SERIE MISCELANEA 233
AGROCRED No. 03
JUNIO 1980

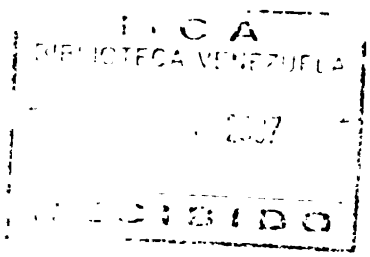
INTRODUCCION AL USO DEL PROGRAMA SAS PARA ANALISIS DE REGRESION

Carlos Pomareda

IICA

INSTITUTO INTERAMERICANO DE CIENCIAS AGRICOLAS – OEA

PROYECTO DE SEGURO AGROCREDITICIO



INTRODUCCION AL USO DEL PROGRAMA SAS PARA ANALISIS DE REGRESION

Carlos Pomareda

00000359

PRESENTACION

Las relaciones causa-efecto son comunes en la agricultura. Se pueden presentar por ejemplo en los procesos productivos, si consideramos que una cantidad dada de un producto se puede obtener con diferentes combinaciones de factores. Este tipo de relaciones, denominado la función de producción explica la relación entre cantidades de fertilizante, agua, tierra, maquinaria y mano de obra, para producir granos u otros cultivos y desde luego permite evaluar el grado en que estos factores pueden substituirse entre sí. Puede tratarse de la combinación de alimentos, concentrados, agua, forraje, sales minerales, para la alimentación de ganado y en esa forma alcanzar diferentes niveles de producción de leche o de ganancia de peso.

Otro tipo de relación causa-efecto se presenta en la producción agrícola sobre el tiempo,en donde las expectativas sobre los precios de los productos considerados como alternativas a nivel de la finca, modifican las decisiones de los agricultores y por consiguiente las áreas sembradas y los niveles de uso de insumos que en última instancia, dependiendo de las condiciones de suelo, clima y agua, afectan los volúmenes producidos. Este tipo de relaciones se denomina la función de oferta de los productos. El conocimiento de los parámetros de este tipo de relaciones funcionales es necesario para la formulación de políticas de precio.

El comportamiento de los consumidores se refleja en las relaciones causa-efecto entre precios e ingresos y el consumo. Este tipo de análisis provee en esencia un conocimiento de los cambios que puedan ocurrir en el consumo de un producto cuando cambien su precio y/o los de sus substitutos y complementos mas cercanos y los ingresos.

El análisis de regresión es la técnica estadística matemática que nos permite establecer la magnitud de los parámetros en los diferentes tipos de relaciones funcionales y nos da pautas sobre el grado de confiabilidad que se puede tener en dichos parámetros. Un estudio profundo de la econometría demanda un texto completo en sí; sin embargo es posible introducir al estudiante a los conceptos básicos, lo cual se ha propuesto el autor en su texto "Métodos Cuantitativos para la Investigación en Economía Agrícola", en donde se discute en primera instancia el modelo de regresión simple y los supuestos básicos en él; luego se progresa al modelo de regresión múltiple y al tipo de problemas que generalmente se encuentran para su estimación. Con el objeto de presentar en la forma más comprensible posible una discusión de la teoría y la estimación empírica de las relaciones funcionales mas comunes (funciones de producción, de oferta y de demanda) se

discute la teoría y luego se presentan casos, cada uno con particularidades en la formulación y en la estimación empírica y en el tipo de análisis económico que se hace de los resultados.

La estimación empírica de relaciones funcionales entre variables usando métodos econométricos, se facilita en gran medida por la disponibilidad de centros de cómputo y de programas apropiados. SAS es uno de los programas mas versátiles y que entre sus muchos procedimientos tiene varios para el análisis de regresión. El propósito de este manual es el de presentar algunas de las características mas importantes de los procedimientos de SAS para análisis de regresión e ilustrar su uso con algunos ejemplos.

El IICA, en la Sede Central, San José, dispone de un centro de cómputo dotado de varios programas ad hoc para el manejo y análisis de información. Entre estos programas se tiene el SAS. Los técnicos del IICA en los países y en la Sede Central y otros técnicos e investigadores de los problemas del desarrollo rural en América Latina están invitados a utilizar estos servicios que ahora provee el IICA en su Sede Central.

C O N T E N I D O

INTRODUCCION AL USO DEL PROGRAMA 'SAS'

PARA ANALISIS DE REGRESION

1. INTRODUCCION

El desarrollo científico demanda el manejo y análisis de cantidades masivas de información; lo cual no podría hacerse sin el acceso a Centros de computación dotados de los programas apropiados.

Existen muchos programas ad-hoc para el manejo y análisis estadístico de información y uno de ellos, el SAS (Statistical Analysis System) es sumamente versátil, de fácil utilización y de una riqueza extraordinaria en procedimientos.[1/] SAS fue desarrollado a inicios de la década de 1970 por A. J. Barr y J. H. Goodnight y la primera versión comercial conocida apareció en el año 1972. Después de eso y con la contribución de varios cientos de profesionales se han publicado las ediciones de 1976 y 1979.[2/]

SAS es un sistema de computación para el análisis de datos, que provee los medios necesarios para el almacenamiento y recuperación de información; modificación y programación de los datos; preparación de reportes, análisis estadístico y manejo de archivos.

1/ El uso de SAS es posible en máquinas IBM 360 y 370 (y otras compatibles como Amdahl, Itel, CDC Omega, Magnuson, Ryad) bajo los sistemas OS y OS/VS.

2/ Todos los procedimientos que aquí se discuten se basan en la versión mas reciente en 1979.

Estas funciones son realizables a través del uso de varias docenas de procedimientos. En este capítulo se hace referencia a los de mayor uso en el análisis estadístico y se da particular énfasis a la discusión de los procedimientos para el análisis de regresión.

Quienes realizan investigación fundamentada en información tienen en términos generales una cantidad de datos y una serie de preguntas a los que quieren encontrar respuestas mediante el análisis de esos datos. La organización de los datos y su incorporación en un set accesible por los procedimientos de SAS se consiguen mediante el paso (step) **DATA**. A través de este paso SAS permite manejar los datos y modificarlos, crear archivos y preparar reportes; es decir el paso **DATA** le dice a SAS cómo procesar los datos. Las preguntas que se desea hacer acerca de los datos se manejan a través del paso **PROC**, el cual le dice a SAS que hacer para obtener lo que el investigador desea. Al darse la instrucción de uno de los procedimientos en **PROC**, SAS llama el programa referido de su biblioteca para analizar los datos.

2. GENERALIDADES SOBRE EL PASO **DATA**

La Guía de uso de SAS, Edición 1979, presenta en su parte II, diez capítulos que describen las instrucciones para lectura e impresión de datos, el tratamiento de datos faltantes, almacenamiento y recuperación de

datos, creación y ordenación de sets de datos a partir de input data, elaboración de subsets de datos dentro del set creado, elaboración de reportes de los datos (listados), modificación de los datos que originalmente aparecen en la fuente de datos (input data) o seleccionar solo algunas de las observaciones.

En esta sección se pretende dar sólo una visión muy general de las instrucciones mas usadas dentro del paso **DATA**, para garantizar el conocimiento mínimo acerca del manejo de información, como condición previa y necesaria para la utilización de las instrucciones para el análisis de regresión.

El primer paso necesario es la creación del set de datos; es decir los datos deben ser leídos desde algún registro para conformar el set de datos correspondiente. Para tal propósito SAS provee al usuario un conjunto de instrucciones dentro de las cuales se destacan las siguientes:

1) **DATA**: usualmente es la primera instrucción de un programa SAS y tiene por función inicar la creación del set de datos y asignarle un nombre. Su expresión simbólica es: **DATA** nombre de los datos;

El caracter (;) es el delimitador de final de la instrucción **DATA**.

Por Ejemplo:
```
DATA DEMANDA;
DATA JORGITO;
DATA FABI;
```

Cada set de datos generado debe llevar un nombre único; sin embargo cuando el usuario no asigna dicho nombre, el sistema asigna nombres a los sets de datos con la palabra DATA, seguida de un numeral consecutivo **DATAØØØ1**, **DATAØØØ2**, etc.

2) **INPUT**: describe al sistema SAS la presentación de una línea de datos, entendiéndose por 'línea' una tarjeta perforada, una hilera de terminal, disco o cinta magnética. En todo caso la máxima longitud de la línea es de 32,000 caracteres. La instrucción **INPUT** para datos números puede adquirir diferentes formas.[3/]

a. Lista de variables numéricas, cuando los datos están perforados en hileras ignorando las columnas; pero los valores numéricos de las variables requieren estar separados por lo menos por un espacio en blanco. Por ejemplo:

```
INPUT    AÑO    QM    QA    QF    QS    I    Y;
```

b. Lista de variables numéricas con subíndice, cuando se especifican únicamente la primera y última variable y quedan sobreentendidas las que están en el intervalo. Por ejemplo.

```
INPUT  X1 - X6;
INPUT QM -  Y;
```

c. Variables numéricas en columnas, que se usa para describir datos que se encuentran arreglados en columnas fijas. En este caso se especifica el campo que ocupa el primer dígito y el campo del último, separados por un guión. Por ejemplo:

```
INPUT    AÑO    1-2    QM    4-9    QA    12-15    etc ;
```

3/ Se puede manejar además variables alfanuméricas.

lo cual indica que la variable AÑO toma valores que se encuentran perforados en las columnas 1 al 2; la variable QM toma los datos perforados en las columnas 4 al 19, etc.

d. Variables numéricas en columnas con decimales, que se usa cuando la información lleva decimales y los mismos fueron perforados en tarjetas, pero sin el punto decimal. Se especifica el número de decimales en la instrucción **INPUT**. Por ejemplo:

INPUT PESO 10-12 3 EDAD 15-18 2 ... etc.

indicándose que a la variable PESO el sistema le agregará un punto decimal antes de los tres últimos dígitos.

e. Varias tarjetas o líneas por observación, que se usa cuando los datos de una observación están contenidos en mas de una tarjeta. En este caso se utiliza el caracter '#' seguido de un numeral para especificar la tarjeta que contiene tales datos. Por ejemplo:

INPUT X 1-4 Y 6-7 #2 A 2-4 B 6-8

según lo cual los datos de la variable X se encuentran en los espacios 1 a 4 de la primera tarjeta; los de Y en los espacios 6-7 de la primera tarjeta y los datos de A y B en la segunda tarjeta en los espacios 2 a 4 y 6 a 8 respectivamente.

Con frecuencia se usa el caracter '#' al final de la instrucción input. Por ejemplo: **INPUT PESO 6-8 #2 CANTIDAD 4-7 #3 INGRESO 2-6 #5;**

en donde el caracter '**#5**' al final de **INPUT** indica que la observación está formada por cinco tarjetas y los datos se leen unicamente en las primeras tres.

f. Transformaciones. Inmediatamente después de la sección **INPUT** se puede insertar cualquier transformación de las variables originales; es decir se puede crear nuevas variables. Por ejemplo:

PMY = PM/Y;	PM÷ Y
Q2 = Q * Q;	Q x Q
LQM = LOG QM;	Logaritmo neperiano de QM

etc.

3. **CARDS**: indica al sistema que los datos están perforados en tarjetas. De otro modo, si los datos estuviesen en cinta magnética o en disco se substituyen por la instrucción 'INFILE'. La instrucción se escribe

CARDS ;

e inmediatamente después se ponen las tarjetas de datos según el formato seleccionado en la sección **INPUT**.

Para el uso de los procedimientos para el análisis de regresión que se describen mas adelante, se considera que estas instrucciones son las fundamentales.

3. GENERALIDADES SOBRE EL PASO **PROC**

Una vez que se ha creado un set de datos, se puede utilizar los múltiples procedimientos de SAS para analizar los datos. Los procedimientos de SAS son programas que pueden leer los datos, realizar varias operaciones e imprimir los resultados de esas operaciones. Algunos procedimientos pueden también crear sets de datos que contienen los resultados de las operaciones realizadas. Las instrucciones que solicitan a SAS la aplicación de determinado procedimiento están compuestas de la instrucción **PROC** seguida del procedimiento específico. Por ejemplo:

```
PROC ANOVA ;
PROC AUTOREG;
etc...
```

En algunos casos la instrucción **PROC** es seguida de otras instrucciones que proveen información adicional acerca del análisis.

Algunos de los procedimientos mas utilizados, en forma previa o acompañando a un análisis de regresión son los que se refieren a continuación.

PROC ANOVA; el cual permite realizar un análisis de variancia para un set balanceado de datos. El procedimiento **GLM** también puede realizar análisis de variancia para datos balanceados o no balanceados pero **ANOVA** es más rápido y usa menos memoria. Con el **PROC ANOVA** se pueden usar algunas opciones.

PROC AUTOREG; estima los parámetros de un modelo lineal cuyo término de error se asume que sigue un proceso autoregresivo de un orden determinado (q) denotado por AR (q). **AUTOREG** debe ser usado sólo para datos de series históricas ordenadas e igualmente espaciadas sin datos faltantes. También con el **PROC AUTOREG** se pueden usar algunas opciones.

PROC CANCORR; ejecuta el análisis de correlación canóniga; la cual es de utilidad cuando se desea investigar la relación entre dos tipos de variables. Típicamente un grupo consiste de variables independientes y el otro de variables dependientes. Para cada uno de los grupos de variables, **CANCORR** halla una combinación lineal de variables, llamada la 'variable canóniga'tal que la correlación entre las dos variables canónigas es maximizada.

PROC CHART; que produce diagramas verticales y horizontales (histogramas), los cuales son muy útiles para ilustrar relaciones entre variables o para mostrar la evolución de una o mas variables.

PROC CORR; que permite estimar los coeficientes de correlación entre variables. Tiene además varias opciones como SPEARMAN para calcular coeficientes SPEARMAN que indican las correlaciones del rango de las variables; RANK, que solicita que los coeficientes de correlación sean impresos en orden de magnitud del mas alto al mas bajo; **BEST (BEST=n)** imprime para cada variable solo los n coeficientes de correlación mas altos.

PROC MEANS; que produce cuadros de las estadísticas simples ya sea para un set completo de datos o para un subset específico. Tiene también varias opciones. Si solo se especifica **PROC MEANS**; entonces imprime la media, desviación standard, valor mínimo, valor máximo, desviación standard de la media, suma, variancia y coeficiente de variación. Puede además incluir opcionales como **NHISS** (que da el número de valores que faltan); **RANGO** (rango de las variables); **USS** (suma de cuadrados sin corregir), **CSS** (suma de cuadrados corregida); **SKEWNESS** (una medida de dispersión), **KURTOSIS** (una medida de kurtosis); **T** (la estadística 't' para probar la hipótesis que la media de la población es cero y finalmente **PRT** (la probabilidad de valores mayores de 't').

PROC PLOT; uno de los procedimientos de mayor uso práctico para relacionar dos variables, grafica una variable contra otra. Las coordenadas de cada punto en el gráfico corresponden a los dos valores de las variables para una misma observación. **PLOT** selecciona en forma automática la escala del gráfico. El uso de **PLOT** requiere especificar:

PROC PLOT;

PLOT variable en el eje de las Y * variable en el eje de las X

Por ejemplo:

```
PROC PLOT;
PLOT Y * X
```

Se puede además especificar un caracter específico para representar los puntos. Por ejemplo

```
PROC PLOT
PLOT Y * X = '*'
```

Se puede solicitar también que varias relaciones de variables aparezcan ploteadas en la misma figura, para lo cual se usa **OVERLAY**. Por ejemplo

```
PROC PLOT;
PLOT PM * AÑO = '*' PA * AÑO = 'O' PF * AÑO = '*' / OVERLAY;
```

También se puede plotear los valores actuales usados en un modelo de regresión y los valores predichos por el modelo. Por ejemplo

modelo de regresión

```
PROC GLM;
MODEL QM = PM
OUTPUT OUT = BOTH   P = PREDHT

PROC PLOT DATA = BOTH;
PLOT QM * PM PREDHT * PM = '*' /OVERLAY;
TITLE PREDICCION VS ACTUAL;
```

y los valores actuales serán impresos con A's y los predichos con *'s. El procedimiento **PLOT** tiene además otras opciones.

PROC PRINT; que imprime un listado de algunas o de todas las variables en el set de datos. La instrucción básica es

```
PROC PRINT;
```

con lo cual se listarán todas las variables, incluyendo aquellas que no están en el set original pero que han sido creadas.

Además de estos procedimientos de uso complementario a un análisis de regresión, existen muchos otros que el investigador puede encontrar en un manual de SAS y que deben con plenitud satisfacer sus necesidades. En cuanto a los procedimientos para el análisis de regresión en sí, éstos son varios, con características similares pero con algunas peculiaridades para cada uno y por consiguiente aplicables a situaciones particulares condicionadas por la naturaleza de las investigaciones. En resumen estos procedimientos son:

GLM ; el cual permite estimar parámetros de regresión simple y múltiple, análisis de variancia y de covariancia, funciones de respuesta, regresión compensada (weigthed), regresión polinomial, correlación parcial y análisis multivariado de variancia.

. **NLIN**; para análisis de regresión no linear.

. **RSQUARE**; que permite estimar todas las regresiones posibles, y;

. **SYSREG**; para estimar regresiones por el método de los mínimos cuadrados ordinarios; mínimos cuadrados en dos etapas; máxima verosimilitud con información limitada; mínimos cuadrados en tres etapas y regresiones aparentemente no relacionadas.

STEPWISE; para estimar regresiones por pasos.

A continuación se describe en mayor detalle algunos de los procedimientos y se ilustra su utilización con ejemplos.

4. ANALISIS DE REGRESION SIMPLE Y MULTIPLE USANDO GLM

Este procedimiento es de suma versatilidad para el análisis de regresión múltiple y debe usarse en determinados casos particulares como por ejemplo cuando no se anticipa que existan problemas de simultaneidad en las ecuaciones, lo cual hará preciso recurrir a modelos de mínimos cuadrados en dos etapas y que por consiguiente deben ser resueltos usando el proceso **SYSREG**, que se discute en la sección 6.

A continuación se presenta las instrucciones básicas para el uso de GLM y luego se ilustra su manejo con un ejemplo.

4.1. PROC GLM;

El procedimiento **GLM** es uno de los mas eficientes para resolver modelos de regresión lineal simple y múltiple por el método de los mínimos cuadrados ordinarios.

La instrucción básica debe ir acompañada de las características del modelo, es decir especificar la variable dependiente y la (s) variable (s) independiente (s). Por ejemplo

```
PROC GLM;
MODEL Q1 = P1;
```

ó

```
PROC GLM;
MODEL Q1 = P1 P2 P3;
```

Si además se desease obtener las predicciones y residuos de la variable dependiente será preciso indicar:

```
PROC GLM;
MODEL Q1 = P1  P2  P3;
PREDICTED = Q1PRED
RESIDUAL = Q1RES
```

En el análisis de problemas de producción es preciso recurrir a modelos no lineales de regresión. Las funciones cuadráticas, cúbicas, de exponente 1.5 y raíz cuadrada, así como las funciones logarítmicas son comunes en el análisis de funciones de producción para cultivos y ganadería y a nivel de las empresas o del sector.

El procedimiento **GLM** puede usarse para regresión no lineal si previamente se definen las variables no lineales. Por ejemplo en la sección de transformaciones podemos especificar

```
X2 = X * X
```

y entonces el modelo puede especificarse como

```
MODEL Y = X X2 ;
```

indicando así que entre **X** e **Y** existe una relación cuadrática.

SAS provee además, a través del proceso **NLIN**, los medios para estimar relaciones no lineales entre variables; sin embargo este no será discutido aquí.

Los resultados del procedimiento **GLM** se presentan en cuatro secciones:

Análisis de variancia

Estadísticas misceláneas

Resultados de las pruebas I y IV

Reporte de los estimados de los parámetros

(1.) Análisis de Variancia, el cual muestra la suma total de los cuadrados de la variable dependiente, en las porciones atribuibles al modelo y al término de error.

(2.) Entre las estadísticas misceláneas se tienen:
'F' el cual es el cuociente entre la media cuadrada (MEAN SQUARE) del modelo y del error y es una prueba de como el modelo en conjunto explica el comportamiento de la variable dependiente. Si el valor de F es mayor que el de las tablas (de distribución de F) indica significancia.

(3.) **R-SQUARE**. EL coeficiente de determinación (R^2) mide cuanta de la variación de la variable dependiente es explicada por las variables independientes, pudiendo alcanzar un valor máximo de 1, lo cual indica que el 100 por ciento de la variación de la variable dependiente es explicado. El R^2 se obtiene de dividir la suma de cuadrados del modelo (SUM OF SQUARES-MODEL) entre la suma de cuadrados del total corregido (SUM OF SQUARES-CORRECTED TOTAL).

(4.) CV. El coeficiente de variación, usado como una medida de la variación de la población. Es igual a la desviación standard de la variable dependiente, dividida por la media, por 100.

(5.) STD DEV. La desviación standard de la variable dependiente.

(6.) 'dependiente' MEAN. La media de la variable dependiente.

(7.) Resultados de la prueba TYPE I. Esta prueba se usa principalmente en el análisis de variancia y mide el incremento en la suma de cuadrados del modelo a medida que cada variable es añadida. El correspondiente valor de F (comparado contra los valores en las tablas) indica si esta contribución es o no significativa.

(8.) Resultados de la prueba TYPE IV. Mide la suma de cuadrados debida a que la variable en referencia sea añadida como última en el modelo.

(9.) Resultados de los estimados de los Parámetros (coeficientes de regresión). Esta es la sección que en última instancia es de mayor importancia y da el valor de los parámetros (ESTIMATE) para el intercepto y para cada uno de los coeficientes de regresión. Asi mismo da el valor de la estadística 't' (para probar la hipótesis de que los coeficientes son iguales a cero, la probabilidad de que los coeficientes sean iguales a cero y el error standard del coeficiente de regresión.

Como se verá mas adelante, aunque los resultados en términos de R^2, magnitud de los coeficientes, valores de 't', etc.,obtenibles con los procedimientos **GLM** y **SYSREG** para modelos lineales resueltos por el método de los mínimos cuadrados ordinarios son escencialmente los mismos; los resultados son presentados en formatos diferentes.

Como se indicó en la sección sobre el uso del **PROC PLOT**, éste puede usarse con GLM para graficar valores actuales y predichos de la variable dependiente.

4.2. Ejemplo Ilustrativo para Regresión Lineal Múltiple

Con el fin de ilustrar la utilización del procedimiento **GLM** para el análisis de regresión, se presentan a continuación una serie de cuadros que no necesitan mas que una breve explicación adicional.

SAS fue utilizado en este caso para analizar la demanda de granos básicos en Honduras. Aquí se presenta solamente parte de la información original y de los resultados del análisis (ver Pomareda, 1980).

1. En el Cuadro 1 se listan las primeras tarjetas utilizadas en este ejemplo y el proceso PRINT.

- La tarjeta 1 (DATA) da el nombre a los datos (DEMANDA).
- La tarjeta 2 (INPUT) identifica las variables dadas como datos.
- Las tarjetas 3 a 7 describen las transformaciones solicitadas. Aquí podría también haberse solicitado transformación logarítmica, digamos

  LPMI = LOG PMI (logaritmo neperiano)

. La tarjeta 8 anuncia los datos (**CARDS**;)

. Las tarjetas 9 a 19 (no mostradas) contienen 11 observaciones de datos.

. La tarjeta 20 solicita (**PROC PRINT**;) que se impriman todos los datos incluyendo las transformaciones.

El listado de los datos se presenta en la parte mas baja de este cuadro.

2. En el Cuadro 2, se presenta la tarjeta 21que solicita la realización del proceso **MEANS** y los resultados de dicho análisis. El cuadro lista las variables como filas y en el lado de las columnas provee el número de observaciones, el promedio (MEAN), la desviación standard (STANDARD DEVIATION), el valor mínimo (MINIMUM VALUE), el valor máximo (MAXIMUM VALUE), el error standard de la media (STD ERROR MEAN), la suma (SUM), la variancia (VARIANCE) y el coeficiente de variación (C.V.).

3. El Cuadro 3 presenta el proceso **CORR**, solicitado a través de la tarjeta 22. Para cada par de variables los coeficientes de correlación aparecen como el número superior y la probabilidad de que dicho valor sea cero, aparece como el número inferior.

4. Las figuras 1 y 2, solicitadas a través de las tarjetas 23, 24 y 25 por medio del proceso **PLOT** presentan los gráficos superpuestos de las variables, cantidad demandada (QM, QA, QF, y QS) contra tiempo y precio (PM, PA, PF y PS) contra tiempo. Las líneas que unen los puntos de las observaciones han sido trazadas a mano.

5. El Cuadro 4 muestra los resultados del análisis de regresión múltiple usando GLM para el modelo lineal.

$$QM = f (PMI, PAI, PFI, IY)$$

La descripción de este cuadro fue hecha en la sección anterior. En términos de una interpretación estadística (resumida) de los resultados, se podrían emitir los siguientes juicios.

> "Según lo sugiere el R^2, todas las variables en conjunto explican la casi totalidad de la variación de la demanda de maíz (R^2 = 0.993) y según lo indican los valores, de F, todas y cada una de las variables (excepto PMI) hacen una contribución significativa a la explicación de la variable dependiente. La menor contribución (R^2 parcial = 0.006, 15399.3/99.5) y la menos significativa es hecha por el precio del propio producto (PMI). En cuanto a los parámetros del modelo todos aparecen con el signo esperado y excepto el del precio del propio producto, todos los coeficientes son significativamente diferentes de cero al 99 por ciento de confiabilidad."

6. En el Cuadro 5 se muestran los resultados del análisis de regresión usando el proceso **SYSREG**, los cuales aparecen en un formato diferente al de **GLM**. En la sección 6 se hace referencia a la utilización de **SYSREG**. Se puede observar que en principio los estimados de los parámetros y las otras estadísticas son idénticas porque tanto **GLM** como **SYSREG** usan el método de los mínimos cuadrados ordinarios.

CUADRO 1

```
1          DATA DEMANDA;
2          INPUT ANO QM QA QF QS PM PA PF PS I Y;
3          PMI=PM/I;
4          PAI=PA/I;
5          PFI=PF/I;
6          PSI=PS/I;
7          IY=Y/I;
8             CARDS;

NOTE: DATA SET WORK.DEMANDA HAS 11 OBSERVATIONS AND 16 VARIABLES, 45 OBS/TRK.
NOTE: THE DATA STATEMENT USED 5.13 SECONDS AND 118K.

20         PROC PRINT;
```


STATISTICAL ANALYSIS SYSTEM 15:03 WEDNESDAY, MAY 21, 1980

OBS	ANO	QM	QA	QF	QS	PM	PA	PF	PS	I	Y	PMI	PAI	PFI	PSI	IY
1	60	219.9	8.3	26.6	47.8	4.88	17.27	13.07	5.61	83.3	525.5	0.0585834	0.207323	0.156903	0.067347	6.30852
2	61	225.3	8.6	29.5	50.9	6.58	16.16	12.58	6.02	88.1	545.8	0.0746879	0.183428	0.142792	0.068331	6.19523
3	62	236.3	7.1	29.0	48.6	6.07	21.40	12.52	5.27	89.5	576.5	0.0678212	0.239106	0.139888	0.058883	6.44134
4	63	243.1	8.5	20.0	49.6	6.49	20.05	13.98	5.41	91.0	631.2	0.0713187	0.220330	0.153626	0.059451	6.93626
5	64	248.2	8.0	32.8	50.1	6.42	18.68	15.15	5.40	94.0	700.3	0.0682579	0.198723	0.161170	0.057447	7.45000
6	65	269.3	8.3	35.4	51.5	5.85	17.93	15.00	4.80	96.7	762.4	0.0604964	0.185419	0.155119	0.049638	7.88452
7	66	290.1	10.8	32.1	53.6	6.03	23.07	16.73	6.68	98.9	828.7	0.0609707	0.233266	0.169161	0.067543	8.37917
8	67	293.8	9.9	32.6	49.8	7.96	27.53	18.29	6.73	102.2	868.4	0.0778865	0.269374	0.178963	0.065851	8.49706
9	68	301.2	10.5	34.2	43.7	7.40	19.88	17.68	9.62	103.8	944.7	0.0712909	0.191522	0.170328	0.092678	9.10116
10	69	323.5	12.2	43.0	46.2	6.06	18.08	15.49	10.74	104.8	971.2	0.0578244	0.172519	0.147805	0.102481	9.26718
11	70	333.3	13.6	43.7	48.2	7.55	20.07	15.94	10.78	110.2	1051.3	0.0685118	0.182123	0.144646	0.097822	9.53993

NOTE: THE PROCEDURE PRINT USED 5.33 SECONDS AND 122K AND PRINTED PAGE 1.

CUADRO 2

21 PROC MEANS;

S T A T I S T I C A L A N A L Y S I S S Y S T E M 15:03 WEDNESDAY, MAY 21, 1980 2

VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
AMO	11	65.CCCCCOCO	3.31662479	60.00000000	70.00000000	1.00000000	715.0000000	11.000000	5.102
QA	11	271.45454545	39.37269012	219.90000000	333.30000000	11.87131274	2986.0000000	1550.208727	14.504
QA	11	5.65454545	2.00E663C6	7.10000000	13.60000000	0.60563470	106.2000000	4.034727	20.805
AF	11	32.62727273	6.77777115	20.00000000	43.70000000	2.04357489	358.9000000	45.938182	20.773
AS	11	48.5C909091	2.77324883	43.70000000	53.60000000	0.83616598	538.0000000	7.690909	5.670
PM	11	6.4EC9C9C9	0.87745603	4.88000000	7.96000000	0.26456295	71.2900000	0.769929	13.539
PA	11	20.C1C9C9C9	3.16257C01	16.16000000	27.53000000	0.95355074	220.1200000	10.001849	15.804
PF	11	15.13CCC00C	1.96803963	12.52000000	18.29000000	0.59338628	166.4300000	3.873180	13.008
PS	11	7.CC545455	2.26116545	4.80000000	10.78000000	0.68176824	77.0600000	5.112887	32.277
I	11	96.55C9C9C9	8.23631648	83.30000000	110.20000000	2.48334286	1062.5000000	67.836909	8.527
Y	11	764.27272727	182.90814138	525.50000000	1051.30000000	55.14887964	8407.0000000	33455.388182	23.932
PAI	11	0.CE7C6271	0.00673C25	0.05782443	0.07788650	0.00202925	0.7376898	0.000045	10.036
PAI	11	0.2C755757	0.02987104	0.17251908	0.26937378	0.00900646	2.2831332	0.000892	14.392
PFI	11	0.15E40C13	0.01252224	0.13986827	0.17896282	0.00377560	1.7204014	0.000157	8.007
PSI	11	0.C715E836	0.01776452	0.04963806	0.10248092	0.00535620	0.7874720	0.000316	24.815
IY	11	7.81912493	1.23342548	6.19523269	9.53992740	0.37189298	86.0103743	1.521348	15.775

NOTE: THE PROCEDURE MEANS USED 6.23 SECONDS AND 122K AND PRINTED PAGE 2.

CUADRO 3

22 PROC CORR; CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 11

	ANO	QM	QA	QF	QS	PM	PA	PF	PS	I	Y	PMI	PAI	PFI	PSI	IY
ANO	1.00000 0.0000	0.98524 0.0001	0.86511 0.0065	0.75896 0.0032	-0.23484 0.4870	0.61199 0.0454	0.34770 0.2948	0.79160 0.0037	0.82273 0.0019	0.99389 0.0001	0.99532 0.0001	-0.00714 0.9834	-0.18144 0.5934	0.26185 0.4367	0.70141 0.0162	0.98941 0.0001
QM	0.98524 0.0001	1.00000 0.0000	0.90750 0.0001	0.81855 0.0021	-0.22296 0.5099	0.55082 0.0791	0.34664 0.2963	0.75881 0.0068	0.84876 0.0010	0.98138 0.0001	0.99002 0.0001	-0.08013 0.8148	-0.17363 0.6096	0.22356 0.5067	0.73838 0.0095	0.98320 0.0001
QA	0.86511 0.0065	0.90750 0.0001	1.00000 0.0000	0.75438 0.0073	-0.20762 0.5402	0.46504 0.1495	0.15278 0.6538	0.59482 0.0536	0.92610 0.0001	0.86591 0.0006	0.89205 0.0002	-0.09919 0.7717	-0.31643 0.3431	0.08803 0.7969	0.87157 0.0005	0.87531 0.0004
QF	0.75856 0.0032	0.81855 0.0021	0.75438 0.0073	1.00000 0.0000	-0.20825 0.5389	0.30384 0.3637	-0.02037 0.9526	0.44836 0.1666	0.75928 0.0067	0.80368 0.0029	0.81002 0.0025	-0.25745 0.4447	-0.46789 0.1467	-0.08804 0.8014	0.67553 0.0214	0.79248 0.0036
QS	-0.23484 0.4870	-0.22296 0.5099	-0.20762 0.5402	-0.20825 0.5389	1.00000 0.0000	-0.16364 0.6307	0.13837 0.6849	-0.04370 0.8985	-0.53136 0.0926	-0.20169 0.5520	-0.23971 0.4777	-0.03604 0.9162	0.23234 0.4918	0.14384 0.6731	-0.59023 0.0556	-0.22863 0.4989
PM	0.61199 0.0454	0.55082 0.0791	0.46504 0.1495	0.30384 0.3637	-0.16364 0.6307	1.00000 0.0000	0.55234 0.0781	0.64849 0.0309	0.45008 0.1648	0.66920 0.0243	0.65289 0.0546	0.78240 0.0044	0.19954 0.5564	0.33305 0.3169	0.34677 0.2961	0.54969 0.0798
PA	0.34770 0.2948	0.34664 0.2963	0.15278 0.6538	-0.02037 0.9526	0.13837 0.6849	0.55234 0.0781	1.00000 0.0000	0.63806 0.0346	0.02064 0.9520	0.36674 0.2673	0.32682 0.3266	0.40985 0.2106	0.85468 0.0008	0.62386 0.0402	-0.08158 0.8115	0.32630 0.3274
PF	0.79160 0.0037	0.75881 0.0068	0.59482 0.0536	0.44836 0.1666	-0.04370 0.8985	0.64849 0.0309	0.63806 0.0346	1.00000 0.0000	0.48728 0.1285	0.78771 0.0040	0.80288 0.0029	0.18960 0.5766	0.22122 0.5133	0.79056 0.0038	0.36692 0.2670	0.82678 0.0017
PS	0.82273 0.0019	0.84876 0.0010	0.92610 0.0001	0.75928 0.0067	-0.53136 0.0926	0.45008 0.1648	0.02064 0.9520	0.48728 0.1285	1.00000 0.0000	0.80884 0.0026	0.83890 0.0012	-0.06745 0.8438	-0.41779 0.2010	-0.03341 0.9223	0.98108 0.0001	0.81822 0.0021
I	0.99389 0.0001	0.98138 0.0001	0.86591 0.0006	0.80368 0.0029	-0.20169 0.5520	0.66920 0.0243	0.36674 0.2673	0.78771 0.0040	0.80884 0.0026	1.00000 0.0000	0.98959 0.0001	0.06471 0.8501	-0.16661 0.6244	0.24711 0.4638	0.68166 0.0209	0.97628 0.0001
Y	0.99532 0.0001	0.99002 0.0001	0.89205 0.0002	0.81002 0.0025	-0.23971 0.4777	0.65289 0.0546	0.32682 0.3266	0.80288 0.0029	0.83890 0.0012	0.98959 0.0001	1.00000 0.0000	-0.03425 0.9204	-0.20247 0.5505	0.28600 0.3939	0.72381 0.0118	0.99588 0.0001
PMI	-0.00714 0.9834	-0.08013 0.8148	-0.09919 0.7717	-0.25745 0.4447	-0.03604 0.9162	0.78240 0.0044	0.40985 0.2106	0.18960 0.5766	-0.06745 0.8438	0.06471 0.8501	-0.03425 0.9204	1.00000 0.0000	0.37915 0.2501	0.20134 0.5527	-0.09966 0.7706	-0.08269 0.8090
PAI	-0.18144 0.5934	-0.17363 0.6056	-0.31643 0.3431	-0.46789 0.1467	0.23234 0.4918	0.19954 0.5564	0.85468 0.0008	0.22122 0.5133	-0.41779 0.2010	-0.16661 0.6244	-0.20247 0.5505	0.37915 0.2501	1.00000 0.0000	0.45663 0.1502	-0.45452 0.1602	-0.19642 0.5627
PFI	0.26185 0.4367	0.22356 0.5087	0.08803 0.7969	-0.08804 0.8014	0.14384 0.6731	0.33305 0.3169	0.62386 0.0402	0.79056 0.0038	-0.03341 0.9223	0.24711 0.4638	0.28600 0.3939	0.20134 0.5527	0.49563 0.1202	1.00000 0.0000	-0.09660 0.7775	0.33665 0.3083
PSI	0.70141 0.0162	0.73838 0.0095	0.87157 0.0005	0.67953 0.0214	-0.59023 0.0556	0.34677 0.2961	-0.08158 0.8115	0.36692 0.2670	0.98108 0.0001	0.68166 0.0209	0.72381 0.0118	-0.09966 0.7706	-0.45452 0.1602	-0.09660 0.7775	1.00000 0.0000	0.70394 0.0156
IY	0.98941 0.0001	0.98320 0.0001	0.87531 0.0004	0.79248 0.0036	-0.22863 0.4989	0.54969 0.0798	0.32630 0.3274	0.82678 0.0017	0.81822 0.0021	0.97628 0.0001	0.99588 0.0001	-0.08269 0.8090	-0.19642 0.5627	0.33665 0.3083	0.70394 0.0156	1.00000 0.0000

FIGURA 1

```
23    PROC PLOT;
24    PLOT PM*ANO='*' PA*ANC='O' PF*ANC='X' PS*ANC='A'/OVERLAY;
25    PLOT QM*ANO='*' QA*ANC='O' QF*ANC='X' QS*ANC='A'/OVERLAY;
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M 15:03 WEDNESDAY, MAY 21, 1980 6

PLOT OF PM*ANC SYMBOL USED IS *
PLOT OF PA*ANO SYMBOL USED IS O
PLOT OF PF*ANO SYMBOL USED IS X
PLOT OF PS*ANO SYMBOL USED IS A

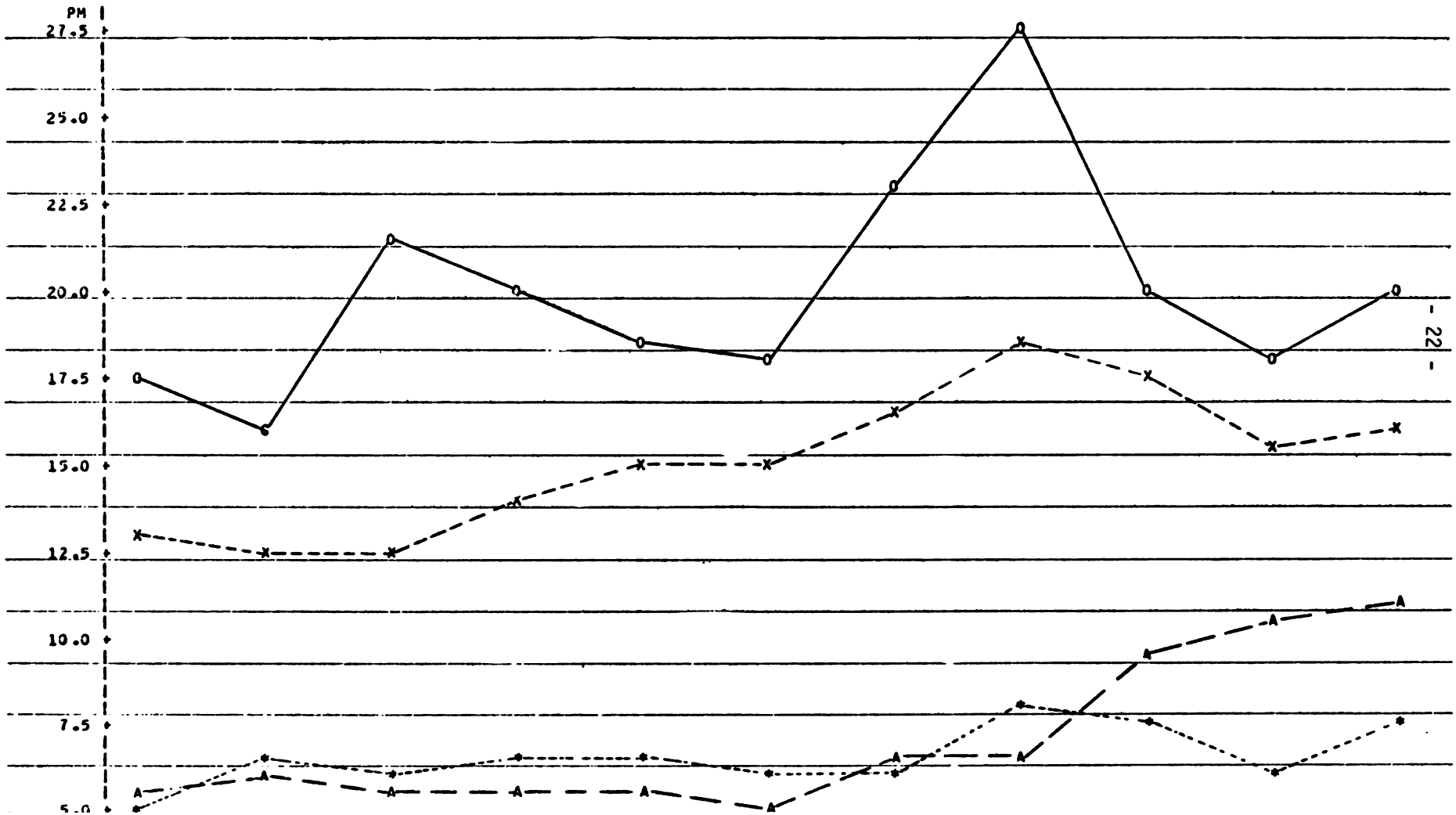
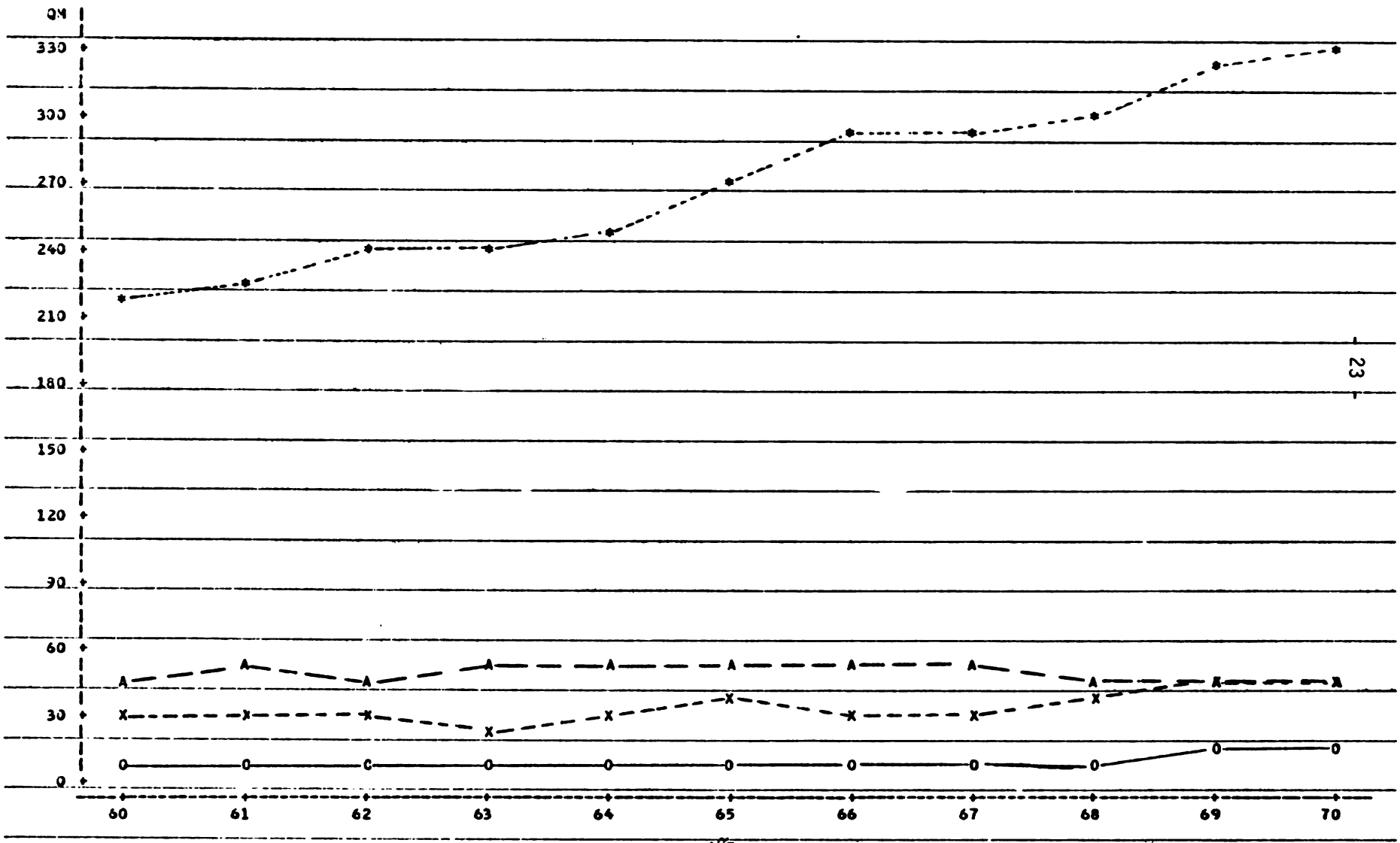


FIGURA 2

PLOT OF QM*ANO SYMBOL USED IS *
PLOT OF QA*ANO SYMBOL USED IS O
PLOT OF QF*ANO SYMBOL USED IS X
PLOT OF QS*ANO SYMBOL USED IS A



NOTE: THE PROCEDURE PLOT USED 11.58 SECONDS AND 148K AND PRINTED PAGES 6 TO 7.

26 PROC GLM; MODEL QM=PMI PAI PFI IY;

CUADRO 4

```
26   PROC GLM;  MODEL QM = PMI  PAI  PFI  IY;
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M 15:03 WEDNESDAY, MAY 21, 1980 8

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: QM

SOURCE	DF	① SUM OF SQUARES	MEAN SQUARE	② F VALUE	PR > F	③ R-SQUARE	④ C.V.
MODEL	4	15399.33280650	3849.83320163	224.80	0.0001	0.993372	1.5245
ERROR	6	102.75446622	17.12574437	⑤ STD DEV			⑥ QM MEAN
CORRECTED TOTAL	10	15502.08727273		4.13832628			271.45454545

SOURCE	DF	⑦ TYPE I SS	F VALUE	PR > F	DF	⑧ TYPE IV SS	F VALUE	PR > F
PMI	1	99.53771074	5.81	0.0525	1	0.11703517	0.01	0.9368
PAI	1	371.52890184	21.69	0.0035	1	189.61854591	11.07	0.0159
PFI	1	1978.07404391	115.50	0.0001	1	406.85596703	23.76	0.0028
IY	1	12950.19215001	756.18	0.0001	1	12950.19215001	756.18	0.0001

PARAMETER	⑨ ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	67.83354661	3.44	0.0138	19.73050360
PMI	-17.37700347	-0.08	0.9368	210.20421216
PAI	196.07644661	3.33	0.0159	58.92637843
PFI	-683.66796036	-4.87	0.0028	140.26509530
IY	34.66052197	27.50	0.0001	1.26043868

NOTE: THE PROCEDURE GLM USED 8.54 SECONDS AND 144K AND PRINTED PAGE 8.

CUADRO 5

```
27        PROC SYSREG; MODEL QM=PMI PAI PFI   IY/STB DW;

                                  S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

MODEL:    MODEL01                 SSE      102.754466      F RATIO          224.80
                                  DFE               6      PRCB>F           0.0001
DEP VAR: QM                       MSE       17.125744      R-SQUARE         0.9934

DURBIN-WATSON D STATISTIC    =    1.9C09
FIRST ORDER AUTOCORRELATION =     C.C516

                          PARAMETER      STANDARD                                  VARIABLE
VARIABLE        DF         ESTIMATE       ERRCR        T RATIO     PROB>|T|        LABEL

INTERCEPT        1        67.833549     19.730504       3.4380       0.0138
PMI              1       -17.377003    210.204212      -0.0827       0.9368
PAI              1       196.C76447     58.926378       3.3275       0.0159
PFI              1      -683.667960    140.265C95      -4.8741       0.0028
IY               1        34.66C522      1.260439      27.4988       0.0001

                                  STANDARDIZED E VALUES

                                                                        QM

                                  INTERCEPT                              0
                                  PMI                        -0.00297037
                                  PAI                         0.14875812
                                  PFI                        -0.21743644
                                  IY                          1.08581124

NOTE: THE PROCEDURE SYSREG USED 6.41 SECONDS AND 126K AND PRINTED PAGE 9.

NOTE: SAS USED 148K MEMORY.

NOTE: SAS INSTITUTE INC.
      P.O. BOX 1CC66
      RALEIGH, N.C. 27CC5
```

5. ANALISIS DE REGRESION MULTIPLE POR ETAPAS USANDO **STEPWISE**

5.1. **PROC STEPWISE;**

El proceso **STEPWISE** provee cinco métodos alternativos para seleccionar de un número grande de variables independientes, aquellas que tienen mayor relevancia desde el punto de vista estadístico y que por consiguiente deben ser incluidas en el modelo. **STEPWISE** es así un proceso útil en un análisis exploratorio, pero no garantiza 'la obtención del mejor modelo', sólo aquel con el R^2 más alto[4/]. La selección del mejor modelo, además de fundamentarse en criterios estadísticos, depende en gran medida de los conocimientos que el investigador tiene de la situación real y su habilidad y experiencia en el manejo de la econometría.

Existen cinco opciones básicas que puede usarse con **STEPWISE** y ellas son:

- **FORWARD** (Forward Selection), la cual comienza por encontrar la variable que produce el R^2 mas alto. Para cada una de las otras variables el FORWARD calcula la "F" indicando así la contribución que dicha variable haría si fuese incluida.

- **BACKWARD** (Backward Elimination), comienza por calcular el modelo incluyendo todas las variables y luego se van eliminando variables una a una hasta que todas las variables que permanecen producen una 'F' significativa al nivel especificado (SLSTAY) o al 0.10 si éste no es especificado [5/]. BACKWARD se puede abreviar B en el MODEL.

[4/] Debe anotarse que STEPWISE difiere del proceso RSQUARE (no discutido en este documento) en que este último permite obtener todas las posibles combinaciones de las variables independientes sin tratar de seleccionar el mejor modelo. Además cuando STEPWISE evalua un modelo, imprime un reporte completo del análisis de regresión mientras que RSQUARE imprime solamente el R^2 para cada modelo.

[5/] SLSTAY indica el nivel de significancia requerido para cada una de las variables.

- **STEPWISE** es una modificación de FORWARD y difiere de este último en que las variables que están en el modelo no necesariamente permanecen allí. Igual que en el procedimiento FORWARD las variables son añadidas una a una y la variable que se añade debe hacer una contribución significativa (indicada por F) al nivel especificado en **SLENTRY**.

- **MAXR** (maximum R^2 improvement), se considera superior a la técnica STEPWISE y tan buena como RSQUARE. En contraste con las tres técnicas arriba referidas MAXR busca el mejor modelo con una variable; el mejor modelo con dos variables y así sucesivamente. MAXR se inicia buscando el modelo con una variable que produce el R^2 mas alto.

- **MINR** (minimum R^2 improvement) parecido a MAXR pero la selección es en base al menor incremento en R^2.

En su forma mas general **PROC STEPWISE** puede presentarse:

PROC STEPWISE;

MODEL dependiente = independientes/opciones;

o para un caso específico: 6/

PROC STEPWISE;

MODEL QM = PMI PAI PFI PFI PSI IY / MAXR;

TITLE REGRESION POR PASOS;

Además de las opciones que se pueden usar con el modelo y que se refirieron antes (FORWARD; BACKWARD; STEPWISE; MAXR y MINR), existen otras como:

- **NOINT** que se usa si no se desea que el modelo incluya un término de intercepto;

- **SLENTRY** =.—(abreviado **SLE**=.—) especifica el nivel de significancia requerido para el ingreso de las variables en el proceso FORWARD. Si no se especifica, el modelo usará .50.

6/ Este caso específico es analizado en la próxima sección y los datos corresponden al problema resuelto con **PROC GLM** en la sección 4.2.

- **SLSTAY** =.— (abreviado **SLS**=.—) es el nivel de significancia requerido para que una variable permanezca en el modelo cuando se usa FORWARD. Si no se especifica, el proceso asume .10.

- **INCLUDE**=— que se usa cuando se desea incluir una variable en cada uno de los modelos que STEPWISE considera. Se lista la(s) variable(s) deseada(s) en primer y subsiguiente(s) lugar(es) después del signo igual en el statement MODEL y se especifica **INCLUDE**=— después de la diagonal (/). En el espacio — se indica el número de variables que se desea forzar en el modelo. En este caso STEPWISE no realiza pruebas de significancia para la(s) variable(s) forzada(s). Por ejemplo 7/

```
PROC STEPWISE;
MODEL QM = PMI  PAI  PFI  PSI  IY/INCLUDE = 1;
```

Existen algunas limitaciones en el uso de **STEPWISE** y la principal se refiere al número de variables independientes; no siendo recomendable incluir más de 20.

La interpretación de los resultados impresos por el proceso **STEPWISE** es la siguiente, (ver el Cuadro 7)

① La fuente de variación atribuible a las variables independientes (REGRESSION);

② La fuente de variación atribuible al error (ERROR);

7/ Este caso se ilustra en la próxima sección.

(3) La fuente de variación total (TOTAL);

(4) Los grados de libertad;

(5) Suma de cuadrados (SUM OF SQUARES);

(6) Media al cuadrado (MEAN SQUARE);

(7) Estadística F (F)

(8) Probabilidad que el modelo no explique significativamente la variación de la variable dependiente (PROB>F)

(9) R^2 (RSQUARE)

(10) Valores de los coeficientes de regresión del intercepto (INTERCEPT) y de las variables.

(11) Error standard del coeficiente (STD ERROR)

(12) Error tipo II (Type II ERR SS)

(13) Estadística F (F) de los coeficientes de las variables independientes.

(14) Probabilidad de que los coeficientes de las variables sean iguales a cero.

5.2. Ejemplo Ilustrativo

Como se ilustró en el caso de la demanda de maíz en Honduras, estimada por el procedimiento **GLM** en la Sección 4.2 la menor contribución y la menos significativa para explicar la cantidad demandada, es hecha por el propio precio del producto.[8/] En esta sección se utiliza el proceso **STEPWISE** para ilustrar la selección del modelo con el R^2 mas alto.

Los datos corresponden a aquellos en el Cuadro 2 y no se han listado nuevamente. En el Cuadro 6 se muestra las tarjetas usadas. Las primeras 19 no han cambiado. La tarjeta 20 solicita que el modelo sea resuelto por el método de regresión por pasos y las tarjetas 21 y 22 presentan el mismo modelo con dos opciones.

En el primer caso se usa la opción **MAXR** y los resultados se presentan en el Cuadro 7. Como lo requiere el proceso **MAXR**, el modelo incluye primero la variable que contribuye a explicar la mayor parte (96.67%) de la variación en la cantidad demandada. En forma sucesiva se van incorporando las otras variables.

"Un análisis estadístico de estos resultados es indicativo, mas no puede ser suficiente para arribar a conclusiones sobre el mejor modelo. Solo en el **STEP 5** se incluye la variable **PMI**, sin embargo esta variable (el propio precio del producto) es desde el punto de vista del análisis económico, la mas importante. Esta situación nos lleva a intuir que en la estimación de la demanda de maíz pudieran estarse presentando

---

8/ No entramos aquí a discutir la relevancia económica de este estimado, pero el lector debe estar conciente de la naturaleza de los problemas que se presentan en el análisis económico y estar preparado para compatibilizar la teoría y lo que la información estadística le permita inferir.

problemas de simultaneidad; es decir que lo que se aprecia en términos de las relaciones QM vs PMI son puntos de equilibrio de oferta y demanda, por lo que se hace necesario recurrir a un modelo de mínimos cuadrados en dos etapas, si se dispusiera de la información necesaria para formular las funciones de oferta y de demanda.9/

Con el objeto de forzar la variable PMI en el modelo, se usó el procedimiento INCLUDE y los resultados del análisis econométrico se muestran en el Cuadro 8. Se puede apreciar que en este caso se presenta además el STEP 0 el cual incluye solo la variable forzada (PMI). La contribución de esta variable con explicar la variación de QM es insignificante (0.006) y el coeficiente aunque con el signo esperado (-) no es significativamente diferente de cero a ningún nivel estadísticamente aceptable. En STEP 1 y STEP 2 aunque se produce un incremento notable en el R^2 (debido a la incorporación de IY), el signo del coeficiente de PMI aparece positivo. El mejor modelo pareciera ser STEP 3 (coincidente con STEP 5) ya que todos son altamente significativos. Puede notarse que este modelo es el mismo que aparece en los Cuadros 4 y 5, resueltos con GLM y SYSREG respectivamente. 10/

9/ Lamentablemente no se pudo disponer de la información para formular la función de oferta i.e. datos climáticos y márgenes de comercialización. Ello hubiera permitido usar este mismo ejemplo para ilustrar el uso de SYSREG para resolver modelos de ecuaciónes simultáneas, discutido en la próxima sección.

10/ El argumento del autor para no incluir el precio del sorgo es que en el consumo humano de maíz para tortillas en Honduras, el sorgo no es un substituto aceptado del maíz. Ambos productos sin embargo se usan en dietas para alimentación animal y en Honduras el sorgo es para ese propósito mas importante que el maíz. (Pomareda. 1980, Volúmen I, Capítulo 12).

CUADRO 6

```
1           DATA DEMANDA;
2           INPUT ANO QM QA QF QS PM PA PF PS I Y;
3           PMI=PM/I;
4           PAI=PA/I;
5           PFI=PF/I;
6           PSI=PS/I;
7           IY=Y/I;
8              CARDS;

NOTE: DATA SET WORK.DEMANDA HAS 11 OBSERVATIONS AND 16 VARIABLES. 45 OBS/TR
NOTE: THE DATA STATEMENT USED 5.14 SECONDS AND 118K.

20          PROC STEPWISE;
21          MODEL QM = PMI PAI PFI PSI IY/MAXR;
22          MODEL QM = PMI PAI PFI PSI IY/INCLUDE=1;

NOTE: THE PROCEDURE STEPWISE USED 19.36 SECONDS AND 122K AND PRINTED PAGES

NOTE: SAS USED 122K MEMORY.

NOTE: SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

CUADRO 7

S T A T I S T I C A L A N A L Y S I S S Y S T E M 10:19 FRIDAY, F

MAXIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE Q

STEP 1 VARIABLE IY ENTERED R SQUARE = 0.96668647 C(P) = 23.54687068

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	14985.65809904	14985.65809904	261.16	0.0001
ERROR	9	516.42917368	57.38101930		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	26.05048842				
IY	31.38510500	1.94209203	14985.65809904	261.16	0.0001

THE ABOVE MODEL IS THE BEST 1 VARIABLE MODEL FOUND.

STEP 2 VARIABLE PFI ENTERED R SQUARE = 0.98020707 C(P) = 13.14914464

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	15195.25558879	7597.62779440	198.09	0.0001
ERROR	8	306.83168393	38.35396049		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	76.37607785				
PFI	-388.56334137	166.21540879	209.59748975	5.46	0.0476
IY	32.72103029	1.68749216	14420.51021274	375.58	0.0001

THE ABOVE MODEL IS THE BEST 2 VARIABLE MODEL FOUND.

STEP 3 VARIABLE PAI ENTERED R SQUARE = 0.99336402 C(P) = 3.08486625

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	15399.21577133	5133.07192378	349.29	0.0001
ERROR	7	102.87150140	14.69592877		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	66.98818248				
PAI	194.66775923	52.25403755	203.96018254	13.88	0.0074
PFI	-683.94983489	129.89578001	407.43229675	27.72	0.0012
IY	34.66263041	1.16736476	12957.07315282	881.68	0.0001

THE ABOVE MODEL IS THE BEST 3 VARIABLE MODEL FOUND.

STEP 4 VARIABLE PSI ENTERED R SQUARE = 0.99448616 C(P) = 4.05592449

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	15416.61118794	3854.15279699	270.54	0.0001
ERROR	6	85.47608479	14.24601413		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	58.73014907				
PAI	207.25235774	52.69335894	220.38425477	15.47	0.0077
PFI	-636.62838974	134.87122043	317.41452709	22.28	0.0033
PSI	123.72550287	111.97022121	17.39541661	1.22	0.3115
IY	33.30535986	1.68216538	5584.49944866	392.00	0.0001

THE ABOVE MODEL IS THE BEST 4 VARIABLE MODEL FOUND.

STEP 5 VARIABLE PMI ENTERED R SQUARE = 0.99454715 C(P) = 6.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	15417.55665433	3083.51133087	182.39	0.0001
ERROR	5	84.53061840	16.90612368		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	60.87C70337				
PMI	-49.94402147	211.19444113	0.94546639	0.06	0.8224
PAI	211.74177140	60.46029243	207.35601550	12.27	0.0172
PFI	-634.16130874	147.29454387	313.37922701	18.54	0.0077
PSI	128.06181926	123.34502802	18.22384782	1.08	0.3467
IY	33.25177586	1.84645562	5482.71899328	324.30	0.0001

THE ABOVE MODEL IS THE BEST 5 VARIABLE MODEL FOUND.

CUADRO 8

STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE QM

THE FIRST 1 VARIABLES IN EACH MODEL ARE INCLUDED VARIABLES.

STEP 0 INCLUDED VARIABLE ENTERED R SQUARE = 0.00642092 C(P) = 904.06334329

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	99.53771074	99.53771074	0.06	0.8148
ERROR	9	15402.54956199	1711.39439578		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	3C2.89172931				
PMI	-468.77295193	1943.76507106	99.53771074	0.06	0.8148

STEP 1 VARIABLE IY ENTERED R SQUARE = 0.96668785 C(P) = 25.54560886

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	14985.67943141	7492.83971570	116.08	0.0001
ERROR	8	516.40784132	64.55098016		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	25.56438885				
PMI	6.88617502	378.79999805	0.02133237	0.00	0.9859
IY	31.38821200	2.06693549	14886.14172067	230.61	0.0001

STEP 2 VARIABLE PFI ENTERED R SQUARE = 0.98113976 C(P) = 14.29351182

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	15209.71426060	5069.90475353	121.38	0.0001
ERROR	7	292.37301213	41.76757316		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	66.65854651				
PMI	184.89160719	314.24806975	14.45867180	0.35	0.5748
PFI	-414.30579414	178.88867932	224.03482919	5.36	0.0537
IY	32.89295775	1.78506748	14181.94570749	339.54	0.0001

STEP 3 VARIABLE PAI ENTERED R SQUARE = 0.99337157 C(P) = 5.07794360

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	15399.33280650	3849.83320163	224.80	0.0001
ERROR	6	102.75446622	17.12574437		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	67.83354861				
PMI	-17.37700347	210.20421216	0.11703517	0.01	0.9368
PAI	196.07644661	58.92637843	189.61854591	11.07	0.0159
PFI	-683.66796036	140.26509530	406.85596703	23.76	0.0028
IY	34.66C52197	1.26043868	12950.19215001	756.18	0.0001

STEP 4 VARIABLE PSI ENTERED R SQUARE = 0.99454715 C(P) = 6.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	15417.55665433	3083.51133087	182.39	0.0001
ERROR	5	84.53061840	16.90612368		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	60.87070337				
PMI	-49.94402147	211.19444113	0.94546639	0.06	0.8224
PAI	211.74177140	60.46029243	207.35601550	12.27	0.0172
PFI	-634.16136874	147.29454387	313.37922701	18.54	0.0077
PSI	128.06181926	123.34502802	18.22364782	1.08	0.3467
IY	33.25177586	1.84645562	5482.71895328	324.30	0.0001

STEP 5 VARIABLE PSI REMOVED R SQUARE = 0.99337157 C(P) = 5.07794360

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	15399.33280650	3849.83320163	224.80	0.0001
ERROR	6	102.75446622	17.12574437		
TOTAL	10	15502.08727273			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	67.83354861				
PMI	-17.37700347	210.20421216	0.11703517	0.01	0.9368
PAI	196.07644661	58.92637843	189.61854591	11.07	0.0159
PFI	-683.66796036	140.26509530	406.85596703	23.76	0.0028
IY	34.66C52197	1.26043868	12950.19215001	756.18	0.0001

NO OTHER VARIABLES MET THE 0.5000 SIGNIFICANCE LEVEL FOR ENTRY INTO THE MODEL.

6. ANALISIS DE REGRESION MULTIPLE CON PROBLEMAS DE SIMULTANEIDAD USANDO SYSREG 11/

La razón para discutir **GLM** y **SYSREG** en una forma comparativa es que ambos procedimientos permiten en primera instancia obtener resultados de análisis de regresión por el método de los mínimos cuadrados. Pero **SYSREG** permite además estimar los parámetros usando los métodos de los mínimos cuadrados en dos etapas (two stage least squares, 2SLS), mínimos cuadrados en tres etapas (three stage least squares, 3SLS), máxima verosimilitud con información limitada (Limited Information Maximum Likelihood, LIML) y regresiones aparentemente no relacionados ('seemingly unrelated' regressions).

6.1. Método de los mínimos cuadrados ordinarios

Para resolver modelos de regresión por el método de los mínimos cuadrados, usando SYSREG, las instrucciones mas simples son:

```
PROC  SYSREG;
MODEL  Y = X;
```

ó

```
PROC  SYSREG;
MODEL Y = X1  X2  X3  X4;
```

Sin embargo **SYSREG** tiene además algunas opciones como:

- **SIMPLE** (también denotada por S), por medio del cual **SYSREG** imprime la suma, promedios, suma de cuadrados no corregida, variancia y desviación standard de todas las variables en los datos primarios.

11/ Para una discusión exhaustiva de estas técnicas el lector es referido a textos de Econometría como Kmenta [1977], Johnston [1975] y otros.

- **USSCP**, que solicita la impresión de la suma de cuadrados no corregida y la matriz de productos cruzados (cross products) para las variables en los datos primarios.

- **USSCP2**, que solicita la impresión de la suma de cuadrados no corregida y la matriz de productos cruzados para todas las variables usadas en el análisis, incluyendo aquellas generadas.

- **DFNO**, que permite que cuando se estiman las desviaciones standard de los coeficientes (en la segunda y tercera etapa, 2SLS y 3SLS) no se haga la corrección por el número de grados de libertad.

- **NOPRINT**, que equivale a poner **NOPRINT** en cada statement de MODEL y solicita que no se imprima la tabla de análisis de variancia ni los estimados de los parámetros.

Estas y algunas otras opciones pueden usarse con **SYSREG** en la siguiente forma,

PROC SYSREG opciones;

Nótese que en estos casos se trata de opciones de uso con **PROC SYSREG**. Existen además opciones que pueden usarse al momento de especificar el modelo dentro del siguiente formato general,

MODEL dependiente = independiente/opciones;

Algunas de estas opciones son:

- **NOPRINT**, que solicta que no se imprima el cuadro de análisis de variancia ni los estimados de los coeficientes.

- **NOINT**, por medio del cual se demanda que el modelo de regresión no incluya un término de intercepto.

- **XPX**, solicita que se imprima la matriz de productos cruzados (X'X) usados en el modelo.

- **COVB**, solicita que se imprima la matriz de covariancia de los estimados de los parámetros.

- **CORRB**, usado para imprimir la matriz de correlación entre los estimados de los parámetros.

- **DW**, demanda al sistema realizar la prueba de Darwin-Watson e imprimir el valor de la estadística D-W y los coeficientes de autocorrelación de los residuos. Si a las observaciones al inicio del set de información les falta datos, éstos serán descartados y los cálculos serán basados en la primera serie de datos completos.

- **STB**,solicita que se impriman los valores estandarizados de los estimados de los parámetros también conocidos como coeficientes parciales standard de regresión, (standard partial regression coeficient) el cual se obtiene de multiplicar el coeficiente en regresión por su correspondiente desciación standard y dividido por la desviación standard de la variable usada.

6.2. Métodos de 2SLS y LIML

Existen problemas específicos en el análisis de regresión para lo que **SYSREG** provee las opciones apropiadas. El problema mas importante es el de 'sesgo por simultaneidad en las ecuaciones' (simultaneous equations bias') Este problema se presenta cuando una o mas de las variables explicativas (o predictivas) responden a cambios en la variable dependiente; lo cual conlleva a la existencia de modelos simultáneos con varias ecuaciones interdependientes. En otras palabras, existe un problema porque la variable independiente está correlacionada con el término de error, violando así una de las condiciones del método de los mínimos cuadrados y por consiguiente el método no proveería resultados satisfactorios.

Una de las situaciones mas comunes donde se presenta este tipo de problema es al estimar los parámetros de funciones de demanda, en donde puede ocurrir que el precio del producto que equilibra el mercado es en realidad una función de la cantidad producida y de la cantidad demandada en forma simultánea; siendo preciso estimar la demanda y la oferta.

Para usar la opción que permite resolver modelos de ecuaciones simultáneas por **2SLS**, es preciso recurrir al statement **BLOCK**, que en términos generales se formula de la siguiente forma,

BLOCK variables conjuntamente dependientes = variables predeterminadas /opciones;

El statement **BLOCK** declara un set de variables, inicia el análisis y genera un primer paso (first step) en la regresión, el cual consiste en regresar las variables conjuntamente dependientes como función de las variables predeterminadas. Las variables conjuntamente dependientes son también llamadas variables endógenas. Las variables predeterminadas incluyen tanto variables exógenas y variables endógenas rezagadas (lags of endogenous variables). Cuando se usa **BLOCK** las variables conjuntamente dependientes se listan al lado izquierdo del signo igual y las variables predeterminadas a la derecha.

Un statement **BLOCK** puede considerarse como una forma especial del statement **MODEL**. Cada opción que se usa con MODEL puede usarse con BLOCK y tiene el mismo efecto y también deben ir precedidas por una diagonal (/). Algunas de estas opciones son:

- **LIML**; que requiere que los parámetros sean estimados por el método de información limitada y máxima verosimilitud. Si LIML no se especifica, el análisis se hará por el método de mínimos cuadrados en dos etapas (2SLS).

- **LIMLP**; el cual solicita que se impriman los resultados de los cálculos intermedios para LIML.

- **PRINT**; solicita que los resultados del análisis de regresión sean impresos. Si se omite, los resultados no serán impresos.

Para ilustrar la formulación completa de un caso específico asumamos un modelo de oferta y demanda, donde,

Demanda: $Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + u_{1t}$

Oferta: $Q_t = \beta_1 + \beta_2 P_t + \beta_3 F_t + \beta_4 A_t + u_{2t}$

en donde se tiene que

Q_t = cantidad demandada = cantidad ofrecida

P_t = precio de equilibrio del mercado

Y_t = ingreso

F_t = relación entre precios recibidos por los productores y precios pagados por los consumidores

A_t = tiempo en años

En este caso las variables P_t y Q_t son endógenas (conjuntamente dependientes) y las variables D_t, F_t y A_t son predeterminadas.

Las instrucciones para resolver este modelo, omitiendo los subíndices de tiempo, son,

```
PROC SYSREG;
BLOCK P Q = Y F A/PRINT;
```

Si además se requiere que el proceso provea las funciones estructurales éstas deben indicarse en el statement MODEL inmediatamente después de **BLOCK**;

```
DEMANDA:  MODEL  Q  =  P  Y;

OFERTA:   MODEL  Q  =  P  F  A;
```

En la próxima sección se discute un ejemplo ilustrativo de un problema semejante.

6.3. Ejemplo Ilustrativo

Se presenta a continuación un caso de estimación de oferta y demanda cuya formulación obedece a aquella presentada en la sección anterior. Con el propósito de que el lector pueda comprobar los resultados del análisis computacional con los aspectos teóricos, se ha tomado el ejemplo de estimación simultánea de oferta y demanda del texto de Kmenta [1977].

Las tarjetas de instrucciones usadas se presentan en el Cuadro 9. Las variables descritas en el input son:

A = tiempo en años

P = precio de equilibrio del mercado

Q = cantidad demandada = cantidad ofrecida

Y = ingreso

F = relación entre precios recibidos por los productores y precios pagados por los consumidores.

Se ha solicitado también en este caso que los datos sean impresos (**PROC PRINT**) y que se realicen los análisis estadísticos básicos (**PROC MEANS**). En el Cuadro 10 se muestran los datos utilizados. Se ha solicitado también que se grafiquen los datos de precio contra cantidad que, como se puede observar en la Figura 3, haría muy difícil trazar a través de estos puntos una ecuación de demanda o una de oferta.

Los resultados del análisis estadístico de las ecuaciones de oferta y demanda estimadas por el método de mínimos cuadrados, se muestran en el Cuadro 11 (ver las instrucciones 28, 29 y 30 en el Cuadro), y comprenden de ecuación de la forma:

Oferta $Q = f(P,F,A)$

Demanda $Q = g(P,Y)$

Como se puede observar en el caso de la función de oferta, el R^2 es razonable (0.6548), pero el coeficiente del precio no es significativamente diferente de cero ni siquiera al 90% de confiabilidad.

Para estimar los parámetros de las funciones de oferta y demanda, aduciendo que existen problemas de simultaneidad, se usó el método de los mínimos cuadrados en dos etapas (ver instrucciones 31, 32, 33 y 34 en el Cuadro 9). En el proceso **BLOCK** se formulan las variables conjuntamente dependientes en función de las variables independientes:

(32) **BLOCK PQ = Y F A / PRINT;**

y los resultados de esta primera etapa se presentan en las primeras dos secciones del Cuadro 12.

La segunda etapa consiste en formular cada uno de los modelos de oferta y demanda a ser estimados a partir de las primeras dos ecuaciones:

(33) **MODEL Q = P F A;**

(34) **MODEL Q = P Y ;**

y los resultados de esta segunda etapa se muestran en las dos últimas secciones del Cuadro 12. Comparando los Cuadros 11 y 12 puede observarse que existe una mejora en el R^2 de los modelosy que el precio del producto en la función de oferta,que aparece ahora con un coeficiente negativo mayor, es significativamente diferente de cero al 99% de confiabilidad.

Los procedimientos aquí presentados son los de mayor uso práctico para el análisis de regresión simple y múltiple. La observación cuidadosa de los datos originales, el análisis detenido del resúmen estadístico y el examen de los gráficos de las variables son requisitos indispensables para el análisis previo de la formulación de los modelos cuyos parámetros se desea estimar.

Es de esperarse que este pequeño manual sirva como una guía introductoria al uso de SAS para el análisis de regresión y estimule su uso para el análisis de problemas de la agricultura que pueden ser mejor explicados con la ayuda de las técnicas estadísticas.

CUADRO 9

```
1          DATA SIMUL;
2          INPUT A P Q Y F;
3          CARDS;

NOTE: DATA SET WORK.SIMUL HAS 20 OBSERVATIONS AND 5 VARIABLES. 138 OBS/TRK.
NOTE: THE DATA STATEMENT USED 4.21 SECONDS AND 118K.

24         PROC PRINT;

NOTE: THE PROCEDURE PRINT USED 4.68 SECONDS AND 122K AND PRINTED PAGE 1.

25         PROC MEANS;

NOTE: THE PROCEDURE MEANS USED 4.48 SECONDS AND 122K AND PRINTED PAGE 2.

26         PROC PLOT;
27         PLOT P*Q='*';

NOTE: THE PROCEDURE PLOT USED 6.73 SECONDS AND 124K AND PRINTED PAGE 3.

28         PROC SYSREG;
29         MODEL Q = P F A;
30         MODEL Q = P Y;

NOTE: THE PROCEDURE SYSREG USED 6.33 SECONDS AND 128K AND PRINTED PAGE 4.

31         PROC SYSREG;
32         BLOCK P Q = Y F A / PRINT;
33         MODEL Q = P F A ;
34         MODEL Q = P Y ;
```

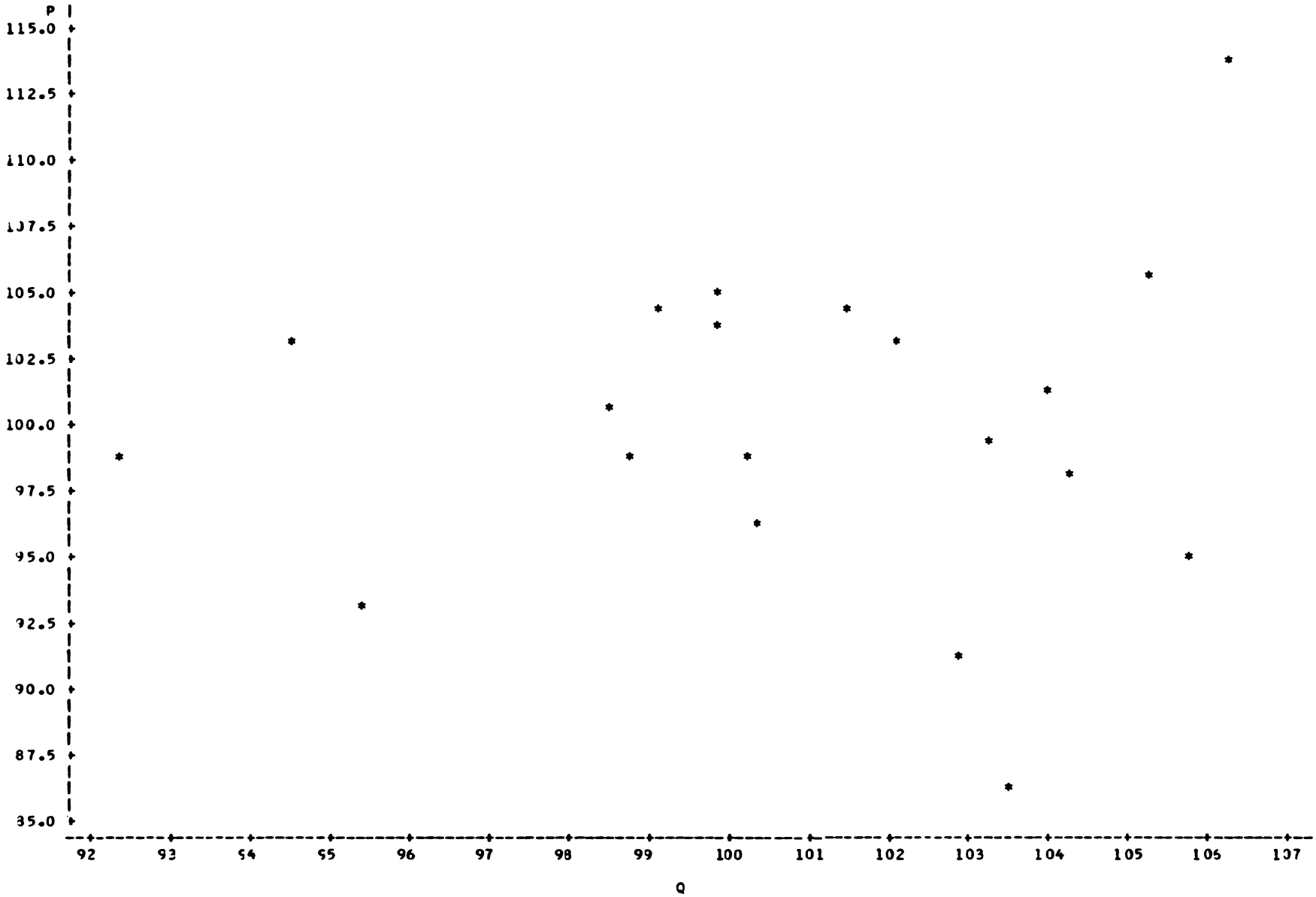
CUADRO 10

S T A T I S T I C A L A N A L Y S I S S Y S T E M

OBS	A	P	Q	Y	F
1	1	100.323	98.485	87.4	98.0
2	2	104.264	99.187	97.6	99.1
3	3	103.435	102.163	96.7	99.1
4	4	104.506	101.504	98.2	98.1
5	5	98.001	104.240	99.8	110.8
6	6	99.456	103.243	100.5	108.2
7	7	101.066	103.993	103.2	105.6
8	8	104.763	99.900	107.8	109.8
9	9	96.446	100.350	96.6	108.7
10	10	91.228	102.820	88.9	100.6
11	11	93.085	95.435	75.1	81.0
12	12	98.801	92.424	76.9	68.6
13	13	102.908	94.535	84.6	70.9
14	14	98.756	98.757	90.6	81.4
15	15	95.119	105.797	103.1	102.3
16	16	98.451	100.225	105.1	105.0
17	17	86.498	103.522	96.4	110.5
18	18	104.016	99.929	104.4	92.5
19	19	105.769	105.223	110.7	89.3
20	20	113.490	106.232	127.1	93.0

FIGURA 3

PLOT OF P*Q SYMBOL USED IS *



CUADRO 11

Estimación de los Parámetros por el Método de los Mínimos Cuadrados Ordinarios

MODEL: MODEL01		SSE	92.551058	F RATIO	10.12
		DFE	16	PROB>F	0.0006
DEP VAR: Q		MSE	5.784441	R-SQUARE	0.6548
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|
INTERCEPT	1	58.275431	11.462910	5.0838	0.0001
P	1	0.160367	0.094884	1.6901	0.1104
F	1	0.246133	0.046188	5.3723	0.0001
A	1	0.248302	0.097518	2.5462	0.0216

MODEL: MODEL02		SSE	63.331650	F RATIO	27.48
		DFE	17	PROB>F	0.0001
DEP VAR: Q		MSE	3.725391	R-SQUARE	0.7638
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|
INTERCEPT	1	99.695423	7.519362	13.2851	0.0001
P	1	-0.316299	0.090677	-3.4882	0.0028
Y	1	0.234636	0.045422	7.3673	0.0001

CUADRO 12

Estimación los Parámetros por el Método de los Mínimos Cuadrados en Dos Etapas

BLOCK: B001		SSE	37.746682	F RATIO	88.94
		DFE	16	PROB>F	0.0001
DEP VAR: P		MSE	2.359168	R-SQUARE	0.9434

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|
INTERCEPT	1	90.267764	3.299306	27.3596	0.0001
Y	1	0.663213	0.041423	16.0107	0.0001
F	1	-0.488448	0.038020	-12.8471	0.0001
A	1	-0.737040	0.075266	-9.7924	0.0001

BLOCK: B001		SSE	74.218588	F RATIO	13.93
		DFE	16	PROB>F	0.0001
DEP VAR: Q		MSE	4.638662	R-SQUARE	0.7232

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|
INTERCEPT	1	71.203546	4.626362	15.3908	0.0001
Y	1	0.159221	0.058084	2.7412	0.0145
F	1	0.136341	0.053313	2.5949	0.0195
A	1	0.075979	0.105540	0.7199	0.4820

MODEL: MODEL02		SSE	96.633244	F RATIO	10.70
		DFE	16	APPROX PR>F	0.0004
DEP VAR: Q		MSE	6.039578	R-SQUARE	0.6674

SECOND STAGE STATISTICS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD EPROP	T RATIO	APPROX PROB>|T|
INTERCEPT	1	49.532442	12.010526	4.1241	0.0008
B001.P	1	0.240076	0.099934	2.4023	0.0288
F	1	0.255606	0.047250	5.4096	0.0001
A	1	0.252924	0.099655	2.5380	0.0219

MODEL: MODEL03		SSE	65.729088	F RATIO	23.81
		DFE	17	APPROX PR>F	0.0001
DEP VAR: Q		MSE	3.866417	R-SQUARE	0.7369

SECOND STAGE STATISTICS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|
INTERCEPT	1	94.633304	7.920838	11.9474	0.0001
B001.P	1	-0.243557	0.096484	-2.5243	0.0218
Y	1	0.313992	0.046944	6.6887	0.0001

REFERENCIAS

Johnston, J. Métodos de Econometría. Editorial Vicens-Vives, Tercera Edición, Barcelona, 1975.

Kmenta, J. Elementos de Econometría. Editorial Vicens-Vives, Primera Edición, Barcelona, 1977.

Pomareda, C. Métodos Cuantitativos para la Investigación en Economía Agrícola, Volúmen I y II, San José, 1980.

Quiroga, V. "Manual para Estimar Parámetros de Seis Modelos Aplicados a Fenómenos Sociales, Económicos y Biológicos", Publicación Miscelánea No. 145, IICA, San José, Costa Rica, Abril 1977.

Quiroga, V. "Manual de Introducción al SAS", Publicación Miscelánea No. 218 IICA, San José, Costa Rica, Diciembre 1979.

SAS INSTITUTE INC., SAS USERS GUIDE, 1979 Edition, Raleigh, N.C., 1979.

FECHA DE DEVOLUCION			