0 (X_{i0}, X_{i1})$$

$$Z_{i1} = t_1 (X_{i0}, X_{i1})$$

$$Z_{i2} = t_2 (X_{i0}, X_{i1})$$

de manera mas explícita:

$$Z_{i0} = X_{i0}$$

$$Z_{i1} = X_{i1}$$

$$Z_{i2} = X_{i0} \cdot X_{i1}$$

en lenguaje ALGOL y en la forma como deberá introducirse al programa es:

$$Z[I,0] := X[I,0];$$

$$Z[I,1] := X[I,1]; \quad (2)$$

$$Z[I,2] := X[I,0] * X[I,1];$$

A continuación el programa necesitará de las transformaciones en lenguaje ALGOL, estas se darán tomando en cuenta el número de tarjetas que se reservaron para uso; si se necesitaran menos de las señaladas se podría escribir en las tarjetas no utilizadas el símbolo % o dejarla(s) en blanco, en caso contrario, se tendrían que agrupar transformaciones en una misma tarjeta. Continuando con el ejemplo, se tendría que teclear (2) cada una por renglón, esperando hasta que aparezca un mensaje en el que pide cada una de las tarjetas. En el momento en que este proceso llega a su fin aparecerá un mensaje especificando el principio de la compilación; junto con esto vendrá una señal del tipo "#BOT" (el significado de BOT es "BEGIN OF TASK") lo cual vendrá a indicar que se comenzó a compilar las transformaciones. La señal "#BOT" aparecerá acompañada de información irrelevante para el usuario. Este mensaje da origen a dos posibilidades: compilación correcta ó error en el proceso de compilación. Si se supone que compiló bien, esto traerá como consecuencia una señal del tipo "#EOT" (donde EOT significa "END OF TASK", al igual que el BOT, esta señal aparecerá acompañada de información irrelevante), también aparecerá un mensaje donde se informa de la compilación correcta. Se considera ahora la segunda posibilidad, aparecerá enton-

ces una señal de error especificando los errores (uno o --
mas) por los cuales no compiló correctamente, así como --
las líneas posteriores o las líneas donde ocurrieron es--
tos. Por tal motivo aparecerá el siguiente mensaje - -
"#SNTX", lo cual indicará que se cometió(eron) error(es)-
de sintaxis. A continuación de esto aparecerán las ins--
trucciones necesarias para poder corregir la sintaxis. -
Esto es de gran utilidad si se tiene en cuenta que no se--
pierde el programa en el caso en que se hayan cometido --
errores de sintaxis en las transformaciones (esto es para
facilidad del usuario). Estas instrucciones son:

- 1.- Mandar cual línea se va a repetir
- 2.- Mandar la línea.

Antes de esto, el programa preguntará cuantas tarjetas se
desean corregir. Estas tienen un límite inferior de 1 y
superior o igual al número de tarjetas usadas. Habiéndo
le contestado esto, se le tendrá que dar las nuevas lí- -
neas según las instrucciones especificadas. Continuando-
con el ejemplo, si se desea corregir sólo una tarjeta, --
debido a que hubo error de sintaxis supóngase en la se- -
gunda tarjeta, se procederá de la siguiente forma: se da- -
rá el dígito 1 cuando pregunte el número de tarjetas a --
corregir, luego al preguntar el número de la tarjeta se -

le dará el 2, y mas tarde, no sin antes esperar a que pida la tarjeta 2, se le dá:

$$Z[I,1] := X[I,1];$$

(La tarjeta supuestamente antes errónea, ahora correcta)

habiendo hecho ésto, comenzará a compilar una vez mas las transformaciones, mandando los mensajes que se mencionaron anteriormente. Este proceso de repite hasta que la compilación sea correcta.

Habiendo compilado sin error las transformaciones se está en capacidad de obtener la matriz Z , es decir, de efectuar las transformaciones; esto ocurrirá una vez que X haya sido leída o simulada. Se notará que este fenómeno ocurre en el momento en que aparezca "#BOT", lo cual denotará el principio de la ejecución de las transformaciones y "#BOT" el fin. En caso de haber error en la ejecución de éstas, no se tendrá la posibilidad de hacer nada mas, debido a que son errores no controlables, como por ejemplo logaritmo natural de cero.

En algunos casos aparecerán mensajes como "#AAAA - - (CLAV) REGRESION/AUXILIAR REMOVED ON PACK" y/o "#BBBB - - (CLAV) FCOMP REMOVED ON PACK", donde AAAA y BBBB son núme



BIBLIOTECA
INSTITUTO DE ECOLOGIA

ros de cuatro dígitos. Dichos mensajes indicaron que --
REGRESION/AUXILIAR y/o FCOMP han sido removidos de la --
clave del usuario, tales archivos son creados por el pro-
grama para ser usados en la compilación y ejecución de --
las transformaciones, por lo que si están presentes el --
programa procede a removerlos. Esto no implica error.

Si $Z=X$ no aparecieran los mensajes mencionados, --
que surgen cuando se hace uso de las transformaciones.

III.3 La Instrucción FOR

La instrucción FOR es una de las mas usadas del lenguaje ALGOL y además una de las mas fluidas con respecto a las formas que puede tomar. Su importancia reside en el hecho de que es de gran utilidad en las partes en donde son necesarias las iteraciones, es decir en procesos con repeticiones sobre variables. Ejemplos de esta instrucción son los siguientes:

```
FOR I:=0 DO
```

lo que indica es que hará la siguiente instrucción para cuando I toma el valor de cero.

```
FOR J:=1 STEP 1 UNTIL 255 DO
```

hará la próxima instrucción para cuando J sea igual a 1, y se incrementará de 1 en 1 hasta 255, es decir hará 255 veces lo que a continuación viene.

```
FOR K:= 0,1,2,5,7,37,56 DO
```

efectuará la instrucción siguiente para K= 0, 1, 2, 5, 7, 37, 56 .

Se pueden hacer combinaciones de éstas como la siguiente:

```
FOR L:=0 STEP 1 UNTIL 5,29,47 STEP 3 UNTIL  
58 DO
```

lo que equivale a

```
FOR L:=0,1,2,3,4,5,29,47,50,53,56 DO
```

es decir primero hace L igual a 0 se incrementa de 1 en 1 hasta 5, después L toma el valor de 29, después el 47 y lo incrementa de 3 en 3 hasta el 58; como el último incremento de 3 a partir de 47 no mayor que 58 es 56, --- hasta aquí llega.

Esto es de gran utilidad si se llegase a presentar - por ejemplo el siguiente caso:

Z[I,0] := X[I,0] + 5;

Z[I,1] := X[I,1] + 5;

: :

Z[I,12] := X[I,12] + 5;

Z[I,13] := X[I,0] * X[I,1];

esto podría ser reducido con el uso del FOR de la siguiente manera:

```
FOR J:=0 STEP 1 UNTIL 12 DO
```

```
  Z[I,J] := X[I,J] + 5;
```

```
  Z[I,13] := X[I,0] * X[I,1];
```

representa esta forma un ahorro de 11 tarjetas, debido a que solo se usan 3 tarjetas, y de otra manera 14.

Cualquiera de las otras formas de la instrucción FOR podrían ser usadas tomando en consideración que sólo son válidas iteraciones sobre **J** y ésta debe representar columnas.

Se pueden hacer tantas iteraciones sobre J como se deseen, ejemplo de ésto se tienen cuando las variables pares en Z son iguales a las pares en X y las impares en Z son iguales a las impares en X mas 5, en este caso se podrán usar 2 FOR.

Para mayor detalle y conocimientos de esta instrucción consultar Burroughs, Languaje Reference Manual.

III.4 Ejemplos

En esta sección se van a mostrar dos ejemplos para facilidad de comprensión de las transformaciones, en el primer caso se tiene $r=p$, pero $Z \neq X$, en el segundo $r > p$. En los ejemplos considerados aparecen errores de compilación, se incurrieron en estos para mostrar de manera mas detallada todo lo que en este capítulo se trata.

Ejemplo 1

Se supone $r=p=7$, con $Z \neq X$, se va a mostrar en este ejemplo también el uso de la instrucción FOR. La relación entre Z y X es la siguiente:

$$Z_0 = e^{x_0}$$

$$Z_1 = \text{Seno}(x_1)$$

$$Z_2 = e^{x_2}$$

$$Z_3 = \text{Seno}(x_3)$$

$$Z_4 = e^{x_4}$$

$$Z_5 = x_0 + x_3 + x_5 + x_6$$

$$Z_6 = \text{Seno}(x_6)$$

Las transformaciones que se tienen son:

$$Z_{i0} = t_0(x_{i0}, x_{i1}, \dots, x_{i6})$$

$$Z_{i1} = t_1(x_{i0}, x_{i1}, \dots, x_{i6})$$

⋮

$$Z_{i6} = t_6(x_{i0}, x_{i1}, \dots, x_{i6})$$

de manera mas explícita se tiene:

$$Z_{i0} = e^{x_{i0}}$$

$$Z_{i1} = \text{Seno}(x_{i1})$$

$$Z_{i2} = e^{x_{i2}}$$

$$Z_{i3} = \text{Seno}(X_{i3})$$

$$Z_{i4} = e^{X_{i4}}$$

$$Z_{i5} = X_{i0} + X_{i3} + X_{i5} + X_{i6}$$

$$Z_{i6} = \text{Seno}(X_{i6})$$

en términos de lenguaje ALGOL se tiene:

$$Z[I,0] := \text{EXP}(X[I,0]);$$

$$Z[I,1] := \text{SIN}(X[I,1]);$$

$$Z[I,2] := \text{EXP}(X[I,2]);$$

$$Z[I,3] := \text{SIN}(X[I,3]);$$

$$Z[I,4] := \text{EXP}(X[I,4]);$$

$$Z[I,5] := X[I,0] + X[I,3] + X[I,5] + X[I,6];$$

$$Z[I,6] := \text{SIN}(X[I,6]);$$

con la instrucción FOR se tendría:

```
FOR J:= 0,2,4 DO
    Z[I,J]:= EXP(X[I,J]);
FOR J:= 1,3,6 DO
    Z[I,J]:= SIN(X[I,J]);
Z[I,5]:= X[I,0]+X[I,3]+X[I,5]+X[I,6];
```

En las siguientes 2 páginas aparecen 2 formas ilustrando como se pueden introducir las transformaciones al programa, una usando la instrucción FOR y otra sin usarla. Como se notará en este caso se necesitan de menos tarjetas cuando se usa la instrucción FOR. Notar que aparecen errores de sintaxis y la forma en que se corrigen.

R SORA

#RUNNING 9755

@ DAME TITULO DEL PROBLEMA

#?

EJEMPLO R=P, Z DIFERENTE DE X, SIN FOR.

@ LA MATRIZ X ES LA MATRIZ DE DATOS

@ LA MATRIZ Z ES LA MATRIZ DE TRANSFORMACIONES

@ DAME EL NUMERO DE OBSERVACIONES

COLUMNAS DE LA MATRIZ X

NUMERO DE BETA-S EN EL MODELO

10,7,7

@ LAS MATRICES X Y Z SON IGUALES ?

NO

#9755 (RH85)REGRESION/AUXILIAR REMOVED ON PACK PK117.

@ AHORA ME TIENES QUE DAR :

LAS TRANSFORMACIONES Z= F(X-MATRIZ)

@ CUANTAS TARJETAS NECESITAS PARA DEFINIR LAS TRANSFORMACIONES ?

7

@ DAME LA TARJETA 1

Z(I,0)=EXP(X(I,0));

@ DAME LA TARJETA 2

Z(I,1)=SIN(X(I,1));

@ DAME LA TARJETA 3

Z(I,2)=EXP(X(I,2));

@ DAME LA TARJETA 4

Z(I,3)=SIN(X(I,3));

@ DAME LA TARJETA 5

Z(I,4)=EXP(X(I,4));

@ DAME LA TARJETA 6

Z(I,5)=X(I,0)+X(I,3)+X(I,5)+X(I,6);

@ DAME LA TARJETA 7

Z(I,6)=SIN(X(I,6));

@ COMIENZO A COMPILAR LAS TRANSFORMACIONES

#BOT 9756

#9756 (RH85)FCOMP REMOVED ON PACK PK117.

@ TU FUNCION COMPILO BIEN

@ VAS A DARME LA MATRIZ X Y/O Y

#EOT 9756

```

R SORA
#RUNNING 9023
@ DAME TITULO DEL PROBLEMA
#?
EJEMPLO R=P, Z DIFERENTE DE X
@ LA MATRIZ X ES LA MATRIZ DE DATOS
@ LA MATRIZ Z ES LA MATRIZ DE TRANSFORMACIONES
@ DAME EL NUMERO DE OBSERVACIONES
      COLUMNAS DE LA MATRIZ X
      NUMERO DE BETA-S EN EL MODELO

50,7,7
@ LAS MATRICES X Y Z SON IGUALES ?
NO
#9023 (RH85)REGRESION/AUXILIAR REMOVED ON PACK PK117.
@ AHORA ME TIENES QUE DAR :
      LAS TRASFORMACIONES Z= F(X-MATRIZ)
@ CUANTAS TARJETAS NECESITAS PARA DEFINIR LAS TRANSFORMACIONES ?
5
@ DAME LA TARJETA 1
FOR J:=0,2,4 DO
@ DAME LA TARJETA 2
  ZCI,JJ:=EXP(XCI,JJ);
@ DAME LA TARJETA 3
FOR J:=1,3,6 DO
@ DAME LA TARJETA 4
  ZCI,JJ:=SIN(XCI,JJ);
@ DAME LA TARJETA 5
ZCI,5:=XCI,0]+XCI,3]+XCI,5]+XCI,6]
@ COMIENZO A COMPILAR LAS TRANSFORMACIONES
#BOT 9025
ZCI,5:=XCI,0]+XCI,3]+XCI,5]+XCI,6]
      ERROR-RIGHT BRACKET EXPECTED. = X
#SNTX 9025

@ PARA CORREGIR SINTAXIS NECESITAS
      1. MANDAR CUAL LINEA VAS A REPETIR
      2. MANDAR LA LINEA

@ CUANTAS TARJETAS QUIERES CORREGIR ?
1
@ DAME EL NUMERO DE TARJETA
5
@ DAME LA TARJETA 5
ZCI,5J:=XCI,0]+XCI,3]+XCI,5]+XCI,6];
@ COMIENZO A COMPILAR LAS TRANSFORMACIONES
#BOT 9026
#9026 (RH85)FCOMP REMOVED ON PACK PK117.
@ TU FUNCION COMPILO BIEN
@ VAS A DARME LA MATRIZ X Y/O Y
#EOT 9026

```

En este ejemplo se incurre en un error en la tarjeta 5, a dicha tarjeta le falta: un paréntesis cuadrado después del primer 5, en vez de decir $X[16]$ debe decir -- $X[1,6]$ y le falta punto y coma (;) después del último paréntesis cuadrado. Se muestra cuales son los pasos a seguir para corregir dicho error.

Ejemplo 2

En este ejemplo se tiene que $r=5$ y $p=3$, por lo que como máximo se utilizan 3 tarjetas. También en este caso se incurrió en errores de compilación por el motivo señalado anteriormente.

La relación entre las columnas de Z y X son las siguientes:

$$Z_0 = X_0 + X_1^2 + X_2^3 + X_3^4 + X_4^5$$

$$Z_1 = X_0 \text{ MOD } X_1$$

$$Z_2 = \text{ENTIER}(X_2)$$

haciendo pasos similares a los del ejemplo anterior se --
obtiene:

$$Z[I,0] := X[I,0] + X[I,1]**2 + X[I,2]**3 + X[I,3]**4 \\ + X[I,4]**5;$$

$$Z[I,1] := X[I,0] \text{ MOD } X[I,1];$$

$$Z[I,2] := \text{ENTIER}(X[I,2]);$$

en el siguiente listado aparecen cada uno de los pasos --
efectuados para hacer estas transformaciones.

R SORA
#RUNNING 9764
@ DAME TITULO DEL PROBLEMA
#?

EJEMPLO R MAYOR QUE P.
@ LA MATRIZ X ES LA MATRIZ DE DATOS
@ LA MATRIZ Z ES LA MATRIZ DE TRANSFORMACIONES
@ DAME EL NUMERO DE OBSERVACIONES
COLUMNAS DE LA MATRIZ X
NUMERO DE BETA-S EN EL MODELO

10,5,3

#9764 (RH85)REGRESION/AUXILIAR REMOVED ON PACK PK117.

@ AHORA ME TIENES QUE DAR :

LAS TRANSFORMACIONES Z= F(X-MATRIZ)

@ CUANTAS TARJETAS NECESITAS PARA DEFINIR LAS TRANSFORMACIONES ?

3

@ DAME LA TARJETA 1

ZCII,0J:=XCI,0J+XCI,1J**2+XCI,2J**3+XCI,3J**4+XCI,4J**5;

@ DAME LA TARJETA 2

ZCII,1J:=XCI,0J MOD XCI,1J;

@ DAME LA TARJETA 3

ZCII,2J:=ENTIER(XCI,2J);

@ COMIENZO A COMPILAR LAS TRANSFORMACIONES

#BOT 9769

ZCII,0J:=XCI,0J+XCI,1J**2+XCI,2J**3+XCI,3J**4+XCI,4J**5;

ERROR-ARITHMETIC PRIMARY CANNOT START WITH THIS. = +

#SNTX 9769

@ PARA CORREGIR SINTAXIS NECESITAS

1. MANDAR CUAL LINEA VAS A REPETIR
2. MANDAR LA LINEA

@ CUANTAS TARJETAS QUIERES CORREGIR ?

1

@ DAME EL NUMERO DE TARJETA

1

@ DAME LA TARJETA 1

ZCII,0J:=XCI,0J+XCI,1J**2+XCI,2J**3+XCI,3J**4+XCI,4J**4

@ COMIENZO A COMPILAR LAS TRANSFORMACIONES

#BOT 9770

ZCII,1J:=XCI,0J MOD XCI,1J;

ERROR-SEMICOLON EXPECTED. = Z

#SNTX 9770

@ PARA CORREGIR SINTAXIS NECESITAS

1. MANDAR CUAL LINEA VAS A REPETIR
2. MANDAR LA LINEA

@ CUANTAS TARJETAS QUIERES CORREGIR ?

1

@ DAME EL NUMERO DE TARJETA

1

@ DAME LA TARJETA 1

ZCII,0J:=XCI,0J+XCI,1J**2+XCI,2J**3+XCI,3J**4+XCI,4J**5;

@ COMIENZO A COMPILAR LAS TRANSFORMACIONES

#BOT 9771

#9771 (RH85)FCOMP REMOVED ON PACK PK117.

#EOT 9771

@ TU FUNCION COMPILO BIEN

@ VAS A DARMELA MATRIZ X Y/O Y

CAPITULO IV

LECTURA DE DATOS

IV.1 Introducción

Para obtener la solución de cualquier problema de -- Regresión es necesario conocer numéricamente Z (Matriz del Modelo) y Y (vector de observaciones de la variable dependiente). Como se cuenta con el uso de las transformaciones resulta ser equivalente lo antes enunciado a conocer X (matriz de las variables observadas) y Y . -- Las formas de obtención de X y Y pueden variar dependiendo de si se dan los valores o se simulan; resultando por consiguiente tres alternativas:

- 1) Leer X, Y
- 2) Leer X , Simular Y
- 3) Simular X, Y

la posibilidad de leer Y y simular X no se consideró debido a que no es de interés desde el punto de vista -- didáctico.

El programa efectuará las preguntas necesarias con -- tal de saber la alternativa deseada. En caso de ser ésta alguna de las dos primeras, se preguntará si la lectura se va a efectuar por terminal, de no ser así los datos -- tendrán que estar en un archivo en disco, el cual debe -- cumplir con ciertos requisitos, tales como: el tamaño del

registro (MAXRECSIZE) debe ser imagen de tarjetas (o sea 14 palabras), no podrá tener el nombre "REGRESION/AUXILIAR" ni "FCOMP" debido a que son archivos usados por el programa para otros procesos y estar grabado en FORMATO LIBRE. Si los datos van a leerse de disco, se preguntará al usuario el nombre del archivo que contiene los datos, esta pregunta se repetirá hasta que el nombre dado sea un archivo de disco, que esté presente en la clave del usuario.

En este momento el programa está capacitado para iniciar las lecturas. Si se van a introducir los datos por terminal se especificará: qué variables se van a leer, el número de valores que serán leídos y la observación que se leerá; en caso de estar ejecutándose la lectura por disco, no se especificará lo antes mencionado pero deberá tenerse en cuenta en el momento de la creación del archivo que contiene los datos.

Ya terminada las lecturas se cuenta con la posibilidad de modificar renglones, en caso de quererlo así; esta parte sólo es posible usarla una vez. Si se supone que se desea modificar renglones, primero se preguntará al usuario cuantas modificaciones hay. Este número deberá-

ser mayor o igual que 1 y menor o igual que el número de observaciones. Habiendo dado el valor deseado, el programa pasará a preguntar cuales son los renglones que se desean modificar. El programa pasará ahora a pedir cada uno de los renglones que se van a modificar, con esto - - concluye la parte de Lectura de Datos.

El tener los datos almacenados en disco es de gran utilidad si se tiene en cuenta que los datos se pueden -- volver a usar las veces que se deseen sin necesidad de -- volverlos a introducir. Además se cuenta con la facilidad de reemplazar renglones, es decir, se puede leer de - un archivo en disco todas las observaciones, ya realizado ésto, se pueden reemplazar las no deseadas. Considerando esta gran facilidad a manera de ayuda, en el anexo 1 - aparece un programa que crea un archivo en disco, con las restricciones que se mencionó debía de cumplir. Este -- programa solo es utilizable desde una terminal remota.

IV.2 Ejemplos

En esta sección se muestran 2 ejemplos de lectura de datos. En el primero se leen X, Y de un archivo en -- disco; también se muestra la forma como se creó el archivo utilizando el programa que aparece en la parte de los-

anexos. En el segundo ejemplo se lee X solamente y los datos son dados desde una terminal remota.

Ejemplo 1

El programa que se utiliza para crear un archivo en disco se llama "AUXILIO". Se ejecuta poniendo "R AUXILIO", como se muestra en el listado (el cual aparece en la página 78); este programa pregunta por el nombre del archivo a crear; se le llamó "DATOS/REGRESION". Se pide ahora el número de observaciones y el número de variables; se dieron 16 y 3 respectivamente. Tenga en cuenta que al dar como número de variables el 3 entonces sí se va a leer Y y X esto implica que la matriz X tiene 2 columnas.

Como se observará el programa pedirá cada uno de los renglones que pertenecen al archivo de datos. Después del listado en donde se muestra la creación del archivo de datos aparece otro que muestra cada una de las preguntas que hace el SORA. Como se podrá notar, en el momento en que el programa pregunte por el nombre del archivo de disco que contiene los datos, en caso de no estar presente dicho archivo en la clave del usuario, el programa continuará preguntando al usuario por el archivo que con-

tiene dichos datos. Como se mencionó a lo largo del capítulo, aunque se haya leído por disco se cuenta con la posibilidad de modificar renglones, dejando el archivo -- intacto, pero si modificando los valores dentro del programa.

R AUXILIO
#RUNNING 9773

#7

DAME EL NOMBRE DEL ARCHIVO A CREAR
DATOS/REGRESION

DAME EL NUMERO DE: OBSERVACIONES, VARIABLES

16,3

ME TIENES QUE DAR RENGLONES DE 3 VALORES

DAME EL RENGLON 1

39,130,190

DAME EL RENGLON 2

81,7,174,176

DAME EL RENGLON 3

42,5,134,205

DAME EL RENGLON 4

98,3,191,210

DAME EL RENGLON 5

52,7,165,230

DAME EL RENGLON 6

82,194,192

DAME EL RENGLON 7

34,5,143,220

DAME EL RENGLON 8

95,4,186,235

DAME EL RENGLON 9

56,7,139,240

DAME EL RENGLON 10

84,4,198,230

DAME EL RENGLON 11

94,3,175,200

DAME EL RENGLON 12

44,3,156,218

DAME EL RENGLON 13

83,3,190,220

DAME EL RENGLON 14

91,4,178,210

DAME EL RENGLON 15

43,5,132,208

DAME EL RENGLON 16

51,7,148,225

#ET=14:43.3 PT=0.8 IO=0.7

@ VAS A DARME LA MATRIZ X Y/O Y
SI
@ VAS A DARME LA MATRIZ X SOLAMENTE ?
NO
@ LOS DATOS ME LOS VAS A DAR POR TERMINAL ?
NO
@ COMO SE LLAMA EL ARCHIVO EN DISCO ?
DATOS/ESTA/MAL
@ ARCHIVO NO PRESENTE
@ COMO SE LLAMA EL ARCHIVO EN DISCO ?
DATOS/REGRESION
@ DESEAS MODIFICAR RENGLONES ?
NO

Como se notará en este ejemplo, el programa pregunta al usuario "VAS A DARME LA MATRIZ X Y/O Y", esto es interpretado como: dar la matriz X y Y ó dar la matriz X solamente. En caso de ser dicha pregunta respondida afirmativamente se preguntará si se leerá únicamente X . A continuación el programa sigue con el proceso de Lectura y/o Simulación de Datos.

Ejemplo 2

En este ejemplo como se mencionó, se va a leer la matriz X únicamente, la cual es de 18×4 , también aparece en este ejemplo la forma de modificar renglones, resultando que la matriz X , después de que han sido modificados los renglones 10 y 14, tenga en dichos renglones los últimos valores que se dan. El listado de este ejemplo aparece a continuación.

R SORA

#RUNNING 6452

@ DAME TITULO DEL PROBLEMA

#?

EJEMPLO DE LECTURA: MATRIZ X SOLAMENTE

@ LA MATRIZ X ES LA MATRIZ DE DATOS

@ LA MATRIZ Z ES LA MATRIZ DE TRANSFORMACIONES

@ DAME EL NUMERO DE OBSERVACIONES

COLUMNAS DE LA MATRIZ X

NUMERO DE BETA-S EN EL MODELO

18,4,4

@ LAS MATRICES X Y Z SON IGUALES ?

SI

@ VAS A DARME LA MATRIZ X Y/O Y

SI

@ VAS A DARME LA MATRIZ X SOLAMENTE ?

SI

@ LOS DATOS ME LOS VAS A DAR POR TERMINAL ?

SI

@ ME TIENES QUE DAR LAS X-S

@ DAME UN VECTOR RENGLON DE 4 COORDENADAS

@ DAME EL RENGLON 1

20,50,75,15

@ DAME EL RENGLON 2

27,55,60,20

@ DAME EL RENGLON 3

22,62,68,16

@ DAME EL RENGLON 4

27,55,60,20

@ DAME EL RENGLON 5

24,75,72,8

@ DAME EL RENGLON 6

30,62,73,18

@ DAME EL RENGLON 7

32,79,71,11

@ DAME EL RENGLON 8
24,75,72,8
@ DAME EL RENGLON 9
22,62,68,16
@ DAME EL RENGLON 10
27,50,60,20
@ DAME EL RENGLON 11
40,90,78,32
@ DAME EL RENGLON 12
32,79,71,11
@ DAME EL RENGLON 13
50,84,72,12
@ DAME EL RENGLON 14
40,90,70,32
@ DAME EL RENGLON 15
20,50,75,15
@ DAME EL RENGLON 16
50,84,72,12
@ DAME EL RENGLON 17
30,62,73,18
@ DAME EL RENGLON 18
27,55,40,20
@ DESEAS MODIFICAR RENGLONES ?
SI
@ CUANTOS ?
2
@ CUALES SON LOS RENGLONES ?
10,14
@ DAME EL RENGLON 10
27,55,60,20
@ DAME EL RENGLON 14
40,90,78,32
@ DAME LA SEMILLA PARA GENERAR LOS ERRORES.

CAPITULO V

SIMULACION DE DATOS

Habiendo el usuario optado por alguna de las alternativas del capítulo IV, si se seleccionó la 2 o la 3, se tendrá que simular Y ó X y Y respectivamente. Se tratará primero la simulación de X debido a que para la simulación de Y sólo es necesario conocer la matriz Z , sin importar si Z se obtuvo de una matriz X simulada o no.

La simulación de X se hace por columnas, en donde los elementos de la columna j ($j = \overline{1, n}$) siguen una misma distribución, esta distribución es la UNIFORME en el intervalo $[a_j, b_j)$, donde los valores a_j , b_j son dados por el usuario, de modo tal que $a_j < b_j$. Además de estos valores, se tendrá que dar una semilla para generar una distribución UNIFORME en el intervalo $[0, 1)$; esto se hace debido a que la distribución UNIFORME en el intervalo $[a_j, b_j)$ se obtiene como una transformación de una variable aleatoria, que se distribuye UNIFORME en el intervalo $[0, 1)$. En términos matemáticos se tiene

$$X_{ij} \sim U[a_j, b_j) \quad i = \overline{1, n}, \quad j = \overline{1, r}$$

donde $U[a_j, b_j)$ indica distribución UNIFORME en el in-

tervalo $[a_j, b_j)$ tal que $a_j < b_j$.

Entonces, sea $x \sim U[0,1)$

$$\text{si } T(x) = (b_j - a_j)x + a_j$$

$$\curvearrowright T(x) \sim U[a_j, b_j) \text{ de donde } T(x) = X_{ij}$$

En la simulación de la primera columna de X se le preguntará al usuario si todas las componentes son igual al número uno. Esto se hace con el fin de poder plantear regresiones con término independiente y siendo $Z = X$ (sin transformaciones). En caso de ser ésto lo deseado, el programa se encargará de meter números 1 en la primera columna. Si lo antes mencionado no cumple con las necesidades del usuario, se le preguntará entonces por los valores a_j , b_j y una semilla. Ya habiendo terminado con la primera columna, se continuará con las siguientes hasta terminar la simulación de X .

Antes de comenzar a efectuar los pasos necesarios -- para simular a Y se efectúan las transformaciones o se iguala Z a X , ya finalizado este proceso, se da principio a la simulación de Y , pidiéndole al usuario una semilla para genera el vector de errores (ϵ). La-

generación de este vector se puede hacer de dos formas:

- i) Distribución Gaussiana (Normal)
- ii) Alguna distribución no Normal.

Estas dos alternativas fueron consideradas debido a que, desde el punto de vista didáctico es importante tener la capacidad de simular problemas, en los que estuvieran presentes tanto la normalidad como la no normalidad de los errores; para así poder ilustrar diferentes aspectos tanto teóricos como prácticos.

Las variables normales se generan por el método de Box y Muller, ver Neumann y Odell (1971). Esta selección da lugar a 2 opciones

$$1) \underline{y} \sim N(\underline{z}\beta, \sigma^2 \mathbf{I})$$

$$2) \underline{y} \sim N(\underline{z}\beta, \mathbf{V})$$

donde \mathbf{I} es la matriz idéntica y \mathbf{V} es una matriz diferente de \mathbf{I} , con determinante de \mathbf{V} diferente de 0 (cero). Para la selección de alguna de las 2 opciones se tendrá que dar el número correspondiente.

Si la opción 1 es la seleccionada se pedirá al usua-

rio el valor de σ^2 , en este momento se tendrá que hacer una transformación de los errores, debido a que se encuentran distribuidos normal $(\underline{0}, \underline{I})$, y es necesario que tengan matriz de varianza-covarianza igual a $\sigma^2 \underline{I}$. Como se sabe

$$\underline{\epsilon} \sim N(\underline{0}, \underline{I}) \quad , \text{ entonces}$$

$$\text{sea } \underline{\epsilon}^* = \sigma \underline{\epsilon} \quad \rightsquigarrow \quad \underline{\epsilon}^* \sim N(\underline{0}, \sigma^2 \underline{I})$$

donde $\underline{\epsilon}^*$ es el nuevo vector de errores.

$$\text{Si } \underline{y} = \underline{Z} \underline{\beta} + \underline{\epsilon}^* \quad , \text{ entonces:}$$

$$E(\underline{y}) = \underline{Z} \underline{\beta} \quad \text{y} \quad V(\underline{y}) = \sigma^2 \underline{I}$$

$$\text{por lo tanto } \underline{y} \sim N(\underline{Z} \underline{\beta}, \sigma^2 \underline{I}) \quad .$$

Si en vez de la opción 1 se desea la 2, se pedirá al usuario la matriz V (varianzas-covarianzas de los errores), con la restricción de que esta sea simétrica y definida positiva. La restricción de ser definida positiva se hace con el fin de efectuar la descomposición espectral: Sean P y Λ matrices de dimensión $n \times n$ tales que

$$V = P' \Lambda P$$

donde P matriz ortogonal que contiene a los eigenvectores de V .

Λ matriz diagonal, donde los elementos de la diagonal son los eigenvalores asociados a V .

Por ser V definida positiva, P y Λ existen y como además $\text{rango}(Z) = p$, se tiene que los elementos de la diagonal de Λ son todos positivos; entonces:

$$V = P' \Lambda^{1/2} \Lambda^{1/2} P$$

donde $\Lambda^{1/2}$ es una matriz diagonal y cada elemento de $\Lambda^{1/2}$ es la raíz cuadrada del respectivo elemento de Λ . Por ser P matriz ortogonal, $\Lambda^{1/2} P$ no es singular, por lo que puede escribirse:

$$V = T'T \quad (3)$$

donde $T = \Lambda^{1/2} P$

Para facilidad del usuario hay dos modos de introdu-

cir la matriz \sqrt{V} , una para cuando las variables no están correlacionadas (matriz diagonal) y otra cuando si hay correlación entre ellas. Cuando se está en el caso en que es diagonal, sólo se le pedirán al usuario los elementos de ésta, especificando el programa el número de valores deseados. En caso de no ser \sqrt{V} diagonal se necesitará sólo la triangular inferior, el programa se encargará de completar la matriz. En este momento se pasará a checar si \sqrt{V} es definida positiva; en caso de no ser así, el programa regresará a proponer al usuario las formas de lectura de \sqrt{V} .

A continuación el programa pasará a obtener la distribución de Y . Se sabe que

$$\underline{\epsilon} \sim N(0, I) \quad , \text{ entonces}$$

$$T\underline{\epsilon} \sim N(0, T'T) \quad \text{donde } T \text{ es obtenida de (3)}$$

$$\text{si } Y = Z\underline{\beta} + T\underline{\epsilon} \quad \text{entonces}$$

$$E(Y) = Z\underline{\beta} \quad \text{y} \quad \sqrt{V(Y)} = T'T$$

por lo tanto

$$Y \sim N(2\beta, T^2) \text{ como } T^2 = V$$
$$\leadsto Y \sim N(2\beta, V)$$

Considerando el caso en donde los errores se generan con distribución diferente a la normal, para simularlos se seleccionó a la distribución Ji-Cuadrada entre todas las posibles no normales de una manera arbitraria. Además se consideró que si el parámetro de la Ji-Cuadrada h (grados de libertad) se escoge de la siguiente manera: $h = \sqrt{3}$, la Ji-Cuadrada resultante sería suficientemente diferente a la distribución Normal.

Este parámetro es seleccionado aleatoriamente con la semilla dada por el usuario para la generación de los errores. Entre los resultados que se obtienen en el estudio de la distribución Ji-Cuadrada figuran:

$$\text{Si } \omega \sim \chi^2(h) \text{ se tiene}$$
$$\text{y } E(\omega) = h$$
$$V(\omega) = 2h$$

No se consideró mas a fondo esta simulación debido a que-

cumple con proporcionar errores distribuidos en forma normal.

Como se ha observado, para la simulación de Y se necesita tener β , por lo que se pedirá al usuario dar los valores de las componentes de β .

CAPITULO VI

COMANDOS DEL PROGRAMA

El programa cuenta con órdenes que son dadas por el usuario en función de lo que desea, como por ejemplo: intervalos de confianza, pruebas de hipótesis, gráficas, -- etc., a estas órdenes se les dió el nombre de comandos. Existen en el programa 8 comandos, tres de los cuales son dinámicos en su estructura por lo que en realidad se cuenta con más de 8. Estos comandos se pueden ejecutar en el momento en que el programa lo señale y sin guardar -- prioridad ni orden alguno. A continuación se explica el uso de cada uno de ellos.

EXP

Este comando sirve para explicar al usuario cada uno de los comandos del programa, la forma y clave que tienen y resultados que se obtienen al ejecutarlos.

INT.<EA>

Con el comando INT.<EA> se obtienen intervalos de -- confianza para: BETA, COMBINACIONES LINEALES DE BETA, COMBINACIONES LINEALES SIMULTANEAS DE BETA, PRONOSTICO DE Y DADO EL VECTOR X , ESPERANZA DE Y DADO EL VECTOR X , - VARIANZA Y VARIANZA SEGUN TATE Y KLETT. Estos intervalos se obtienen dependiendo de la clave seleccionada para <EA> , <EA> se substituye por cualquiera de las siguien

tes 7 claves: IND, COM, SIM, PRE, EST, VARAP y VARTK. Estas dan lugar respectivamente a los intervalos de confianza enunciados anteriormente. Una característica general de este comando es que preguntará al usuario, en el momento en que finalice el cálculo de un intervalo de confianza, si desea volver a calcular otro intervalo de confianza - - (dentro de la misma clave para $\langle EA \rangle$); en caso de contestar "si" no se tendrá necesidad de teclear nuevamente dicho comando. A continuación se explica cada una de las formas que puede tomar este comando.

INT.IND

Intervalo de confianza para β_i . Primero el programa leerá el subíndice i , valor que tendrá la condición $i = \overline{1, p}$, en caso contrario se volverá a leer i ; a continuación se leerá la confiabilidad con que se desea el intervalo de confianza para β_i , ésta deberá ser un valor entre 0 y 1. En este momento se dará a conocer al usuario el intervalo de confianza escogido, el programa pasará a preguntar si se desea otro intervalo de confianza (para β_i). En caso afirmativo comenzará con la lectura de i y así sucesivamente.

INT.COM

Intervalo de confianza para combinaciones lineales de beta. Como primer paso para obtener dicho intervalo se le pedirá al usuario un vector de $k \times p$, que es el que define la combinación lineal; siguiendo esto se pedirá la confiabilidad con la que se desea el intervalo. Aparecerá entonces el intervalo de confianza deseado.

INT.SIM

Intervalo de confianza para combinaciones lineales simultáneas de beta. Primero se pasará a leer la confiabilidad con la que se desean los intervalos, esta confiabilidad deberá pertenecer al intervalo $[.7, .995]$ en caso contrario se leerá nuevamente, se continuará con la lectura del número de vectores que se desea tengan una misma confiabilidad en sus intervalos simultáneos, dicho número deberá ser menor o igual que p y mayor o igual a 2. A continuación se leerán cada uno de los vectores. Ya finalizada esta lectura aparecerán cada uno de los intervalos de confianza deseados.

INT.PRE

Intervalo de confianza para el valor de Y dado el vector X (además da el centro del intervalo que es la predicción). Como primera lectura a efectuarse figura la de un vector de $k \times p$, pasará más tarde a leer la con--

fiabilidad del intervalo, con lo cual se obtiene el intervalo de confianza deseado.

INT.EST

Intervalo de confianza para la esperanza de Y dado el vector X , dará también el centro del intervalo (el valor esperado de Y dado X). Al igual que INT.PRE efectuará las mismas lecturas.

INT.VARAP

Intervalo de confianza para la varianza, sólo necesitará la confiabilidad con la cual se desea el intervalo.

INT.VARTK

Intervalo de confianza para la varianza según TATE y KLETT. Necesitará de los valores a_n, b_n de las tablas de TATE y KLETT y la confiabilidad con que se dieron dichos valores, el programa especificará al usuario los grados de libertad de la prueba. Ver anexo 3.

HIP

Comando que tiene como finalidad realizar pruebas de hipótesis, como se mencionó en I.4, la prueba que realiza

es $C\beta = \gamma$. La matriz C es de dimensión $q \times p$ con la condición que $q \leq p$, y el vector γ es de $q \times 1$. Como primer paso para realizar la prueba de hipótesis se tendrá que conocer el valor de q , después la matriz C (la cual se lee por renglones) y el vector γ (se lee su transpuesto que es de $1 \times q$). Ya habiendo terminado el programa de realizar todas estas lecturas en el orden y forma mencionada, pasará a calcular la tabla de Análisis de Varianza, la que aparecerá tanto en terminal como en impresora. El número de renglones de dicha tabla varía dependiendo si es posible probar carencia de ajuste.

Este comando presenta la posibilidad de realizar la prueba $C\beta = \gamma$ para los tamaños de error tipo I (alfa) que se deseen. La manera de hacerlo es proporcionando el valor de alfa cada vez que se pida. Para terminar con este proceso en el momento en que se pregunte si se quiere probar esta hipótesis con otra alfa se deberá contestar "no".

Siguiendo con la prueba, en caso de que hubiesen repeticiones de las observaciones, y por lo tanto posibilidades de probar carencia de ajuste se comparará la F calculada contra la F teórica, si la F calculada es menor aparece un mensaje indicando: "RECHAZAMOS CARENCIA", esto es-

no hay falta de ajuste; si es mayor aparece "NO RECHAZAMOS CARENCIA", esto es, hay falta de ajuste y el modelo propuesto no sirve, hay que proponer otro.

Se podrá realizar esta prueba con diferentes niveles de significancia y siguiendo el procedimiento indicado para probar $C\beta = \underline{\gamma}$. Para hacer mejor uso de este comando léase Nota Importante.

El programa preguntará ahora si se desea o no realizar otra prueba de hipótesis, con el fin de empezar con otra hipótesis o finalizar la ejecución del comando.

Entre otras de las informaciones que se obtienen al ejecutar este comando está la del coeficiente de correlación múltiple al cuadrado.

NOTA IMPORTANTE

Se recomienda, para el caso en que existan repeticiones, usar únicamente un tamaño de alfa para la prueba $C\beta = \underline{\gamma}$, con el objeto de probar la hipótesis de carencia de ajuste; ya que si esta no es rechazada no tendrá significado práctico la realización de la primera prueba ($C\beta = \underline{\gamma}$).

RAC

Este comando da la sucesión de signos de los residuales, número de : + (mas), -(menos) y corridas de ellos.

GRA

Este comando sirve para graficar cualquier columna de X o Z , vector $YE (\hat{Y})$ o T (tiempo) contra residuales estandarizados. El programa preguntará contra que matriz se desean graficar los residuales pudiendo dar el usuario X, Z, YE o T .

En estos momentos el programa necesitará saber si se desea banda de confianza en la gráfica; en caso afirmativo pasará entonces a preguntar la confiabilidad con la que se desea dicha banda.

Si se desean graficar los residuales contra X o Z , el programa tendrá la necesidad de saber la columna contra la cual se desea hacer la gráfica; pudiendo dar el usuario un valor de $\bar{1,r}$ si es contra X o $\bar{1,p}$ si es contra Z .

El programa enviará un mensaje con tal de saber si -

la gráfica se desea por terminal, en caso que se responda negativamente ésta saldrá por impresora; ya contestada la pregunta anterior, aparecerá un letrero en la terminal el cual indica que el programa necesita que se fije un número de renglones para hacer la gráfica, este número deberá ser mayor o igual a 10. Uno de los consejos que se dá al usuario de este programa, es que el número de renglones no debe de ser muy grande (<70); además de tener en consideración que si se está ejecutando el programa en una terminal que no trabaje con papel, este número no deberá exceder de 35; y en caso de ser impresora, de 60 (para impresora ha dado buenos resultados el 40).

Otra de las características de este comando es, que se cuenta con la posibilidad de regresar ha hacer otra gráfica.

Lista.<AE>

Con el comando LISTA.<AE> se listan o imprimen¹ vectores, matrices o valores en terminal o impresora. Las matrices, vectores o valores que se pueden listar o imprimir con este comando se consideran de interés en el estudio del Análisis de Regresión.

¹En lenguaje de computación se usa listar para terminal e imprimir para impresora.

Al igual que el comando INT.<EA>, dependiendo de la clave que se le dé a <AE> se listarán γ , β , Z , X , etc. A continuación se dá cada una de las claves para -- <AE> y se menciona la matriz, vector o valores que lista o imprime.

LISTA.YE.	Lista o imprime	γ	estimada (vector)
LISTA.Y.	Lista o imprime	γ	(vector)
LISTA.BE.	Lista o imprime	β	estimada (vector)
LISTA.X.	Lista o imprime	X	(matriz)
LISTA.Z.	Lista o imprime	Z	(matriz)
LISTA.R.	Lista o imprime	Residuales	(vector)
LISTA.RS.	Lista o imprime	Residuales Estandarizados	(vector)
LISTA.SM1.	Lista o imprime	$(2'2)^{-1}$	(matriz)
LISTA.DW.	Lista o imprime	Estadística de Durbin-Watson	(valor)
LISTA.DV.	Lista o imprime	Varianza y desviación standar estimada, $\underline{\gamma}\underline{\gamma}$, $\underline{\gamma}$	(valores)
LISTA.*.	Equivale a listar o imprimir todas las -- matrices, vectores y valores que se han -- mencionado, es decir, equivale a -- LISTA.YE., LISTA.Y.,..., LISTA.DV.		

Como se ha observado no se ha preguntado al usuario por -- donde desea que aparezca la matriz, vector o valor(es); --

en el momento en que esto suceda, se podrá contestar TERM (para terminal) ó IMPRE (para impresora) si así se desea, o TERMINAL o IMPRESORA.

En este momento el programa listará o escribirá en la terminal o impresora, en caso tal que las columnas de la matriz que se desea listar o imprimir no quepan en la terminal o en el papel, se listarán o imprimirán todos -- los renglones de la matriz con las columnas que quepan a partir de la 1; a continuación se pone una marca (asterisco) la cual indicará que se terminaron de listar o imprimir todos los renglones de la matriz. Se pasará ahora a imprimir todas las columnas siguientes hasta terminar con la matriz.

Al terminar de imprimir o listar cada matriz, vector o valores se pondrá un asterisco como marca para indicar que se terminó de listar o imprimir.

NUE

El comando NUE inicia la ejecución del programa, es decir, se pasará a leer n, r, p nuevamente, sin importar lo realizado anteriormente, se usa sobre todo cuando se desea analizar o simular otro problema.

FIN

Con este comando finaliza la ejecución del programa, se utiliza cuando no se desea estudiar más el modelo que se tiene como tampoco considerar otros problemas.

CAPITULO VII

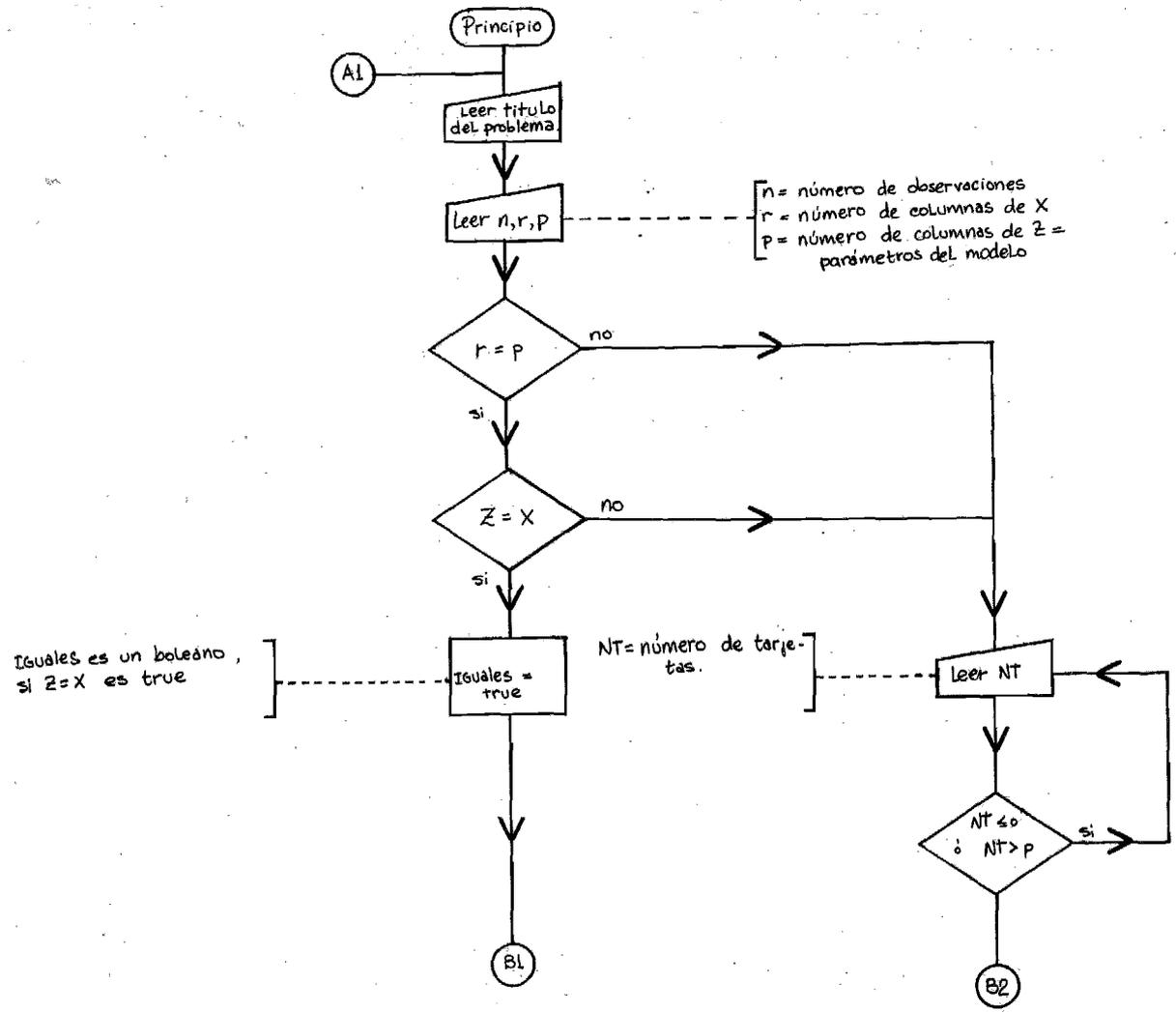
DIAGRAMAS DE FLUJO

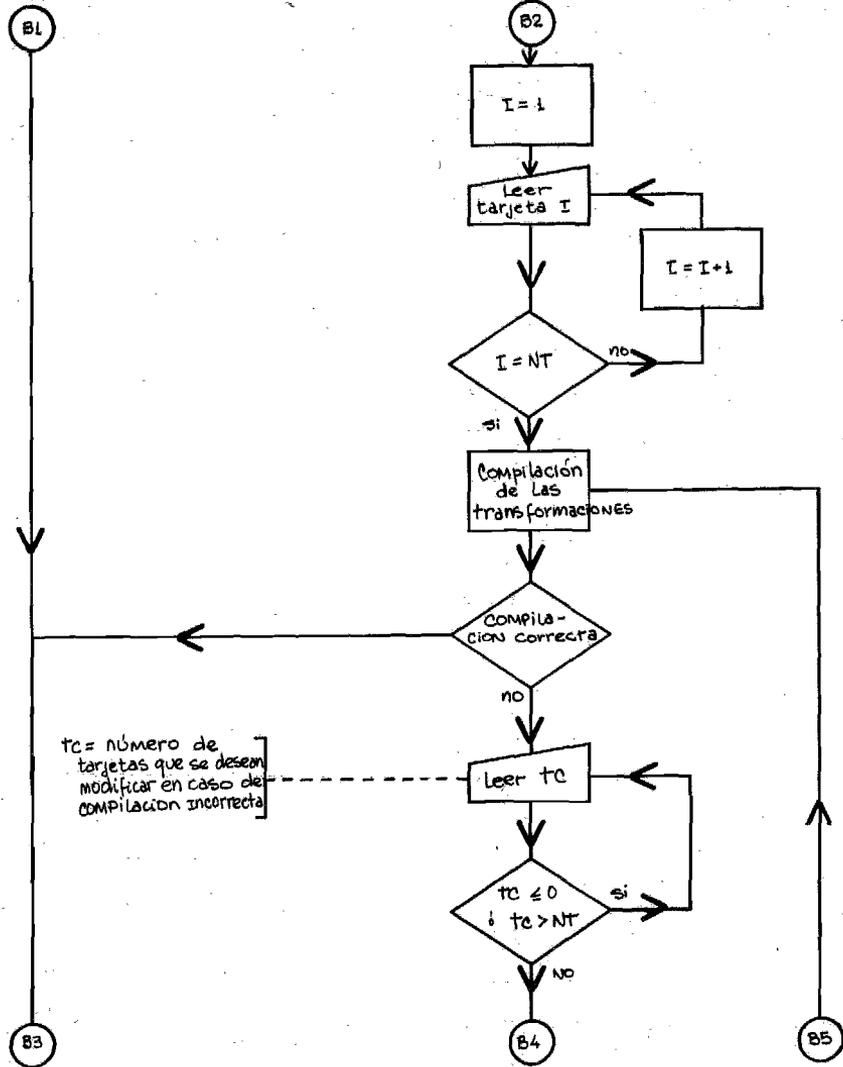
En este capítulo aparecen diagramas de flujo con el fin de dar al usuario una visión más general del programa y a la vez facilitarle la entrada de los datos. Estos diagramas podrían no considerarse de flujo, sino diagramas de bloque extendidos, porque el hacer un diagrama de flujo de las partes que se explican resultaría muy extenso, de aquí que sólo se consideren las partes que involucran decisiones por parte del usuario, como también las lecturas a efectuarse.

Entre los diagramas figuran i) inicio del programa - hasta fin de estimación puntual (pasando por Transformaciones, Lectura y/o Simulación), ii) comando HIP (Hipótesis), iii) comando GRA (Graficación de Residuales Estandarizados).

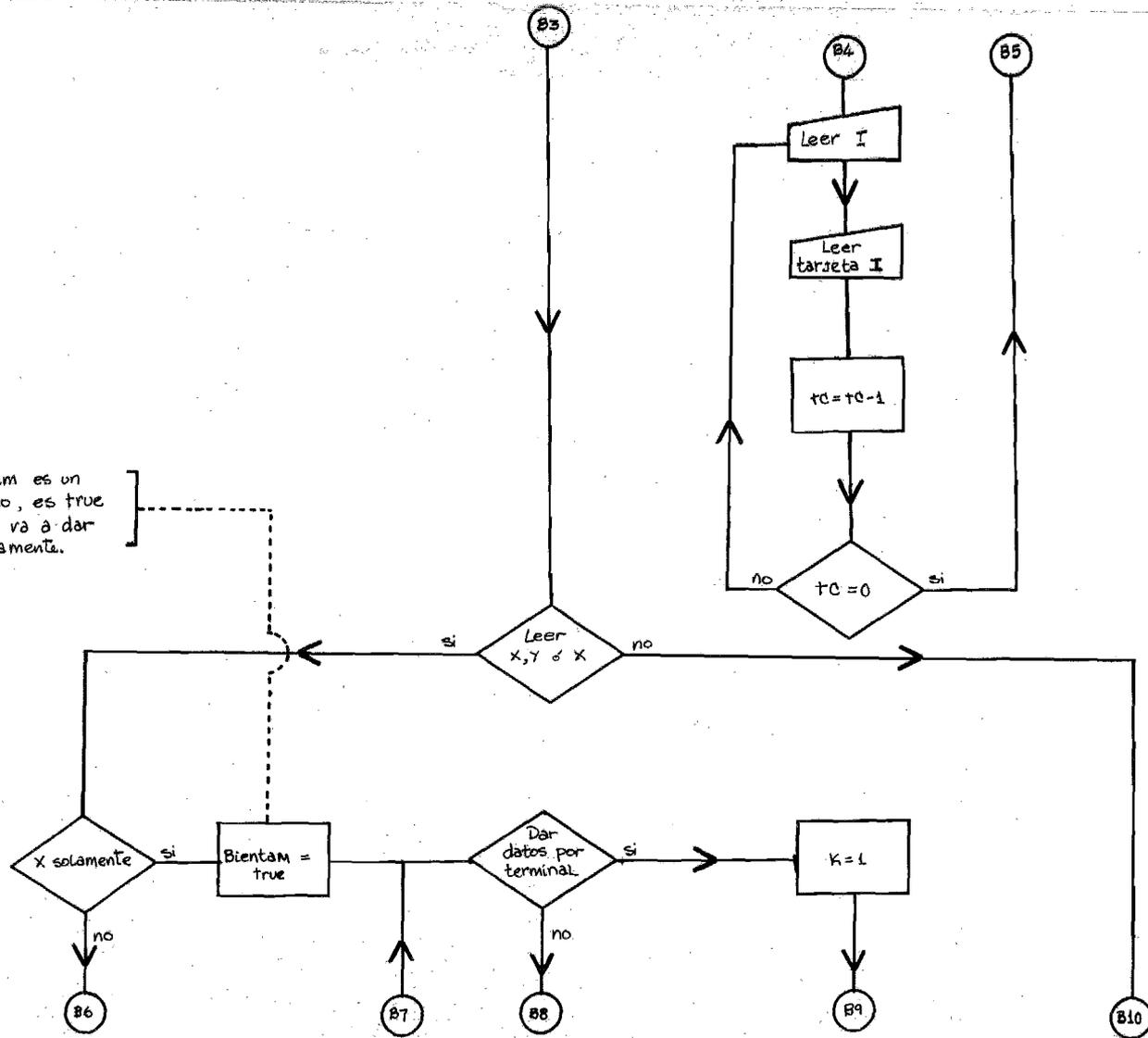


BIBLIOTECA
INSTITUTO DE ECOLOGIA
UNAM

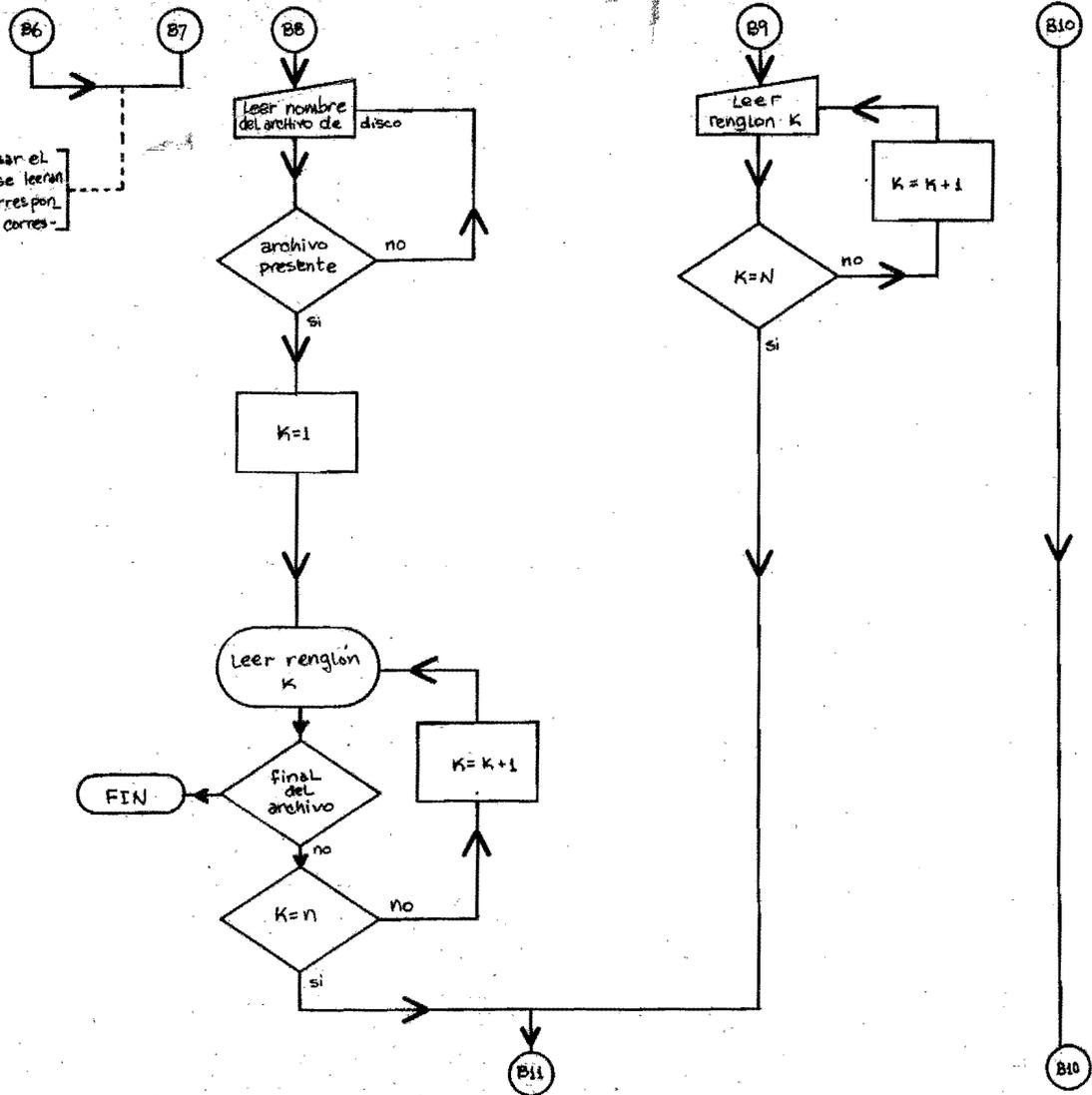


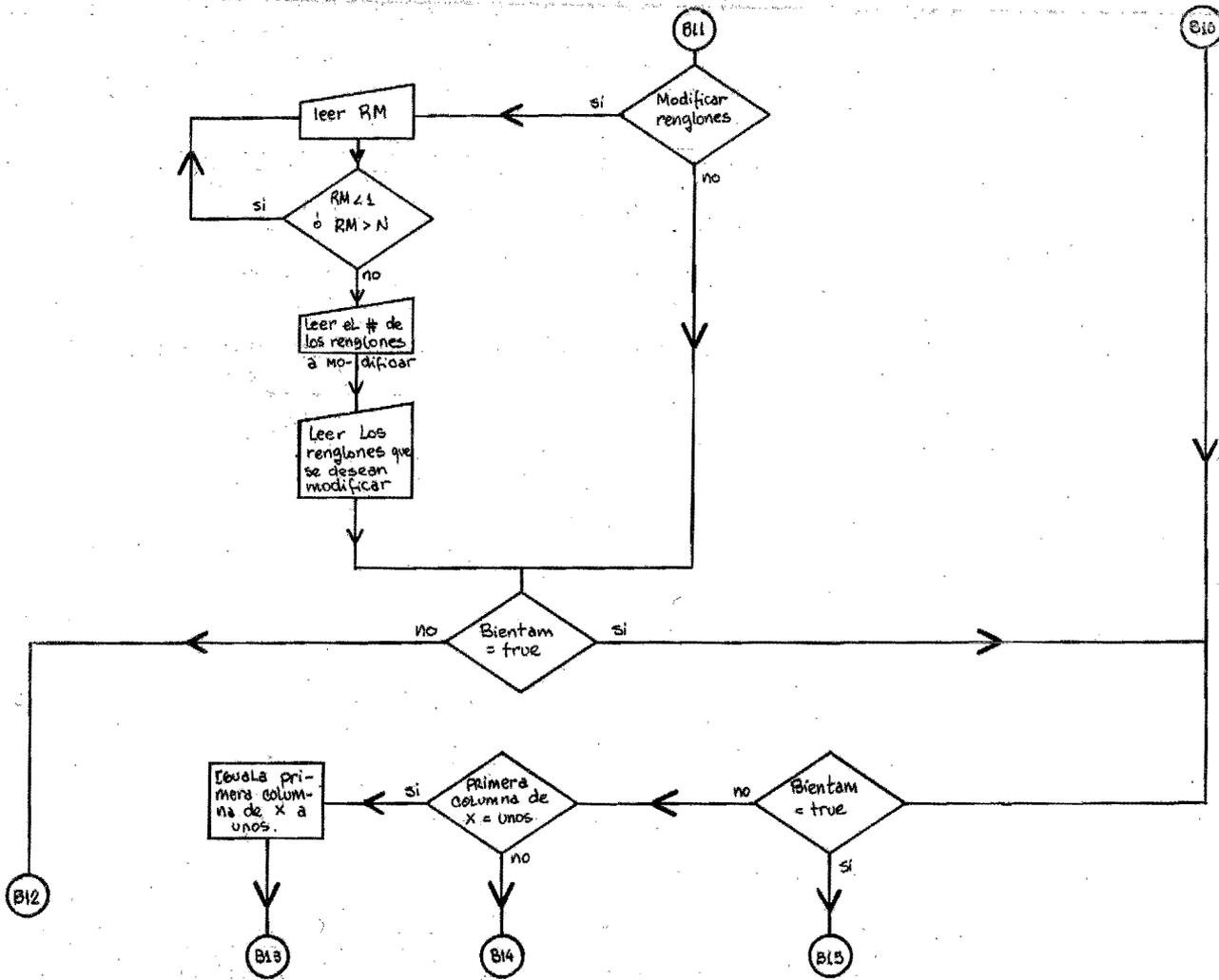


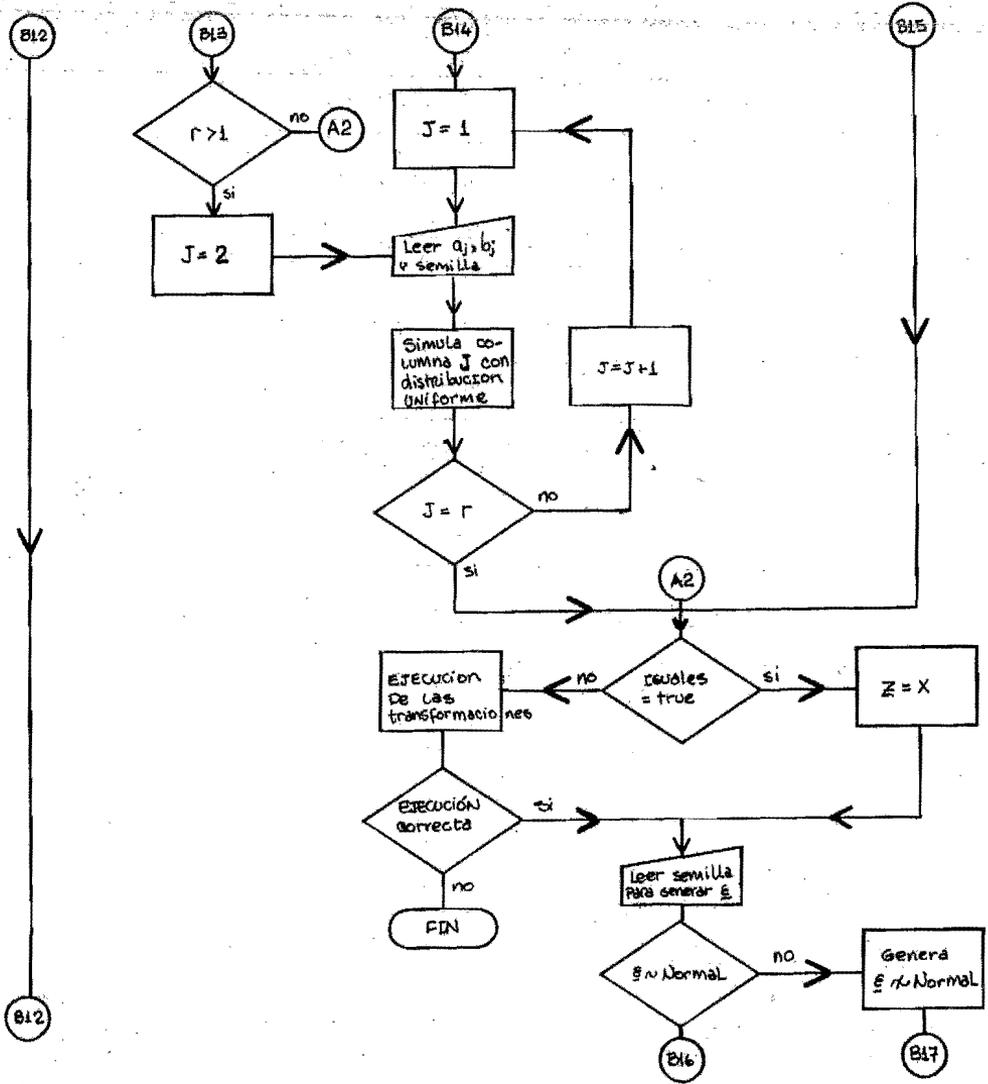
Bientam es un
booleano, es true
si se va a dar
X solamente.

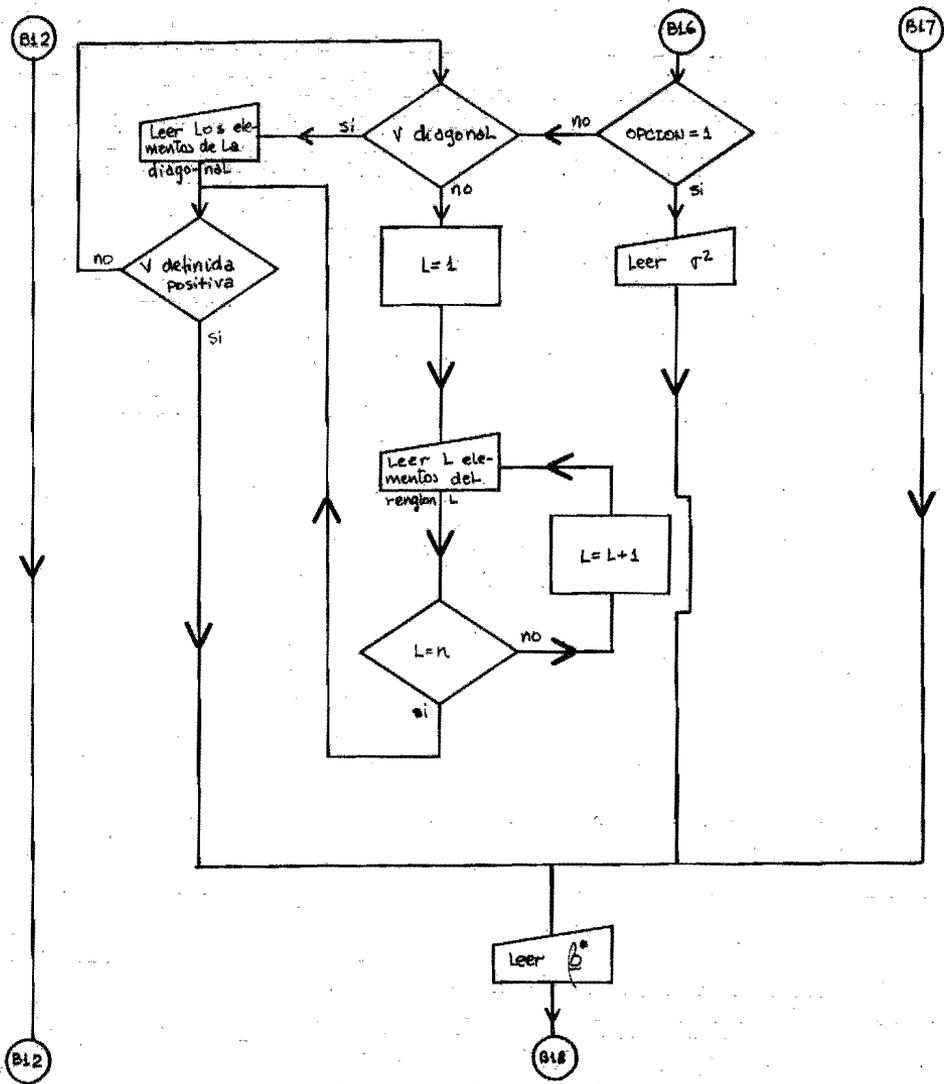


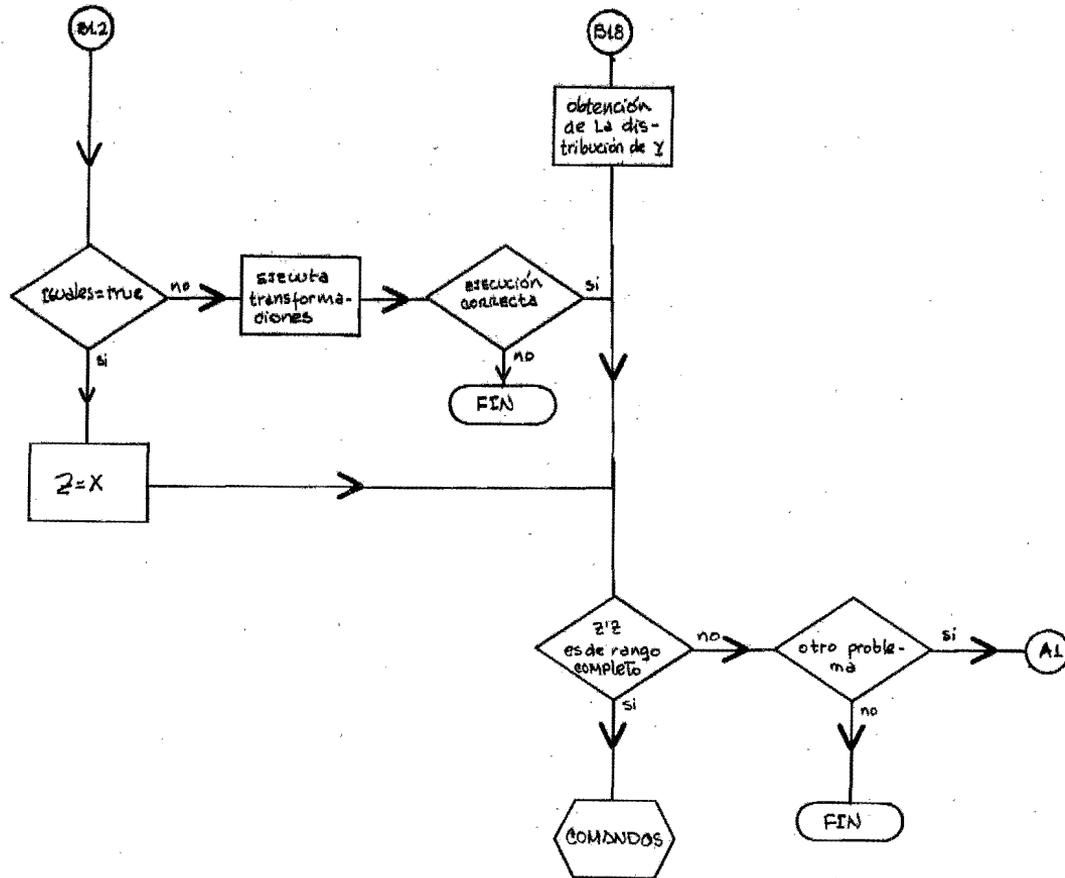
En caso de pasar el flujo por aquí se leerán n+1 valores, correspondientes a X y L correspondiente a Y



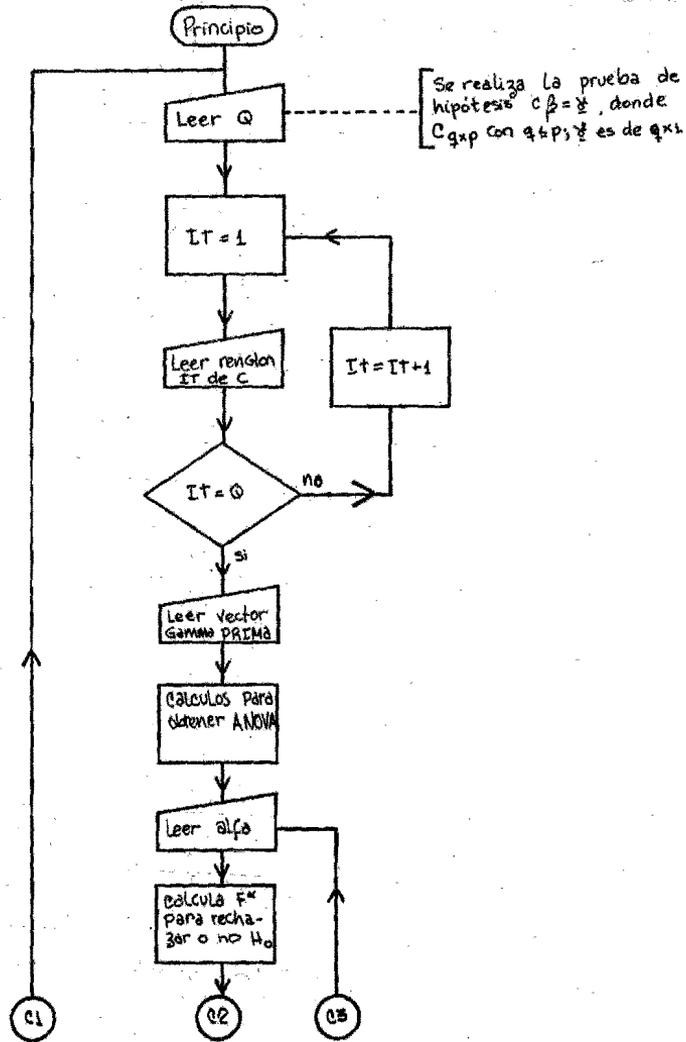


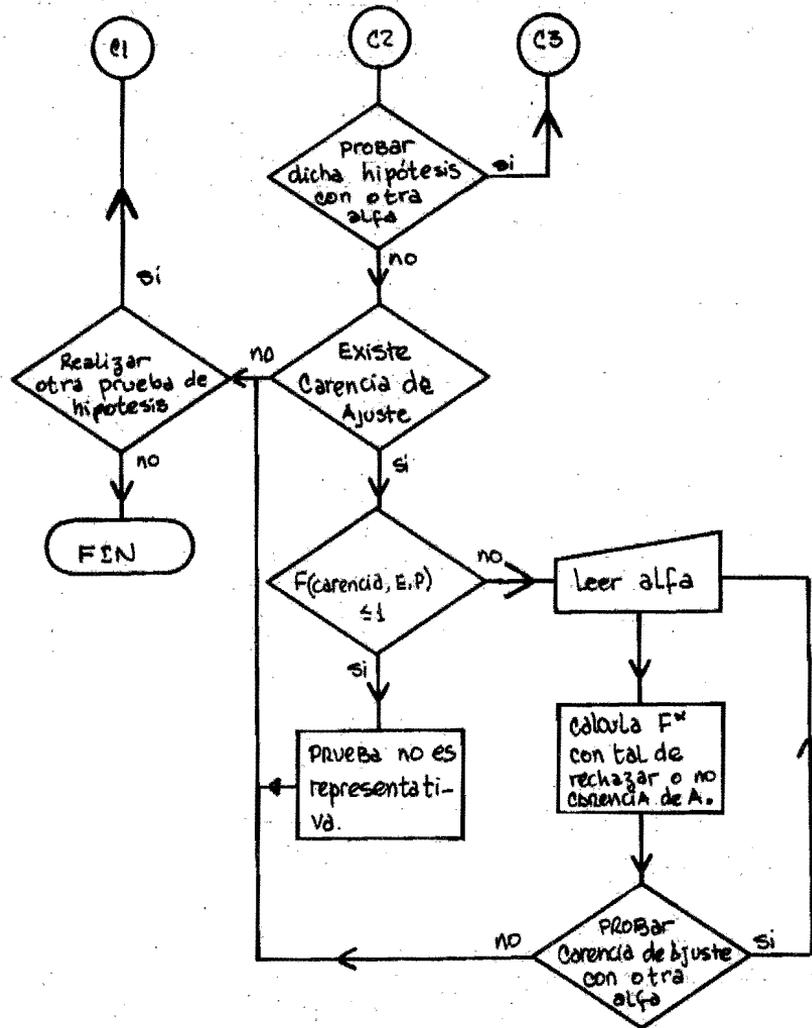




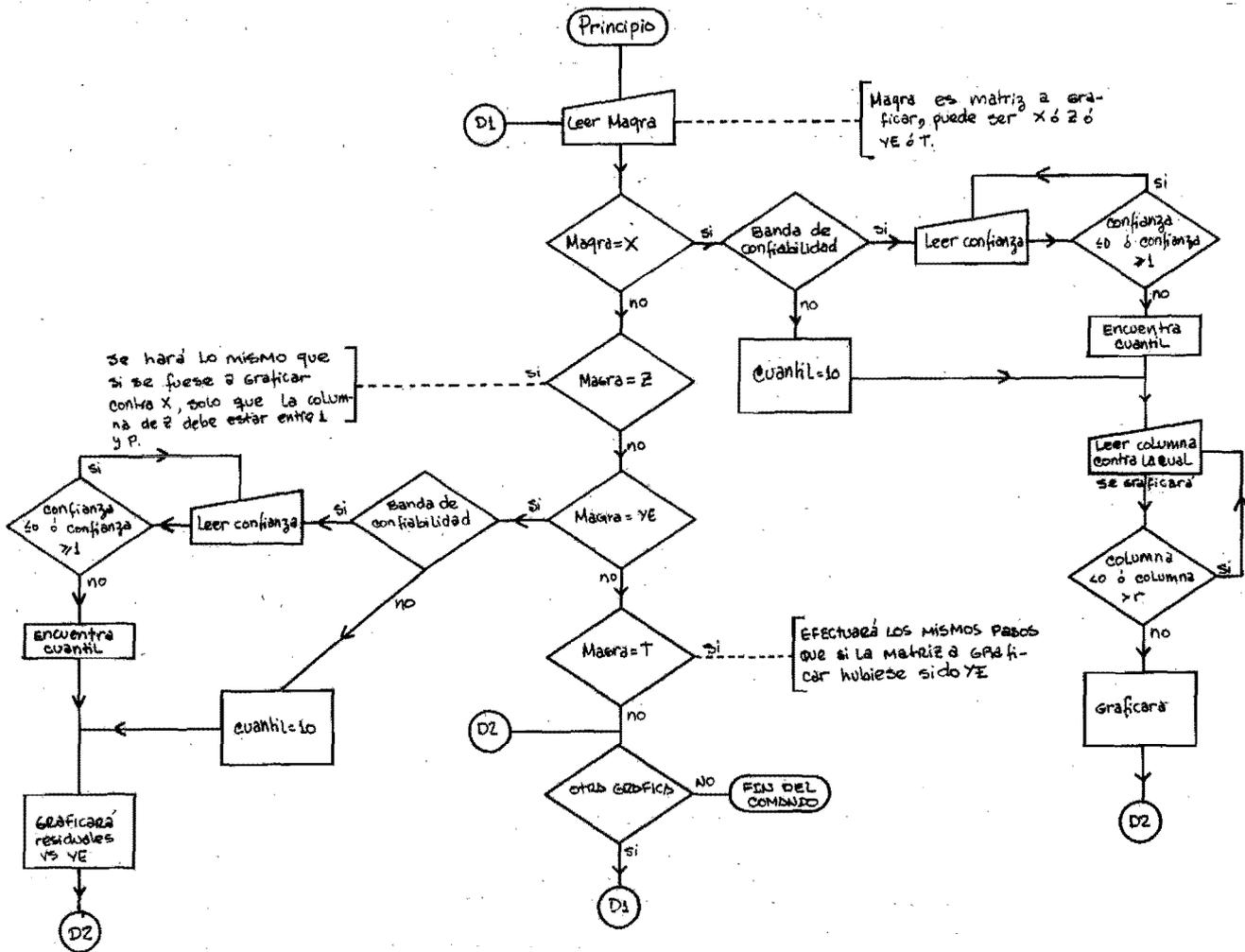


COMANDO HIP (HIPOTESIS)





COMANDO GRA (GRAFICACION)



CAPITULO VIII

EJEMPLO

En este capítulo se mostrará un ejemplo del uso del programa. En el se utilizan todos los recursos del programa (intervalos de confianza, pruebas de hipótesis, graficación, etc.). El ejemplo tiene como finalidad el -- mostrar explícitamente el manejo del programa.

Se hace la aclaración antes de dar principio al ejemplo, que los niveles de significancia (alfa), tanto como vectores, etc. se han escogido de manera arbitraria.

Primero se procede a simular los datos; ya simulados estos, se obtienen cada uno de los intervalos de confianza que se pueden obtener. Se ejecutan también los comandos HIP, RAC, GRA. Se muestra también el comando FIN y-EXP.

Se observará a lo largo del ejemplo, la propiedad -- que se señaló tenían la mayoría de los comandos; dicha -- propiedad es la iteración dentro del mismo comando. Tam-- bién se puede notar que cuando se deseó un intervalo para β_3 , como no existía tal, el programa mandó un mensaje-- indicando el error cometido. En la prueba de hipótesis, en el renglón de la tabla correspondiente a total, el es-

pacio asignado a la suma de cuadrados resultó ser muy pequeño, por tal motivo dicho lugar aparece lleno con asteriscos; esto es un gran problema en caso que no se pudiese obtener dicho valor, por tal motivo se consideró de -- utilidad que el comando LISTA.<AE>., cuando <AE> sea VD -- (como también *) se obtuviese dicho valor, por mas grande que fuese, como puede notarse cuando se pidió LISTA.*.

```

R SORA
#RUNNING 9009
@ DAME TITULO DEL PROBLEMA
#?
CAPITULO VIII
@ LA MATRIZ X ES LA MATRIZ DE DATOS
@ LA MATRIZ Z ES LA MATRIZ DE TRANSFORMACIONES
@ DAME EL NUMERO DE OBSERVACIONES
      COLUMNAS DE LA MATRIZ X
      NUMERO DE BETA-S EN EL MODELO

30:2:2
@ LAS MATRICES X Y Z SON IGUALES ?
SI
@ VAS A DAME LA MATRIZ X Y/O Y
NO
@ VOY A GENERAR LA MATRIZ X
@ LA MATRIZ X TIENE LA PRIMERA COLUMNA DE UNO-S ?
SI
@ DAME EL LIMITE INFERIOR Y SUPERIOR PARA EL PROXIMO NUMERO ALEATORIO
      A GENERAR ASI COMO SU SEMILLA
-1000.432,524287
@ LOS VALORES QUE DISTE SON :
      -1000.000      432.000      524287.000
@ DESEAS CAMBIARLOS ?
NO
@ DAME LA SEMILLA PARA GENERAR LOS ERRORES
524285
@ EL VALOR QUE ME DISTE ES :
      524285.000
@ DESEAS CAMBIARLO ?
NO
@ LOS ERRORES, DESEAS GENERARLOS CON DISTRIBUCION NORMAL ?
SI
@ SE TIENEN 2 OPCIONES

@ 1. Y DISTRIBUIDA N( Z * BETA, SIGNA**2 * I )
@ 2. Y DISTRIBUIDA N( Z * BETA, V )

@ CUAL DE LAS OPCIONES DESEAS
1
@ DAME LA VARIANZA DEL ERROR
11
@ DAME UN VECTOR BETA-PRIMA PARA CALCULAR Y
      DE 1 X 2
12:17
@ DESEAS CONOCER EL DETERMINANTE DE Z-PRIMA*Z ?

```

SI
@ EL DETERMINANTE DE Z-PRIMA*Z ES

166704322.8656

@ LA VARIANZA ESTIMADA ES 1.040
@ EXISTE EL COMANDO EXPLICA
@ DAME COMANDO DE EJECUCION
EXPLICA
@ TENGO 8 COMANDOS, LOS CUALES SON :

INT.<EA> DA INTERVALOS DE CONFIANZA
SI <EA> = IND PARA BETA(I)
COM COMBINACIONES LINEALES DE BETA
SIM COM. LINEALES SIMULTANEAS DE BETA
PRE VALOR DE Y DADO X, Y EL CENTRO DEL
INTERVALO
EST ESPERANZA DE Y DADO X, Y EL CENTRO
DEL INTERVALO
VARAP VARIANZA
VARTK VARIANZA SEGUN TATE & KLETT
HIP PRUEBA DE HIPOTESIS
RAC OBTENCION DE LAS RACHAS PARA RESIDUALES
GRA GRAFICACION DE RESIDUALES ESTANDARIZADOS
LISTA.<AE> LISTA O IMPRIME EN TERMINAL O IMPRESORA
SI <AE> = YE Y-ESTIMADA
Y Y
BE BETA-ESTIMADA
X X
Z Z
R RESIDUALES
RS RESIDUALES ESTANDARIZADOS
SM1 INVERSA DE Z-PRIMA * Z
DW ESTADISTICA DURBIN-WATSON
VD VARIANZA Y DESVIACION ESTANDAR ESTIMADA,
Y-PRIMA * Y, Y-BARRA
* TODAS LAS ANTES MENCIONADAS
NUE INICIALIZA LA EJECUCION DEL PROGRAMA
FIN FINALIZA LA EJECUCION DEL PROGRAMA
EXP EXPLICA LOS COMANDOS EXISTENTES
@ DAME COMANDO DE EJECUCION
INT.IND
@ DAME EL # DE LA BETA PARA LA CUAL DESEAS EL INT. DE CONF.
1
@ DAME LA CONFIABILIDAD
.975
@ EL INTERVALO DE CONFIANZA PARA LA BETA- 1 AL .975 ES
I 12.320, 13.5231
@ QUIERES OTRO INTERVALO DE CONFIANZA ?

SI
@ DAME EL # DE LA BETA PARA LA CUAL DESEAS EL INT. DE CONF.
2
@ DAME LA CONFIABILIDAD
.997
@ EL INTERVALO DE CONFIANZA PARA LA BETA- 2 AL .997 ES
E [16.999, 17.002]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
SI
@ DAME EL # DE LA BETA PARA LA CUAL DESEAS EL INT. DE CONF.
3
@ DAME UN NUMERO ENTRE 1 Y 2
@ DAME EL # DE LA BETA PARA LA CUAL DESEAS EL INT. DE CONF.
2
@ DAME LA CONFIABILIDAD
.5
@ EL INTERVALO DE CONFIANZA PARA LA BETA- 2 AL .500 ES
E [17.000, 17.001]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
NO
@ DAME COMANDO DE EJECUCION
INT.COM
@ ME TIENES QUE DAR UN VECTOR R-PRIMA DE 1 X 2
1,1
@ DAME LA CONFIABILIDAD
.95
@ EL INTERVALO DE CONFIANZA PARA R-PRIMA*BETA AL .950
ES E [29.401, 30.442]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
SI
@ ME TIENES QUE DAR UN VECTOR R-PRIMA DE 1 X 2
0,1
@ DAME LA CONFIABILIDAD
.975
@ EL INTERVALO DE CONFIANZA PARA R-PRIMA*BETA AL .975
ES E [16.999, 17.001]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
NO
@ DAME COMANDO DE EJECUCION
INT.SIN
@ DAME LA CONFIABILIDAD CON QUE QUIERES LOS INTERVALOS
.975
@ DAME EL NUMERO DE VECTORES (VER MANUAL)

2
@ DAME VECTORES DE 1 X 2
@ DAME EL VECTOR 1
1,1
@ DAME EL VECTOR 2
0,1
@ LOS INTERVALOS SIMULTANEOS DE CONFIANZA .995 SON :

VECTOR	INTERVALO
1	[29.009 , 30.834]
2	[16.999 , 17.002]

@ QUIERES OTRO INTERVALO DE CONFIANZA ?

NO

@ DAME COMANDO DE EJECUCION
INT.PRE
@ ME TIENES QUE DAR UN VECTOR XO-PRIMA DE 1 X 2

1,1
@ DAME LA CONFIABILIDAD

.99
@ EL INTERVALO DE CONFIANZA PARA YIX-CERO AL .990
ES [27.018, 32.825] CENTRADO EN 29.921

@ QUIERES OTRO INTERVALO DE CONFIANZA ?

SI

@ ME TIENES QUE DAR UN VECTOR XO-PRIMA DE 1 X 2
2,5

@ DAME LA CONFIABILIDAD
.97

@ EL INTERVALO DE CONFIANZA PARA YIX-CERO AL .970
ES [108.238, 113.450] CENTRADO EN 110.844

@ QUIERES OTRO INTERVALO DE CONFIANZA ?

NO

@ DAME COMANDO DE EJECUCION
INT.EST
@ ME TIENES QUE DAR UN VECTOR XO-PRIMA DE 1 X 2

1,1
@ DAME LA CONFIABILIDAD

.975
@ EL INTERVALO DE CONFIANZA PARA E(YIX-CERO) AL .975
ES [29.319, 30.524] CENTRADO EN 29.921

@ QUIERES OTRO INTERVALO DE CONFIANZA ?

SI

@ ME TIENES QUE DAR UN VECTOR XO-PRIMA DE 1 X 2
2,5

@ DAME LA CONFIABILIDAD

```

.97
@ EL INTERVALO DE CONFIANZA PARA E(YIX-CERO) AL .970
  ES [ 109.679, 112.009] CENTRADO EN 110.844
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
  NO
@ DAME COMANDO DE EJECUCION
INT.VARAP
@ DAME LA CONFIABILIDAD
.995
@ EL INTERVALO DE CONFIANZA PARA SIGMA**2 AL .995 ES
  [ 0.543, 2.532]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
  SI
@ DAME LA CONFIABILIDAD
.999
@ EL INTERVALO DE CONFIANZA PARA SIGMA**2 AL .999 ES
  [ 0.491, 3.020]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
  NO
@ DAME COMANDO DE EJECUCION
INT.VARTK
@ DAME LOS VALORES A-N Y B-N DE LAS TABLAS DE TATE & KLETT PARA GL= 28
  Y LA CONFIABILIDAD RESPECTIVA
18.3095,45.8446,.9
@ EL INTERVALO DE CONFIANZA PARA SIGMA**2 AL .900 ES
  [ -0.636, 1.592]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
  SI
@ DAME LOS VALORES A-N Y B-N DE LAS TABLAS DE TATE & KLETT PARA GL= 28
  Y LA CONFIABILIDAD RESPECTIVA
12.3211,59.623,.995
@ EL INTERVALO DE CONFIANZA PARA SIGMA**2 AL .995 ES
  [ 0.489, 2.366]
@ QUIERES OTRO INTERVALO DE CONFIANZA ?
  NO
@ DAME COMANDO DE EJECUCION
HIP
@ HAGO LA PRUEBA C*ETA=GAMMA
@ LA MATRIZ C ES DE Q X 2
@ DAME EL VALOR DE Q (DEBE SER MENOR O IGUAL QUE 2)
1
@ DAME LA MATRIZ C
@ DAME EL RENGLON 1
1.1

```


GRA
@ CONTRA QUE MATRIZ DESEAS GRAFICAR LOS RESIDUALES ?

T

@ QUIERES BANDA DE CONFIABILIDAD EN LA GRAFICA ?

SI

@ DAME LA CONFIABILIDAD

.975

@ QUIERES LA GRAFICA POR TERMINAL ?

SI

@ DAME EL NUMERO DE RENGLONES DE LA GRAFICA

30

I 4.00

I

I

I

I 2.72

I

I

I 1.44 *

I

I

I 0.16

I

I

I -1.11

I

I

I -2.39

I

I

I -3.67

I

.....
LOS LIMITES EN EL EJE X SON: INFERIOR 1
SUPERIOR 30

LAS ESCALAS EN LOS EJES SON: X 1: 0.460 Y 1: 0.320

LA BANDA DE CONFIANZA SE ENCUENTRA EN LOS PUNTOS [-2.241, 2.241]

@ QUIERES HACER OTRA GRAFICA ?

SI
 @ CONTRA RUE MATRIZ DESEAS GRAFICAR LOS RESIDUALES ?
 YE
 @ QUIERES BANDA DE CONFIABILIDAD EN LA GRAFICA ?
 NO
 @ QUIERES LA GRAFICA POR TERMINAL ?
 SI
 @ DAME EL NUMERO DE RENGLONES DE LA GRAFICA

35

I 4.00

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

I

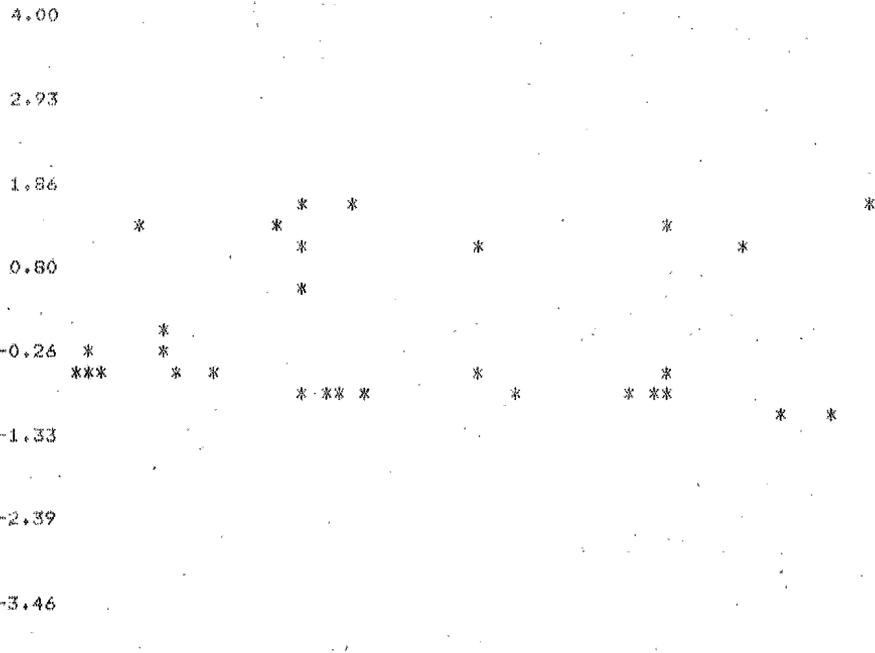
I

I

I

I

I



LOS LIMITES EN EL EJE X SON: INFERIOR -16618
 SUPERIOR 7353

LAS ESCALAS EN LOS EJES SON: X 1:380.492 Y 1: 0.267

@ QUIERES HACER OTRA GRAFICA ?

LOS LIMITES EN EL EJE X SON: INFERIOR
SUPERIOR

-979
432

LAS ESCALAS EN LOS EJES SON: X 1: 22.397 Y 1: 0.229

@ QUIERES HACER OTRA GRAFICA ?

NO

@ DAME COMANDO DE EJECUCION

LISTA.*

@ POR DONDE DESEAS LA IMPRESION DE LA MATRIZ ?

TERM

@ VECTOR Y-ESTIMADA

-14890.194

-16617.461

3400.190

1115.249

-4306.178

-13597.169

-9811.549

-7961.183

1290.826

-15942.946

1411.992

-8838.461

7352.648

728.546

-9948.738

-12551.939

-13954.826

-3357.763

-10696.175

298.054

-16303.536

-8738.965

-14008.898

-16274.624

-9735.265

6216.391

-3231.562

-9883.639

-4283.254

4760.997

*

@ VECTOR Y

-14888.709

-16618.042

3401.238

1114.306

-4306.723

-13597.811

-9811.139

-7961.941

1290.342

-15943.540

1413.408

-8839.201

7354.403

727.610

-9947.056

-12552.603

-13953.086

-3358.615

-10694.771

297.127

-16303.682

-8739.707

-14008.932

-16275.211

-9734.142

6215.343

-8229.832

-9884.358

-4282.089

4759.979

*

@ VECTOR BETA-ESTIMADA

12.921

17.000

*

@ MATRIZ X

1.000 -876.636

1.000 -978.238

1.000 199.247

1.000 64.842

1.000 -254.059

1.000 -800.577

1.000 -577.898

1.000 -469.055

1.000 75.149

1.000 -938.561

1.000 82.297

1.000 -520.659

1.000 431.740

1.000 42.095

1.000 -585.968

1.000 -739.094

1.000 -821.615

1.000 -198.271

1.000 -629.934

1.000 16.772

1.000 -959.772

1.000 -514.806

1.000 -824.796

1.000 -958.071

1.000 -573.411

1.000 364.903

1.000 -484.960

1.000 -582.139

1.000 -252.711

1.000 279.293

*

@ MATRIZ Z

1.000 -876.636

1.000 -978.238

1.000 199.247

1.000 64.842

1.000 -254.059

1.000 -800.577

1.000 -577.898

1.000 -469.055

1.000 75.169

1.000 -938.561

1,000	82,297
1,000	-520,659
1,000	431,740
1,000	42,095
1,000	-595,968
1,000	-739,094
1,000	-821,615
1,000	-198,271
1,000	-629,934
1,000	16,772
1,000	-959,772
1,000	-514,806
1,000	-824,796
1,000	-958,071
1,000	-573,411
1,000	364,903
1,000	-484,960

1.000 -582.139

1.000 -252.711

1.000 279.293

*

@ VECTOR DE RESIDUALES

1.484

-0.581

1.048

-0.944

-0.544

-0.642

0.410

-0.758

-0.483

-0.594

1.414

-0.740

1.755

-0.936

1.682

-0.664

-0.260

-0.852

1.404

-0.927

-0.146

-0.742

-0.033

-0.588

1.123

-1.048

1.730

-0.719

0.010

1.165

-1.018

*

@ VECTOR DE RESIDUALES ESTANDARIZADOS

1 1.455

-0.569

1.027

-0.925

-0.533

-0.630

0.402

-0.743

-0.474

-0.593

1.388

-0.725

1.720

-0.917

1.648

-0.651

-0.255

-0.835

1.376

-0.908

-0.143

-0.727

-0.032

-0.576

1.101

-1.027

1.695

-0.704

1.142

-0.998

*

@ MATRIZ INVERSA DE Z-PRIMA * Z

0.062 0.000

0.000 0.000

*

@ DURBIN-WATSON = 3.108
@ VARIANZA ESTIMADA = 1.040
DESV. ESTAN. ESTIMA.= 1.020
Y-PRIMA * Y = 2984486094.248
Y-BARRA = -6778.648
@ DAME COMANDO DE EJECUCION
FIN
#ET=32:02.5 PT=8.7 IO=0.8

BIBLIOTECA
INSTITUTO DE ECOLOGIA
UNAM



ANEXO 1

BURROUGHS B670) ALGOL COMPILER, VERSION 2.8.060, WEDNESDAY, 02/15/78, 03:25 AM.

(H H B S) C A N C E / C O D E 7 8 0 = O N P A C K
= = = = =

```
BEGIN
                                0"000100 000:0000:0
                                E."000 IS SEGMENT 00003
FILE ARCH1(KIND=RENTE,"YLSE=10) 1 0"000200 003:0000:1
                                ,SAL(KIND=DISK,MAXRECSIZE=14,FLEXIBLE,SAVEFACTOR=99)
                                0"000300 003:0000:1
                                DATA IS 0005 LONG
                                0"000400 003:0000:1
                                DATA IS 0006 LONG
INTEGER I
                                0"000500 003:0000:1
                                ,J
                                0"000600 003:0000:1
                                ;
                                0"000700 003:0000:1
ARRAY TI(0:31)
                                0"000800 003:0000:1
POINTER P;
                                0"000900 003:0000:1
WRITE(ARCH1,<"DAME EL NOMBRE DEL ARCHIVO A CREAR">);
                                0"001000 003:0000:1
READ(ARCH1,4,TI(+1));
                                0"001100 003:0005:2
SCAN P:POINTER(TI) UNTIL EOL " ";
                                0"001200 003:0008:2
REPLACE P BY ".";
                                0"001300 003:0000:4
REPLACE SAL.TITLE BY TI;
                                0"001400 003:000E:5
WRITE(ARCH1,<"DAME EL NUMERO DE: OBSERVACIONES, VARIABLES">);
                                0"001500 003:0010:5
                                DATA IS 0009 LONG
READ(ARCH1,/,M,N);
                                0"001600 003:0015:2
BEGIN
                                0"001700 003:0021:2
INTEGER I
                                2 0"001800 003:0021:2
                                E."001 IS SEGMENT 00008
                                0"001900 008:0000:1
                                ,J
                                0"002000 008:0000:1
                                ;
                                0"002100 008:0000:1
ARRAY A(0:N-1);
                                0"002200 008:0003:3
WRITE(ARCH1,<"ME TIENES QUE DAR RENGLONES DE ",I2," VALORES">,N);
```

```

FOR I:=0 STEP 1 UNTIL N-1 DO
  BEGIN
    WRITE(ARCHI, <"DAME EL REGLON ", I2>, I+1);
    READ(ARCHI, //, FOR J:=0 STEP 1 UNTIL N-1 DO A[J]);
    WRITE(SAL, //, FOR J:=0 STEP 1 UNTIL N-1 DO A[J]);
  END;
END;

LOCK(SAL);
END.

```

```

DATA IS 000B LONG
0002300 008:000A:2
0002400 008:000F:1
3 0002500 008:000F:1
0002600 008:0017:2
0002700 008:0027:2
0002800 008:0034:2
3 0002900 008:0034:5
R,0001(008) IS 0041 LONG
2 0003000 003:0022:0
0003100 003:0023:3
R,0000(003) IS 0033 LONG
DATA IS 0014 LONG

```

```

=====
NUMBER OF ERRORS DETECTED = 0.
NUMBER OF SEGMENTS = 8. TOTAL SEGMENT SIZE = 167 WORDS. CORE ESTIMATE = 2415 WORDS. STACK ESTIMATE = 21
PROGRAM SIZE = 31 CARDS, 190 SYNTACTIC ITEMS, 21 DISK SEGMENTS.
PROGRAM FILE NAME: (RM35)CANDE/CODE780 OR PACK.
COMPILATION TIME = 97.147 SECONDS ELAPSED; 0.828 SECONDS PROCESSING; 0.560 SECONDS I/O.
=====

```

ANEXO 2

La siguiente tabla enlista las características de las funciones externas utilizables en las transformaciones, aparte de las básicas - se usará <ae> para indicar una expresión aritmética.

NOMBRE SIMBOLICO	PARAMETROS	RESULTADO
ABS(<ae>)		Valor absoluto de <ae>
ARCCOS(<ae>)		Valor principal del -- arco-coseno de <ae>, -1 < <ae> < 1
ARCSIN(<ae>)		Valor principal del -- arco-seno de <ae>, -1 < <ae> < 1
ARCTAN(<ae>)		Valor principal del -- arco-tangente de <ae>
ARCTAN2(<ae1>, <ae2>)	Real, Real	Valor principal del -- arco-tangente de <ae1> / <ae2>
ATANH(<ae>)		Arco-tangente hiperbólica de <ae>
COS(<ae>)		Coseno de <ae>
COSH(<ae>)		Coseno hiperbólico de <ae>
COTAN(<ae>)		Cotangente de <ae>
(<ae1>) DIV (<ae2>)		Cociente entero de -- <ae1> / <ae2>, donde -- <ae1> y <ae2> son expresiones aritméticas (no necesariamente enteros), tal que <ae2> ≠ 0.
EXP(<ae>)		Exponencial de <ae>
INTEGER(<ae>)		ENTIER (<ae> + .5)
INTEGERT(<ae>)		Parte entera de <ae>
LN(<ae>)		Logaritmo natural de - <ae>

NOMBRE SIMBOLICO	PARAMETROS	RESULTADO
LOG(<ae>)		Logaritmo base 10 de <ae>
MAZ(<ae1>, ..., <aeN>)	Real, ..., Real	Máximo de los valores <ae1>, ..., <aeN>, tal que $N > 1$
MIN(<ae1>, ..., <aeN>)	Real, Real	Mínimo de los valores <ae1>, ..., <aeN>, tal que $N > 1$
<ae1> MOD <ae2>		<ae1> modulo <ae2> -- donde <ae1> y <ae2> son expresiones aritméticas (no necesariamente enteras) tal que <ae2> $\neq 0$
NABS(<ae>)		Valor absoluto negativo de <ae>
RANDOM(<ae>)	Real Llamado por nombre	Genera un número aleatorio entre 0 y 1. Más detalle ver capítulo 2.
SIGN(<ae>)		1 si <ae> > 0 , 0 si <ae> = 0, -1 si <ae> < 0
SIN(<ae>)		Seno de <ae>
SINH(<ae>)		Seno hiperbólico de <ae>
SORT(<ae>)		Raíz cuadrada de <ae>, <ae> ≥ 0 .
TAN(<ae>)		Tangente de <ae>
TANH(<ae>)		Tangente hiperbólica de <ae>

Todas las funciones antes enlistadas son de argumento real y se obtiene como resultado también un número real. Como nota aclaratoria, en los casos en los cuales en la columna de parámetros aparece real querrá decir que el argumento de la función debe ser un número real y podrá utilizarse dicho parámetro por nombre o por valor; -- donde por nombre quiere decir que debe de ser un término literal, es decir no se puede poner RANDOM (544287). Por valor acepta tanto por nombre como el número.

Cuando no se haga uso de la columna parámetros se -- podrá utilizar expresiones aritméticas. Se debe mencionar que las funciones de esta tabla no son todas con la que cuenta el lenguaje ALGOL de la B-6700 (BURROUGHS 6700) y además que DIV y MOD no son consideradas como tal.

ANEXO 3

TABLA DE DIVISORES PARA INTERVALOS DE CONFIANZA DE LARGO MINIMO

Coefficiente de Confianza = ϵ , Grados de Libertad = n ,
 a_n es el número superior y b_n el inferior.

(Para $n \geq 30$ ver Tate y Klett (1959)).

$n \backslash \epsilon$	ϵ				
	.900	.950	.990	.995	.999
2	.2104 18.0077	.1025 21.4812	.0201 29.1362	.0100 32.3240	.0020 39.5708
3	.5821 17.6381	.3513 20.7437	.1148 27.5102	.0717 30.3027	.0244 36.5959
4	1.0561 18.1062	.7082 21.0632	.2969 27.4603	.2069 30.0848	.0908 35.9845
5	1.5938 18.9081	1.1392 21.8001	.5534 28.0269	.4113 30.5697	.2102 36.2654
6	2.1750 19.8739	1.6233 22.7410	.8700 28.8928	.6747 31.3966	.3806 36.9947
7	2.7883 20.9303	2.1473 23.7944	1.2350 29.9229	.9871 32.4106	.5979 37.9541
8	3.4262 22.0405	2.7027 24.9147	1.6397 31.0507	1.3406 33.5358	.8560 39.0631
9	4.0840 23.1844	3.2836 26.0769	2.0775 32.2397	1.7288 34.7308	1.1499 40.2631
10	4.7584 24.3498	3.8855 27.2662	2.5434 33.4685	2.1469 35.9714	1.4755 41.5223
11	5.4467 25.5294	4.5054 28.4733	3.0334 34.7240	2.5906 37.2430	1.8287 42.8238
12	6.1472 26.7180	5.1409 29.6920	3.5447 35.9963	3.0573 38.5330	2.2078 44.1445
13	6.8583 27.9126	5.7899 30.9184	4.0744 37.2809	3.5439 39.8378	2.6086 45.4880
14	7.5788 29.1109	6.4510 32.1497	4.6205 38.5733	4.0483 41.1517	3.0296 46.8441

n	ε				
	.900	.950	.990	.995	.999
15	8.3078	7.1227	5.1813	4.5685	3.4676
	30.3113	33.3842	39.8715	42.4732	48.2150
16	9.0446	7.8043	5.7559	5.1040	3.9248
	31.5125	34.6197	41.1710	43.7951	49.5766
17	9.7883	8.4947	6.3425	5.6523	4.3954
	32.7139	35.8560	42.4728	45.1206	50.9511
18	10.5385	9.1932	6.9402	6.2128	4.8806
	33.9148	37.0919	43.7748	46.4465	52.3245
19	11.2947	9.8991	7.5481	6.7846	5.3786
	35.1148	38.3271	45.0765	47.7723	53.6990
20	12.0563	10.6119	8.1654	7.3666	5.8882
	36.3137	39.5611	46.3772	49.0974	55.0743
21	12.8230	11.3310	8.7915	7.9580	6.4085
	37.5112	40.7936	47.6767	50.4216	56.4507
22	13.5946	12.0561	9.4259	8.5588	6.9406
	38.7070	42.0243	48.9736	51.7426	57.8190
23	14.3706	12.7868	10.0679	9.1679	7.4824
	39.9011	43.2532	50.2686	53.0616	59.1857
24	15.1508	13.5227	10.7169	9.7845	8.0322
	41.0935	44.4802	51.5619	54.3793	60.5545
25	15.9351	14.2636	11.3728	10.4088	8.5919
	42.2840	45.7051	52.8521	55.6935	61.9157
26	16.7230	15.0090	12.0348	11.0396	9.1580
	43.4728	46.9281	54.1407	57.0065	63.2808
27	17.5145	15.7587	12.7024	11.6764	9.7293
	44.6598	48.1491	55.4277	58.3186	64.6514
28	18.3095	16.5128	13.3767	12.3211	10.3146
	45.8446	49.3675	56.7096	59.6230	65.9955
29	19.1076	17.2706	14.0554	12.9699	10.9003
	47.0279	50.5843	57.9914	60.9295	67.3589

ANEXO 4

Distribución Acumulativa del Número Total de Corridas u en Muestras de Tamaño (n_1, n_2)

$(n_1, n_2) \backslash u$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
(2, 3)	0.200	0.500	0.900	1.000															
(2, 4)	0.133	0.400	0.800	1.000															
(2, 5)	0.095	0.333	0.714	1.000															
(2, 6)	0.071	0.286	0.643	1.000															
(2, 7)	0.056	0.250	0.583	1.000															
(2, 8)	0.044	0.222	0.533	1.000															
(2, 9)	0.036	0.200	0.491	1.000															
(2, 10)	0.030	0.182	0.455	1.000															
(3, 3)	0.100	0.300	0.700	0.900	1.000														
(3, 4)	0.057	0.200	0.543	0.800	0.971	1.000													
(3, 5)	0.036	0.143	0.429	0.714	0.929	1.000													
(3, 6)	0.024	0.107	0.345	0.643	0.881	1.000													
(3, 7)	0.017	0.083	0.283	0.583	0.833	1.000													
(3, 8)	0.012	0.067	0.236	0.533	0.788	1.000													
(3, 9)	0.009	0.055	0.200	0.491	0.745	1.000													
(3, 10)	0.007	0.045	0.171	0.455	0.706	1.000													
(4, 4)	0.029	0.114	0.371	0.629	0.886	0.971	1.000												
(4, 5)	0.016	0.071	0.262	0.500	0.786	0.929	0.992	1.000											
(4, 6)	0.010	0.048	0.190	0.405	0.690	0.881	0.976	1.000											
(4, 7)	0.006	0.033	0.142	0.333	0.606	0.833	0.954	1.000											
(4, 8)	0.004	0.024	0.109	0.279	0.531	0.788	0.929	1.000											
(4, 9)	0.003	0.018	0.085	0.236	0.471	0.745	0.902	1.000											
(4, 10)	0.002	0.014	0.068	0.203	0.419	0.706	0.874	1.000											
(5, 5)	0.008	0.040	0.167	0.357	0.643	0.833	0.960	0.992	1.000										
(5, 6)	0.004	0.024	0.110	0.262	0.522	0.738	0.911	0.976	0.998	1.000									
(5, 7)	0.003	0.015	0.076	0.197	0.424	0.652	0.854	0.955	0.992	1.000									
(5, 8)	0.002	0.010	0.054	0.152	0.347	0.576	0.793	0.929	0.984	1.000									
(5, 9)	0.001	0.007	0.039	0.119	0.287	0.510	0.734	0.902	0.972	1.000									
(5, 10)	0.001	0.005	0.029	0.119	0.239	0.455	0.678	0.874	0.958	1.000									
(6, 6)	0.002	0.013	0.067	0.175	0.392	0.608	0.825	0.933	0.987	0.998	1.000								
(6, 7)	0.001	0.008	0.043	0.121	0.296	0.500	0.733	0.879	0.966	0.992	0.999	1.000							
(6, 8)	0.001	0.005	0.028	0.086	0.226	0.413	0.646	0.821	0.937	0.984	0.998	1.000							
(6, 9)	0.000	0.003	0.019	0.063	0.175	0.343	0.566	0.762	0.902	0.972	0.994	1.000							
(6, 10)	0.000	0.002	0.013	0.047	0.137	0.288	0.497	0.706	0.864	0.958	0.990	1.000							
(7, 7)	0.001	0.004	0.025	0.078	0.209	0.383	0.617	0.791	0.922	0.975	0.996	0.999	1.000						
(7, 8)	0.000	0.002	0.015	0.051	0.149	0.296	0.514	0.704	0.867	0.949	0.988	0.998	1.000	1.000					
(7, 9)	0.000	0.001	0.010	0.035	0.108	0.231	0.427	0.622	0.806	0.916	0.975	0.994	0.999	1.000					
(7, 10)	0.000	0.001	0.006	0.024	0.080	0.182	0.355	0.549	0.743	0.879	0.957	0.990	0.998	1.000					
(8, 8)	0.000	0.001	0.009	0.032	0.100	0.214	0.405	0.595	0.786	0.900	0.968	0.991	0.999	1.000	1.000				
(8, 9)	0.000	0.001	0.005	0.020	0.069	0.157	0.319	0.500	0.702	0.843	0.939	0.980	0.996	0.999	1.000	1.000			
(8, 10)	0.000	0.000	0.003	0.013	0.048	0.117	0.251	0.419	0.621	0.782	0.903	0.964	0.990	0.998	1.000	1.000	1.000		
(9, 9)	0.000	0.000	0.003	0.012	0.044	0.109	0.238	0.399	0.601	0.762	0.891	0.956	0.988	0.997	1.000	1.000	1.000		
(9, 10)	0.000	0.000	0.002	0.008	0.029	0.077	0.179	0.319	0.510	0.681	0.834	0.923	0.974	0.992	0.999	1.000	1.000	1.000	
(10, 10)	0.000	0.000	0.001	0.004	0.019	0.051	0.128	0.242	0.414	0.586	0.758	0.872	0.949	0.981	0.996	0.999	1.000	1.000	1.000

ANEXO 5

**Puntos de significancia de 5% y 1% para la razón de la diferencia
sucesiva del cuadrado medio a la varianza**

Valores de $\frac{\delta^2}{s^2}$ para diferentes valores de significancia

Valores de k			Valores de k'		Valores de k			Valores de k'	
n	$P = .01$	$P = .05$	$P = .95$	$P = .99$	n	$P = .01$	$P = .05$	$P = .95$	$P = .99$
4	.8341	1.0406	4.2927	4.4992	31	1.2469	1.4746	2.6587	2.8864
5	.6724	1.0255	3.9745	4.3276	32	1.2570	1.4817	2.6473	2.8720
6	.6733	1.0682	3.7318	4.1262	33	1.2667	1.4885	2.6365	2.8583
7	.7163	1.0919	3.5748	3.9504	34	1.2761	1.4951	2.6262	2.8451
8	.7575	1.1228	3.4486	3.8139	35	1.2852	1.5014	2.6163	2.8324
9	.7974	1.1524	3.3476	3.7025	36	1.2940	1.5075	2.6068	2.8202
10	.8353	1.1803	3.2642	3.6091	37	1.3025	1.5135	2.5977	2.8085
11	.8706	1.2062	3.1938	3.5294	38	1.3108	1.5193	2.5889	2.7973
12	.9033	1.2301	3.1335	3.4603	39	1.3188	1.5249	2.5804	2.7865
13	.9336	1.2521	3.0812	3.3996	40	1.3266	1.5304	2.5722	2.7760
14	.9618	1.2725	3.0352	3.3458	41	1.3342	1.5357	2.5643	2.7658
15	.9880	1.2914	2.9943	3.2977	42	1.3415	1.5408	2.5567	2.7560
16	1.0124	1.3090	2.9577	3.2543	43	1.3486	1.5458	2.5494	2.7466
17	1.0352	1.3253	2.9247	3.2148	44	1.3554	1.5506	2.5424	2.7376
18	1.0566	1.3405	2.8948	3.1787	45	1.3620	1.5552	2.5357	2.7289
19	1.0766	1.3547	2.8675	3.1456	46	1.3684	1.5596	2.5293	2.7205
20	1.0954	1.3680	2.8425	3.1151	47	1.3745	1.5638	2.5232	2.7125
21	1.1131	1.3805	2.8195	3.0869	48	1.3802	1.5678	2.5173	2.7049
22	1.1298	1.3923	2.7982	3.0607	49	1.3856	1.5716	2.5117	2.6977
23	1.1456	1.4035	2.7784	3.0362	50	1.3907	1.5752	2.5064	2.6908
24	1.1606	1.4141	2.7599	3.0133	51	1.3957	1.5787	2.5013	2.6842
25	1.1748	1.4241	2.7426	2.9919	52	1.4007	1.5822	2.4963	2.6777
26	1.1883	1.4336	2.7264	2.9718	53	1.4057	1.5856	2.4914	2.6712
27	1.2012	1.4426	2.7112	2.9528	54	1.4107	1.5890	2.4866	2.6648
28	1.2135	1.4512	2.6969	2.9348	55	1.4156	1.5923	2.4819	2.6585
29	1.2252	1.4594	2.6834	2.9177	56	1.4203	1.5955	2.4773	2.6524
30	1.2363	1.4672	2.6707	2.9016	57	1.4249	1.5987	2.4728	2.6465
					58	1.4294	1.6019	2.4684	2.6407
					59	1.4339	1.6051	2.4640	2.6350
					60	1.4384	1.6082	2.4596	2.6294

Estadística de Durbin-Watson (d). Puntos de Significancia d_L Y d_U

:38

:11

n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5	
	d_L	d_U								
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5	
	d_L	d_U								
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

n = número de observaciones

k' = número de variables independientes
excluyendo el término constante.

BIBLIOGRAFIA

Anton H., Elementary Linear Algebra, John Wiley, 1973.

Burroughs, Input/Output Subsystem, Information Manual,
B6700, 1971.

Burroughs, Languaje Reference Manual, ALGOL, B6700/B77
00, 1974

Burroughs, Mathematical Intrinsic, Information Manual,
B6700, 1971.

Burroughs, System Software, Handbook, B6700/B7700, 1972.

Burroughs, Using Pinter Expressions in ALGOL on B6700/B
7700, 1971.

Collected ACM Algorithms, 1960-1977.

Conover W. J., Practical Nonparametric Statistics, Wiley
Internacional Edition, 1971.

Draper N. y Smith H., Applied Regresion Analysis, John -
Wiley, 1966.

Graybill F. A., An Introduction to Linear Statistical --
Models, Volume I, Mc Graw-Hill, 1961.

Gregory R.T. y Karney D.L., A collection of Matrices for
testing Computacional Algorithms, John Wiley, 1969.

Hart B.E., "Significance Levels for the Ratio of the Mean Square Successive Difference to the Variance", Annals of Mathematical Statistics, 1942, Vol. 13, No. 4, - pp. 445-447.

Johnston N.L. y Kots S., Continuous Univariate Distributions-1, Houghton Mifflin Company, 1970.

Johnston N.L. y Kots S., Continuous Univariate Distributions-2, Houghton Mifflin Company, 1970.

Johnston J., Econometric Methods. Mc Graw-Hill Kogakusha, 1972.

Lang S., Linear Algebra, John Wiley, 1973.

Mood A.M. y Graybill F.A., Introducción a la teoría de la Estadística, Mc Graw-Hill, 1972.

Mood A.M., Graybill F.A. y Boes D.C., Introduction to the theory of Statistics, Mc Graw-Hill Kogakusha, 1963.

Naylor T.H., Balintfy J.L., Burdick D.S. y CHU K., técnicas de Simulación en Computadoras, LIMUSA, 1973.

Newmann T.G. y Odell P.L., the Generation of Random Variables, Griffin's Statistical Monographs & Courses, -- 1971.

Rao R., Linear Statistical Inference and Its Applications,
John Wiley, 1973.

Scheffé H., the Analysis of Variance, John Wiley, 1959.

Searle S.R., Linear Models, John Wiley, 1971.

Tate R. F. y Klett G. W., "Optimum Confidence Intervals -
for the Variance of a Normal Distributions", Journal
of the American Statistical Association, 1959, Vol.
54, p.p. 674-682.