

CIAT
SA
7-
25
NG

C U R S O

"NOCIONES BASICAS SOBRE EL
MANEJO Y ANALISIS DE DATOS CON SAS"

CIAT, Agosto 4 a 8 de 1983

6669

EXPOSITORES: Maria Cristina Amézquita
Jefe Sección de Biometría, U.S.D.

Eduardo Granados
Consultor Estadístico
Sección de Biometría, U.S.D.

Preparado por: Maria Cristina Amézquita y Eduardo Granados
con la colaboración de:

Alfredo Rojas
Rosalba López de Barona
Argemiro Monsalve
Germán Lema



18 OCT. 1983

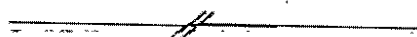
55912

CURSO
NOCIONES BASICAS SOBRE EL MANEJO Y
ANALISIS DE DATOS CON SAS
(Statistical Analysis System)

CONTENIDO

	<u>Pag.</u>
1. INTRODUCCION A SAS	3
- Qué es SAS	4
- Qué permite SAS	5
- Cómo está conformado?	6
- Cómo opera?	7
- En qué equipos de computación puede operar?	8
- Desarrollo histórico de SAS	10
- SAS y sus sub-sistemas opcionales	11
- SAS en el mundo	13
- Manuales disponibles	16
2. EL LENGUAJE SAS	19
2.1 El Lenguaje SAS	21
2.2 Una proposición SAS	22
2.3 Elementos del Lenguaje SAS	24
2.4 Funciones SAS	29
2.5 Un Archivo SAS	36
2.6 El Programa SAS	39
3. NOCIONES BASICAS SOBRE EL MANEJO DE DATOS CON SAS	42
3.1 Crear un Archivo	44
3.2 Ordenar un Archivo	45
3.3 Imprimir un Archivo	46
3.4 Obtener Sub-archivos	47

	<u>Pág.</u>
3.5 Concatenar Archivos	49
3.6 Intercalar Archivos	50
3.7 Mezclar Archivos	51
3.8 Actualizar un Archivo	53
4. SAS PARA EL ANALISIS DESCRIPTIVO	55
4.0 Definiciones	56
4.1 Proc Means	64
4.2 Proc Summary	74
4.3 Proc Freq	80
4.4 Proc Corr	88
4.5 Proc Plot	95
4.6 Proc Chart	103
5. SAS PARA EL ANALISIS ESTADISTICO INFERENCIAL	116
5.1 Conceptos	117
5.2 Análisis de Varianza para diseños balanceados	121
5.2.1 Conceptos sobre Diseño Experimental	122
5.2.2 Procedimientos SAS para Análisis de Varianza	141
5.2.3 Proc Anova	143
5.2.4 Diseño Completamente al Azar	144
5.2.5 Diseño Bloques Completos al Azar	152
5.2.6 Diseño Parcelas Divididas y Sub-divididas	168, 185
5.2.7 Diseño Franjas Divididas	193
5.2.8 Diseños con Factores Anidados	203
5.3 Análisis de Regresión bajo modelos lineales o linealizables	209
5.3.1 Conceptos	210
5.3.2 Procedimientos SAS para Análisis de Regresión	225
5.3.3 Proc Reg	226
5.3.4 Proc Stepwise	277



NOCIONES BASICAS SOBRE EL
MANEJO Y ANALISIS DE DATOS CON
S A S
(STATISTICAL ANALYSIS SYSTEM)

NOCIONES BASICAS SOBRE EL
MANEJO Y ANALISIS DE DATOS CON SAS

1. INTRODUCCION A SAS
2. EL LENGUAJE SAS
3. SAS PARA EL MANEJO DE DATOS
4. SAS PARA ANALISIS DESCRIPTIVO
5. SAS PARA ANALISIS ESTADISTICO (INFERENCIAL)

1. INTRODUCCION A SAS

QUE ES SAS?

ES UN SISTEMA DE COMPUTACION PARA MANEJO DE DATOS, PRODUCCION DE REPORTES, RECUPERACION DE INFORMACION Y ANALISIS ESTADISTICO.

ES UN LENGUAJE DE PROGRAMACION DE ALTO NIVEL CON UNA BIBLIOTECA COMPUESTA POR MAS DE 75 SUB-PROGRAMAS DISPONIBLES.

QUE PERMITE SAS?

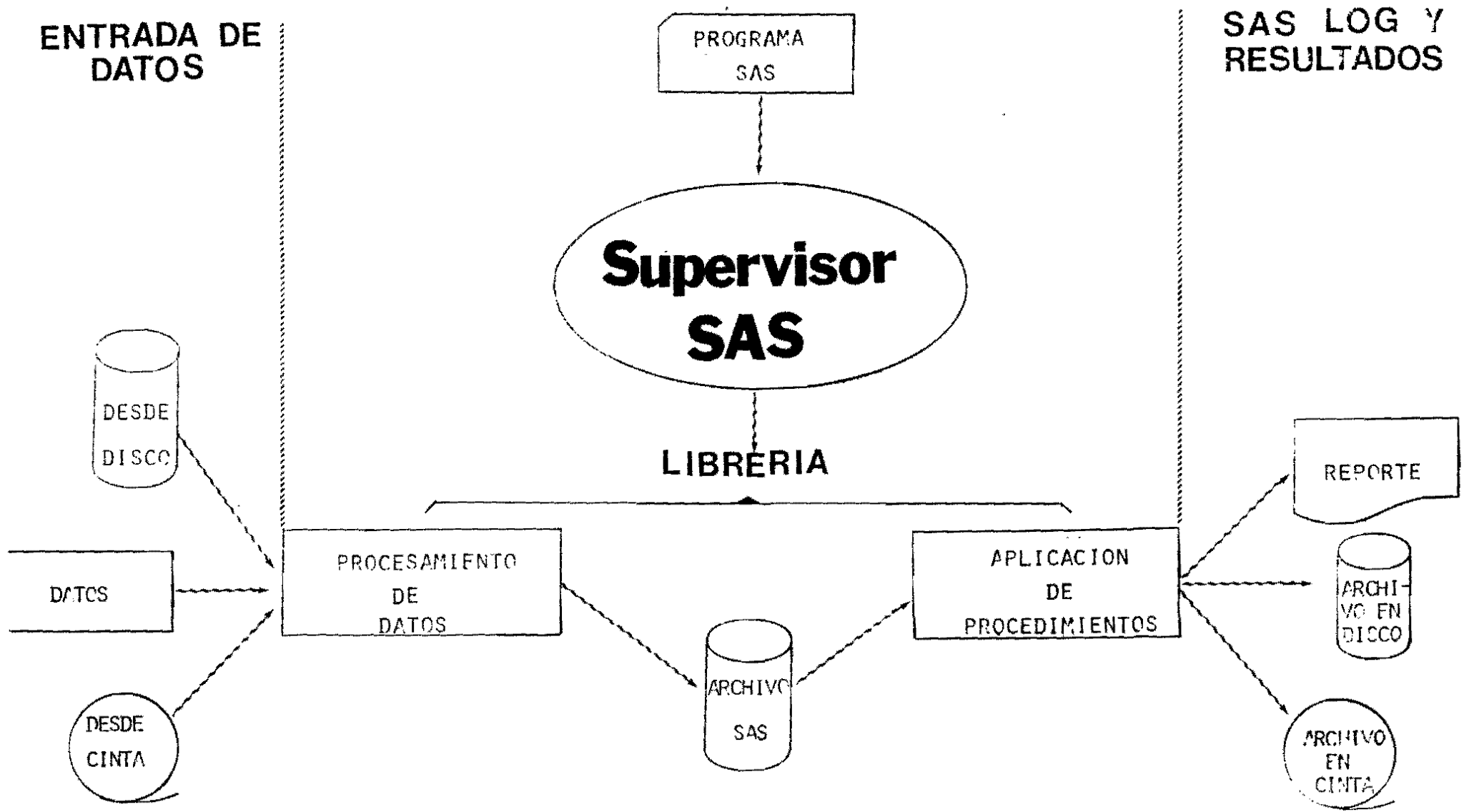
- . MANEJO DE DATOS
- . PRODUCCION DE REPORTES
- . RECUPERACION DE INFORMACION
- . ANALISIS ESTADISTICO
- . ANALISIS ECONOMETRICOS, FINANCIEROS Y DE SERIES DE TIEMPO
- . EVALUACION DEL COMPORTAMIENTO DEL EQUIPO DE COMPUTACION
- . OFRECE VARIOS PROGRAMAS DE UTILIDAD
- . PRODUCE GRAFICAS A COLOR
- . ENTRADA, RECUPERACION Y EDICION DE DATOS EN FORMA INTERACTIVA

COMO ESTA CONFORMADO SAS?

SAS ESTA CONFORMADO POR:

- UN LENGUAJE MANIPULADOR DE DATOS
(ESCRITO EN ASSEMBLER, P11 Y FORTRAN)
- UNA LIBRERIA DE PROCEDIMIENTOS
(> 75)

COMO OPERA SAS ?



EN QUE EQUIPOS DE COMPUTACION
PUEDE TRABAJAR SAS?

EN COMPUTADORES:

→
IBM 360/370/50cc/43xx
Y MAQUINAS COMPATIBLES

EN MODO:

→
INTERACTIVO Y "BATCH"

BAJO SISTEMAS OPERACIONALES:

→
OS, OS/VS, VM/CMS
DOS/VSE, TSO

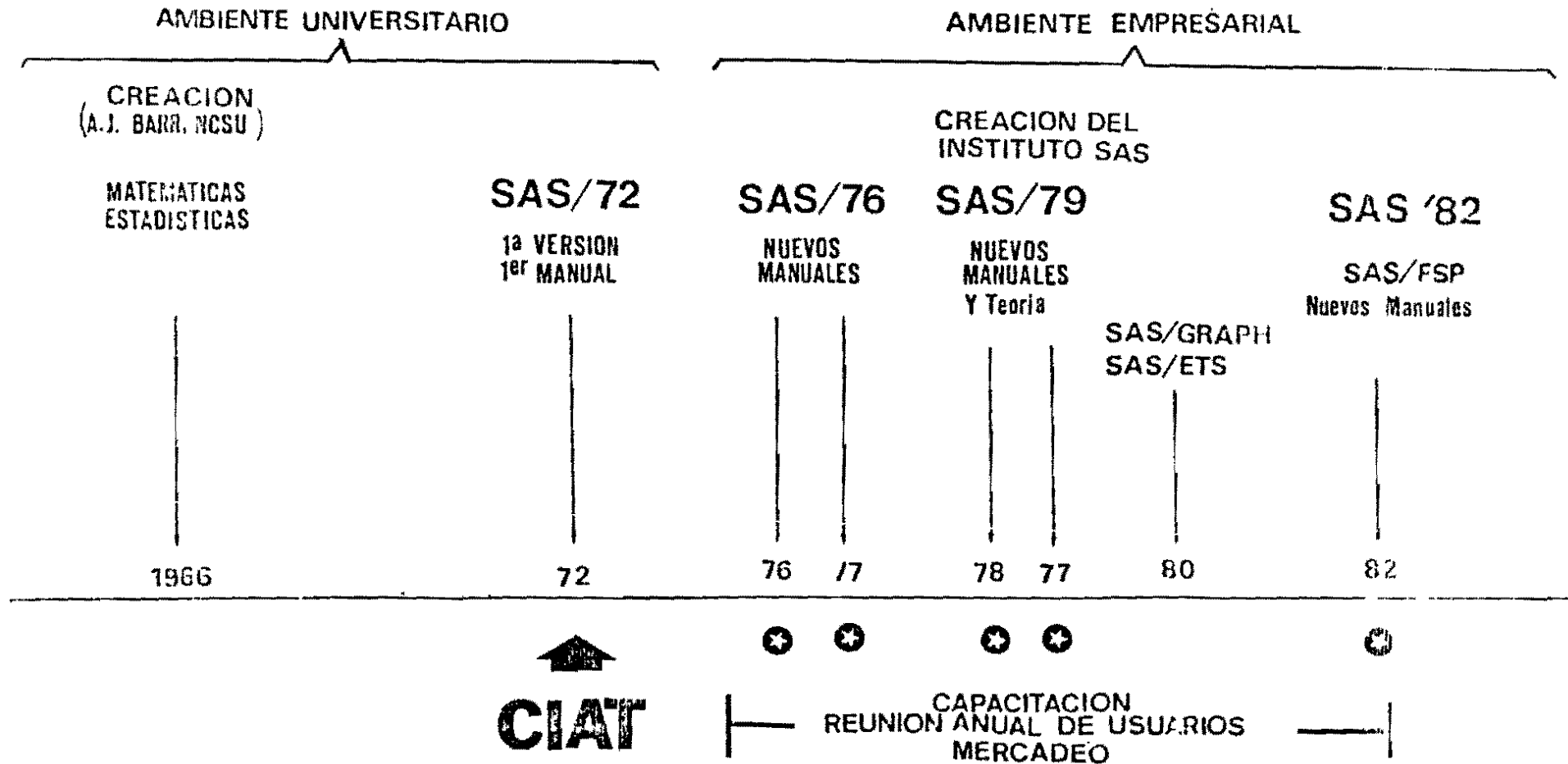
EN QUE EQUIPOS DE COMPUTACION
PUEDE TRABAJAR SAS?

ADEMAS,
DESDE JULIO 1983,

"EL SISTEMA SAS PORTATIL" (QUE CONTIENE UNA GRAN PARTE DE LOS PROGRAMAS
SAS) TRABAJA EN 3 MINICOMPUTADORES

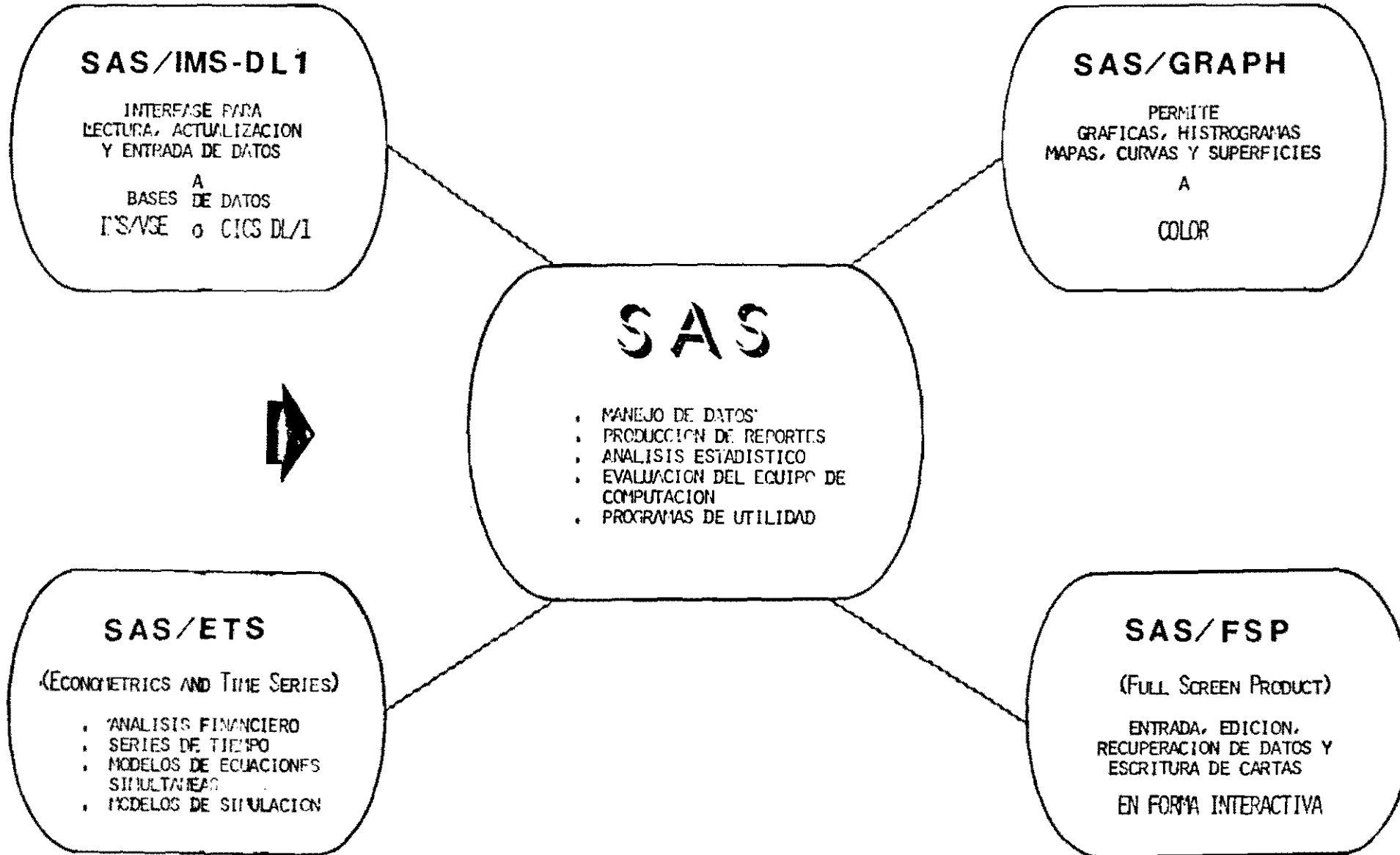
- . DEC (DIGITAL EQUIPMENT CORPORATION)
MODELO VAX 11/780
- . PRIME COMPUTER INC.
MODELO 550
- . DATA GENERAL
MODELO ECLIPSE MV/8000

DESARROLLO HISTORICO DE SAS



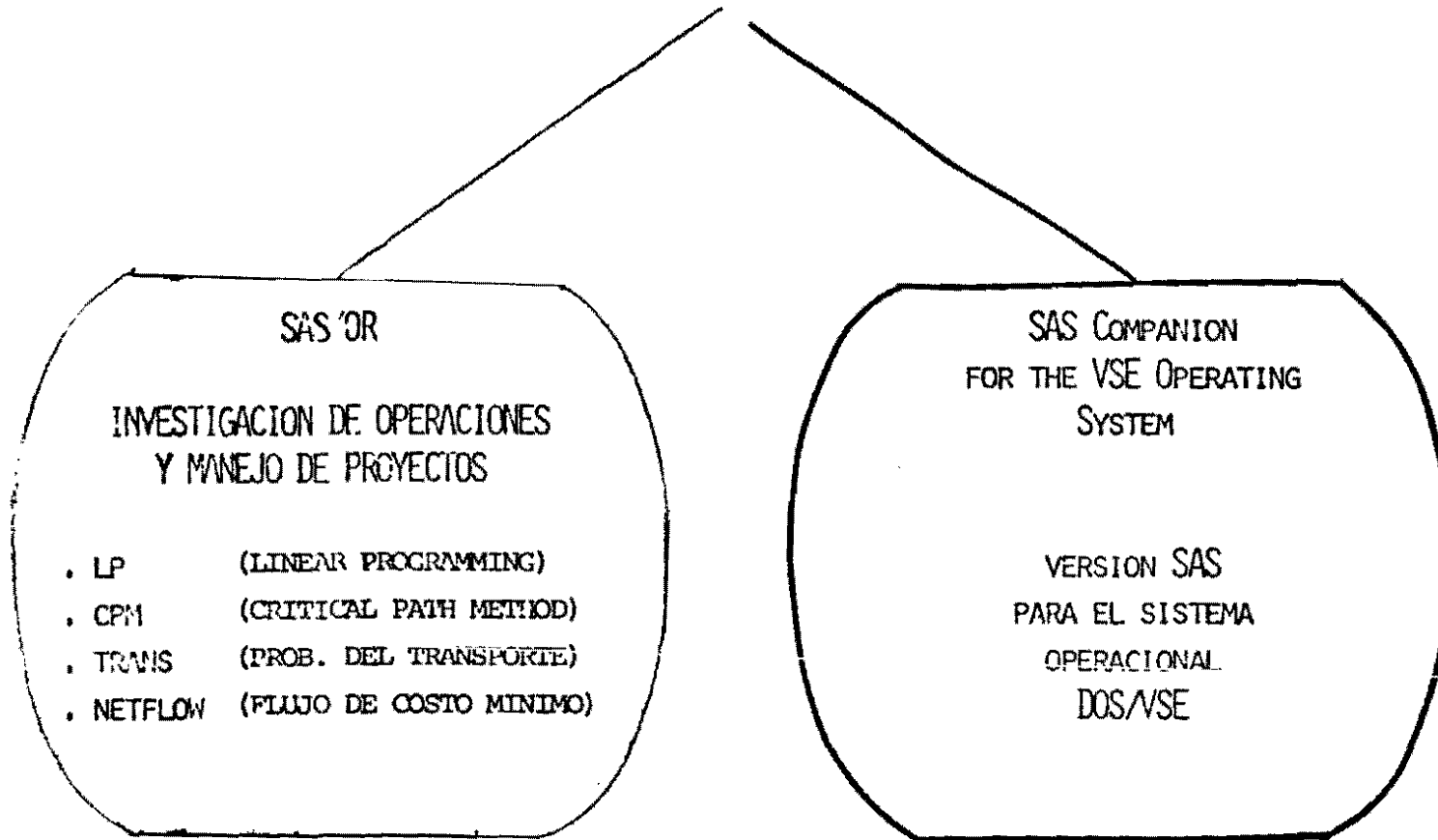
⊛ : PREMIO AL MEJOR SOFTWARE - DATAPRO - USA

SAS Y SUS SUBSISTEMAS OPCIONALES



ULTIMOS PRODUCTOS

SAS *



* Tomado de SAS Communications magazine, verano 1983

SAS EN EL MUNDO

SUBSIDIARIAS

- USA
- ALEMANIA
- INGLATERRA
- AUSTRALIA
- NUEVA ZELANDIA

AGENCIAS

- ITALIA
- SINGAPORE
- ISRAEL
- JAPON

USUARIOS DE SAS

6162 PRODUCTOS INSTALADOS
EN 3101 ORGANIZACIONES ALREDEDOR
DEL MUNDO
A DICIEMBRE 31 DE 1982

SAS EN EL MUNDO EMPRESARIAL
EMPRESAS DE USA QUE TRABAJAN
CON SAS

EXXON
IBM
GENERAL MOTORS
MOBIL
STANDARD OIL OF CALIFORNIA
TEXACO
STANDARD OIL OF INDIANA
GENERAL ELECTRIC
GULF OIL
SEARS ROEBUCK
'
'
'

If 48 of the 50 most
profitable companies
in the U.S. use SAS,
can you afford to
do without it?

A glance at the list below will show you the source of
our pride. These are the 50 most profitable* companies in
the United States, as compiled by *Forbes* magazine.

Rank of Company

1 AT&T	26 BankAmerica
2 Exxon	27 Procter & Gamble
3 IBM	28 Tenneco
4 General Motors	29 Xerox
5 Mobil	30 Occidental Petroleum
6 Standard Oil of California	31 Aetna Life & Casualty
7 Texaco	32 Union Carbide
8 Standard Oil of Indiana	33 R.J. Reynolds
9 General Electric	34 Citicorp
10 Gulf Oil	35 Ashland Oil
11 Standard Oil of Ohio	36 International Paper
12 Ford Motor	37 Weyerhaeuser
13 Atlantic Richfield	38 Philip Morris
14 Shell Oil	39 Amerada Hess
15 Eastman Kodak	40 Boeing
16 E.I. Dupont	41 Alcoa
17 Phillips Petroleum	42 Union Oil of California
18 Conoco	43 Caterpillar Tractor
19 Sears Roebuck	44 Pacific Gas & Electric
20 Dow Chemical	45 Coca-Cola
21 Sun Company	46 American Home Products
22 Schlumberger	47 Travelers
23 3M Company	48 Union Pacific
24 GTE	49 ITT
25 Getty Oil	50 Merck

And SAS is used by 83 of the top 100 companies on the *Forbes* list!

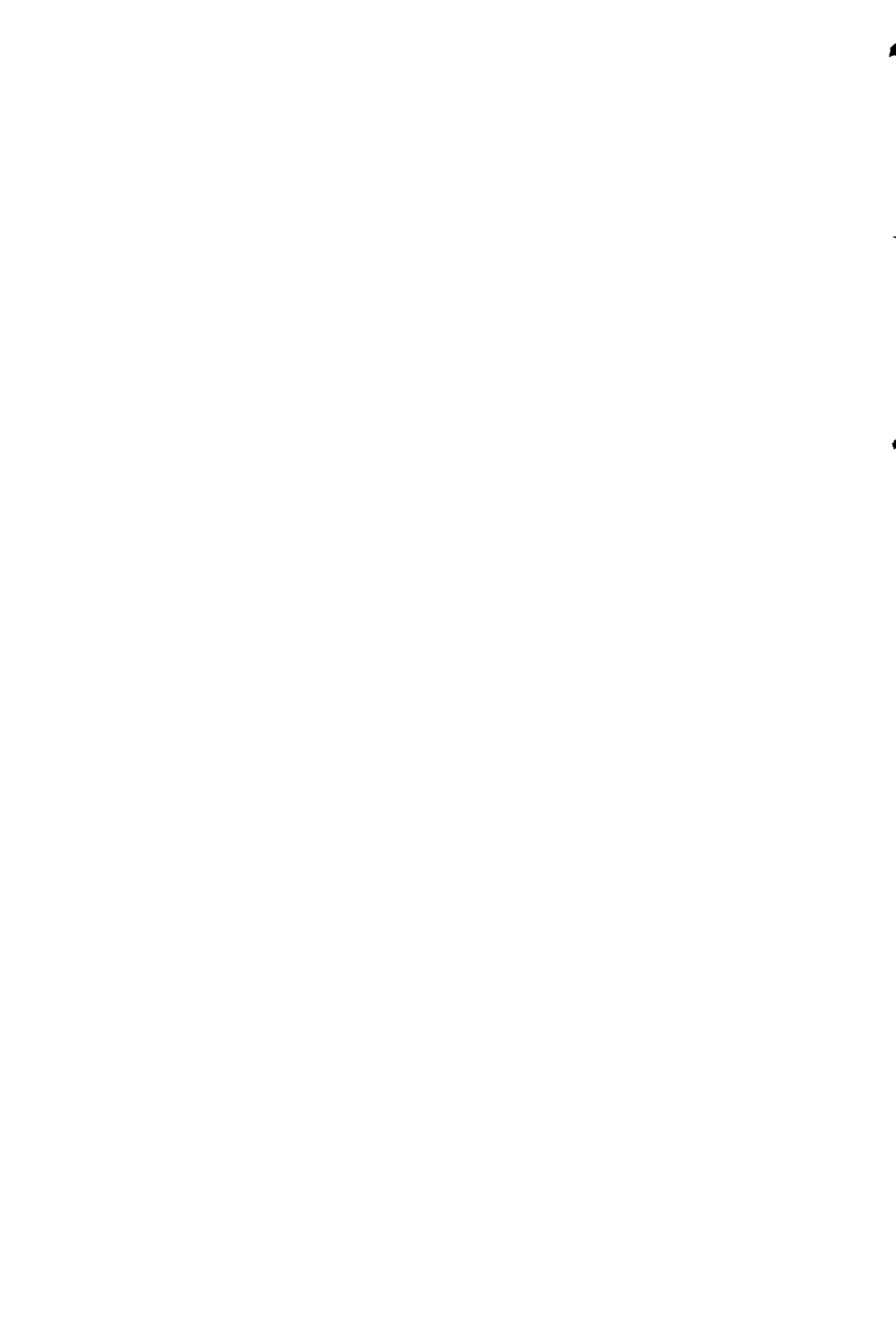
*Based on net profits for 1979.

The SAS logo consists of the letters 'SAS' in a large, bold, stylized, cursive font. The letters are black and have a slight shadow effect, giving them a three-dimensional appearance.

SAS Institute Inc.
Box 8000, Cary, NC 27511
919/467-8000 • Telex 802505

MANUALES DISPONIBLES SAS

1. SAS INTRODUCTORY GUIDE
(INGLES, ESPAÑOL, ALEMÁN Y FRANCÉS)
2. AN INTRODUCTION TO THE STATISTICAL ANALYSIS
SYSTEM 1972
A USER'S GUIDE TO SAS76
SAS USER'S GUIDE 1979 EDITION
SAS USER'S GUIDE: BASICS, 1982 EDITION
SAS USER'S GUIDE: STATISTICS, 1982 EDITION
3. SAS SUPPLEMENTAL LIBRARY USER'S GUIDE 1972
SAS SUPPLEMENTAL LIBRARY USER'S GUIDE 1977
SAS SUPPLEMENTAL LIBRARY USER'S GUIDE 1979
4. SAS PROGRAMMER'S GUIDE 1977
SAS PROGRAMMER'S GUIDE 1979
SAS PROGRAMMER'S GUIDE 1981
5. SAS VIEWS
6. SAS APPLICATIONS GUIDE



MANUALES DISPONIBLES

SAS/ETS:

- SAS/ETS USER'S GUIDE 1980
- SAS/ETS USER'S GUIDE 1982

SAS/GRAPH:

- SAS/GRAPH USER'S GUIDE 1981
- SAS/GRAPH VIEWS

SAS/FSP:

- SAS/FSP USER'S GUIDE 1982

SAS/IMS:

- SAS/IMS-DL/I USER'S GUIDE 1982

MANUALES DISPONIBLES
SAS
PARA DIFERENTES SISTEMAS OPERACIONALES

- SAS/CMS COMPANION 1981
- SAS COMPANION FOR THE VSE OPERATING SYSTEM 1982

2. EL LENGUAJE SAS

EL LENGUAJE SAS

CONCEPTOS

1. QUE ES EL LENGUAJE SAS
2. UNA PROPOSICION SAS
3. ELEMENTOS DEL LENGUAJE SAS
 - . PALABRAS CLAVES
 - . NOMBRES
 - . CARACTERES ESPECIALES
 - . OPERADORES
4. FUNCIONES SAS
5. UN ARCHIVO SAS
 - . DATOS
 - . OBSERVACIONES
 - . VARIABLES
6. EL PROGRAMA SAS
 - LOS 2 "PASOS" SAS
 - . PASO DATA
 - . PASO PROC

2.1 QUE ES EL LENGUAJE SAS

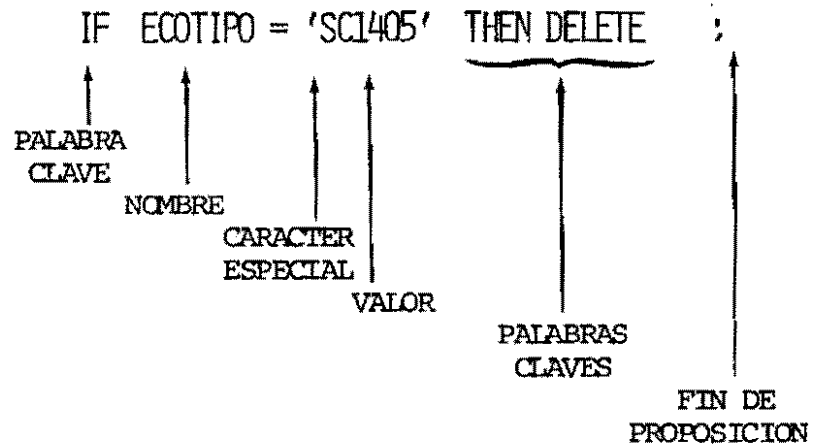
COMO CUALQUIER LENGUAJE, SAS TIENE SU VOCABULARIO Y SINTAXIS PROPIOS Y UN CONJUNTO DE REGLAS QUE INDICAN SU USO APROPIADO.

CON EL LENGUAJE SAS, USTED DEFINE SUS DATOS Y LO QUE USTED QUIERE QUE SAS HAGA CON ELLOS, MEDIANTE UNA SECUENCIA DE "PROPOSICIONES SAS" QUE CONSTITUYEN UN "PROGRAMA SAS",

2.2 UNA PROPOSICION SAS

ES UNA CADENA DE "PALABRAS CLAVES", "NOMBRES", "CARACTERES ESPECIALES", "VALORES" Y "OPERADORES", QUE COMIENZA CON UNA PALABRA CLAVE Y TERMINA CON UN ;

EJEMPLO:



EJEMPLOS DE PROPOSICIONES SAS

1. DATA A ;
2. INPUT X1 1-2 X2 5-6 ;
3. IF (X1 + X2) > = 100 THEN DELETE ;
4. CARDS ;

35	42	} NO SON PROPOSICIONES
50	60	
80	25	
5. PROC MEANS DATA = A ;
6. VARIABLE X1 ;

2.3 ELEMENTOS DEL LENGUAJE SAS

PALABRAS CLAVES:

SON SECUENCIAS DE CARACTERES QUE TIENEN UN SIGNIFICADO PRECISO PARA SAS, POR SI MISMAS O POR EL CONTEXTO EN QUE SE ENCUENTRAN.

EJEMPLOS:

```
DATA A ;  
INPUT X1 1-2 X2 5-6 ;  
IF (X1+X2) > = 100 THEN DELETE ;  
CARDS ;
```

- DATOS -

```
PROC MEANS DATA = A ;  
VARIABLE X1 ;
```

ELEMENTOS DEL LENGUAJE SAS

NOMBRES

SON SECUENCIAS DE 1 A 8 CARACTERES DEFINIDAS POR EL PROGRAMADOR Y UTILIZADAS PARA IDENTIFICAR ARCHIVOS, VARIABLES, FORMATOS, ETC.

EJEMPLO:

```
DATA A ;  
INPUT X1 1-2 X2 5-6 ;  
IF (X1 + X2) > = 100 THEN DELETE;  
CARDS ;  
      - DATOS -  
PROC MEANS DATA=A ;  
VARIABLE X1 ;
```

ELEMENTOS DEL LENGUAJE SAS

CARACTERES ESPECIALES:

SON SIMBOLOS QUE TIENEN SIGNIFICADO MUY PRECISO PARA SAS Y NO FORMAN PARTE DE PALABRAS CLAVES O NOMBRES.

- , : PUNTO DECIMAL
- ; : FIN DE PROPOSICION
- \$: INDICA QUE LA VARIABLE ES DE CARACTERES
- % : FIN DE MACRO
- = : SIGNO IGUAL UTILIZADO PARA IDENTIFICAR ARCHIVOS
- | : INTERACCION DE EFECTOS

EJEMPLO:

```
DATA ALUMNOS ;  
INPUT NOMBRE $ 1-20 SEXO $22 EDAD 25-27;  
CARDS ;  
    _ DATOS _  
PROC PRINT DATA = ALUMNOS ;
```


ELEMENTOS DEL LENGUAJE SAS

OPERADORES

SON SIMBOLOS QUE INDICAN OPERACIONES MATEMATICAS O LOGICAS.

EJEMPLOS:

+	:	SUMA	<	:	MENOR QUE
-	:	RESTA	<=	:	MENOR O IGUAL A
*	:	MULTIPLICACION	>	:	MAYOR QUE
/	:	DIVISION	>=	:	MAYOR O IGUAL A
**	:	EXPONENCIACION	≠	:	NO ES IGUAL A
			=	:	IGUAL A

, ENTRE OTROS.

ELEMENTOS DEL LENGUAJE SAS

EJEMPLO DE USO DE OPERADORES:

```
DATA SUELDOS ;
INPUT NOMBRE $ 1-20 X1 22-25 X2 27-30
      X3 32-35 X4 37-39 ;

A = 2.3 * X1/X2 ;
B = (1.8 * X4) ** 2 ;
C = (A + B)/X2 ;
SALARIO = 2.3 * C - 10000 ;
IF A>3 AND SALARIO > 20000 THEN OUTPUT ;
CARDS;
      - datos -
PROC PRINT DATA=SUELDOS ;
```

2.4 FUNCIONES SAS

FUNCION: ES UNA RUTINA QUE PRODUCE UN VALOR
CALCULADO A PARTIR DE ARGUMENTOS.
EN GENERAL, LA FUNCION SE EJECUTA
TANTAS VECES COMO OBSERVACIONES
HAYA EN EL ARCHIVO.

FORMA DE UNA FUNCION SAS:

NOMBRE (ARGUMENTO1, ARGUMENTO2, ... ARGUMENTO_n)

FUNCIONES SAS

EJEMPLOS DE FUNCIONES SAS

<u>FUNCION</u>	<u>VALOR RESULTANTE</u>
ABS (-2.4)	2.4
INT (2.4)	2
FLOOR (2.4)	2
CEIL (2.4)	3
LOG10 (1)	0
LOG (1)	0
MEAN (8,4,0)	4
NORMAL(entero impar)	Valor de una variable N(0,1)
MDY (01, 01, 61)	366

FUNCIONES SAS

TIPOS DE FUNCIONES

- . ARITMETICAS
- . DE REDONDEO
- . MATEMATICAS
- . TRIGONOMETRICAS
- . PROBABILISTICAS
- . DE ESTADISTICAS MUESTRALES
- . GENERADORAS DE NUMEROS ALEATORIOS
- . DE CARACTERES
- . DE FECHA Y DE TIEMPO
- . DE CODIGO POSTAL Y ESTADO (para USA)
- . DEL SISTEMA OPERACIONAL
- . ESPECIALES

FUNCIONES SAS

ARITMETICAS

. ABS . MOD
. MAX . SIGN
. MIN . SORT

DE REDONDEO
. CEIL . INT
. FLOOR . ROUND

DE ESTADISTICAS MUESTRALES

. N . MEAN
. NMISS . SUM
. RANGE . STD
. MIN . STDERR
. MAX . SKEWNESS
. CV . KURTOSIS
. VAR . USS
. CSS

FUNCIONES SAS

MAIEMATICAS

. LOG
. LOG10
. LOG2
. EXP
. GAMMA
. DIGAMMA
. LGAMMA
. ERF

IRIGONOMETRICAS

. SIN
. COS
. TAN
. ARSIN
. ARCOS
. ATAN
. SINH
. COSH
. TANH

FUNCIONES SAS

PROBABILIDAD

- . POISSON
- . PROBF
- . PROBT
- . PROBNORM
- . PROBCHI
- . PROBYPR
- . PROBNEGB
- . BETAINV
- . GAMINV
- . PROBBETA
- . PROBBNML
- . PROBGAM
- . PROBIT

NUMEROS ALEATORIOS

- . UNIFORM
- . NORMAL
- . RANBIN
- . RANEXP
- . RANGAM
- . RANCAU
- . RANUNI
- . RANTRI
- . RANTBL
- . RANNOR
- . RANPOI

FUNCIONES SAS

CARACTERES

. LENGTH	. COLLATE
. VERIFY	. INDEXC
. LEFT	. REPEAT
. TRANSLATE	. REVERSE
. INDEX	. RIGHT
. COMPRESS	. SCAN
. TRIM	. UPCASE
. SUBSTR	

FECHA Y TIEMPO

. DATE	. DATEPART
. DATEJUL	. DATETIME
. MDY	. TIME
. YEAR	. HOUR
. MONTH	. TODAY
. DAY	. TIMEPART
. JULDATE	. QTR
. WEEKDAY	. YYQ
. Y OTRAS	

2.5 UN ARCHIVO SAS

UN ARCHIVO SAS ES UN CONJUNTO DE DATOS ARREGLADOS
EN FORMA RECTANGULAR,
LA UNIDAD BASICA DE UN ARCHIVO ES

UN DATO

LAS FILAS DEL ARCHIVO SE DENOMINAN

OBSERVACIONES (SUJETOS)

LAS COLUMNAS DEL ARCHIVO SE DENOMINAN

VARIABLES (CARACTERISTICAS DE LOS
SUJETOS)

UN ARCHIVO SAS

EJEMPLO:

-----VARIABLES-----

NOMBRE	SECCION	SALARIO	AÑOINAC
PAZ	07	45000	1958
GONZALEZ	02	40000	1946
VELEZ	07	23500	1962
.			
.			
.			
NIETO	07	46500	1953

- 1.
- 2.
- 3.
- .
- .
- .
- 11.

-----OBSERVACIONES-----

UN ARCHIVO SAS

TIPO DE VARIABLES:

- NUMERICAS: SUS DATOS SON NUMEROS (DE 10^{-73} A 10^{73})
- EJ: - RENDIMIENTO (CUANTITATIVA CONTINUA)
- NO. DE PLANTAS (CUANTITATIVA DISCRETA)
- ESCALA NUMERICA DE DAÑO (CUALITATIVA)
- DE CARACTERES: SUS DATOS SON CARACTERES (HASTA 200 LETRAS Y NUMEROS)
- EJ: - TIPO DE DAÑO (CUALITATIVA)
- NOMBRE (DESCRIPTIVA)

2.6 UN PROGRAMA SAS

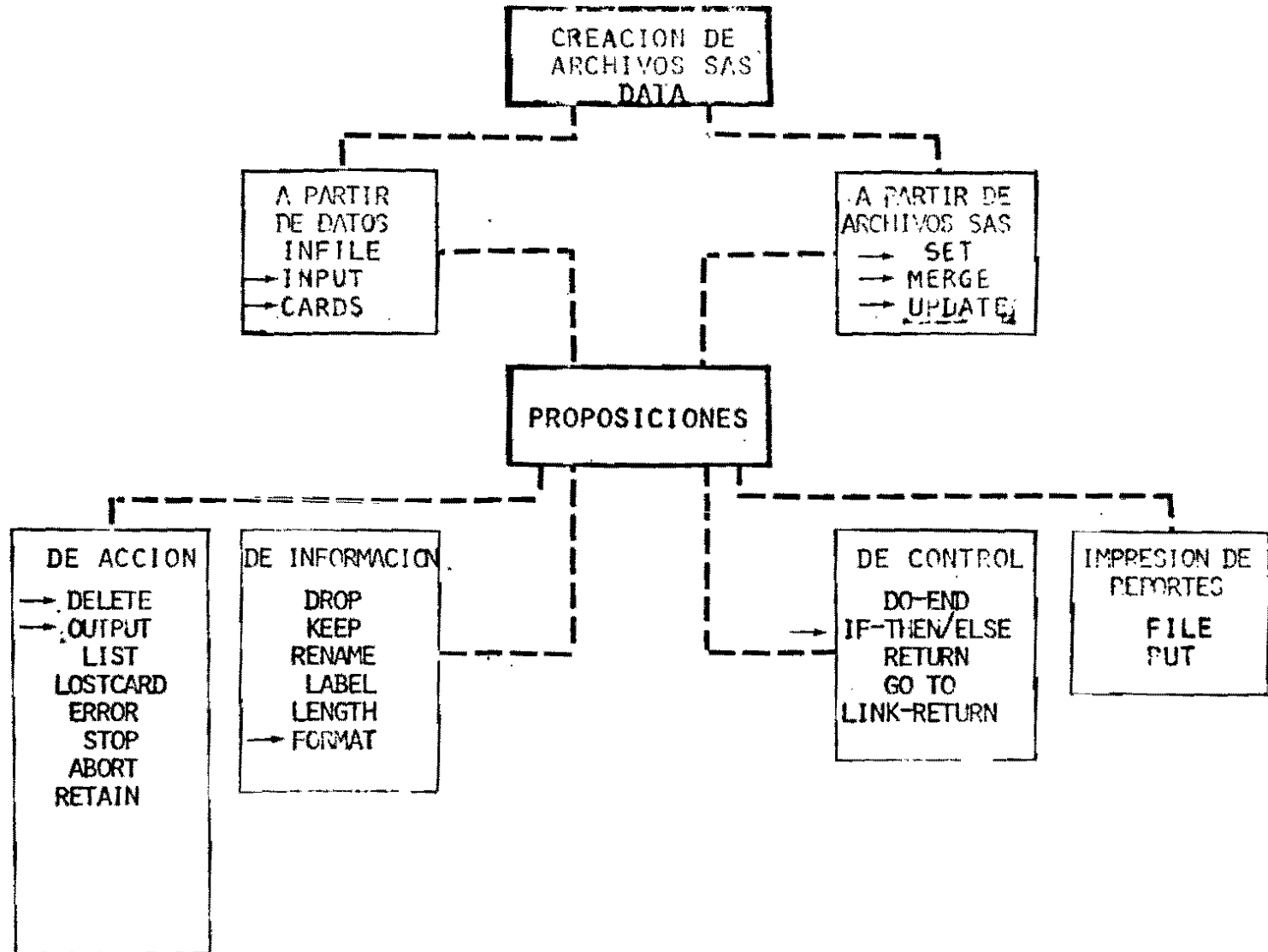
ES UNA SECUENCIA DE PROPOSICIONES SAS AGRUPADAS EN "PASOS" (SAS STEPS)
HAY 2 TIPOS DE PASOS SAS:

- . PASO DATA: PREPARA ARCHIVOS SAS
 - MANEJO DE ARCHIVOS
 - RECUPERACION DE DATOSSE MANEJAN CON PASOS DATA

- . PASO PROC: ANALIZA O PROCESA ARCHIVOS SAS
 - ANALISIS ESTADISTICO
 - ANALISIS ECONOMETRICO
 - ANALISIS FINANCIERO Y
PRODUCCION DE INFORMESSE MANEJAN CON PASOS DATA Y PASOS PROC

PALABRAS CI AVES Y PROPOSICIONES UTILIZADAS

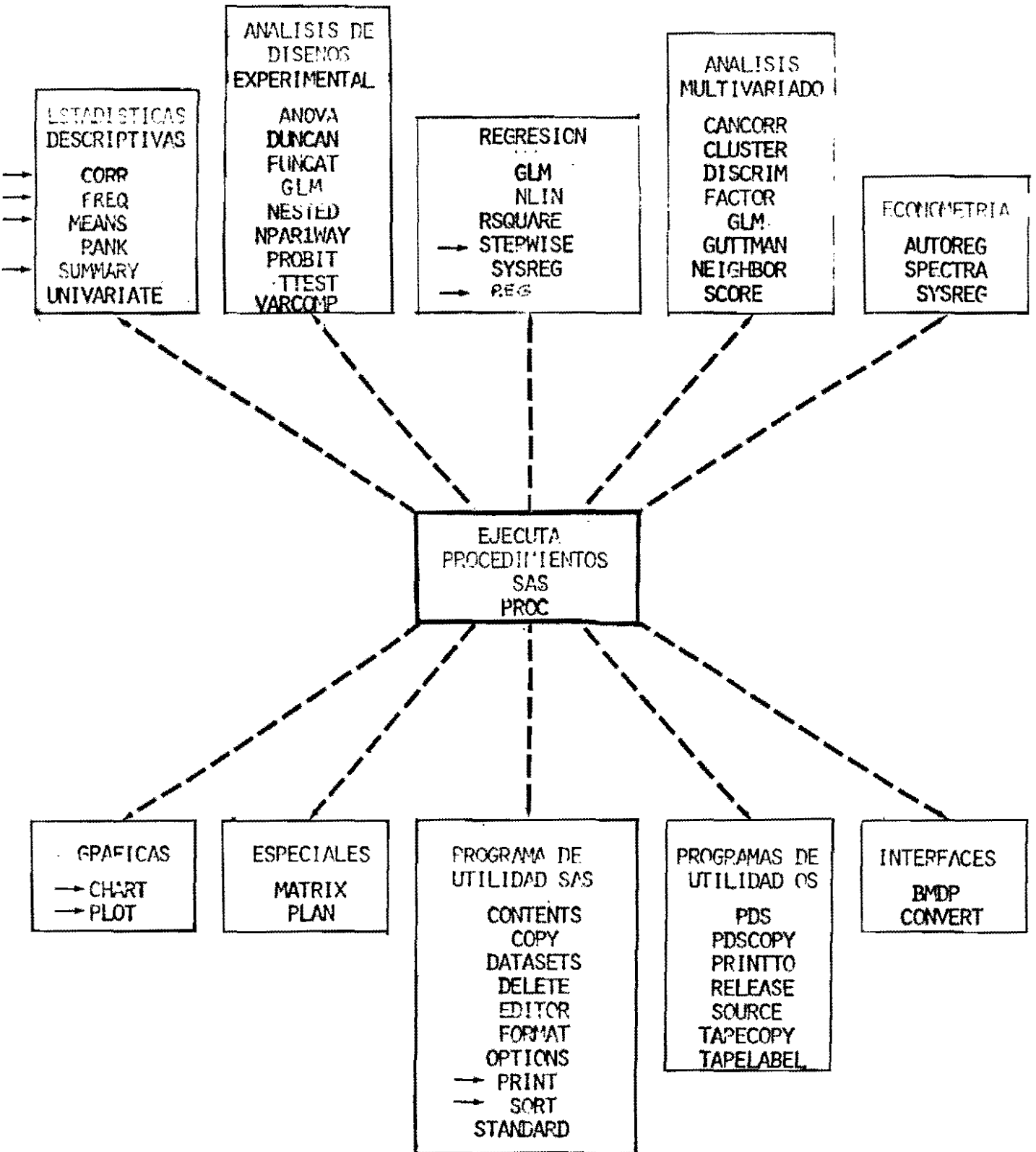
Paso Data



LOS DOS PASOS SAS

PALABRAS C AV Y PROPOSICIONES UTILIZADAS

Paso Proc



3. NOCIONES BASICAS SOBRE
MANEJO DE DATOS CON
SAS

SAS PARA EL MANEJO DE DATOS

FUNCIONES	INSTRUCCIONES SAS
1. CREAR UN ARCHIVO	INPUT
2. ORDENAR UN ARCHIVO	PROC SORT
3. IMPRIMIR UN ARCHIVO	PROC PRINT
4. OBTENER SUB-ARCHIVOS DE UN ARCHIVO	SET
5. CONCATENAR 2 O MAS ARCHIVOS	SET
6. INTERCALAR 2 O MAS ARCHIVOS	SET CON BY
7. MEZCLAR 2 O MAS ARCHIVOS	MERGE
8. ACTUALIZAR UN ARCHIVO	SET, UPDATE

3.1 CREAR UN ARCHIVO

EJEMPLO No. 1:

TOME LOS DATOS DEL ARCHIVO DE PERSONAL
DE LA EMPRESA Y CREE UN ARCHIVO SAS.

DATA PERSONAL;

INPUT NOMBRE \$ 1-10 SECCION 13-14
SALARIO 17-21 AÑONAC 24-27;

CARDS;

3.2 ORDENAR UN ARCHIVO

EJEMPLO No. 2:

ORDENE EL ARCHIVO PERSONAL POR ORDEN
ASCENDENTE DE SALARIOS.

```
PROC SORT DATA=PERSONAL; BY SALARIO;
```

3.3 IMPRIMIR UN ARCHIVO

EJEMPLO No. 3:

IMPRIMA EL ARCHIVO QUE USTED HABIA PREVIAMENTE ORDENADO
POR SALARIOS.

```
PROC PRINT DATA=PERSONAL;
```

OBS	NOMBRE	SECCION	SALARIO	AÑONAC
1	JIMENEZ	02	10000	1945
2	GARCIA	02	12000	1962
3	CASTRO	04	12500	1960
4	LOPEZ	04	15000	1950
5	PEREZ	04	15800	1949
6	ACOSTA	02	21000	1963
7	VELEZ	07	23500	1962
8	REJIREDO	07	23600	1954
9	GONZALEZ	02	40000	1946
10	PAZ	07	45000	1958
11	NIETO	07	46500	1953

3.4A OBTENER SUB-ARCHIVOS DE UN ARCHIVO

EJEMPLO No. 4:

EL JEFE DE LA SECCION 07 NECESITA LAS EDADES DE SUS EMPLEADOS.
DESEA IMPRIMIR ESE SUB-ARCHIVO CON LAS VARIABLES "NOMBRE"
"SECCION" Y "EDAD".

```
DATA B; SET PERSONAL;  
IF SECCION NE 07 THEN DELETE; (*)  
EDAD = 1983 - AÑONAC;  
KEEP NOMBRE SECCION EDAD;  
PROC PRINT;
```

CBS	NOMBRE	SECCION	EDAD
1	VELEZ	07	21
2	RESTREPO	07	29
3	PAZ	07	25
4	NIEYO	07	30

(*)

IF SECCION = 07;
(es instrucción equivalente)

3.4B OBTENER 2 O MAS SUB-ARCHIVOS DE UN ARCHIVO

EJEMPLO No. 5:

CADA JEFE DE SECCION DESEA LA LISTA DE SU PERSONAL PARA VERIFICACION.GENERE ESOS ARCHIVOS Y LISTELOS.

```
DATA sec02 sec04 sec07;  
SET PERSONAL;  
KEEP NOMBRE;  
IF SECCION=02 THEN OUTPUT SEC02;  
IF SECCION=04 THEN OUTPUT SEC04;  
IF SECCION=07 THEN OUTPUT SEC07;  
PROC PRINT DATA=SEC02;  
PROC PRINT DATA=SEC04;  
PROC PRINT DATA=SEC07;
```

ARCHIVO SEC02		ARCHIVO SEC04		ARCHIVO SEC07	
OBS	NOMBRE	OBS	NOMBRE	OBS	NOMBRE
1	JIMENEZ	1	CASTRO	1	VELEZ
2	GARCIA	2	LOPEZ	2	RESTREPO
3	ACOSTA	3	PEREZ	3	PAZ
4	GONZALEZ			4	NIETO

3.5 CONCATENAR 2. O MAS ARCHIVOS

EJEMPLO No. 6:

SE DESEA CREAR UN ARCHIVO QUE CONTENGA LOS EMPLEADOS DE LAS SECCIONES 04 Y 07.

```
DATA UNIDOS;  
SET SEC04 SEC07;  
PROC PRINT;
```

ARCHIVO UNIDOS

OBS	NOMBRE
1	CASTRO
2	LOPEZ
3	PEREZ
4	VELEZ
5	RESTREPO
6	PAZ
7	NIEVO

3.6 INTERCALAR 2 O MAS ARCHIVOS

EJEMPLO No. 7:

SE DESEA CREAR UN ARCHIVO QUE CONTenga LOS EMPLEADOS DE LAS SECCIONES 04 Y 07 PERO ORDENADO POR NOMBRE.

```
PROC SORT DATA=SEC04 ; BY NOMBRE ;  
PROC SORT DATA=SEC07 ; BY NOMBRE ;  
DATA INTERCAL;  
SET SEC04 SEC07 ; BY NOMBRE ;  
PROC PRINT;
```

ARCHIVO INTERCAL

OBS	NOMBRE
1	CASTRO
2	LOPEZ
3	NIETO
4	PAZ
5	PEREZ
6	RESTREPO
7	VELEZ

3.7 MEZCLAR 2 O MAS ARCHIVOS:

EJEMPLO No. 8:

LA EMPRESA TAMBIEN MANTIENE UN ARCHIVO DE EMPLEADOS QUE CONTIENE DIAS DE VACACION ACUMULADAS, DE NOMBRE "VACACION", QUE LUCE ASI:

ARCHIVO "VACACION"

CBS	NOMBRE	DIASVAC
1	JIMENEZ	15
2	NIETO	32
3	PEREZ	28
4	VELEZ	15

MEZCLAR 2 O MAS ARCHIVOS: (CONT.)

EJEMPLO No. 8: (CONT.)

EL JEFE DE LA SECCION 07 DESEA UNA LISTA DE
TODOS SUS EMPLEADOS CON SUS DIAS DE VACACION
ACUMULADOS, SI LOS TIENEN.

```
PROC SORT DATA=SEC07; BY NOMBRE;  
PROC SORT DATA=VACACION; BY NOMBRE;  
DATA VAC07;  
MERGE SEC07 (IN=S7) VACACION (IN=V); BY NOMBRE;  
IF S7=1 THEN OUTPUT;  
PROC PRINT;
```

ARCHIVO "VAC07"

OBS	NOMBRE	DIASVAC
1	NIEYO	32
2	PAZ	.
3	RESTREFO	.
4	VELEZ	15

3.8 ACTUALIZAR UN ARCHIVO

EJEMPLO No. 9:

EL EMPLEADO LOPEZ RECIBIO UN INCREMENTO SALARIAL, SIENDO SU NUEVO SALARIO DE \$20000. SE DESEA ACTUALIZAR EL ARCHIVO PERSONAL.

ALTERNATIVA 1: (SET)

```
DATA PERSONAL; SET PERSONAL;  
IF NOMBRE='LOPEZ' THEN SALARIO=20000;
```

ACTUALIZAR UN ARCHIVO

EJEMPLO No. 9 (CONT.)

DATA CAMBIOS;

INPUT NOMBRE \$1-10 SALARIO 12-17;

CARDS;

LOPEZ 20000 ← DATOS

PROC SORT DATA=CAMBIOS; BY NOMBRE;

PROC SORT DATA=PERSONAL; BY NOMBRE;

DATA PERSONAL;

UPDATE PERSONAL CAMBIOS; BY NOMBRE;

ARCHIVO "PERSONAL" ACTUALIZADO

OBS	NOMBRE	SECCION	SALARIO	ANOVAC
1	ACCOSTA	02	21000	1963
2	CASTRO	04	12500	1960
3	GARCIA	02	12000	1962
4	GONZALEZ	02	40000	1946
5	JIMENEZ	02	10000	1945
6	LOPEZ	04	20000	1950
7	NIEVO	07	46500	1953
8	PAZ	07	45000	1958
9	PEREZ	04	15800	1949
10	RESTREPO	07	23600	1954
11	VELEZ	07	23500	1962

4. SAS PARA EL
ANALISIS DESCRIPTIVO

EL ANALISIS DESCRIPTIVO

EL ANALISIS DESCRIPTIVO CONSISTE EN SINTETIZAR UN CONJUNTO DE DATOS, YA SEA MEDIANTE ESTADISTICAS DESCRIPTIVAS QUE LO CARACTERICEN, O MEDIANTE DISTRIBUCIONES DE FRECUENCIA , RELACIONES GRAFICAS ENTRE VARIABLES Y DIAGRAMAS DIVERSOS.

ESTADÍSTICAS DESCRIPTIVAS

- DEFINICIONES -

1. MEDIDA DE TENDENCIA CENTRAL:

MEDIA ARITMÉTICA :
(O PRIMER MOMENTO)

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

ESTADÍSTICAS DESCRIPTIVAS

- DEFINICIONES -

2. MEDIDAS DE DISPERSION:

VARIANZA
(O SEGUNDO MOMENTO)

:

$$\sigma^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$

DESVIACION ESTANDAR
POBLACIONAL

:

$$\sigma = \sqrt{\sigma^2}$$

DESVIACION ESTANDAR
MUESTRAL

:

$$s = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1}}$$

COEFICIENTE DE
VARIACION

:

$$CV = \frac{s}{\bar{X}} \times 100$$

ERROR ESTANDAR O
DESVIACION ESTANDAR
DE LA MEDIA

:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

ESTADÍSTICAS DESCRIPTIVAS
- DEFINICIONES -

3. MEDIDAS DE SIMETRÍA DE LA DISTRIBUCIÓN:

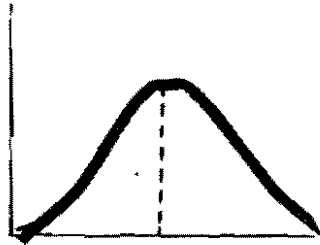
TERCER MOMENTO :

$$M_3 = \frac{(X_1 - \bar{X})^3 + (X_2 - \bar{X})^3 + \dots + (X_n - \bar{X})^3}{n}$$

SKEWNESS :

$$\text{SKEWNESS} = \frac{M_3}{(\sigma^2)^{3/2}}$$

-59-



\bar{X}

SIMETRICA

$$M_3 = 0$$

$$-0.5 < \text{SKEWNESS} < 0.5$$

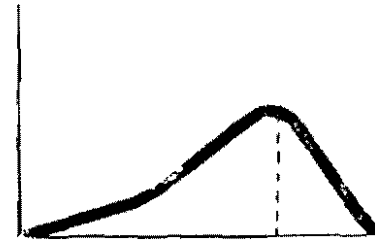


\bar{X}

POSITIVAMENTE SESGADA

$$M_3 > 0$$

$$\text{SKEWNESS} \geq 0.5$$



\bar{X}

NEGATIVAMENTE SESGADA

$$M_3 < 0$$

$$\text{SKEWNESS} \leq -0.5$$

ESTADISTICAS DESCRIPTIVAS

- DEFINICIONES -

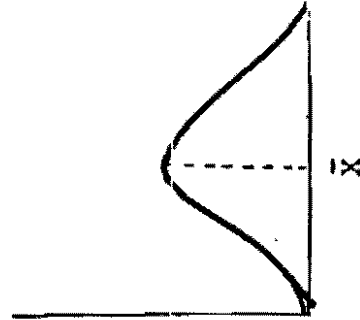
4. MEDIDAS DE CUAN PLANA ES LA DISTRIBUCION:

CUARTO MOMENTO :

$$M_4 = \frac{(X_1 - \bar{X})^4 + (X_2 - \bar{X})^4 + \dots + (X_n - \bar{X})^4}{n}$$

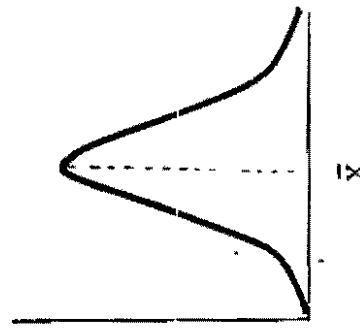
KURTOSIS :

$$\text{KURTOSIS} = \frac{M_4}{(\sigma^2)^2}$$



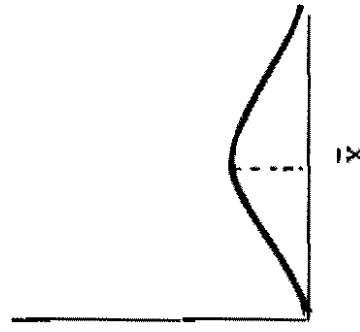
NORMAL

KURTOSIS = 3



MAS "PICUDA" QUE LA NORMAL

KURTOSIS > 3



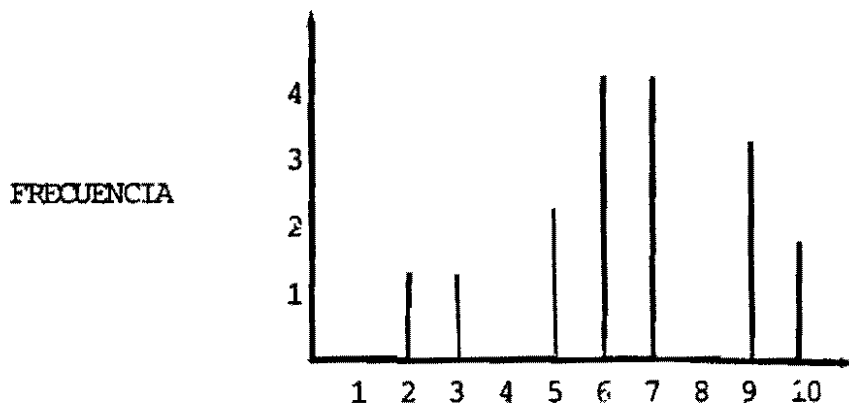
MAS "APLANADA" QUE LA NORMAL

KURTOSIS < 3

EJEMPLO:

CALCULAR LAS ESTADISTICAS DESCRIPTIVAS DEL CONJUNTO DE DATOS
{2, 3, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 9, 9, 9, 10}

DISTRIBUCION DE FRECUENCIAS



NUMERO DE DATOS	N = 16
MEDIA ARITMETICA	$\bar{X} = 6.5$
VARIANZA	$\sigma^2 = 4.38$
DESVIACION ESTANDAR	S = 2.16
ERROP. ESTANDAR	$S_{\bar{X}} = 0.54$
M ₃	= -3.19
SKEWNESS	= -0.35 ASIMETRICA. NEGATIVAMENTE SESGADA
M ₄	= 52.38
KURTOSIS	= 2.73 CASI TAN "PICUDA" COMO LA NORMAL

ESTADISTICAS DESCRIPTIVAS
- DEFINICIONES -

5. MEDIDA DE RELACION LINEAL ENTRE DOS VARIABLES:

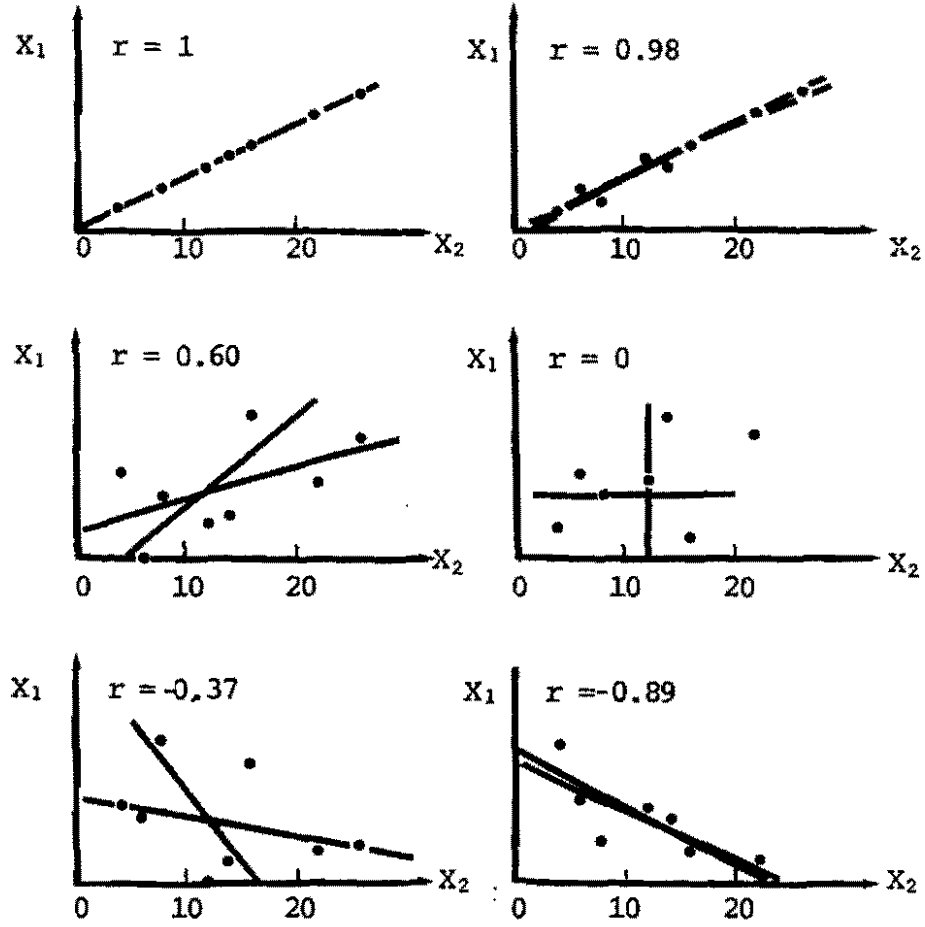
COEFICIENTE DE
CORRELACION MUESTRAL

:

$$r = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

$$-1 \leq r \leq 1$$

VALOR DEL COEFICIENTE
DE CORRELACION



SAS PARA EL ANALISIS DESCRIPTIVO
Y PRODUCCION DE GRAFICOS Y REPORTES

PROCEDIMIENTO	FUNCION
1. PROC PRINT	: IMPRIME UN ARCHIVO
2. PROC MEANS	}
3. PROC SUMMARY	
4. PROC TABULATE	
5. PROC UNIVARIATE	
6. PROC CORR	
7. PROC FREQ	: OBTIENEN ESTADISTICAS DESCRIPTIVAS
8. PROC PLOT	: PRODUCE GRAFICAS EN 2 DIMENSIONES
9. PROC CHART	: HISTOGRAMAS, DISTRIBUCIONES
10. PROC CALENDAR	: ORGANIZA DATOS MES POR MES EN FORMATO CALENDARIO
11. PROC FORMS	: PRODUCE LABELS TIPO CARTA

4.1 PROC MEANS

OBJETIVO: EL PROCEDIMIENTO MEANS PRODUCE ESTADISTICAS DESCRIPTIVAS UNIVARIADAS PARA VARIABLES NUMERICAS. ES EL PROCEDIMIENTO DESCRIPTIVO MAS FACIL DE USAR.

QUE ESTADISTICAS DESCRIPTIVAS PRODUCE?

MEAN	STD	STDERR
RANGE	MIN	MAX
SUM	KURTOSIS	SKEWNESS
VAR	USS	CSS
CV		

ESPECIFICACIONES: EL PROCEDIMIENTO MEANS SE CONTROLA POR LAS SIGUIENTES PROPOSICIONES:

- . PROC MEANS opciones ;
- . VAR variables;
- . ID variables;
- . OUTPUT OUT = nombre palabra=...;
 archivo clave

PROC MEANS

OPCIONES DE LA PROPOSICION "PROC MEANS"

- . T : ESTADISTICA T-STUDENT PARA H_0 : MEDIA=0
- . PRT : PROBABILIDAD DE SIGNIFICANCIA DE T
- . NOPRINT : ELIMINA IMPRESION DE RESULTADOS
- . MAXDEC=N : ADOPTA "N" DECIMALES PARA LOS VALORES

ESTADISTICAS DESCRIPTIVAS POSIBLES(*)

MEAN	STD	STDERR	N
RANGE	MIN	MAX	NMISS
SUM	KURTOSIS	SKEWNESS	
VAR	USS	CSS	
CV			

(*) Pueden ser usados como palabras claves en la proposición OUTPUT.

PROC MEANS

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE VALORES DE RENDIMIENTO (REND, VAINASXP, GRANOSXV Y P100SEM) PARA 2 VARIETADES SEMBRADAS BAJO 6 DENSIDADES EN 4 REPETICIONES.

DATA A ;

INPUT VARIEDAD 1 DENSIDAD 3

REP 5 REND 8-12

VAINASXP 15-16 GRANOSXV 19-21

P100SEM 24-27 ;

CARDS ;

2	1	1	299.9	22	4.3	21.9
2	1	3	246.7	13	4.6	24.9
3	1	1	292.3	26	4.0	28.0
.
3	6	4	204.0	8	3.7	28.7

ARCHIVO DE DATOS originales

URS	VARIEDAD	DENSIDAD	REP	REND	VAINASKp	GRANDSXV	PICOSEM
1	2	1	1	299.9	22	4.3	21.9
2	2	1	2	423.7	23	5.0	22.5
3	2	1	3	246.7	13	4.6	24.9
4	2	1	4	240.2	13	5.1	22.1
5	2	2	1	272.8	11	4.4	27.7
6	2	2	2	479.4	20	3.7	20.4
7	2	2	3	429.5	19	3.7	24.0
8	2	2	4	351.7	20	5.1	23.4
9	2	3	1	597.3	20	4.9	22.4
10	2	3	2	464.9	13	4.4	22.9
11	2	3	3	488.4	10	4.1	25.6
12	2	3	4	371.7	12	5.2	19.6
13	2	4	1	615.3	9	3.6	20.7
14	2	4	2	545.9	11	3.8	35.3
15	2	4	3	491.1	7	5.3	21.5
16	2	4	3	492.5	23	2.6	8.1
17	2	5	1	675.5	15	4.0	22.2
18	2	5	2	577.0	11	4.7	27.3
19	2	5	3	498.2	11	4.4	26.7
20	2	5	4	481.0	8	4.0	23.1
21	2	6	1	571.5	11	3.5	36.7
22	2	6	2	714.4	16	3.6	23.1
23	2	6	3	354.2	8	6.2	62.6
24	2	6	4	315.6	5	4.4	19.6
25	3	1	1	292.3	26	4.0	28.0
26	3	1	2	319.1	25	3.5	26.4
27	3	1	3	179.9	14	4.3	29.6
28	3	1	4	207.1	12	4.5	29.4
29	3	2	1	553.6	12	4.5	26.2
30	3	2	2	311.9	16	4.3	25.3
31	3	2	3	294.3	11	5.4	26.0
32	3	2	4	319.6	9	5.7	26.5
33	3	3	1	379.5	9	3.4	28.5
34	3	3	2	332.6	7	7.6	23.8
35	3	3	3	349.3	12	4.0	23.6
36	3	3	4	199.3	12	2.8	25.4
37	3	4	1	332.2	14	4.3	25.0
38	3	4	2	329.7	16	2.7	39.3
39	3	4	3	288.7	23	4.2	27.5
40	3	4	3	255.0	7	4.2	25.8
41	3	5	1	348.1	6	3.5	25.4
42	3	5	2	452.0	10	3.6	23.7
43	3	5	3	447.2	5	4.7	25.4
44	3	5	4	277.3	10	4.8	23.6
45	3	6	1	484.2	13	4.9	26.2
46	3	6	2	414.5	6	1.8	48.7
47	3	6	3	397.6	7	3.2	26.7
48	3	6	4	204.0	8	3.7	28.7

PROC MEANS

EJEMPLO NO. 1:

```
PROC MEANS N MEAN STD DATA=A ;  
VAR REAO VAINASXP GRANOSXV P100SEM;  
TITLE PROMEDIO Y DESVIACION ESTANDAR ;  
TITLE2 RENDIMIENTO Y SUS COMPONENTES ;
```

PROMEDIO Y DESVIACION ESTANDAR
 RENDIMIENTO Y SUS COMPONENTES

VARIABLE	N	MEAN	STANDARD DEVIATION
KEND	48	395.15408866	129.91828401
VAINASXP	48	12.93750000	5.57122642
GRANOSXV	48	4.25416631	0.95490730
FLUCSEM	48	26.43749446	7.84207428

PROC MEANS

EJEMPLO NO. 2:

```
PROC SORT DATA=A ;  
  BY VARIEDAD DENSIDAD REP ;  
PROC MEANS DATA=A MAXDEC=2 ;  
  BY VARIEDAD ;  
VAR REND VAINASXP GRANOSXV P100SEM ;  
TITLE ESTADISTICOS DESCRIPTIVOS POR VARIEDAD ;
```

ESTADISTICOS DESCRIPTIVOS POR VARIEDAD

VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE
----- VARIEDAD=2 -----					
REND	24	458.27	132.04	240.20	714.40
VAINASXP	24	13.79	5.35	5.00	23.00
GRANOSXV	24	4.36	0.76	2.60	6.20
PIOSSEM	24	25.18	9.60	8.10	62.60

----- VARIEDAD=3 -----					
REND	24	332.04	93.51	179.90	553.60
VAINASXP	24	12.08	5.77	5.00	26.00
GRANOSXV	24	4.15	1.14	1.80	7.60
PIOSSEM	24	27.70	5.49	23.60	48.70

(cont.)

variable	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
----- VARIEDAD=2 -----				
REND	26.95	10993.40	17434.39	28.81
VAINASXP	1.09	331.00	28.61	38.78
GRANOSXV	0.16	104.60	0.59	17.55
PIOSSEM	1.96	604.30	92.25	38.14

----- VARIEDAD=3 -----				
REND	19.09	7969.00	8744.25	28.16
VAINASXP	1.18	290.00	33.33	47.75
GRANOSXV	0.23	99.60	1.29	27.42
PIOSSEM	1.12	664.70	30.12	19.82

PROC MEANS

EJEMPLO NO. 3:

```
PROC MEANS NOPRINT MAXDEC=2 DATA=A;  
  BY VARIEDAD DENSIDAD ;  
  VAR  REND ;  
  OUTPUT OUT=B  
        MEAN=P._REND  
        MAX= MAXREND;  
PROC PRINT DATA=B ;  
  TITLE RENDIMIENTO PROMEDIO Y MAXIMO;
```

RENDIMIENTO promedio y maximo

UBS	VARIEDAD	DENSIDAD	P_RENO	MAXRENO
1	2	1	302.625	423.7
2	2	2	383.350	479.4
3	2	3	480.575	597.3
4	2	4	536.200	615.3
5	2	5	557.925	675.5
6	2	6	488.925	714.4
7	3	1	249.600	310.1
8	3	2	369.850	553.6
9	3	3	315.175	379.5
10	3	4	301.400	332.2
11	3	5	381.150	452.0
12	3	6	375.075	464.2

4.2 PROC SUMMARY

OBJETIVO: EL PROCEDIMIENTO SUMMARY CALCULA ESTADISTICAS DESCRIPTIVAS UNIVARIADAS PARA VARIABLES NUMERICAS Y FORMA UN NUEVO ARCHIVO SAS CON LOS RESULTADOS. CADA OBSERVACION EN EL NUEVO ARCHIVO CONTIENE LAS ESTADISTICAS DESCRIPTIVAS PARA SUBGRUPOS DE OBSERVACIONES DEL ARCHIVO ORIGINAL, REPRESENTADOS POR TODAS LAS POSIBLES COMBINACIONES DE LAS VARIABLES EN LA PROPOSICION CLASS.

ESPECIFICACIONES: EL PROCEDIMIENTO SUMMARY SE CONTROLA POR LAS SIGUIENTES PROPOSICIONES:

- . PROC SUMMARY opciones ;
- . CLASS variables ;
- . VAR variables ;
- . BY variables ;
- . FREQ variables ;
- . ID variables ;
- . OUTPUT OUT = nuevo archivo SAS
Palabra clave = lista de variables ;

PROC SUMMARY

OPCIONES DE LA PROPOSICION "PROC SUMMARY"

- . DATA = : NOMBRE DEL ARCHIVO SOBRE EL CUAL SE APLICA
EL PROCEDIMIENTO
- . MISSING: CONSIDERA UN SUBGRUPO ADICIONAL CON LOS
VALORES "MISSING".

- . ESTADISTICAS
DESCRIPTIVAS: IGUAL QUE EN PROC MEANS.

PROC SUMMARY

EJEMPLO NO. 1:

```
PROC SUMMARY DATA=A ;  
  CLASS VARIEDAD DENSIDAD ;  
  VAR REND ;  
  OUTPUT OUT=C  
      MEAN=P _REND ;  
PROC PRINT DATA=C ;  
  BY _TYPE_ ;  
  TITLE PROMEDIOS DE RENDIMIENTO ;
```

PROMEDIOS DE RENDIMIENTO

----- _TYPE_=0 -----

OBS	VARIEDAD	DENSIDAD	_FREQ_	P_REND
1	.	.	48	395.154

----- _TYPE_=1 -----

OBS	VARIEDAD	DENSIDAD	_FREQ_	P_REND
2	.	1	8	276.112
3	.	2	8	376.600
4	.	3	8	397.875
5	.	4	8	418.800
6	.	5	8	469.537
7	.	6	8	432.000

----- _TYPE_=2 -----

OBS	VARIEDAD	DENSIDAD	_FREQ_	P_REND
8	2	.	24	458.267
9	3	.	24	332.042

----- _TYPE_=3 -----

OBS	VARIEDAD	DENSIDAD	_FREQ_	P_REND
10	2	1	4	302.625
11	2	2	4	383.350
12	2	3	4	480.575
13	2	4	4	536.200
14	2	5	4	557.925
15	2	6	4	488.925
16	3	1	4	249.600
17	3	2	4	369.850
18	3	3	4	315.175
19	3	4	4	301.400
20	3	5	4	381.150
21	3	6	4	375.075

PROC SUMMARY

EJEMPLO NO. 2:

```
PROC SUMMARY DATA=A;  
  BY VARIEDAD ;  
  CLASS DENSIDAD ;  
  VAR  REND ;  
  OUTPUT OUT=D  
        MEAN=P_REND RANGE=RA_REND ;  
PROC PRINT DATA=D ;  
  BY VARIEDAD ;  
  TITLE PROMEDIO Y RANGO DE RENDIMIENTO POR VARIEDAD;
```

PROMEDIO Y RANGO DE RENDIMIENTO POR VARIEDAD

----- VARIEDAD=2 -----

OBS	DENSIDAD	_TYPE_	_FREQ_	P_REND	RA_REND
1	.	0	24	458.267	474.2
2	1	1	4	302.625	183.5
3	2	1	4	383.350	206.6
4	3	1	4	480.575	225.6
5	4	1	4	536.200	124.2
6	5	1	4	557.925	194.5
7	6	1	4	488.925	398.8

----- VARIEDAD=3 -----

OBS	DENSIDAD	_TYPE_	_FREQ_	P_REND	RA_REND
8	.	0	24	332.042	373.7
9	1	1	4	249.600	139.2
10	2	1	4	369.850	259.3
11	3	1	4	315.175	180.2
12	4	1	4	301.400	77.2
13	5	1	4	381.150	174.7
14	6	1	4	375.075	280.2

4.3 PROC FREQ

OBJETIVO: EL PROCEDIMIENTO FREQ PRODUCE TABLAS DE FRECUENCIA DE UNA, DOS O N CLASIFICACIONES

ESPECIFICACIONES: EL PROCEDIMIENTO FREQ SE ESPECIFICA CON LAS SIGUIENTES PROPOSICIONES:

- . PROC FREQ opciones;
- . TABLES tablas requeridas/opciones;
- . WEIGHT variables;
- . BY variables;

PROC FREQ
OPCIONES DE LA PROPOSICION "PROC FREQ"

NOFREQ : SUPRIME LA IMPRESION DE FRECUENCIAS POR CELDA
NOROW : SUPRIME % RESPECTO AL TOTAL FILA
NOCOL : SUPRIME % RESPECTO AL TOTAL COLUMNA
NOPERCENT : SUPRIME % RESPECTO DEL TOTAL
OUT=nombre : CREA UN ARCHIVO SAS CON RESULTADOS

ADICIONALMENTE:

CHISQ	EXPECTED	DEVIATION
CELLCHI2	ALL	SPARSE
LIST	MISSING	

PROC FREQ

EJEMPLO NO. 1:

```
DATA A1;
  SET A ;
  IF 0<REND<= 200 THEN GRUPO=1 ; ELSE
  IF 200<REND<= 400 THEN GRUPO=2 ; ELSE
  IF 400<REND<= 600 THEN GRUPO=3 ; ELSE
  IF REND> 600 THEN GRUPO=4 ;
PROC FREQ DATA=A1;
  TABLES GRUPO;
  TABLES VARIEDAD*GRUPO/NOPERCENT;
  TABLES VARIEDAD*DENSIDAD*GRUPO/NOROW NOCOL NOPERCENT;
  TITLE DISTRIBUCION DE FRECUENCIAS PARA RENDIMIENTO;
```

DISTRIBUCION DE FRECUENCIAS PARA RENDIMIENTO

GRUPO	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	2	2	4.167	4.167
2	25	27	52.083	56.250
3	18	45	37.500	93.750
4	3	48	6.250	100.000

DISTRIBUCION DE FRECUENCIAS PARA RENDIMIENTO

TABLE OF VARIEDAD BY GRUPO

VARIEDAD	GRUPO				TOTAL
FREQUENCY	1	2	3	4	
ROW PCT					
COL PCT					
2	0	8	13	3	24
	0.00	33.33	54.17	12.50	
	0.00	32.00	72.22	100.00	
3	2	17	5	0	24
	8.33	70.83	20.83	0.00	
	100.00	68.00	27.78	0.00	
TOTAL	2	25	18	3	48

DISTRIBUCION DE FRECUENCIAS PARA RENDIMIENTO

TABLE OF DENSIDAD BY GRUPO
CONTROLLING FOR VARIEDAD= 2

DENSIDAD	GRUPO				TOTAL
FREQUENCY	1	2	3	4	
1	0	3	1	0	4
2	0	2	2	0	4
3	0	1	3	0	4
4	0	0	3	1	4
5	0	0	3	1	4
6	0	2	1	1	4
TOTAL	0	8	13	3	24

DISTRIBUCION DE FRECUENCIAS PARA RENDIMIENTO

TABLE OF DENSIDAD BY GRUPO
CONTROLLING FOR VARIEDAD= 3

DENSIDAD	GRUPO				TOTAL
FREQUENCY	1	2	3	4	
1	1	3	0	0	4
2	0	3	1	0	4
3	1	3	0	0	4
4	0	4	0	0	4
5	0	2	2	0	4
6	0	2	2	0	4
TOTAL	2	17	5	0	24

PROC FREQ

EJEMPLO NO. 2:

```
PROC FREQ DATA=A1;  
TABLES VARIEDAD*GRUPO/  
        CHISQ CELLCH2 DEVIATION EXPECTED;
```

DISTRIBUCION DE FRECUENCIAS

TABLE OF VARIEDAD BY GRUPO

VARIEDAD	GRUPO				TOTAL
FREQUENLY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT	1	2	3	4	
2	0	8	13	3	24
	1.0	12.5	9.0	1.5	
	-1.0	-4.5	4.0	1.5	
	1.0	1.6	1.8	1.5	
	0.00	16.67	27.09	6.25	50.00
	0.00	33.33	54.17	12.50	
	0.00	33.00	72.22	100.00	
3	2	17	5	0	24
	1.0	12.5	9.0	1.5	
	1.0	4.5	-4.0	-1.5	
	1.0	1.6	1.8	1.5	
	4.17	35.42	10.42	0.00	50.00
	8.33	70.83	20.83	0.00	
	100.00	68.00	27.78	0.00	
TOTAL	2	25	18	3	48
	4.17	52.08	37.50	6.25	100.00

STATISTICS FOR 2-WAY TABLES

WARNING: OVER 20% OF THE CELLS HAVE EXPECTED COUNTS LESS THAN 5.
TABLE IS SO SPARSE THAT CHI-SQUARE MAY NOT BE A VALID TEST.

CHI-SQUARE 11.796 DF= 3 PROB=0.0081
 PHI 0.496
 CONTINGENCY COEFFICIENT 0.444
 CRAMER'S V 0.496
 LIKELIHOOD RATIO CHISQUARE 13.928 DF= 3 PROB=0.0030

4.4 PROC CORR

OBJETIVO: CALCULA 3 TIPOS DE COEFICIENTES DE CORRELACION ENTRE VARIABLES NUMERICAS.

- PEARSON'S PRODUCT-MOMENT (PARAMETRICO)
- SPEARMAN'S RANK ORDER (NO PARAMETRICO)
- KENDALL'S TAU-B (NO PARAMETRICO)

ESPECIFICACIONES: EL PROCEDIMIENTO CORR SE ESPECIFICA CON LAS SIGUIENTES PROPOSICIONES:

- . PROC CORR opciones;
- . VAR variables;
- . WITH variables;
- . WEIGHT variables;
- . FREQ variables;
- . BY variables;

PROC CORR
OPCIONES DE LA PROPOSICION "PROC CORR"

BEST=N : IMPRIME SOLO LAS MEJORES N CORRELACIONES
NOSIMPLE : SUPRIME ESTADISTICAS DESCRIPTIVAS
NOMISS : ELIMINA OBSERVACIONES CON ALGUN DATO "MISSING"
NOPRINT : SUPRIME IMPRESION DE RESULTADOS
NOCORR : SUPRIME IMPRESION DE CORRELACIONES
NOPROB : SUPRIME IMPRESION DE PROBABILIDADES DE SIGNIFICANCIA
PARA LAS CORRELACIONES
RANK : IMPRIME CORRELACIONES EN ORDEN DESCENDENTE

- ADICIONALMENTE:

SPEARMAN
OUTS=nombre
COV

KENDALL
OUTK=nombre
SSP

PEARSON
OUTP=nombre

PROC CORR

EJEMPLO NO.1:

```
PROC CORR DATA=A;  
VAR REND VAINASXP GRANOSXV P100SEM;  
TITLE CORRELACION DE RENDIMIENTO Y SUS COMPONENTES;
```

CORRELACION DE RENDIMIENTO Y SUS COMPONENTES

VARIABLE	N	MEAN	STD DEV	SUM	MINIMUM	MAXIMUM
REND	48	395.15408866	129.91823401	3967.39625549	179.89999390	714.39990234
VAINASXP	48	12.93750000	5.57122642	621.00000000	5.00000000	26.00000000
GRANOSXV	48	4.25416631	0.96490730	204.19998264	1.79999924	7.59999943
P100SEM	48	26.43749446	7.84207428	1268.99973392	8.09999943	62.59999084

-92-

CORRELATION COEFFICIENTS / PROB > |R| UNDER H0:RHO=0 / N = 48

	REND	VAINASXP	GRANOSXV	P100SEM
REND	1.00000 0.0000	0.01364 0.9267	-0.12965 0.3798	-0.11489 0.4368
VAINASXP	0.01364 0.9267	1.00000 0.0000	-0.13036 0.3772	-0.26048 0.0738
GRANOSXV	-0.12965 0.3798	-0.13036 0.3772	1.00000 0.0000	-0.02150 0.8847
P100SEM	-0.11489 0.4368	-0.26048 0.0738	-0.02150 0.8847	1.00000 0.0000

PROC CORR

EJEMPLO NO. 2:

```
PROC CORR NOSIMPLE COV NOPROB DATA=A;  
  BY VARIEDAD ;  
  VAR READ P100SEM ;  
  WITH VAINASXP GRANOSKV ;  
  FORMAT VARIEDAD IVARI ;  
  TITLE1 CORRELACION DE RENDIMIENTO Y PESO DE 100 SEMILLAS ;  
  TITLE2 VS VAINAS POR PLANIA Y GRANOS POR VAINA ;
```

CORRELACION DE RENDIMIENTO Y PESO DE 100 SEMILLAS
VS VAINAS POR PLANTA Y GRANOS POR VAINA
VARIEDAD=V7982

COVARIANCE MATRIX

REND P100SEM

VAINASXP 15.1189 -19.7871

GRANOSXV -45.3336 3.46562

CORRELATION COEFFICIENTS / N = 24

REND P100SEM

VAINASXP 0.02141 -0.38519

GRANOSXV -0.44883 0.47171

CORRELACION DE RENDIMIENTO Y PESO DE 100 SEMILLAS
VS VAINAS POR PLANTA Y GRANOS POR VAINA
VARIEDAD=V7987

COVARIANCE MATRIX

REND P100SEM

VAINASXP -107.447 -1.22572

GRANOSXV -1.59786 -3.52457

CORRELATION COEFFICIENTS / N = 24

REND P100SEM

VAINASXP -0.19913 -0.03871

GRANOSXV -0.01502 -0.56441

4.5 PROC PLOT

OBJETIVO: EL PROCEDIMIENTO PLOT PRODUCE GRAFICAS DE UNA VARIABLE VS. OTRA EN DOS EJES DE COORDENADAS CARTESIANAS. CADA PUNTO DE LA GRAFICA CORRESPONDE A UNA PAREJA DE VALORES

ESPECIFICACIONES: EL PROCEDIMIENTO PLOT ES CONTROLADO POR LAS SIGUIENTES PROPOSICIONES:

- . PROC PLOT opciones;
- . BY variables;
- . PLOT gráficas requeridas/opciones;

PROC PLOT
OPCIONES DE LA PROPOSICION "PROC PLOT"

UNIFORM : CONSERVA LA MISMA ESCALA EN LOS EJES PARA
TODAS LAS GRAFICAS REQUERIDAS AL UTILIZAR
LA PROPOSICION BY

NOLEGEND : SUPRIME LA IMPRESION DE LA LEYENDA EN LA
PARTE SUPERIOR DE CADA GRAFICA.

PROC PLOT
OPCIONES DE LA PROPOSICION "PLOT"

VAXIS=	}	CONTROLA LAS ESCALAS DE LOS EJES VERTICAL Y
HAXIS=		HORIZONTAL, RESPECTIVAMENTE
VZERO	}	INCLUYE EN LA ESCALA DE LOS EJES VERTICAL Y
HZERO		HORIZONTAL, RESPECTIVAMENTE
OVERLAY	:	SUPERPONE LAS GRAFICAS REQUERIDAS

ADICIONALMENTE:

HREF=	VREF=	VREVERSE
HREFCHAR=	VREFCHAR=	CONTOUR=
HPOS=	VPOS=	VSPACE=
		HSPACE=

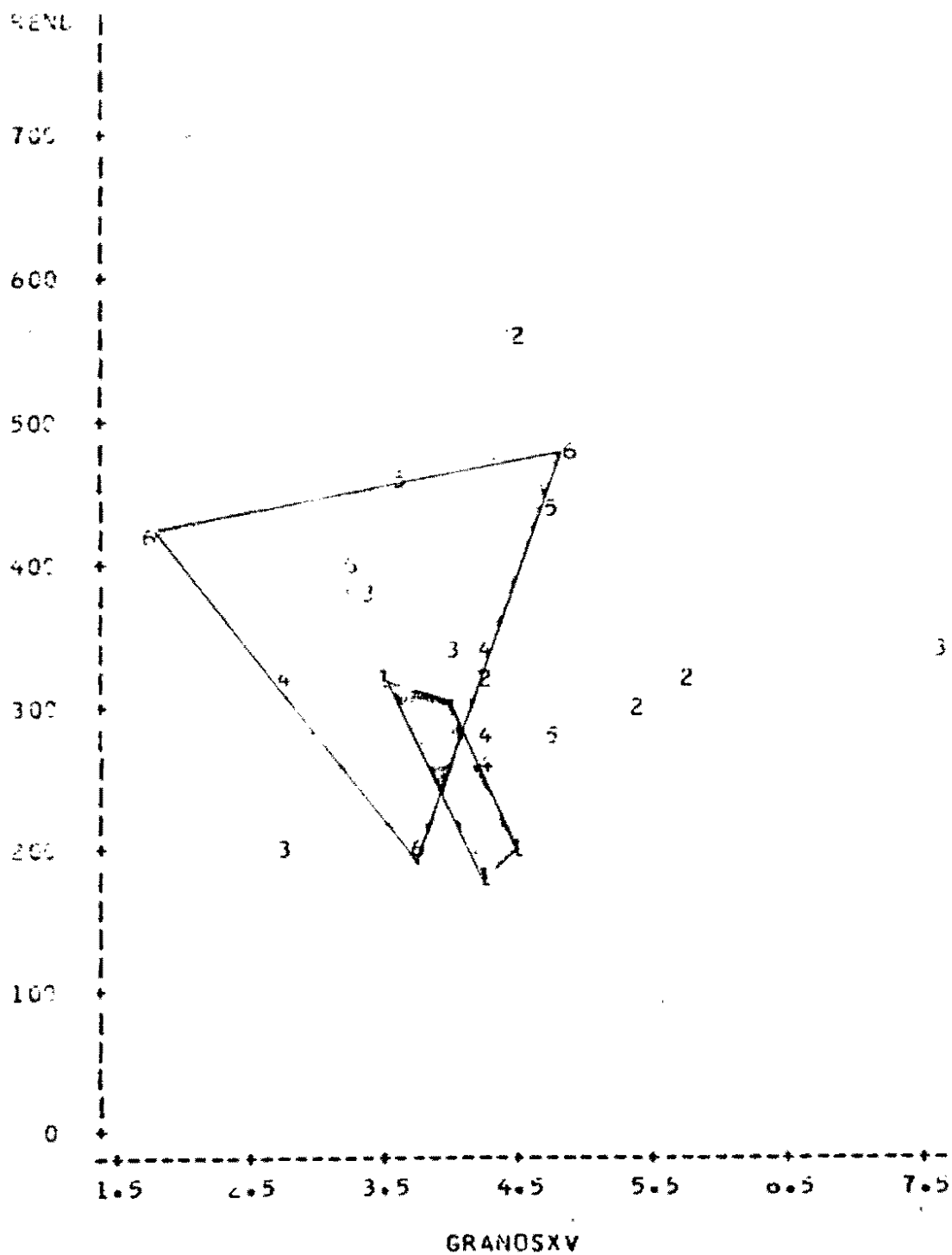
PROC PLOT

EJEMPLO NO. 1:

```
PROC PLOT UNIFORM DATA=A;  
BY VARIEDAD ;  
PLOT REND*GRANOSXV=DENSIDAD / HPOS=50 VIOS=40;  
TITLE RENDIMIENTO VS GRANOS POR VAJINA ;
```

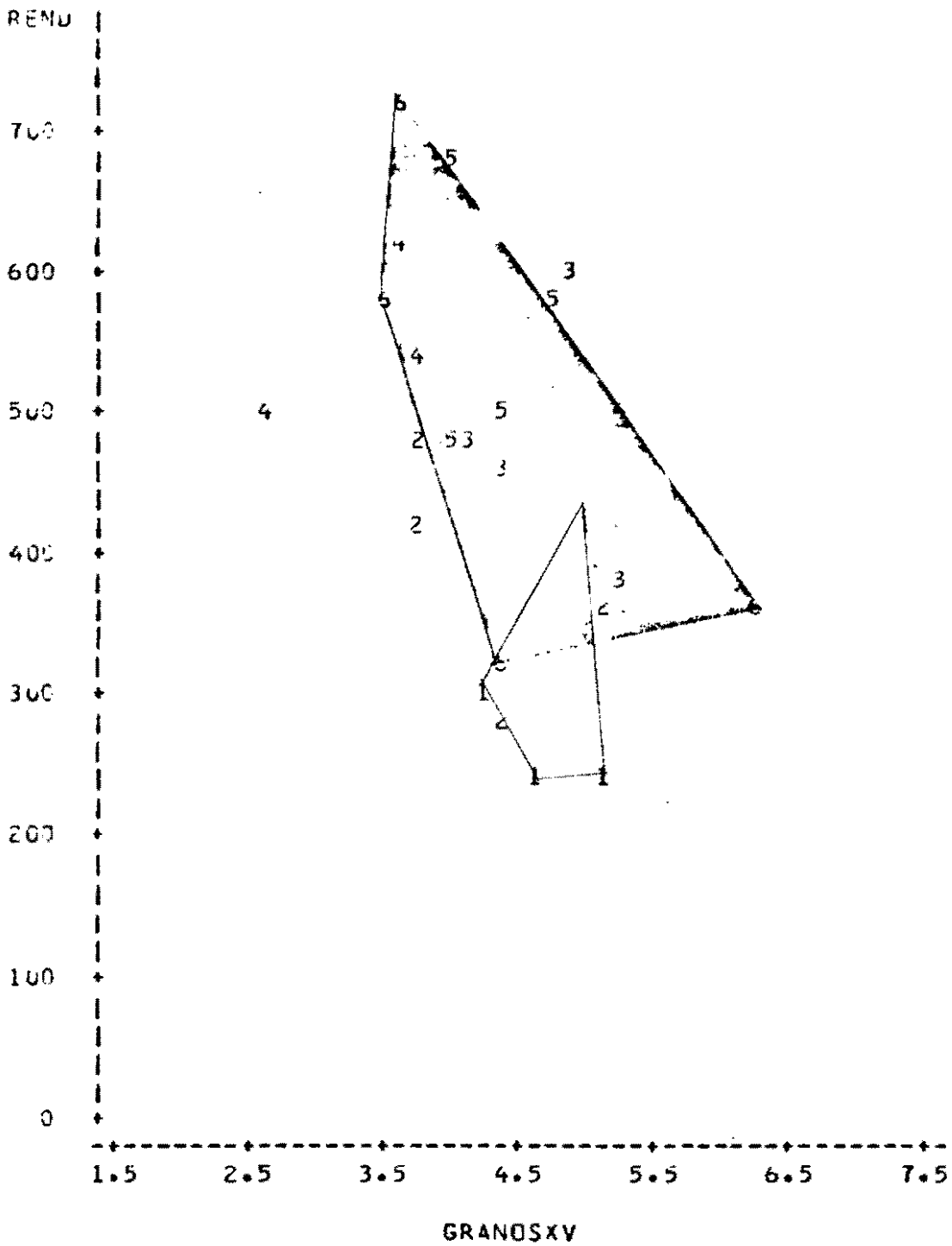
RENDIMIENTO VS GRANOS POR VAINA
 VARIEDAD=J

PLOT OF REND GRANOSXV SYMBOL IS VALUE OF DENSIDAD



RENDIMIENTO VS GRANOS POR VAINA
 VARIEDAD=2

PLOT OF REND-GRANOSXV SYMBOL IS VALUE OF DENSIDAD



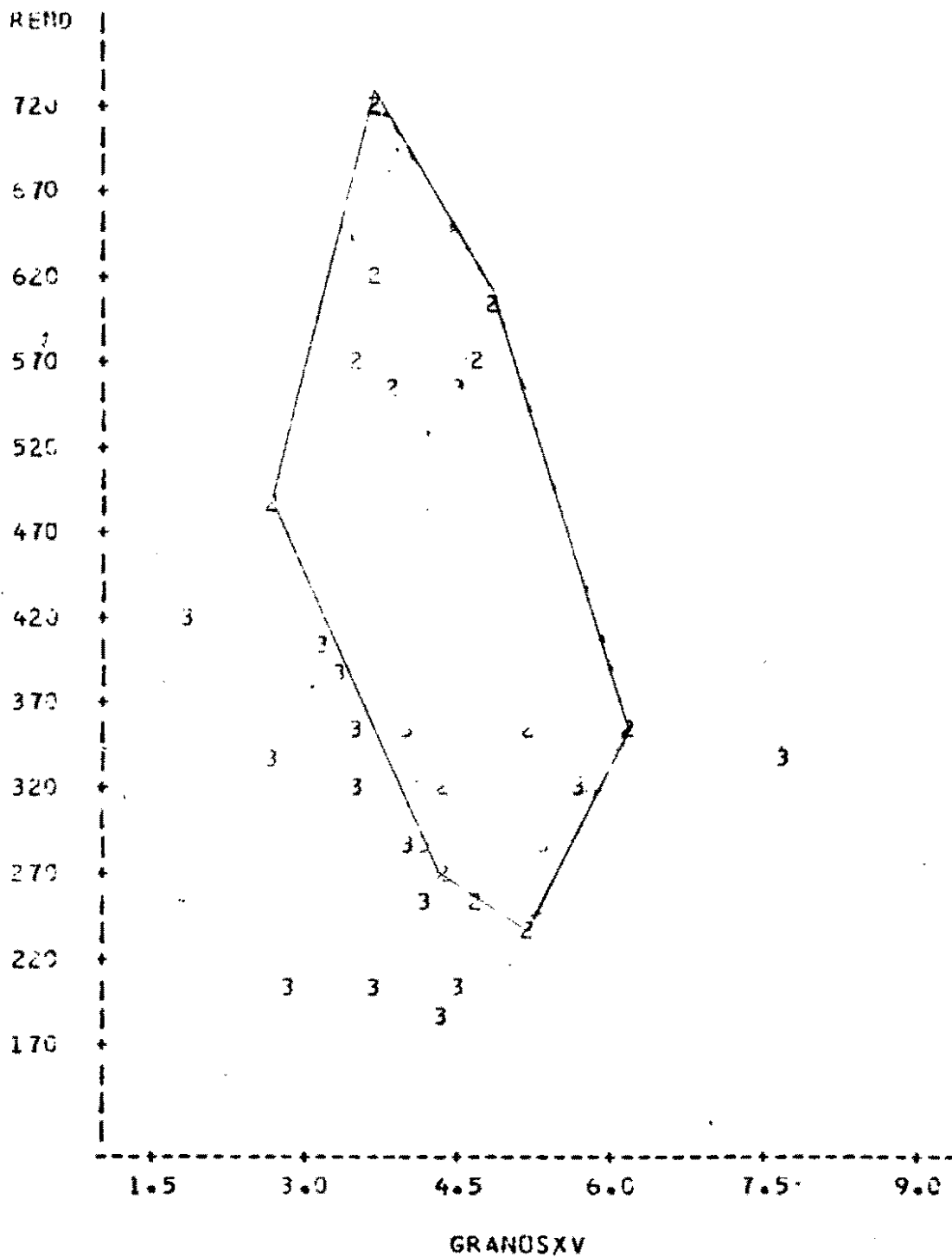
PROC PLOT

EJEMPLO NO. 2:

```
PROC PLOT DATA=A ;  
  PLOT REND*GRANOSXV=VARIEDAD/ HPOS=50 VPOS=40  
      VAXIS=170 TO 720 BY 50  
      HAXIS=1.5 3.0 4.5 6.0 7.5 9.0 ;  
  TITLE RENDIMIENTO VS GRANOS POR VAINA ;
```

RENDIMIENTO VS GRANOS POR VAINA

PLOT OF REND GRANOSXV SYMBOL IS VALUE OF VARIEDAD



NOTE: 1 OBS HIDDEN

4.7 PROC CHART

OBJETIVO: EL PROCEDIMIENTO CHART PRODUCE:

- . HISTROGRAMAS : DIAGRAMAS DE BARRAS HORIZONTALES Y VERTICALES
- . DIAGRAMAS DE BLOQUES
- . DIAGRAMAS CIRCULARES ("PIE DIAGRAMS")
- . DIAGRAMAS EN FORMA DE ESTRELLA ,

PARA VARIABLES NUMERICAS O DE CARACTERES, ILUSTRANDO

- . SU FRECUENCIA/FRECUENCIA ACUMULADA
- . SU PORCENTAJE/PORCENTAJE ACUMULADO
- . SUS TOTALES
- . SUS PROMEDIOS

PROC CHART - ESPECIFICACIONES

ESTE PROCEDIMIENTO SE CONTROLA CON LAS PROPOSICIONES

- . PROC CHART opciones ;
- . BY variables;
- . VBAR variables/opciones;
- . HBAR variables/opciones;
- . HBLOCK variables/opciones;
- . PIE variables/opciones;
- . STAR variables/opciones;

PROC CHART - OPCIONES

- PARA PROPOSICIONES VBAR, HBAR, BLOCK, PIE, STAR:

SUMVAR= VARIABLE	ESPECIFICA LA VARIABLE DEPENDIENTE	
MIDPOINTS=VALORES	ESPECIFICA MARCAS DE CLASE	
FREQ=VARIABLE	ESPECIFICA UNA VARIABLE DE EXPANSION	
AXIS=VALOR	ESPECIFICA EL MAXIMO EN LA ESCALA PARA FREQ, PCT, CFREQ, CPCT, SUM, MEAN.	
TYPE=FREQ	TYPE=CFREQ	TYPE=SUM
TYPE=PCT	TYPE=CPCT	TYPE=MEAN
DISCRETE	MISSING	

PROC CHART - OPCIONES

- PARA PROPOSICIONES	VBAR, HBAR, BLOCK
GROUP=variable	PRODUCE GRAFICAS DESAGREGANDO POR LAS CATEGORIAS DE ESTA VARIABLE
SUBGROUP=variable	SUBDIVIDE CADA BARRA EN SIMBOLOS QUE REFLEJAN LA CONTRIBUCION DE LAS CATEGORIAS DE ESTA VARIABLE.
LEVELS=N	SYMBOL='sim' NOSYMBOL
NOZEROS	G100 REF=valor
ASCENDING	DESCENDING

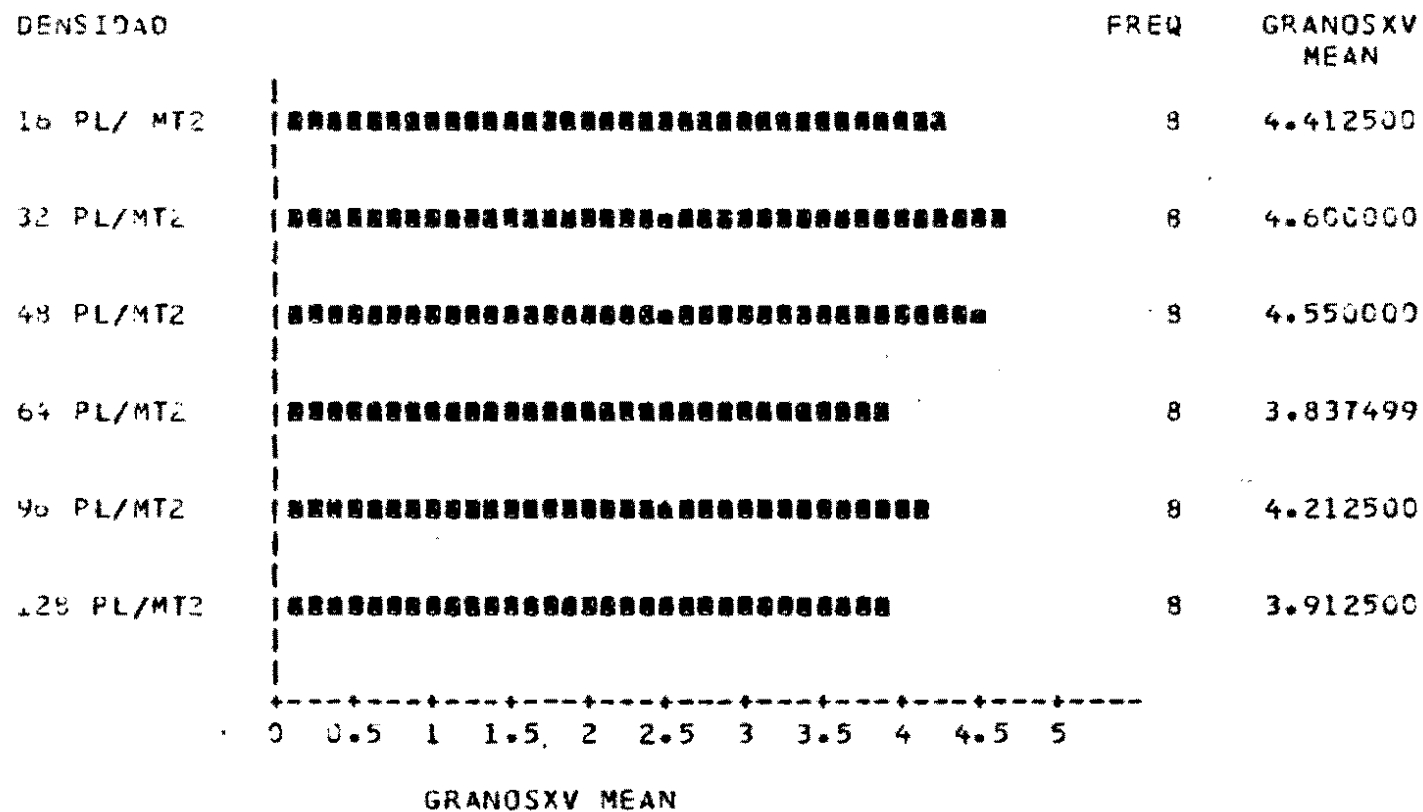
PROC CHART

EJEMPLO NO. 1:

```
PROC FORMAT ;
  VALUE FVARI 2='V7982'
              3='V7987' ;
  VALUE FDENS 1='16 PL/MT2'
              2='32 PL/MT2'
              3='48 PL/MT2'
              4='64 PL/MT2'
              5='96 PL/MT2'
              6='128 PL/MT2' ;
PROC CHART DATA=A ;
  HBAR DENSIDAD/DISCRETE SUMVAR=GRANOSXV TYPE=MEAN
  SYMBOL='XOA' ;
  FORMAT DENSIDAD FDENS. ;
  TITLE DIAGRAMA DE BARRAS HORIZONTALES;
```

DIAGRAMA DE BARRAS HORIZONTALES

BAR CHART OF MEANS



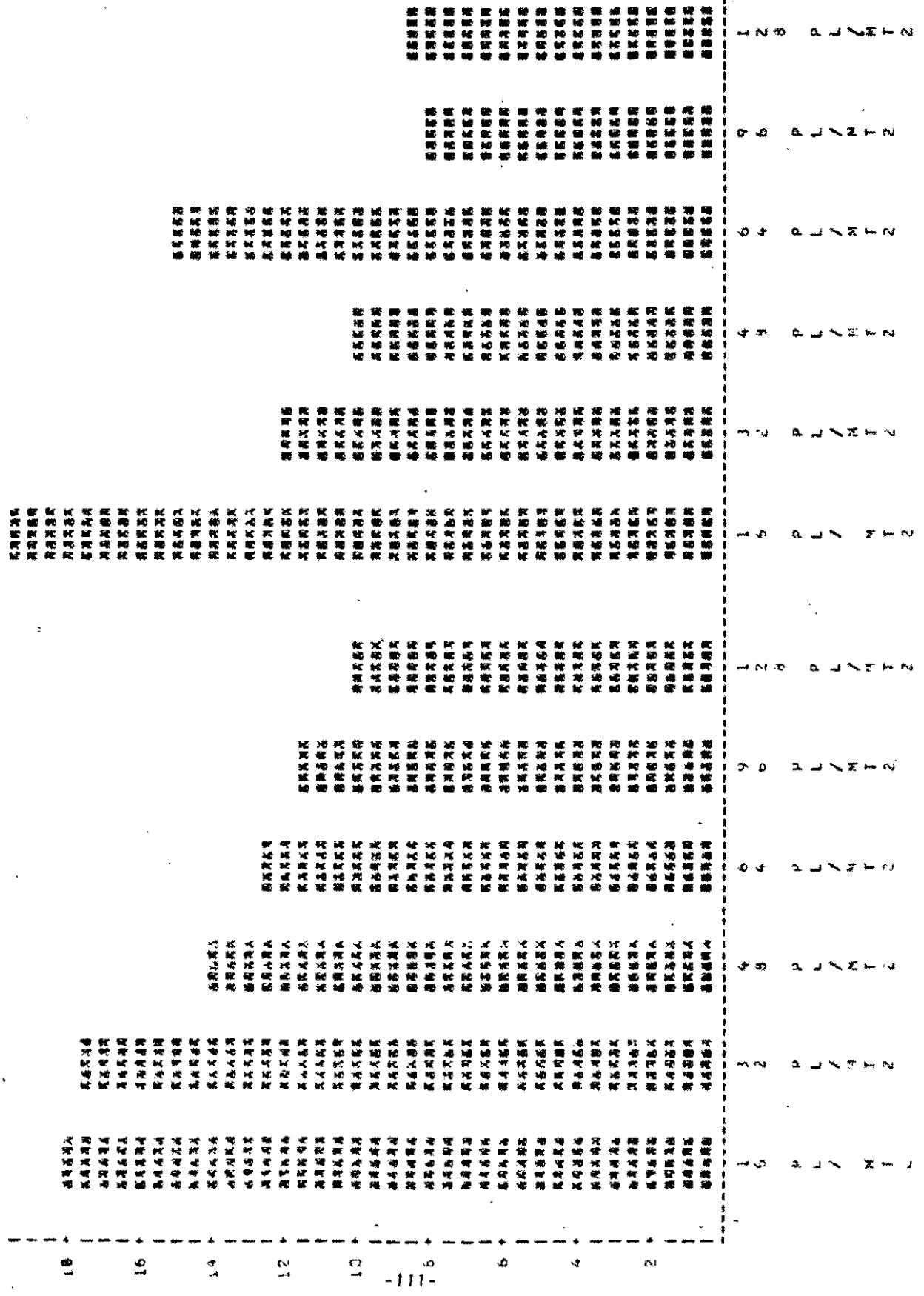
PROC CHART

EJEMPLO NO. 2:

```
PROC CHART DATA=A;  
  VBAR DENSIDAD / DISCRETE GROUP=VARIEDAD  
  SUMVAR=VAINASXP SYMBOL='XOA' TYPE=MIN;  
  FORMAT DENSIDAD FDENS. VARIEDAD FVARI. ;  
  TITLE DIAGRAMA DE BARRAS VERTICALES ;
```

BAR CHART OF MEANS

VAINAS - MPAN

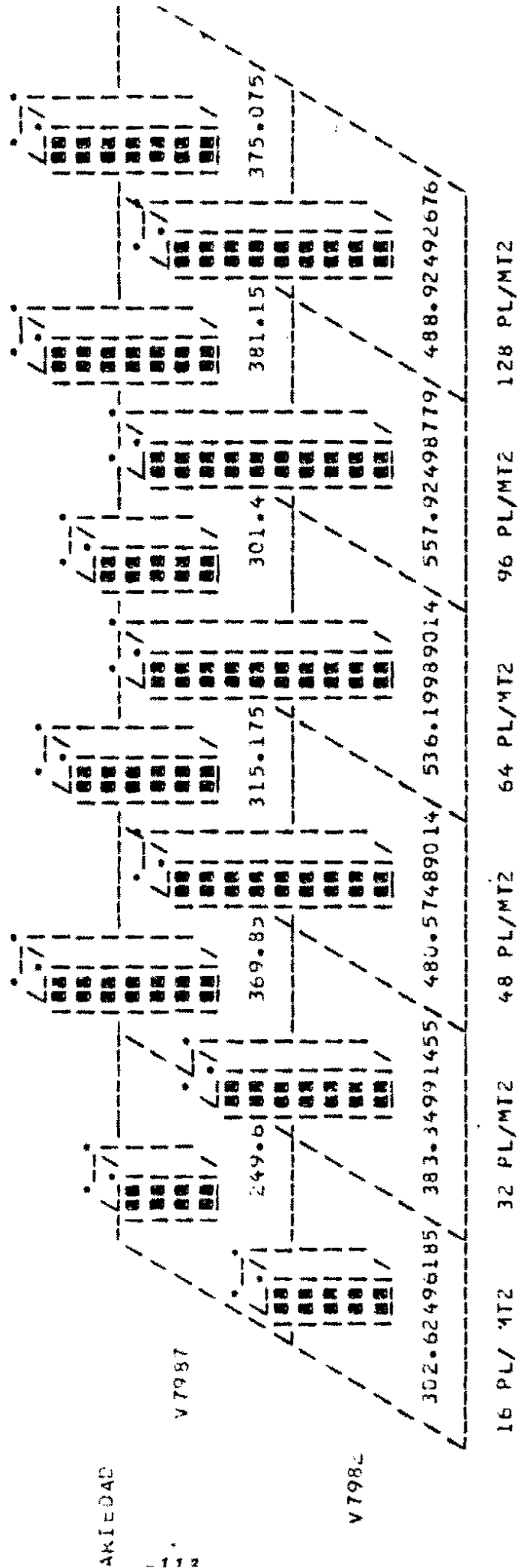


PROC CHART

EJEMPLO NO. 3:

```
PROC CHART DATA=A;  
BLOCK DENSIDAD / DISCRETE GROUP=VARIEDAD  
SUMVAR=REND SYMBOL='XCA' TYPE=MEAN ;  
FORMAT DENSIDAD FDENS. VARIEDAD FVARI. ;  
TITLE DIAGRAMA DE BLOQUES ;
```

DIAGRAMA DE BLOQUES
 BLOCK CHART OF REND MEANS



AKIEDAD

V7987

V7982

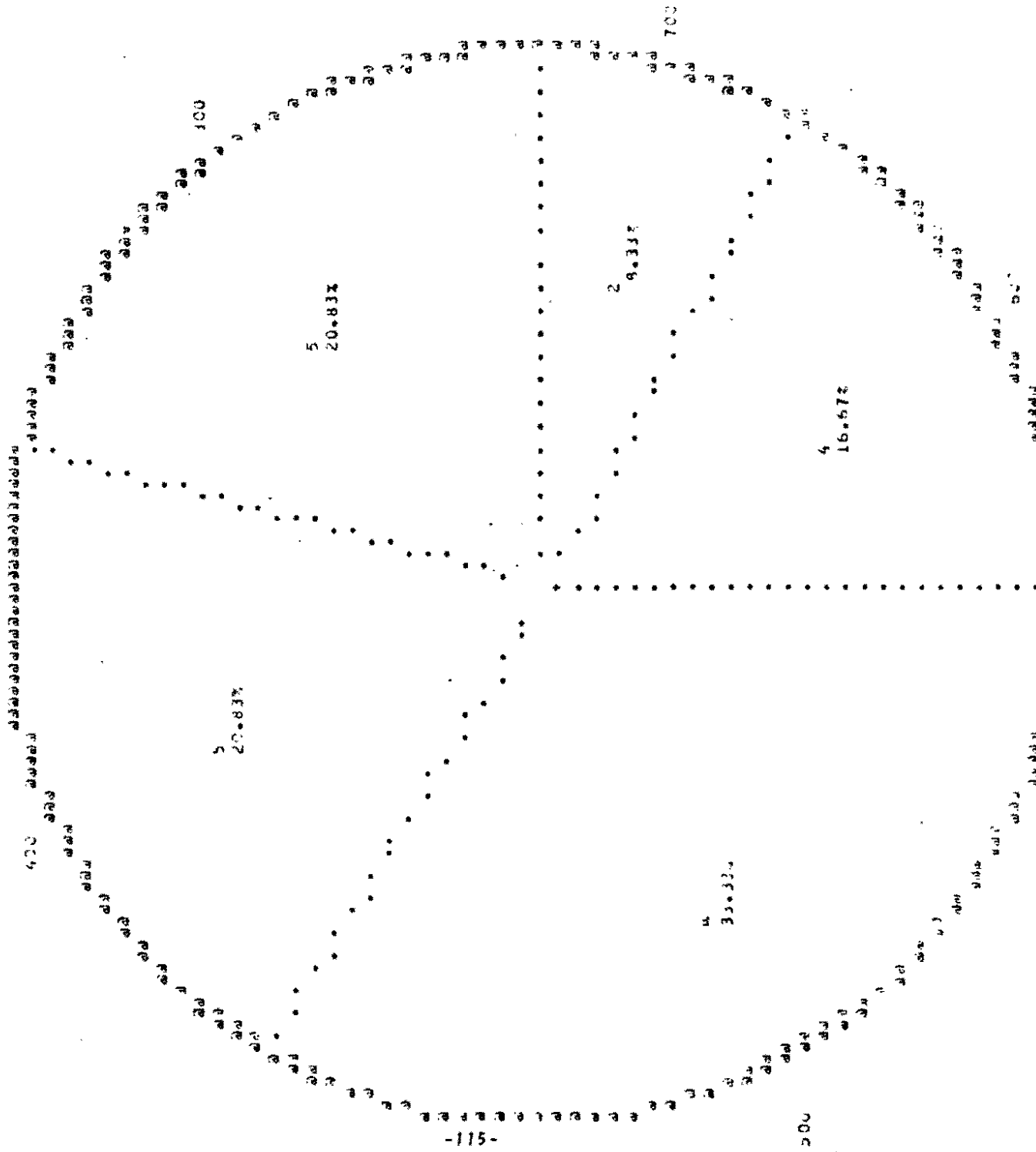
DENSIDAD

PROC CHART

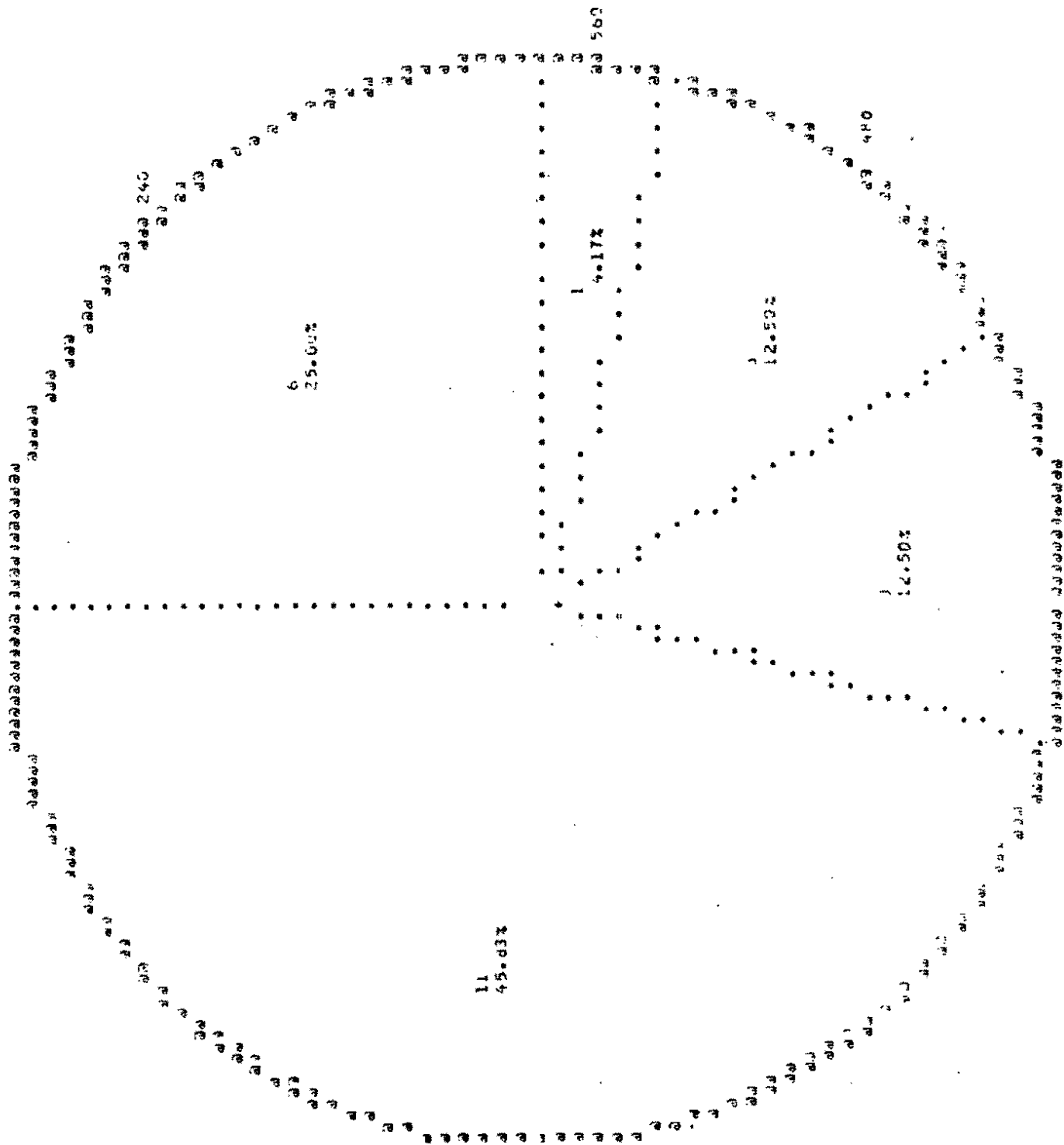
EJEMPLO NO. 4:

```
PROC SORT DATA=A ;  
  BY VARIEDAD ;  
PROC CHART DATA=A ;  
  BY VARIEDAD ;  
  PIE REFD ;  
  FORMAT VARIEDAD FVARI. ;  
  TITLE DIAGRAMA DE CIRCULO ;
```


FIG. 1 PIE CHART OF REND



FREQ PIE CHART OF RESP



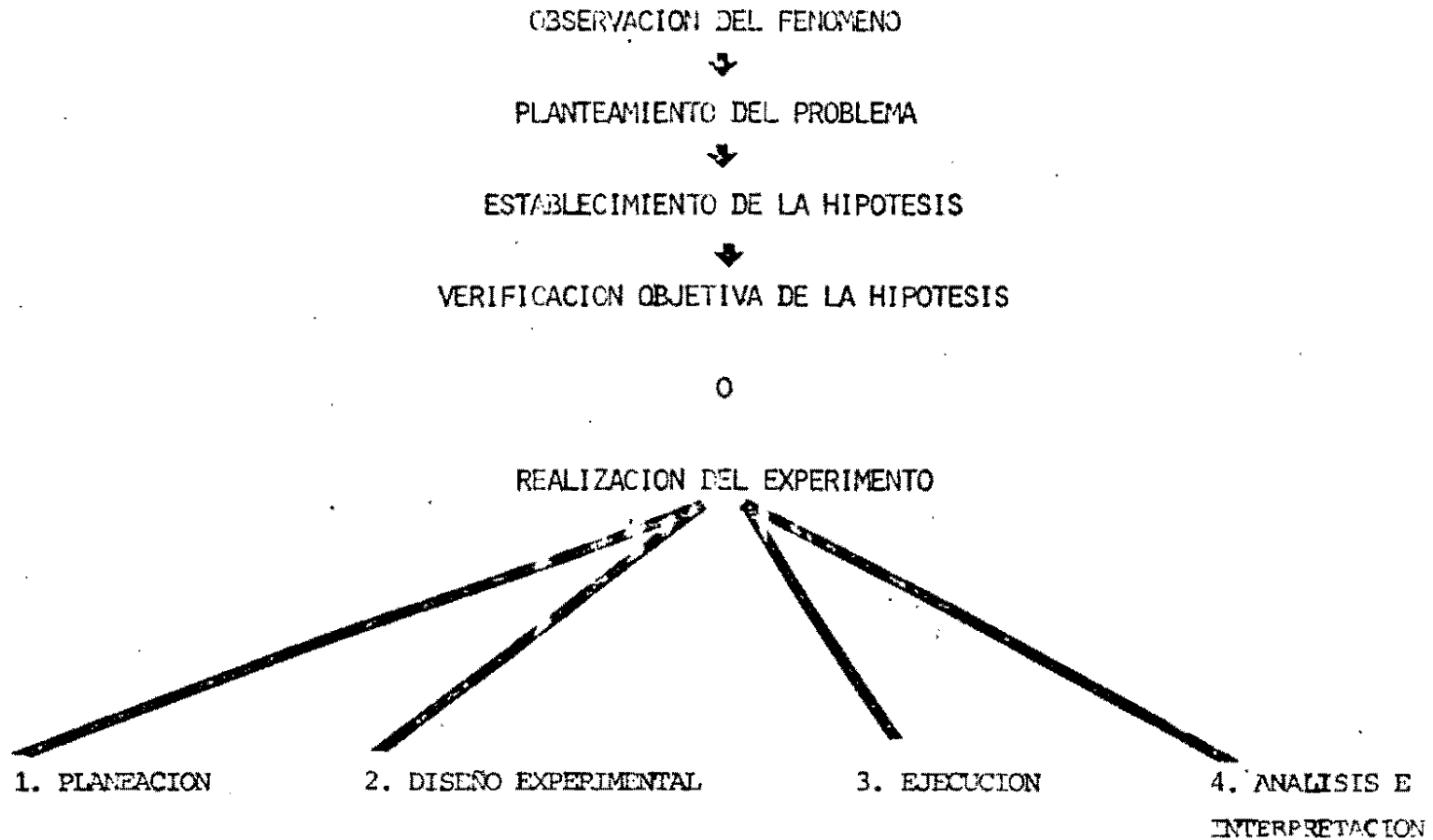
5. SAS PARA
ANALISIS ESTADISTICO
(INFERENCIAL)

"EL PROPOSITO DE LA CIENCIA ESTADISTICA ES PROPORCIONAR
UNA BASE OBJETIVA PARA EL ANALISIS DE PROBLEMAS EN LOS
CUALES LOS DATOS SE APARTAN DE LAS LEYES EXACTAS DE LA
CAUSALIDAD".

EL METODO CIENTIFICO

ES LA APLICACION DE LA LOGICA Y LA
OBJETIVIDAD AL MEJOR ENTENDIMIENTO
DE UN FENOMENO.

ETAPAS DEL METODO CIENTIFICO



EN ESTE CURSO REVISAREMOS:

1. ANALISIS DE VARIANZA PARA
DISEÑOS BALANCEADOS
2. REGRESION:
BAJO MODELOS LINEALES O
LINEALIZABLES

ANALISIS DE
VARIANZA
PARA DISEÑOS BALANCEADOS

- CONCEPTOS -
EL DISEÑO DE EXPERIMENTOS

EL PROCEDIMIENTO GENERAL EN LA INVESTIGACION CIENTIFICA ES FORMULAR HIPOTESIS Y LUEGO VERIFICARLAS DIRECTAMENTE O POR SUS CONSECUENCIAS. ÉSTA VERIFICACION NECESITA DE LA RECOLECCION DE OBSERVACIONES.

EL DISEÑO DE EXPERIMENTOS ES ESENCIALMENTE EL CONJUNTO DE REGLAS PARA PLANEAR EXPERIMENTOS Y RECOLECTAR OBSERVACIONES QUE PERMITEN OBTENER LA MAXIMA INFORMACION POSIBLE CON LOS RECURSOS DISPONIBLES.

EL DISEÑO DE EXPERIMENTOS

INVOLUCRA:

1. DEFINICION DE LAS VARIABLES DE RESPUESTA
(variables dependientes)
2. SELECCION DE TRATAMIENTOS
(variables independientes, controladas)
 - . ESCOGENCIA DE LOS FACTORES Y SUS NIVELES
 - . ESPECIFICACION SOBRE EL TIPO DE FACTORES:
FIJOS: sus niveles son los de interés al investigador
ALEATORIOS: sus niveles son una muestra representativa
de cierta población de niveles
3. CONTROL DEL ERROR
(mediante selección cuidadosa de unidades experimentales y
eliminación máxima de perturbaciones externas)
4. ESPECIFICACION SOBRE EL GRADO DE GENERALIZACION DE LOS RESULTADOS

REQUERIMIENTOS PARA UN
DISEÑO EXPERIMENTAL VALIDO

1. ALEATORIZACION EN LA APLICACION DE TRATAMIENTOS
2. USO DE REPETICIONES
3. MAXIMO CONTROL DEL ERROR EXPERIMENTAL

DISEÑO BALANCEADO U ORTOGONAL:

CADA TRATAMIENTO SE ASIGNA A IGUAL NUMERO DE UNIDADES EXPERIMENTALES.

LA ESTIMACION DE LAS SUMAS DE CUADRADOS POR LAS FORMULAS CONVENCIONALES, ES VALIDA.

DISEÑO DESBALANCEADO O NO ORTOGONAL:

NO TODOS LOS TRATAMIENTOS SE ASIGNAN A IGUAL NUMERO DE UNIDADES EXPERIMENTALES.

CON EXCEPCION DEL DISEÑO "COMPLETAMENTE AL AZAR", LA ESTIMACION DE LAS SUMAS DE CUADRADOS POR LAS FORMULAS CONVENCIONALES ES INVALIDA.

EL ANALISIS DE VARIANZA

UN OBJETIVO PRIMORDIAL DEL ANALISIS DE VARIANZA ES DECIDIR SI LA VARIACION OBSERVADA EN UN CONJUNTO DE RESULTADOS SE DEBE AL EFECTO DEL TRATAMIENTO O ES PURAMENTE ATRIBUIBLE AL AZAR, Y ENTONCES PODER INFERIR SOBRE LAS MEDIAS DE TRATAMIENTOS.

EL ANALISIS DE VARIANZA

- SE USA PARA HACER INFERENCIA ESTADISTICA SOBRE MAS DE 2 MEDIAS.
- PARTICIONA LA VARIACION DE LA VARIABLE DE RESPUESTA ENTRE:

VARIACION ATRIBUIBLE A FACTORES CONTROLADOS : (DEBIDA A "EFECTOS"^{1/})
Y
VARIACION ATRIBUIBLE A FACTORES NO CONTROLADOS : (DEBIDA AL ERROR)

^{1/}Efecto: tratamiento, bloque, etc.

EL ANALISIS DE VARIANZA
 - PARTICION DE LA VARIACION TOTAL -
 DE LA VARIABLE DE RESPUESTA Y

VARIACION EN LA VARIABLE DE RESPUESTA Y	=	VARIACION ATRIBUIBLE A "EFECTOS" A, B, ...	+	VARIACION ATRIBUIBLE AL ERROR
---	---	---	---	-------------------------------------

SCT	=	SCA + SCB + ... +	SCE
GL(TOTAL)	=	GL(A)+GL(B)+...+	GL(E)

EL CUADRADO MEDIO DE
UN EFECTO A

:

$$CMA = \frac{SCA}{GL(A)}$$

ES UN ESTIMADOR DE LA VARIANZA
ATRIBUIBLE A A

LA RAZON F PARA UN
EFECTO A

:

$$F_A = \frac{CMA}{CM_{ERROR}}$$

SE UTILIZA PARA COMPARAR LOS DOS
ESTIMADORES DE VARIANZA: LA
VARIANZA ATRIBUIBLE AL EFECTO A
CON LA VARIANZA ATRIBUIBLE AL
ERROR

$$F_A \sim F_{GL(A), GL(ERROR)}$$

REGLA DE DECISION

:

$$SI F_A > F_{TABLA}$$

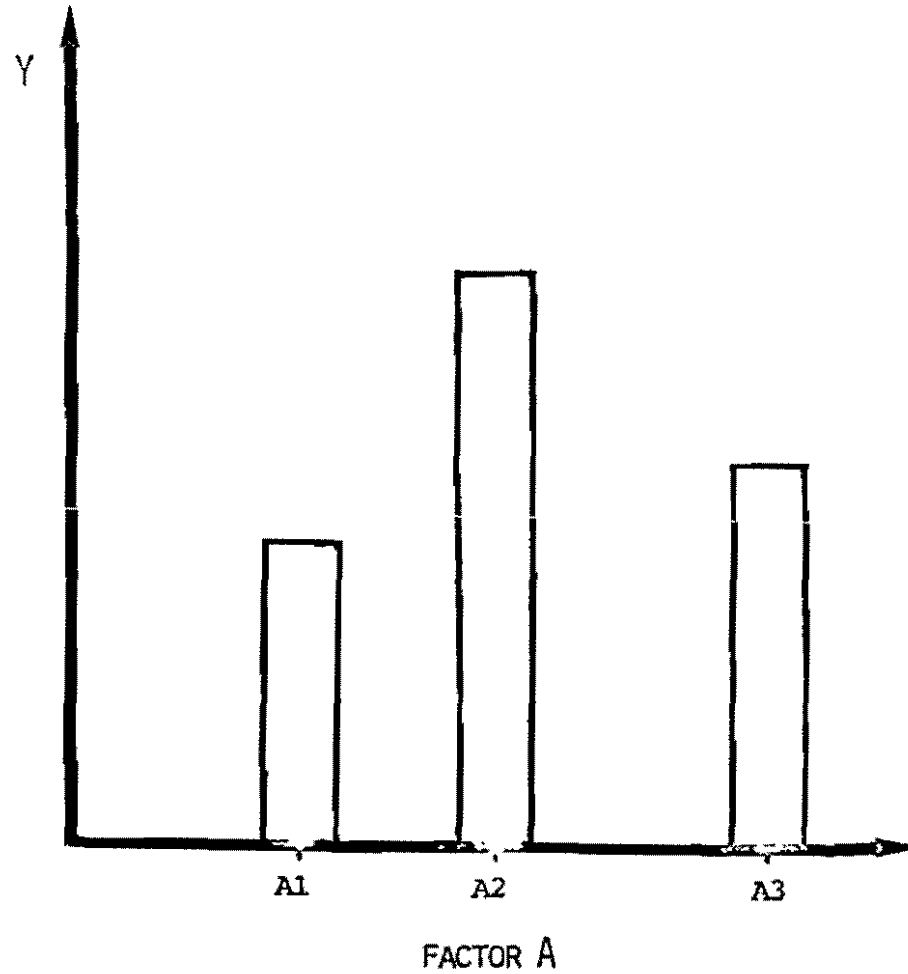
LAS MEDIAS DE Y
EN CADA NIVEL DE
A DIFIEREN SIGNI-
FICATIVAMENTE.

PRESENTACION TIPICA DE UNA TABLA
DE ANALISIS DE VARIANZA

FUENTES DE VARIACION	GL	SC	SM	F	PROB(F)
EFECTOS					
A	GL(A)	SCA	CM _A	F _A	PROB(F _A)
B	GL(B)	SCB	CM _B	F _B	PROB(F _B)
C	GL(C)	SCC	CM _C	F _C	PROB(F _C)
⋮					
⋮					
⋮					
ERROR	GL(ERROR)	SC _{ERROR}	CM _{ERROR}		
TOTAL	GL(TOTAL)	SCT			

CLASES DE EFECTOS EN UN ANALISIS DE VARIANZA

- EFECTO PRINCIPAL A: MIDE LA RESPUESTA PROMEDIO EN LOS NIVELES DEL FACTOR A



CALCULO DE LA
SUMA DE CUADRADOS TOTAL

$$SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

MIDE LA VARIABILIDAD TOTAL DE LA VARIABLE
DE RESPUESTA Y.

CALCULO DE LA SUMA DE CUADRADOS
PARA UN EFECTO PRINCIPAL
- CASO BALANCEADO -

$$SC(A) = \sum_i r(\bar{Y}_i - \bar{Y})^2 = \frac{1}{r} \sum T_i^2 - \frac{\bar{Y}^2}{N}$$

MIDE LA VARIABILIDAD ENTRE MEDIAS PARA LOS NIVELES DEL FACTOR A

donde,

r = número de observaciones en cada nivel del factor A

\bar{Y}_i = valor promedio de Y en nivel i

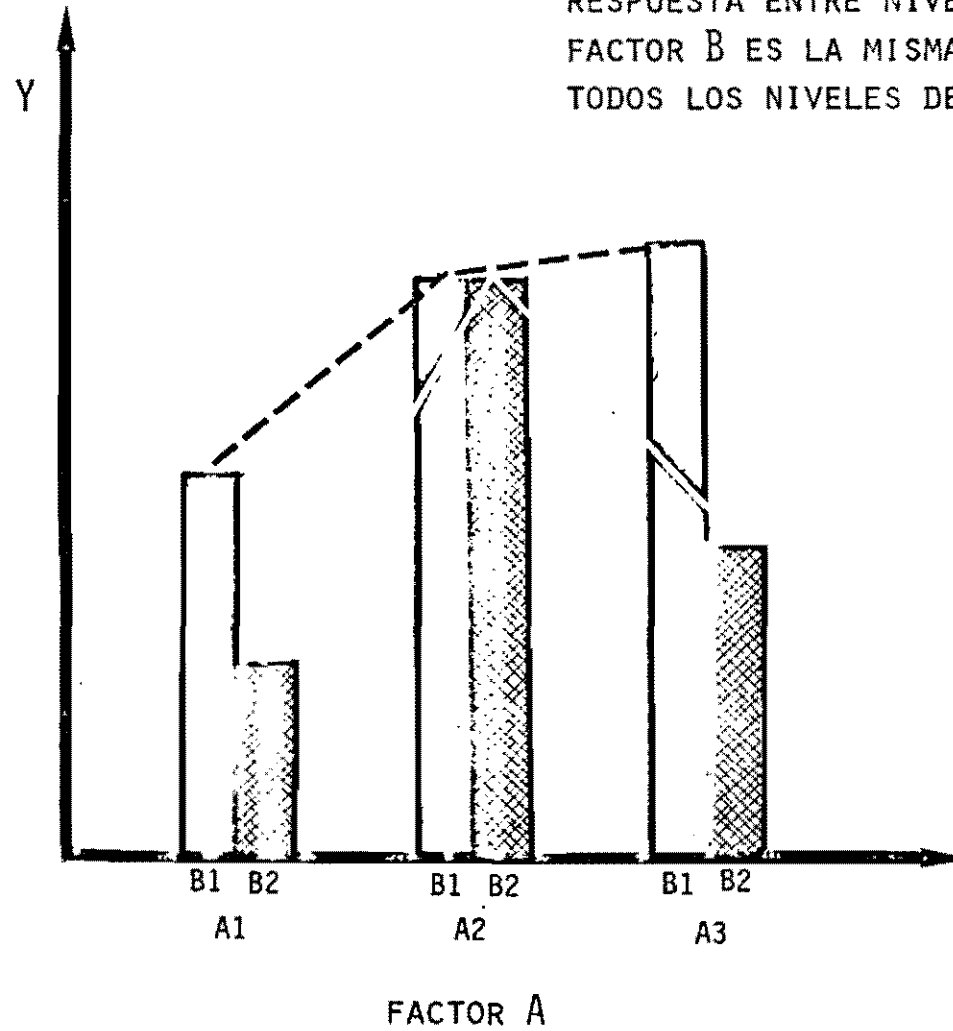
\bar{Y} = media general de Y

T_i = total de Y en el nivel i

N = número total de observaciones

CLASES DE EFECTOS EN UN ANALISIS DE VARIANZA

EFECTO INTERACCION A * B: MIDE SI LA DIFERENCIA DE LA RESPUESTA ENTRE NIVELES DEL FACTOR B ES LA MISMA PARA TODOS LOS NIVELES DEL FACTOR A.



CALCULO DE LA SUMA DE CUADRADOS
PARA UN EFECTO INTERACCION
- CASO BALANCEADO -

$$\begin{aligned}
 SC(A*B) &= \sum_{ij} n_{ij} (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y})^2 \\
 &= \sum_{ij} T_{ij}^2 / n_{ij} - \sum_i T_{i.}^2 / n_{i.} - \sum_j T_{.j}^2 / n_{.j} + \frac{\bar{Y}^2}{N}
 \end{aligned}$$

donde,

n_{ij} = número común de observaciones en cada celda ij

\bar{Y}_{ij} = promedio de Y en nivel i de A y nivel j de B

$\bar{Y}_{i.}$ = promedio de Y en nivel i de A

$\bar{Y}_{.j}$ = promedio de Y en nivel j de B

\bar{Y} = media general de Y

T_{ij} = total de Y en el nivel i de A y nivel j de B

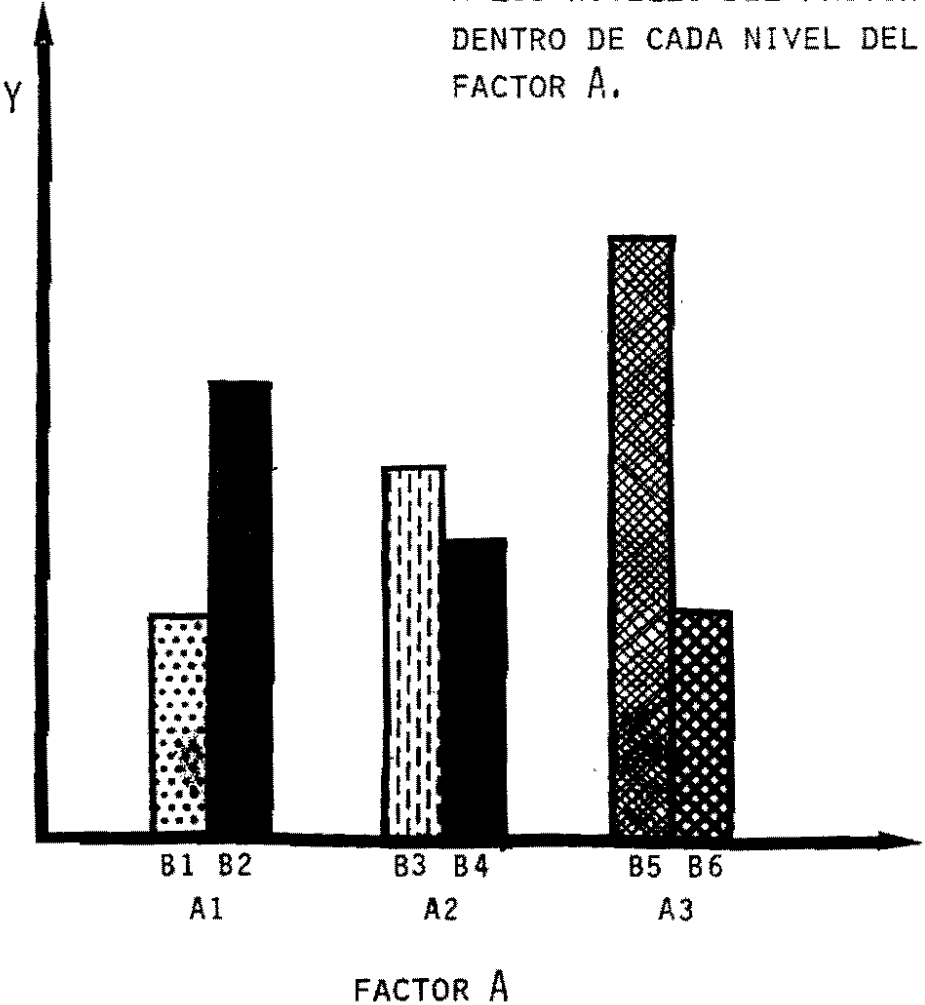
$T_{i.}$ = total de Y en el nivel i de A

$T_{.j}$ = total de Y en el nivel de j de B

N = número total de observaciones

CLASES DE EFECTOS EN UN ANALISIS DE VARIANZA

EFECTO ANIDADO B(A): MIDE LA RESPUESTA PROMEDIO A LOS NIVELES DEL FACTOR B DENTRO DE CADA NIVEL DEL FACTOR A.



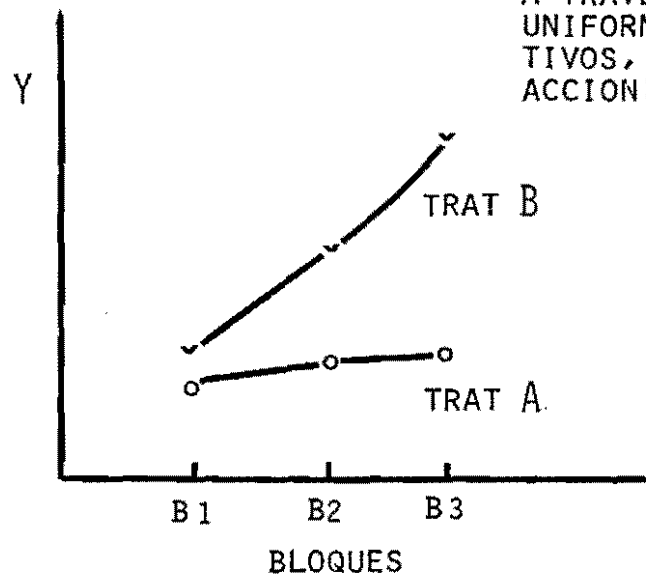
CALCULO DE LA SUMA DE CUADRADOS
PARA UN EFECTO ANIDADO
- CASO BALANCEADO -

$$\begin{aligned} SC(B(A)) &= \sum_{ij} n_{ij} (\bar{y}_{ij} - \bar{y}_{i.})^2 \\ &= \sum_{ij} T_{ij}^2 / n_{ij} - \sum_i T_{i.}^2 / n_{i.} \end{aligned}$$

MIDE LA VARIABILIDAD ENTRE LAS MEDIAS DE NIVELES DEL
FACTOR B DENTRO DE CADA NIVEL DEL FACTOR A

SUPUESTOS BASICOS PARA REALIZAR ANALISIS DE VARIANZA A UNA VARIABLE Y

1. ADITIVIDAD DE EFECTOS: SE VIOLA POR EJ., CUANDO LOS EFECTOS DE TRATAMIENTO A TRAVES DE BLOQUES NO SON UNIFORMES SINO MULTIPLICATIVOS, O CUANDO HAY INTERACCION: BLOQUE X TRAT.



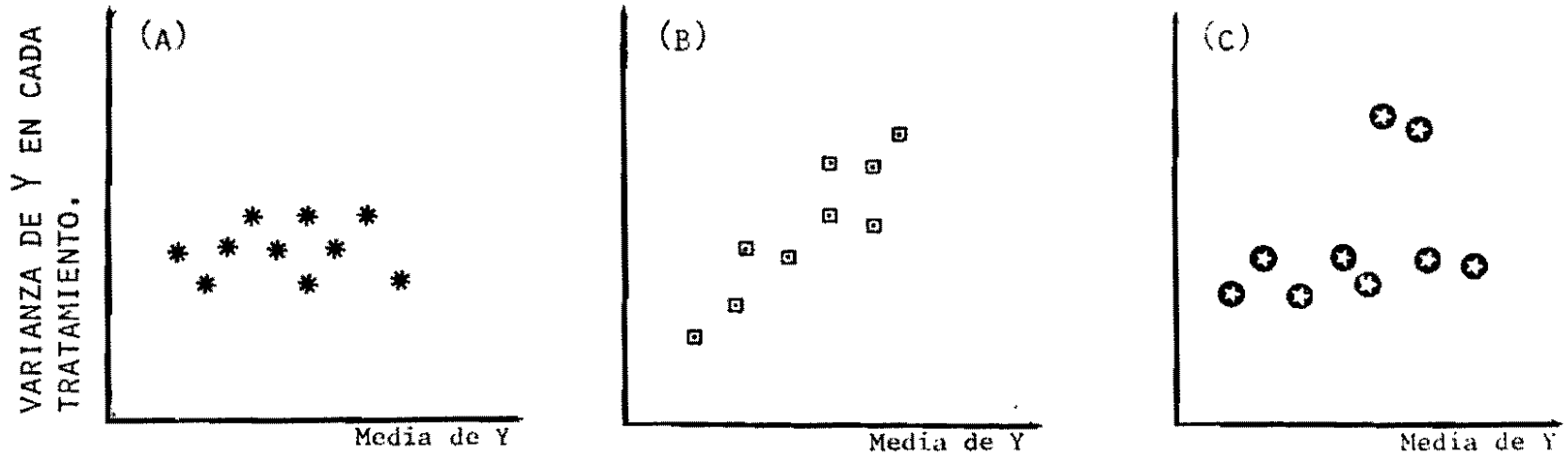
SUPUESTOS BASICOS PARA REALIZAR ANALISIS
DE VARIANZA A UNA VARIABLE Y

2. INDEPENDENCIA DE ERRORES: SE VIOLA
POR EJ., CUANDO LOS TRATAMIENTOS SE
APLICAN A LAS UNIDADES EXPERIMENTALES
EN FORMA SISTEMATICA.

SUPUESTOS BASICOS PARA REALIZAR ANALISIS DE VARIANZA A UNA VARIABLE Y

3. HOMOGENEIDAD DE VARIANZA DE Y A TRAVES DE TRATAMIENTOS:
SE VIOLA, POR EJ., CUANDO LAS VARIANZAS DE TRATAMIENTO
SON PROPORCIONALES A LAS MEDIAS DE TRATAMIENTO.

-141-



(A) HOMOGENEIDAD DE VARIANZA, (B) HETEROGENEIDAD DE VARIANZA,
(C) HETEROGENEIDAD DE VARIANZA SIN RELACION FUNCIONAL.

MEDIDAS REMEDIALES
CUANDO HAY VIOLACION DE SUPUESTOS
PARA EL ANALISIS DE VARIANZA

1. SI EXISTE HETEROGENEIDAD ENTRE VARIANZAS DE TRATAMIENTO

- A) CUANDO HAY RELACION FUNCIONAL ENTRE $VAR(Y)$ Y MEDIA DE Y ➤ TRANSFORME Y
EJ: \sqrt{Y} , $\ln(Y)$
ARCSIN \sqrt{Y}
- B) CUANDO NO HAY RELACION FUNCIONAL ENTRE $VAR(Y)$ Y MEDIA DE Y ➤ CONFORME GRUPOS DE TRATAMIENTOS CON VARIANZAS HOMOGENEAS Y ANALICELOS INDEPENDIENTEMENTE.

MEDIDAS REMEDIALES
CUANDO HAY VIOLACION DE SUPUESTOS
PARA EL ANALISIS DE VARIANZA

2. SI NO HAY INDEPENDENCIA DE
ERRORES ENTRE TRATAMIENTOS
(Causado por asignación sistemática) → NO HAY
REMEDIO
3. SI LOS EFECTOS DE TRATAMIENTO SON
MULTIPLICATIVOS, NO ADITIVOS → TRANSFORME
Y

PROCEDIMIENTOS SAS PARA EL ANALISIS DE VARIANZA

GLM	:	REALIZA ANALISIS DE VARIANZA, REGRESION, ANALISIS DE COVARIANZA Y ANALISIS DE VARIANZA MULTIVARIADO, PARA DISEÑOS BALANCEADOS O DESBALANCEADOS.
▶ ANOVA	:	REALIZA ANALISIS DE VARIANZA PARA DISEÑOS BALANCEADOS.
NESTED	:	REALIZA ANALISIS DE VARIANZA PARA MODELOS ANIDADOS CON EFECTOS ALEATORIOS.
VARCOMP	:	ESTIMA COMPONENTES DE VARIANZA PARA EFECTOS ALEATORIOS.
NPAR1WAY	:	REALIZA ANALISIS NO-PARAMETRICO DE 1 CLASIFICACION
TTEST	:	REALIZA PRUEBAS DE t PARA COMPARACION DE DOS MEDIAS
PLAN	:	GENERA PLANES EXPERIMENTALES ALEATORIOS

CUANDO USAR PROC ANOVA EN DISEÑOS NO BALANCEADOS

1. CUANDO EL DISEÑO ES COMPLETAMENTE AL AZAR,
2. CUANDO SE TRATA DE DISEÑOS CUADRADO LATINO O GRECO LATINO CON POCO GRADO DE DESBALANCE,
3. CUANDO SE TRATA DE DISEÑOS CON INTERACCIONES NO SIGNIFICATIVAS, CON POCO GRADO DE DESBALANCE.

PROC ANOVA

OBJETIVO: REALIZAR ANALISIS DE VARIANZA
PARA CUALQUIER TIPO DE DISEÑO EX-
PERIMENTAL BALANCEADO, SALVO
EXCEPCIONES EXPUESTAS.

ESPECIFICACIONES:

LAS PROPOSICIONES DISPONIBLES EN
PROC ANOVA SON:

PROC ANOVA opción;

CLASS variables;

MODEL variables dep = efectos;

MEANS efectos/opciones;

ABSORB variables;

FREQ variable;

TEST H= efectos = E = efecto;

BY variables;

DISEÑO COMPLETAMENTE AL AZAR

CARACTERISTICAS:

1. EXIGE UNIDADES EXPERIMENTALES HOMOGENEAS
2. PERMITE PROBAR CUALQUIER NUMERO DE TRATAMIENTOS, YA SEA UN FACTOR A VARIOS NIVELES O COMBINACIONES DE DISTINTOS FACTORES
3. PERMITE ESTIMAR LAS SUMAS DE CUADRADOS MEDIANTE LAS FORMULAS TRADICIONALES, AUNQUE SEA DESBALANCEADO.
4. LOS TRATAMIENTOS SE APLICAN A LAS UNIDADES EXPERIMENTALES AL AZAR.

EJEMPLO: DISEÑO COMPLETAMENTE AL AZAR

5 TRATAMIENTOS

3 REPLICACIONES POR TRATAMIENTO

- DISTRIBUCION DE TRATAMIENTOS -

T ₁	T ₃	T ₃	T ₄	T ₂
T ₅	T ₂	T ₅	T ₁	T ₄
T ₂	T ₅	T ₄	T ₁	T ₃

DISEÑO COMPLETAMENTE AL AZAR

ESTRUCTURA DE FUENTES DE VARIACION PARA EL ANOVA

<u>Fv</u>	<u>gl.</u>
TRATAMIENTOS	t-1
Error	(N-1)-(t-1)
<u>TOTAL</u>	<u>N-1</u>

PROC ANOVA

EJEMPLO NO. 1: DISEÑO COMPLETAMENTE AL AZAR

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE DATOS DE RENDIMIENTO (GR/POTE) DE 4 VARIETADES DE ARROZ PROBADAS A NIVEL DE INVERNADERO, USANDO 3 POTES POR VARIETADE, BAJO UN DISEÑO COMPLETAMENTE AL AZAR.

```
DATA CA;  
    INPUT VAR 1 REP 3 REND 7-10 ;  
    CARDS;  
    1 1 3.40  
    1 2 3.30  
    1 3 4.00  
    2 1 4.70  
    2 2 4.70  
    2 3 5.00  
    3 1 4.00  
    3 2 3.10  
    3 3 4.10  
    4 1 6.30  
    4 2 7.20  
    4 3 9.50
```

PROC ANOVA

EJEMPLO NO. 1 (CONT.): ANALISIS DE VARIANZA PARA UN DISEÑO
COMPLETAMENTE AL AZAR

```
PROC ANOVA;
  CLASSES VAR;
  MODEL REND= VAR;
  MEANS VAR/DUNCAN;
  TITLE1***** ;
  TITLE2* ;
  TITLE3* ANALISIS DE VARIANZA PARA UN * ;
  TITLE4* DISENO COMPLETAMENTE AL AZAR * ;
  TITLE5* ;
  TITLE6***** ;
```

ANALISIS DE VARIANZA PARA UN
 * DISEÑO COMPLETAMENTE AL AZAR *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: REND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	3	32.38916667	10.79638889
ERRUR	8	6.40000000	0.80000000
CORRECTED TOTAL	11	38.78916667	

MODEL F = 13.50 PR > F = 0.0017

R-SQUARE	C.V.	STD DEV	REND MEAN
0.835005	18.0997	0.89442719	4.94166667

SOURCE	DF	ANOVA SS	F VALUE	PR > F
VAR	3	32.38916667	13.50	0.0017

 * ANALISIS DE VARIANZA PARA UN *
 * DISEÑO COMPLETAMENTE AL AZAR *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: REND
 ALPHA=0.05 DF=8 MSE=0.8

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	VAR
	A	7.6667	3	4
	B	4.8000	3	2
	B	3.7333	3	3
	B	3.5667	3	1

DISEÑO DE BLOQUES COMPLETOS AL AZAR

CARACTERISTICAS:

1. PERMITE UTILIZAR UNIDADES EXPERIMENTALES HETEROGENEAS PERO POSIBLES DE ESTRATIFICAR EN GRUPOS O BLOQUES HOMOGENEOS.
2. REQUIERE MAXIMA HOMOGENEIDAD DENTRO DE BLOQUE Y MAXIMA HETEROGENEIDAD ENTRE LOS BLOQUES.
3. PERMITE PROBAR UN NUMERO NO MUY GRANDE DE TRATAMIENTOS (≤ 15), QUE PUEDEN CORRESPONDER A VARIOS NIVELES DE UN FACTOR O A COMBINACIONES DE FACTORES.
4. LOS TRATAMIENTOS SE APLICAN AL AZAR DENTRO DE CADA BLOQUE
5. PARA EFECTO DEL ANALISIS ESTADISTICO ES PREFERIBLE QUE SEA BALANCEADO.
6. -SI EXISTE GRADIENTE DEFINIDO, LOS BLOQUES DEBEN SER RECTANGULARES Y LOCALIZARSE EN FORMA PERPENDICULAR A LA GRANDIENTE.
-SI NO HAY GRADIENTE DEFINIDA LOS BLOQUES DEBEN SER CUADRADOS.

EJEMPLO: DISEÑO EN BLOQUES AL AZAR

CON 3 BLOQUES

5 TRATAMIENTOS (NIVELES P)

- DISTRIBUCION DE TRATAMIENTOS -

BLOQUE 1	BLOQUE 2	BLOQUE 3
T ₂	T ₃	T ₁
T ₄	T ₅	T ₃
T ₁	T ₂	T ₂
T ₃	T ₁	T ₄



DISEÑO EN BLOQUES COMPLETOS AL AZAR

ESTRUCTURA DE FUENTES DE VARIACION PARA EL ANOVA

FV	gl.
BLOQUES	b-1
TRATAMIENTOS	t-1
Error	(b-1)(t-1)
TOTAL	N - 1

PROC ANOVA

EJEMPLO NO. 2: DISEÑO EN BLOQUES COMPLETOS AL AZAR

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE DATOS DE RENDIMIENTO DE RAICES (TON/HA) DE 4 VARIETADES DE YUCA, BAJO UN DISEÑO DE BLOQUES COMPLETOS AL AZAR.

```
DATA BCA;
      INPUT VAR 1 BLOQUE 3 REID 5-7 ;
      CARDS;
1 1 5.1
1 2 7.4
1 3 4.2
2 1 0.8
2 2 2.5
2 3 1.3
3 1 2.0
3 2 2.6
3 3 0.6
4 1 5.9
4 2 7.3
4 3 5.8
```

PROC ANOVA.

EJEMPLO NO. 2: ANALISIS DE VARIANZA PARA UN DISEÑO
DE BLOQUES COMPLETOS AL AZAR

-158-

```
PROC ANOVA;
  CLASSES BLOQUE VAR;
  MODEL REND=BLOQUE VAR;
  MEANS VAR/WALLER;
  TITLE1 *****;
  TITLE2 * *;
  TITLE3 * ANALISIS DE VARIANZA PARA UN*;
  TITLE4 * DISEÑO EN BLOQUES AL AZAR *;
  TITLE5 * *;
  TITLE6 *****;
```

* ANALISIS DE VARIANZA PARA UN
* DISEÑO EN BLOQUES AL AZAR *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: REND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	5	64.21083333	12.84216667
ERROR	6	1.87833333	0.31305556
CORRECTED TOTAL	11	66.08916667	

MODEL F = 41.02 PR > F = 0.0001

R-SQUARE	C.V.	STD DEV	REND MEAN
0.971579	14.6918	0.55951368	3.80833333

SOURCE	DF	ANOVA SS	F VALUE	PR > F
BLOQUE	2	8.18166667	13.07	0.0065
VAR	3	55.02916667	59.66	0.0001

* * * * *
 *
 * ANALISIS DE VARIANZA PARA UN *
 * DISEÑO EN BLOQUES AL AZAR *
 *
 * * * * *

b

ANALYSIS OF VARIANCE PROCEDURE

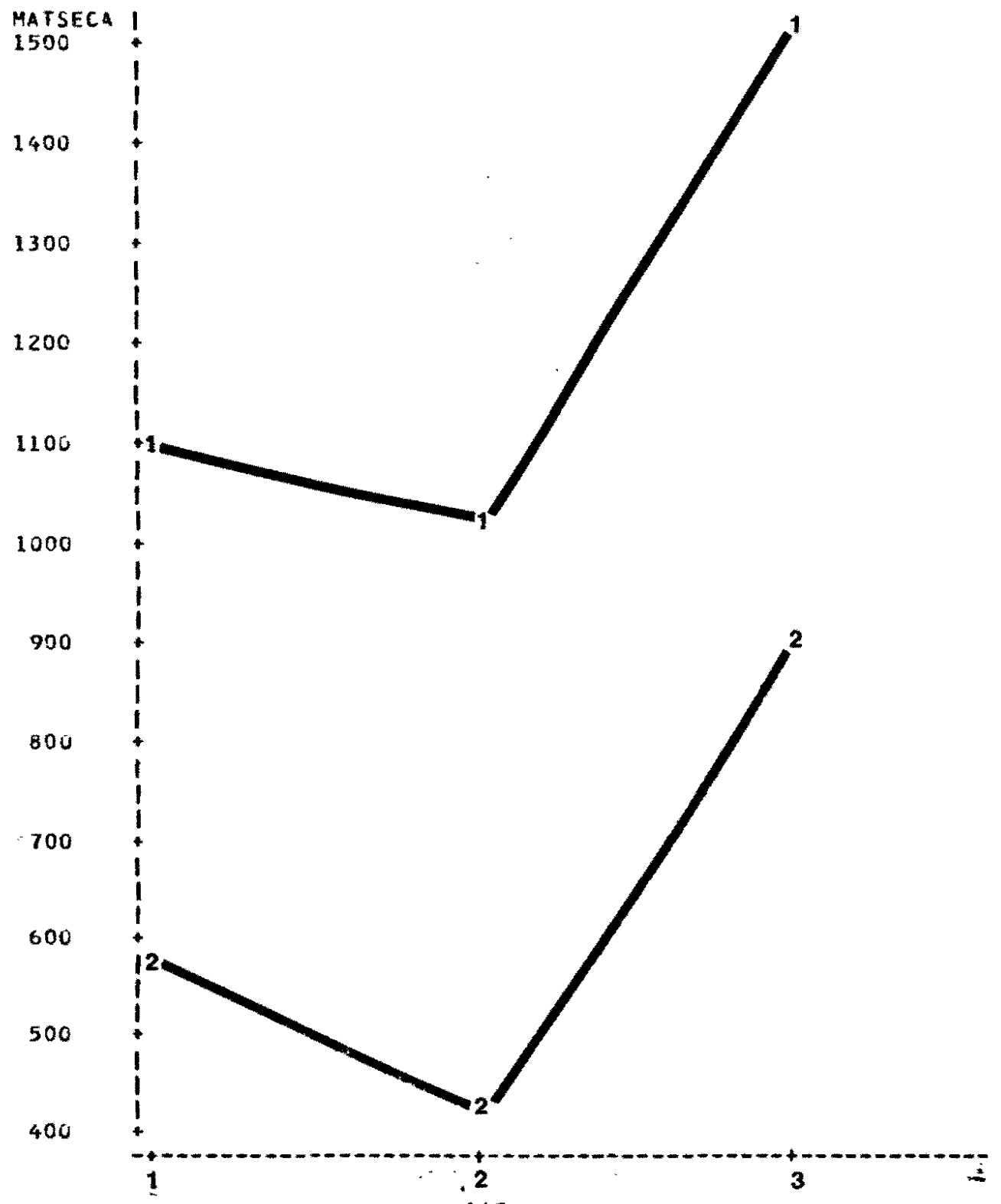
WALLER-DUNCAN K-RATIO T TEST FOR VARIABLE: REND
 KRATIO=100 DF=6 MSE=0.313056 F=59.6584 T=2.35182 LSD=1.07898

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

WALLER	GROUPING	MEAN	N	VAR
	A	6.3333	3	4
	A			
	A	5.5667	3	1
	B	1.8000	3	3
	B			
	D	1.5333	3	2

* ANALISIS DE VARIANZA PARA UN DISEÑO
* EN BLOQUES COMPLETOS AL AZAR CON
* ESTRUCTURA FACTORIAL.
* GRAFICO DE LA INTERACCION N x P

PLOT OF MATSECA'N SYMBOL IS VALUE OF P



DISEÑO EN FRANJAS DIVIDIDAS Y SUB-DIVIDIDAS

CARACTERISTICAS:

1. PERMITE ESTUDIAR DOS O TRES FACTORES, DOS DE LOS CUALES REQUIEREN PARCELAS GRANDES.
2. EL PRIMER FACTOR SE APLICA EN FRANJAS HORIZONTALES AL AZAR DENTRO DE CADA REPETICION, EL SEGUNDO FACTOR SE APLICA EN FRANJAS VERTICALES AL AZAR DENTRO DE CADA REPETICION, EL TERCER FACTOR SE APLICA AL AZAR A LAS SUB-PARCELAS DENTRO DE CADA COMBINACION DE FACTOR 1 X FACTOR 2.
3. EL ORDEN DE PRECISION EN EL ESTUDIO DE LOS FACTORES ES:
MINIMA PRECISION: FACTOR 1 X FACTOR 2
PRECISION MEDIA : FACTOR 1 X FACTOR 2
MAXIMA PRECISION: FACTOR 3 E INTERACCIONES CON EL
4. REQUIERE DOS REPETICIONES COMO MINIMO Y ES PREFERIBLE QUE SEA BALANCEADO.

EJEMPLO: DISEÑO DE PARCELAS DIVIDIDAS

P.PAL: FACTOR 1 A 2 NIVELES

S.PARC: FACTOR 2 A 5 NIVELES

No. REPS: 3

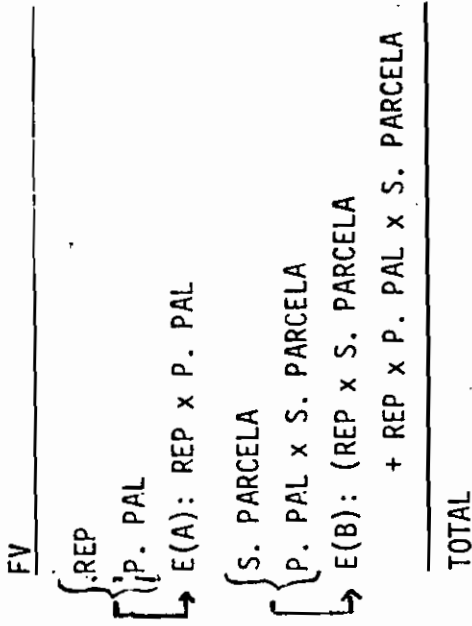
- DISTRIBUCION DE TRATAMIENTOS -

- 69 -

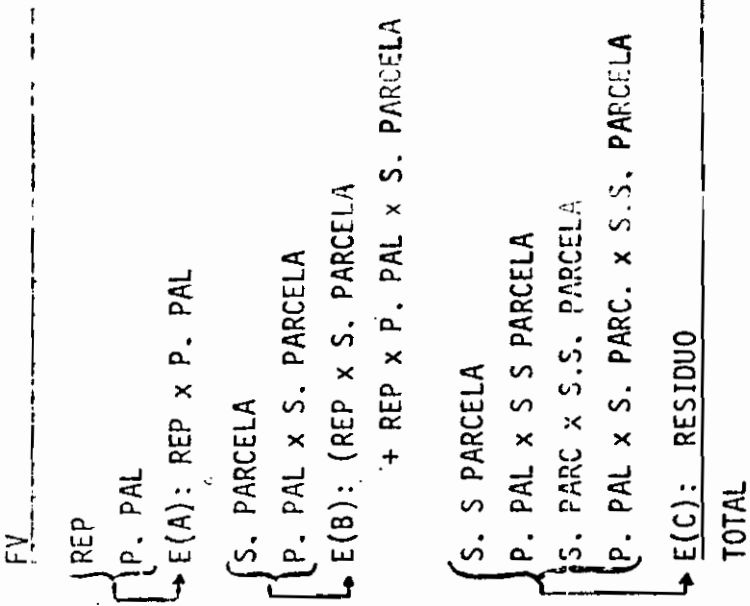
REP I		REP II		REP III	
P.PAL 1	P.PAL 2	P.PAL 2	P.PAL 1	P.PAL 2	P.PAL 1
1	4	4	2	1	3
5	3	5	1	4	2
2	2	1	4	2	1
4	1	3	5	5	4
3	5	2	3	3	5

DISTRIBUCION DE FUENTES DE VARIACION PARA EL ANOVA

A. DISEÑO DE PARCELAS DIVIDIDAS



B. DISEÑO DE PARCELAS SUP-DIVIDIDAS



PROC ANOVA

EJEMPLO NO. 4 : DISEÑO DE PARCELAS DIVIDIDAS

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE DATOS DE GANANCIA DIARIA DE PESO (GR/DIA) PROVENIENTES DE UN ENSAYO BAJO DISEÑO DE PARCELAS DIVIDIDAS, CON PARCELA PRINCIPAL: TIPO DE PASTO (2) Y SUBPARCELA: DOSIS DE SUPLEMENTO MINERAL (3), UTILIZANDO 3 NOVILLOS POR SUBPARCELA. EL EXPERIMENTO SE HIZO CON 2 REPETICIONES.

-171-

```
DATA PD ;  
    INPUT REP 1. PASTO 3 TRAT 5  
    ANIMAL 7 GANPESO 9-11;  
    CARDS;  
    1 1 1 1 155  
    1 1 1 2 170  
    1 1 1 3 150  
    .  
    2 2 3 2 333  
    2 2 3 3 346
```

PROC ANOVA

EJEMPLO NO. 4 (CONT.): ANALISIS DE VARIANZA PARA UN DISEÑO DE PARCELAS
DIVIDIDAS UTILIZANDO COMO VARIABLE DE RESPUESTA
LAS GANANCIAS DE PESO PROMEDIO POR SUBPARCELA
(MODELO 1)

```
PROC SORT;
  BY REP PASTO TRAT;
PROC MEANS NOPRINT;
  BY REP PASTO TRAT;
  VAR GANPESO;
  OUTPUT OUT = ARSALIDA MEAN=GANPESO;
PROC ANOVA DATA=ARSALIDA;
  CLASSES REP PASTO TRAT;
  MODEL GANPESO=REP PASTO REP*PASTO TRAT TRAT*PASTO;
  TEST H= REP PASTO E=REP*PASTO;
  MEANS PASTO ;
  MEANS TRAT TRAT*PASTO / DUNCAN;
TITLE1 *****;
TITLE2 *****;
TITLE3 * ANALISIS DE VARIANZA PARA UN DISEÑO EN PARCELAS DIVIDIDAS*;
TITLE4 * MODELO I - TOMANDO PROMEDIOS POR SUBPARCELA *;
TITLE5 *****;
TITLE6 *****;
```

ANALISIS DE VARIANZA PARA UN DISEÑO EN PARCELAS DIVIDIDAS
 * MODELO I - TOMANDO PROMEDIOS POR SUBPARCELA *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: GANPESO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	7	43293.74074074	6184.82010582
ERRUR	4	540.44444444	135.11111111
CORRECTED TOTAL	11	43834.18518519	

MODEL F = 45.78 PR > F = 0.0012

R-SQUARE	L.V.	STD DEV	GANPESO MEAN
0.987671	4.4412	11.62373052	261.72222222

SOURCE	DF	ANOVA SS	F VALUE	PR > F
REP	1	192.00000000	1.42	0.2991
PASTO	1	15841.33333333	117.25	0.0004
REP*PASTO	1	1008.33333333	7.46	0.0523
TRAT	2	20718.51851852	76.67	0.0006
PASTO*TRAT	2	5533.55555556	20.48	0.0079

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*PASTO AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
REP	1	192.00000000	0.19	0.7381
PASTO	1	15841.33333333	15.71	0.1573

ANALISIS DE VARIANZA PARA UN DISENO EN PARCELAS DIVIDIDAS
 * MODELO I - TOMANDO PROMEDIOS POR SUBPARCELA *
 *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: GANPESU
 ALPHA=0.05 DF=4 MSE=135.111

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	TRAT
	A	311.17	4	3
	B	264.50	4	2
	C	209.50	4	1

PROC ANOVA

EJEMPLO NO. 4 (CONT.): ANALISIS DE VARIANZA PARA DISEÑO DE PARCELAS
DIVIDIDAS UTILIZANDO COMO VARIABLE DE RESPUES-
TA LA GANANCIA DE PESO DE ANIMALES INDIVIDUALES.
(MODELO 2)

```
PROC ANOVA DATA=PD;
  CLASSES REP PASTO TRAT ANIMAL;
  MODEL GANPESO = REP PASTO REP*PASTO;
                TRAT TRAT*PASTO REP*TRAT(PASTO);
  TEST H = REP PASTO E=REP*PASTO;
  TEST H = TRAT TRAT*PASTO E=REP*TRAT(PASTO);
  MEANS PASTO/DUNCAN E=REP*PASTO;
  MEANS TRAT TRAT*PASTO / DUNCAN E=REP*TRAT(PASTO);
  TITLE1 *****;
  TITLE2 *****;
  TITLE3 * ANALISIS DE VARIANZA PARA UN DISEÑO EN PARCELAS DIVIDIDAS*;
  TITLE4 *                MODELO II                *;
  TITLE5 *                                                *;
  TITLE6 *****;
```

ANÁLISIS DE VARIANZA PARA UN DISEÑO EN PARCELAS DIVIDIDAS
 MODELO II

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: GANPESO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	11	131502.55555556	11954.77777778
ERROR	24	24296.66666667	1012.36111111
CORRECTED TOTAL	35	155799.22222222	

MODEL F = 11.81 PR > F = 0.0001

R-SQUARE	C.V.	STD DEV	GANPESO MEAN
0.844051	12.1570	31.91762265	261.72222222

SOURCE	DF	ANOVA SS	F VALUE	PR > F
REP	1	576.00000000	0.57	0.4580
PASTO	1	47524.00000000	46.94	0.0001
REP*PASTO	1	3025.00000000	2.99	0.0967
TRAT	2	62155.55555556	30.70	0.0001
PASTO*TRAT	2	16600.66666667	8.20	0.0019
REP*TRAT(PASTO)	4	1621.33333333	0.40	0.8064

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*PASTO AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
REP	1	576.00000000	0.19	0.7391
PASTO	1	47524.00000000	15.71	0.1573

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*TRAT(PASTO)
 AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRAT	2	62155.55555556	76.67	0.0006
PASTO*TRAT	2	16600.66666667	20.48	0.0079

PROC ANOVA

EJEMPLO NO. 3: DISEÑO EN BLOQUES COMPLETOS AL AZAR, CON ESTRUCTURA FACTORIAL

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE DATOS DE PRODUCCION DE MATERIA SECA (KG/HA) DE *Stylosanthes capitata*, BAJO UN DISEÑO DE BLOQUES COMPLETOS AL AZAR, CON ESTRUCTURA FACTORIAL DE 3 NIVELES DE APLICACION DE N X 2 NIVELES DE APLICACION DE P.

```
DATA BCAF;  
    INPUT N 1 P 3 BLOQUE 5 MATSECA 7=10;  
    CARDS;  
1 1 1 960  
1 1 2 1110  
1 1 3 1204  
1 2 1 936  
1 2 2 325  
1 2 3 476  
2 1 1 1067  
2 1 2 768  
2 1 3 1204  
2 2 1 399  
2 2 2 245  
2 2 3 604  
3 1 1 1460  
3 1 2 1678  
3 1 3 1456  
3 2 1 1235  
3 2 2 625  
3 2 3 876
```

PROC ANOVA

EJEMPLO NO. 3 (CONT.): ANALISIS DE VARIANZA PARA UN DISEÑO DE BLOQUES
COMPLETOS AL AZAR, CON ESTRUCTURA FACTORIAL:

-162-

```
PROC ANOVA;
  CLASSES N P BLOQUE;
  MODEL MATSECA=BLOQUE N P N*P ;
  MEANS N P/DUNCAN;
  TITLE1 *****;
  TITLE2 * *;
  TITLE3 * ANALISIS DE VARIANZA PARA UN DISEÑO*;
  TITLE4 * EN BLOQUES COMPLETOS AL AZAR CON *;
  TITLE5 * ESTRUCTURA FACTORIAL. *;
  TITLE6 * *;
  TITLE7 *****;
```

* ANALISIS DE VARIANZA PARA UN DISEÑO *
 * EN BLOQUES COMPLETOS AL AZAR CON *
 * ESTRUCTURA FACTORIAL. *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: MATSECA

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	7	2507388.55555555	358198.36507936
ERROR	10	456064.55555556	45606.45555556
CORRECTED TOTAL	17	2963453.11111111	

MODEL F = 7.85 PR > F = 0.0022

R-SQUARE	C.V.	STD DEV	MATSECA MEAN
0.846104	23.1178	213.55667996	923.77777778

SOURCE	DF	ANOVA SS	F VALUE	PR > F
BLOQUE	2	161364.77777778	1.77	0.2199
N	2	842321.44444444	9.23	0.0053
P	1	1494144.22222222	32.76	0.0002
N.P	2	9558.11111111	0.10	0.9015

* * * * *
 * ANALISIS DE VARIANZA PARA UN DISEÑO *
 * EN BLOQUES COMPLETOS AL AZAR CON *
 * ESTRUCTURA FACTORIAL. *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: MATSECA
 ALPHA=0.05 DF=10 MSE=45605.5

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	N
	A	1221.7	6	3
	B	835.2	6	1
	B	714.5	6	2

```

* ANALISIS DE VARIANZA PARA UN DISEÑO
* EN BLOQUES COMPLETOS AL AZAR CON
* ESTRUCTURA FACTORIAL.

```

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: MATSECA
 ALPHA=0.05 DF=10 MSE=45606.5

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	P
	A	1211.9	9	1
	B	655.7	9	2

PROC ANOVA

EJEMPLO NO. 3 (CONT.): GRAFICO DE LA INTERACCION (N X P)

-166-

```
PROC SORT DATA=BCAF ;
      BY N P ;
PROC MEANS NOPRINT ;
      BY N P ;
      VAR MATSECA;
      OUTPUT OUT=MED MEAN=MATSECA ;
PROC PLOT DATA=MED ;
      PLOT MATSECA*N=P;
TITLE1 *****;
TITLE2 * *;
TITLE3 * ANALISIS DE VARLANZA PARA UN DISENO*;
TITLE4 * EN BLOQUES COMPLETOS AL AZAR CON *;
TITLE5 * ESTRUCTURA FACTORIAL. *;
TITLE6 * GRAFICO DE LA INTERACCION N*P *;
TITLE7 * *;
TITLE8 *****;
```

ANALISIS DE VARIANZA PARA UN DISEÑO EN PARCELAS DIVIDIDAS
 MODELO II

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: GANPESO
 ALPHA=0.05 DF=4 MSE=405.333

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	TRAT
	A	311.17	12	3
	B	264.50	12	2
	C	209.50	12	1

ANALISIS DE VARIANZA PARA UN DISEÑO EN PARCELAS DIVIDIDAS
 MODELO II

ANALYSIS OF VARIANCE PROCEDURE

MEANS

PASTO	TRAT	N	GANPESO
1	1	6	163.166667
1	2	6	258.000000
1	3	6	255.000000
2	1	6	255.833333
2	2	6	271.000000
2	3	6	367.333333

PROC ANOVA

EJEMPLO NO. 5: DISEÑO DE PARCELAS SUB-DIVIDIDAS

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE DATOS DE RENDIMIENTO DE 3 VARIETADES DE FRIJOL (SUB-SUB-PARCELA), EVALUADAS CON Y SIN APLICACION DE INSECTICIDA (S/E-PARCELA), CON Y SIN APLICACION DE RIEGO (PARCELA PRINCIPAL), EN UN DISEÑO DE PARCELAS SUBDIVIDIDAS CON 2 REPETICIONES.

```
DATA PSD;
      INPUT REP 1 RIEGO 3 INSECT 5
            VAR 7 REND 9-14;

CARDS;
1 0 0 1 800.5
1 0 0 2 1050.8
1 0 0 3 700.0
1 0 1 1 1300.0
1 0 1 2 1000.0
1 0 1 3 1250.0
1 1 0 1 1500.0
.
.
.
2 1 1 3 1370.0
```

PROC ANOVA

EJEMPLO NO. 5 (CONT.): ANALISIS DE VARIANZA PARA UN DISEÑO DE
PARCELAS SUBDIVIDIDAS

```
PROC ANOVA DATA = PSD ;
  CLASSES REP RIEGO INSECT VAR ;
  MODEL REJD = REP RIEGO REP*RIEGO
              INSECT INSECT*RIEGO REP*INSECT(RIEGO)
              VAR VAR*INSECT VAR*RIEGO VAR*INSECT*RIEGO;
  TEST H = RIEGO E = REP*RIEGO;
  TEST H = INSECT INSECT*RIEGO E = REP*INSECT(RIEGO) ;
  MEANS RIEGO / DUNCAN E=REP*RIEGO;
  MEANS INSECT INSECT*RIEGO / DUNCAN E=REP*INSECT(RIEGO) ;
  MEANS VAR VAR*INSECT VAR*RIEGO VAR*INSECT*RIEGO / DUNCAN;
  TITLE1 *****;
  TITLE2 * *;
  TITLE3 * DISEÑO EN PARCELAS SUBDIVIDIDAS*;
  TITLE4 * *;
  TITLE5 *****;
```

* * * * *
 * DISEÑO EN PARCELAS SUBDIVIDIDAS *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: REND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	15	610762.50000000	40717.50000000
ERROR	8	109033.33333333	13629.16666667
CORRECTED TOTAL	23	719795.83333333	

MODEL F = 2.99 PR > F = 0.0611

R-SQUARE	L.V.	STD DEV	REND MEAN
0.848522	25.1288	116.74402197	464.58333333

SOURCE	DF	ANOVA SS	F VALUE	PR > F
REP	1	1204.16666667	0.09	0.7739
RIEGO	1	185504.16666667	13.61	0.0061
REP*RIEGO	1	104.16666667	0.01	0.9325
INSECT	1	104.16666667	0.01	0.9325
RIEGO*INSECT	1	28704.16666667	2.11	0.1848
REP*INSECT (RIEGO)	2	8308.33333333	0.30	0.7455
VAR	2	229658.33333333	8.43	0.0107
INSECT*VAR	2	45108.33333333	1.65	0.2504
RIEGO*VAR	2	97008.33333333	3.56	0.0784
RIEGO*INSECT*VAR	2	15058.33333333	0.55	0.5960

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*RIEGO AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
RIEGO	1	185504.16666667	1780.84	0.0151

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*INSECT (RIEGO) AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
INSECT	1	104.16666667	0.03	0.8887
RIEGO*INSECT	1	28704.16666667	6.91	0.1194

DISEÑO EN PARCELAS SUBDIVIDIDAS

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: REND
ALPHA=0.05 DF=1 MSE=104.167

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GRUPO	MEAN	N	RIEGO
	A	552.50	12	1
	B	376.67	12	0

* * * * *
 * DISEÑO EN PARCELAS SUBDIVIDIDAS *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

MEANS

RIEGO	INSECT	N	RFND
0	0	6	340.000000
0	1	6	413.333333
1	0	6	585.000000
1	1	6	520.000000

* DISEÑO EN PARCELAS SUBDIVIDIDAS *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: REND
 ALPHA=0.05 DF=8 MSE=13529.2

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	VAR
	A	535.00	8	2
	A	532.50	8	3
	B	326.25	8	1

* * * * *

DISENO EN PARCELAS SUBDIVIDIDAS

* * * * *

ANALYSIS OF VARIANCE PROCEDURE

MEANS

INSECT	VAR	N	REND
0	1	4	315.000000
0	2	4	590.000000
0	3	4	482.500000
1	1	4	337.500000
1	2	4	480.000000
1	3	4	582.500000

RIEGO	VAR	N	REND
0	1	4	262.500000
0	2	4	510.000000
0	3	4	357.500000
1	1	4	390.000000
1	2	4	560.000000
1	3	4	707.500000

RIEGO	INSECT	VAR	N	REND
0	0	1	2	235.000000
0	0	2	2	495.000000
0	0	3	2	290.000000
0	1	1	2	290.000000
0	1	2	2	525.000000
0	1	3	2	425.000000
1	0	1	2	395.000000
1	0	2	2	685.000000
1	0	3	2	675.000000
1	1	1	2	385.000000
1	1	2	2	435.000000
1	1	3	2	740.000000

DISEÑO DE PARCELAS DIVIDIDAS Y SUB-DIVIDIDAS

CARACTERISTICAS

1. PERMITE ESTUDIAR DOS O TRES FACTORES, UNO DE LOS CUALES REQUIERE PARCELAS MAS GRANDES QUE LOS OTROS. EL PRIMERO SE ASIGNA A LAS "PARCELAS PRINCIPALES" AL AZAR DENTRO DE CADA REPETICION; EL SEGUNDO SE ASIGNA A LAS "SUB-PARCELAS" AL AZAR DENTRO DE CADA PARCELA PRINCIPAL Y EL TERCERO A LAS "SUB-PARCELAS" AL AZAR DENTRO DE CADA SUB-PARCELA.
2. EL ORDEN DE PRECISION EN EL ESTUDIO DE LOS FACTORES ES:
 - MINIMA PRECISION: FACTOR 1
 - PRECISION MEDIA: FACTOR 2 Y FACTOR 1 X FACTOR 2
 - PRECISION MAXIMA: FACTOR 3 E INTERACCIONES CON EL.
3. REQUIERE 2 REPETICIONES COMO MINIMO Y ES PREFERIBLE QUE SEA BALANCEADO.

EJEMPLO: DISEÑO EN FRANJAS SUB-DIVIDIDAS

FRANJAS HORIZONTALES: FACTOR 1 A 3 NIVELES

FRANJAS VERTICALES : FACTOR 2 A 4 NIVELES

SUB PARCELA : FACTOR 3 A 3 NIVELES

No. DE REPLICACIONES: 2

- DISTRIBUCION DE TRATAMIENTOS -

		REP I			
		FV ₂	FV ₃	FV ₁	FV ₄
FH ₁		1	2	3	1
		3	1	2	2
		2	3	1	3
		1	2	3	1
FH ₃		3	1	2	2
		2	3	1	3
		3	1	2	2
FH ₂		2	2	1	3
		1	3	3	1

		REP II			
		FV ₄	FV ₂	FV ₁	FV ₃
FH ₃		1	3	1	2
		2	2	3	1
		3	1	2	3
		3	2	1	1
FH ₁		1	1	3	2
		2	3	2	3
		3	2	2	3
FH ₂		2	1	3	1
		1	3	1	2

DISTRIBUCION DE FUENTES DE VARIACION PARA EL ANOVA

A. DISEÑO EN FRANJAS DIVIDIDAS

FV
REP
[FACTOR 1
→ E(A): REP x FACTOR 1
[FACTOR 2
→ E(B): REP x FACTOR 2
[FACTOR 1 x FACTOR 2
→ E(C): REP x FACTOR 1 x FACTOR 2
TOTAL

B. DISEÑO EN FRANJAS SUB-DIVIDIDAS

FV
REP
[FACTOR 1
→ E(A): REP x FACTOR 1
[FACTOR 2
→ E(B): REP x FACTOR 2
[FACTOR 1 x FACTOR 2
→ E(C): REP x FACTOR 1 x FACTOR 2
[FACTOR 3
[FACTOR 3 x FACTOR 1
[FACTOR 3 x FACTOR 2
[FACTOR 3 x FACTOR 2 x FACTOR 1
→ E(D): RESIDUO
TOTAL

PROC ANOVA

EJEMPLO NO. 6: DISEÑO DE FRANJAS DIVIDIDAS

CONSIDERE EL SIGUIENTE ARCHIVO SAS, EL CUAL CONTIENE DATOS DEL EFECTO DE LA APLICACION DE RIEGO CON APLICACION DE NITROGENO A 3 NIVELES SOBRE EL RENDIMIENTO DE UNA VARIEDAD DE FRIJOL. EL DISEÑO UTILIZADO FUE DE FRANJAS DIVIDIDAS CON 3 REPETICIONES, CON RIEGO Y N COMO FACTORES EN FRANJAS.

```
DATA FD;
      INPUT REP 8 RIEGO 10
            N 12-14 REND 16-20;
CARDS;
1 0 0 600
1 0 50 700
1 0 100 1000
1 1 0 800
1 1 50 1000
1 1 100 1500
2 0 0 750
2 0 50 700
2 0 100 950
2 1 0 900
2 1 50 1050
2 1 100 1200
3 0 0 680
3 0 50 850
3 0 100 700
3 1 0 1200
3 1 50 1100
3 1 100 1400
```

PROC ANOVA

EJEMPLO NO. 6 (CONT.): ANALISIS DE VARIANZA PARA
UN DISEÑO DE FRANJAS DI-
VIDIDAS

```
PROC ANOVA;
  CLASSES REP RIEGO N ;
  MODEL REND = REP RIEGO REP*RIEGO
           N REP*N
           RIEGO*N
  TEST H = RIEGO E = REP*RIEGO;
  TEST H = N E = REP*N;
  MEANS RIEGO / DUNCAN E = REP*RIEGO;
  MEANS N / DUNCAN E = REP*N;
  MEANS RIEGO*N;
  TITLE1 *****;
  TITLE2 * *;
  TITLE3 * DISEÑO EN FRANJAS DIVIDIDAS*;
  TITLE4 * *;
  TITLE5 *****;
```

* DISEÑO EN FRANJAS DIVIDIDAS *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: REND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	13	628529.50000000	48348.42307692
ERROR	4	19392.00000000	4848.00000000
CORRECTED TOTAL	17	647921.50000000	

MODEL F = 9.97 PR > F = 0.0196

R-SQUARE	C.V.	STD DEV	REND MEAN
0.970070	16.1987	69.62758074	429.83333333

SOURCE	DF	ANOVA SS	F VALUE	PR > F
REP	2	7167.58333333	0.74	0.5331
RIEGO	1	275282.00000000	56.78	0.0017
REP*RIEGO	2	18254.25000000	1.88	0.2653
N	2	230295.58333333	23.75	0.0060
REP*N	4	12647.83333333	0.65	0.6555
RIEGO*N	2	84882.25000000	8.75	0.0346

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*RIEGO AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
RIEGO	1	275282.00000000	30.16	0.0316

TESTS OF HYPOTHESES USING THE ANOVA MS FOR REP*N AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
N	2	230295.58333333	36.42	0.0027

 * DISEÑO EN FRANJAS DIVIDIDAS *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: REND
 ALPHA=0.05 DF=2 MSE=9127.13

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	RIEGO
	A	553.50	9	1
	B	306.17	9	0

* * * * *
 * DISEÑO EN FRANJAS DIVIDIDAS *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: REND
 ALPHA=0.05 DF=4 MSE=3161.96

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	N
	A	553.42	b	100
	B	456.00	b	50
	C	280.08	b	0

* * * * *
 * DISEÑO EN FRANJAS DIVIDIDAS *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

MEANS

RIEGO	N	N	REND
0	0	3	253.333333
0	50	3	278.500000
0	100	3	386.666667
1	0	3	306.833333
1	50	3	633.500000
1	100	3	720.166667

PROC ANOVA

EJEMPLO NO. 7: DISEÑO CON EFECTOS ANIDADOS

SE DESEA COMPARAR ECOTIPOS PERTENECIENTES A DOS ESPECIES DE STYLOSANTHES, CON RELACION A SU PRODUCCION DE MATERIA SECA, EN UN EXPERIMENTO AGRONOMICO CONDUCTO BAJO UN DISEÑO DE BLOQUES COMPLETOS AL AZAR CON DOS BLOQUES.

DATA EA;

INPUT ESPECIES \$ 1-2 ECOTIPO 3-7 BLOQUE 9

MSECA 11-14;

CARDS;

SC	1405	1	1800
SG	191	1	772
SC	1019	1	1312
SG	136	1	986
SC	1315	1	1506
	.		
	.		
	.		
SC	1405	2	1300
SG	191	2	1016

PROC ANOVA

EJEMPLO NO. 7 (CONT.): ANALISIS DE VARIANZA PARA UN DISEÑO DE EFECTOS ANIDADOS

```
PROC ANOVA;
  CLASSES BLOQUE ESPECIE ECOTIPO;
  MODEL MSECA=BLOQUE ESPECIE ECOTIPO(ESPECIE);
  TEST H=ESPECIE      E=ECOTIPO(ESPECIE);
  MEANS ESPECIE/DUNCAN  E=ECOTIPO(ESPECIE);
  MEANS ECOTIPO/DUNCAN ;
  TITLE1 *****;
  TITLE2 * *;
  TITLE3 * DISEÑO DE EFECTOS ANIDADOS *;
  TITLE4 * *;
  TITLE5 *****;
```

* * * * *
 * DISEÑO DE EFECTOS ANIDADOS *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: MSECA

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	10	1127205.83333333	112720.583333333
ERROR	1	115705.166666667	115705.166666667
CORRECTED TOTAL	11	1242911.00000000	

MODEL F = 0.97 PR > F = 0.6651

R-SQUARE	C.V.	STD DEV	MSECA MEAN
0.906908	28.3344	340.15462170	1200.50000000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
BLOQUE	1	10561.33333333	0.09	0.8132
ESPECIE	1	453963.00000000	3.92	0.2976
ECOTIPO(ESPECIE)	8	662681.50000000	0.72	0.7288

TESTS OF HYPOTHESES USING THE ANOVA MS FOR ECOTIPO(ESPECIE)
 AS AN ERROR TERM

SOURCE	DF	ANOVA SS	F VALUE	PR > F
ESPECIE	1	453963.00000000	5.48	0.0473

```

* * * * *
*          *
*  DISEÑO DE EFECTOS ANIDADOS  *
*          *
* * * * *

```

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: MSECA
 ALPHA=0.05 DF=8 MSE=82835.2

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	ESPECIE
	A	1395.0	6	SC
	B	1006.0	6	SG

* * * * *
 * DISEÑO DE EFECTOS ANIDADOS *
 * * * * *

ANALYSIS OF VARIANCE PROCEDURE

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE: MSECA
 ALPHA=0.05 DF=1 MSE=115705
 HARMONIC MEAN OF CELL SIZES=1.11111

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

DUNCAN	GROUPING	MEAN	N	ECOTIPO
	A	1670.0	1	519
	A	1551.0	2	1405
	A	1506.0	1	1315
	A	1312.0	1	1019
	A	1296.0	1	584
	A	1016.0	1	1677
	A	1003.0	1	864
	A	974.5	2	136
	A	780.0	1	455
	A	772.0	1	191

ANALISIS DE REGRESION
BAJO MODELOS
LINEALES O LINEALIZABLES

ANALISIS DE REGRESION - CONCEPTOS -

- MODELO - REPRESENTACION MATEMATICA DE UN FENOMENO
- MODELO ESTADISTICO - REPRESENTACION MATEMATICA DE UN FENOMENO ESTOCASTICO O NO DETERMINISTICO
- MODELO LINEAL - ES UNA ECUACION QUE EXPRESA A LA VARIABLE DEPENDIENTE EN TERMINOS DE UNA COMBINACION LINEAL DE LOS PARAMETROS, SIN SER NECESARIAMENTE FUNCION LINEAL DE LAS VARIABLES INDEPENDIENTES.
- MODELO NO LINEAL - ES UNA ECUACION DONDE LA VARIABLE DEPENDIENTE NO ESTA EXPRESADA COMO FUNCION LINEAL DE LOS PARAMETROS.
- MODELO LINEALIZABLE - ES UN MODELO NO-LINEAL QUE POR UNA TRANSFORMACION PUEDE SER EXPRESADO COMO UNA FUNCION LINEAL DE LOS PARAMETROS.

ANALISIS DE REGRESION - CONCEPTOS -

- ECUACION ES LA IGUALDAD ENTRE DOS EXPRESIONES MATEMATICAS
- REGRESION ES EL AJUSTE DE UNA ECUACION A UN CONJUNTO DE DATOS
- ECUACION DE REGRESION ES LA ECUACION AJUSTADA MEDIANTE REGRESION A UN COJUNTO DE DATOS
- ELEMENTOS
 - ECUACION =
 - VARIABLE DEPENDIENTE O RESPUESTA (Observable) Y
 - VARIABLE(S) INDEPENDIENTE(S) O REGRESORES (Fijas u Observables) X
 - PARAMETROS β 's
 - DESVIACION RESPECTO DE LA ECUACION DE REGRESION (o error aleatorio no observable) e

VALOR DE LA ECUACION DE REGRESION-ESTIMADA PARA EL
SUJETO I:

$$\hat{\text{PESO}}_i = b_0 + b_1 \cdot \text{TALLA}_i$$

↑
VALOR ESTIMADO
DE LA VARIABLE
DEPENDIENTE

↑
ESTIMADORES
DE LOS
PARAMETROS

↑
VALOR DE LA
VARIABLE
INDEPENDIENTE

EJEMPLOS DE MODELOS LINEALES:

- MODELO DE REGRESION LINEAL SIMPLE:

$$Y = \beta_0 + \beta_1 X_1 + e$$

- MODELO DE REGRESION LINEAL MULTIPLE:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + e$$

- MODELO DE REGRESION POLINOMIAL DE ORDEN P:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_p X^p + e$$

- MODELO DE SUPERFICIE DE RESPUESTA DE ORDEN 2:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_1 \cdot X_2 + e$$

EJEMPLOS DE MODELOS NO LINEALES:

* EXPONENCIAL:

$$Y = \alpha \cdot \beta^x \cdot e \quad \leftarrow \text{LINEALIZABLE}$$

* EXPONENCIAL MODIFICADO:

$$Y = k + \alpha\beta^x \cdot e \quad \leftarrow \text{NO LINEALIZABLE}$$

*

GOMPERTZ:

$$Y = k \cdot (\alpha)^{e^x} \cdot e \quad \leftarrow \text{NO LINEALIZABLE}$$

MODELO LINEAL



LA VARIABLE DEPENDIENTE ES
FUNCION LINEAL DE LA(S)
INDEPENDIENTE(S)

$$Y = \beta_0 + \beta_1 X + e$$

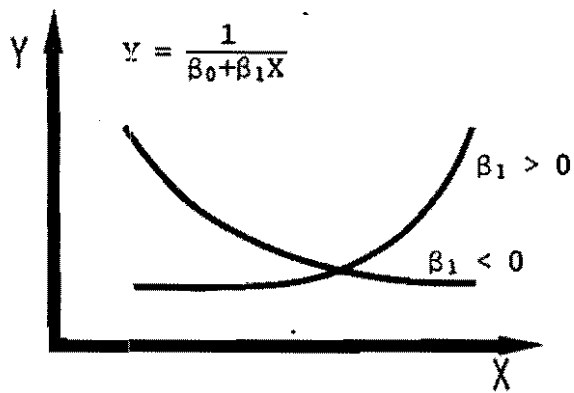
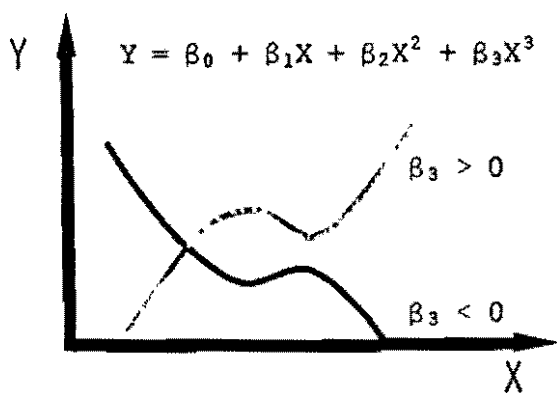
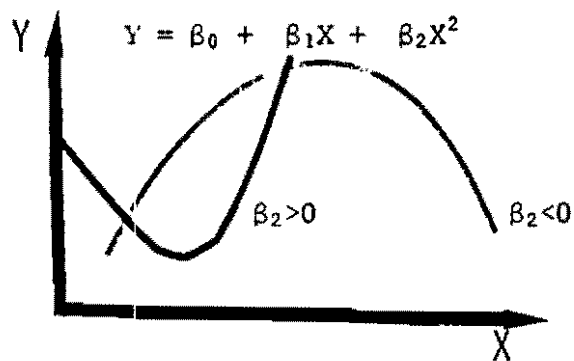
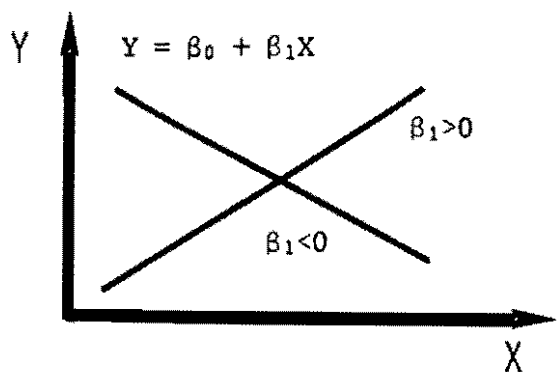
LA VARIABLE DEPENDIENTE NO
ES FUNCION LINEAL DE LA(S)
INDEPENDIENTE(S)

$$Y = \beta_0 + \beta_1 \ln X + e$$

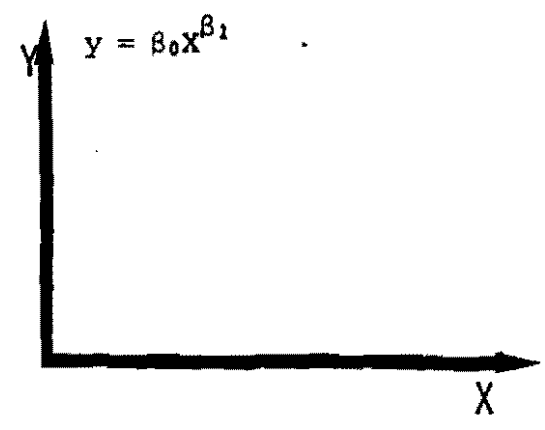
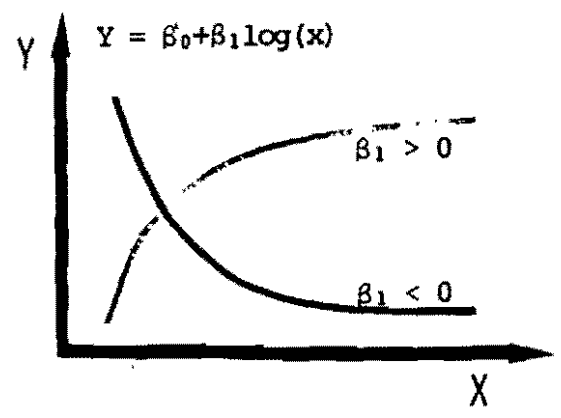
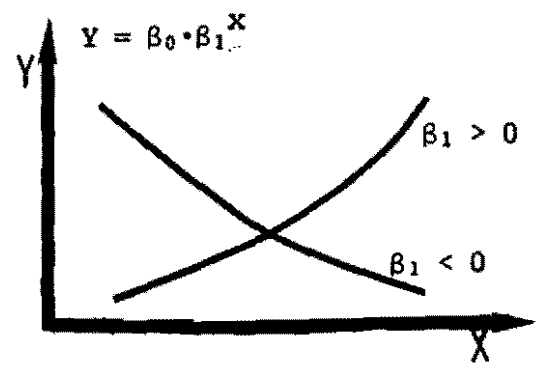
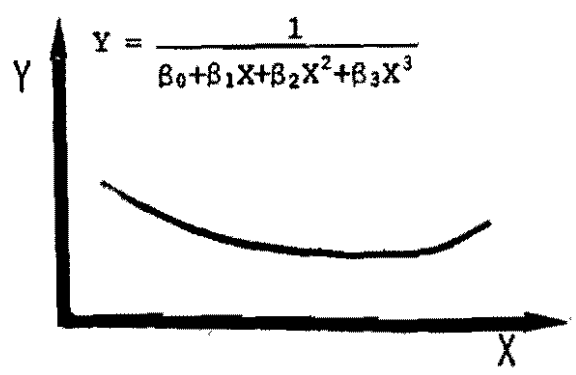
$$\ln Y = \beta_0 + \beta_1 \ln X + e$$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$$

EJEMPLOS DE ECUACIONES DE REGRESION



EJEMPLOS DE ECUACIONES DE REGRESION



EJEMPLO DE LA LINEALIZACION DE UN MODELO NO-LINEAL:

- MODELO ORIGINAL :	$Y = \alpha \cdot \beta^X \cdot e$
- TRANSFORMACION :	LOGARITMO DECIMAL
- MODELO TRANSFORMADO:	$\text{Log } Y = \text{Log}(\alpha) + \text{Log}(\beta) \cdot X + \text{Log}(e)$
- MODELO EQUIVALENTE :	$Z = \theta_0 + \theta_1 \cdot X + v$

USO DEL ANALISIS
DE REGRESION

1. DE CARACTER
PREDICTIVO
(LOS VALORES DE LOS
REGRESORES SON FIJOS)

2. DE CARACTER
EXPLORATORIO
(LOS VALORES DE LOS
REGRESORES SON ALEATORIOS)

- PROPOSITOS GENERALES DEL ANALISIS DE REGRESION:
 1. ESTIMAR LOS PARAMETROS
 2. ESTIMAR LA VARIANZA DEL ERROR
 3. "ESTIMAR" LA VARIANZA DE LOS ESTIMADORES DE LOS PARAMETROS
 4. HACER PROYECCIONES
 5. DOCIMAR HIPOTESIS ACERCA DE LOS PARAMETROS
 6. EVALUAR EL AJUSTE O FALTA DE AJUSTE

- DATOS Y SUPUESTOS BASICOS:

- . UN CONJUNTO DE N SUJETOS
- . EN CADA SUJETO SE EVALUAN TANTO LAS VARIABLES INDEPENDIENTES COMO LAS DEPENDIENTES.
- . LOS VALORES DE LAS VARIABLES INDEPENDIENTES PUEDEN SER:
 - FIJADOS DE ANTEMANO POR EL INVESTIGADOR
 - OBTENIDOS ALEATORIAMENTE A PARTIR DE UNA MUESTRA
- . LA VARIABLE DEPENDIENTE ES ALEATORIA
- . LAS DESVIACIONES RESPECTO DE LA ECUACION DE REGRESION SON ALEATORIAS Y TIENEN VARIANZA CONSTANTE Y MEDIA CERO.

CRITERIO PARA LA ESTIMACION
DE PARAMETROS

METODO DE MINIMO CUADRADOS

MINIMIZAR LOS CUADRADOS DE LAS DIFERENCIAS ENTRE LOS
VALORES OBSERVADOS (y) Y LOS VALORES PREDICHOS POR LA
REGRESION (\hat{y})

$$\text{MINIMIZAR } \sum_{i=1}^n e_i^2$$

CRITERIO PARA LA ESTIMACION DE PARAMETROS

SU INTERPRETACION:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

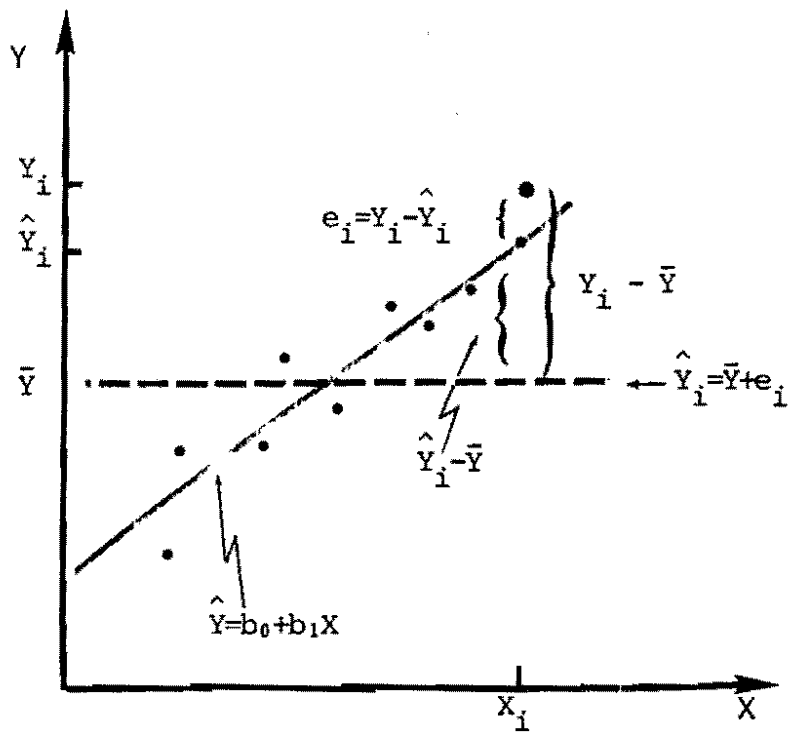
↑
VARIACION
TOTAL EN Y

↑
VARIACION EN
Y NO EXPLICADA
POR LA REGRESION

↑
VARIACION EN Y
EXPLICADA POR
LA REGRESION

$$SCT = SCE + SCR$$

SU INTERPRETACION GRAFICA EN EL
CASO DE UNA REGRESION LINEAL SIMPLE



PROCEDIMIENTOS SAS PARA ANALISIS DE REGRESION

- ➡ - REG PROPOSITOS GENERALES CON MUCHAS FACILIDADES
Y DIAGNOSTICOS
- RSQUARE OBTIENE MEDIDAS DE AJUSTE DE POSIBLES MODELOS
A CONSTRUIR
- ➡ - STEPWISE SELECCION DEL MEJOR MODELO PASO A PASO
- NLIN AJUSTA MODELOS NO LINEALES
- RSREG AJUSTA MODELO DE SUPERFICIE DE RESPUESTA DE
ORDEN CUADRATICO
- GLM MODELO LINEAL GENERAL. INCLUYE VARIABLES
CUALITATIVAS Y TERMINOS POLINOMIALES. UTILIZADO
EN ANALISIS DE REGRESION, DE VARIANZA Y DE CO-
VARIANZA.
- AUTOREG UTILIZA DATOS DE SERIES DE TIEMPO CON RESIDUOS
AUTOCORRELACIONADOS (SAS/ETS).
- SYSREG MODELOS CON SISTEMAS DE ECUACIONES SIMULTANEAS
(SAS/ETS)
- SYSNLIN MODELOS NO LINEALES CON SISTEMAS DE ECUACIONES
SIMULTANEAS (SAS/ETS).

PROC REG

OBJETIVO:

REALIZAR UN ANALISIS DE REGRESION AJUSTANDO, POR EL METODO DE MINIMOS CUADRADOS, UN MODELO LINEAL A UN CONJUNTO DE DATOS.

PERMITE:

- AJUSTAR VARIOS MODELOS SIMULTANEAMENTE
- OBTENER UN ARCHIVO SAS ESPECIAL CON CORRELACIONES O SUMAS DE CUADRADOS Y PRODUCTOS CRUZADOS COMO ENTRADA O SALIDA.
- IMPRIMIR Y ALMACENAR EN UN ARCHIVO SAS LOS ESTIMADOS DE LOS PARAMETROS, LOS VALORES PREDICHOS, RESIDUOS E INTERVALOS DE CONFIANZA.
- IMPRIMIR UNA ESTADISTICA DE INFLUENCIA
- DOCIMAR HIPOTESIS SOBRE LOS PARAMETROS
- DIAGNOSTICAR COLINEALIDAD ENTRE REGRESORES

PROC REG

ESPECIFICACIONES:

```
PROC REG opciones;  
MODEL var.dep=regresores/opciones;  
VAR variable;  
FREQ variable;  
WEIGHT variable;  
ID variable;  
OUTPUT OUT= nombre de archivo  
      palabra clave = nombres ...;  
RESTRICT ecuación lineal;  
TEST ecuación lineal;  
MTEST ecuación lineal;  
BY variables;
```

PROC REG

OPCIONES DE LA PROPOSICION "PROC REG":

OUTEST = nombre	GUARDA LAS ESTIMACIONES DE LOS PARAMETROS EN EL ARCHIVO SAS QUE SE ESPEFIQUE.
OUTSSCP = nombre	GUARDA LA MATRIZ DE SUMAS DE CUADRADOS Y PRODUCTOS CRUZADOS EN EL ARCHIVO ESPECIAL SAS QUE SE ESPECIFIQUE.
ALL	IMPRIME TODAS LAS OPCIONES DISPONIBLES EN TODAS LAS PROPOSICIONES MODEL

ADICIONALMENTE:

DATA=	NOPRINT	SIMPLE
COVOUT	USSCP	EPSILON=N

PROC REG

PROPOSITION MODEL:

MODEL VAR.DEPENDIENTES = REGRESORES/OPCIONES;

OPCIONES:

NOPRINT	NOINT	ALL
SS1	SS2	STB
TOL	VIF	COVB
CORRB	SEQB	COLLIN
COLLINNOINT	P	R
CLM	CLI	DW
INFLUENCE	PARTIAL	XPX I

PROC REG

PALABRAS CLAVES PARA LA PROPOSICION "OUTPUT"

PREDICTED =	RESIDUAL =
L95M =	U95M =
L95 =	U95 =
STDP =	STDR =
STUDENT =	COORD =
H =	PRESS =
RSTUDENT =	DIFFITS
COVRATIO =	

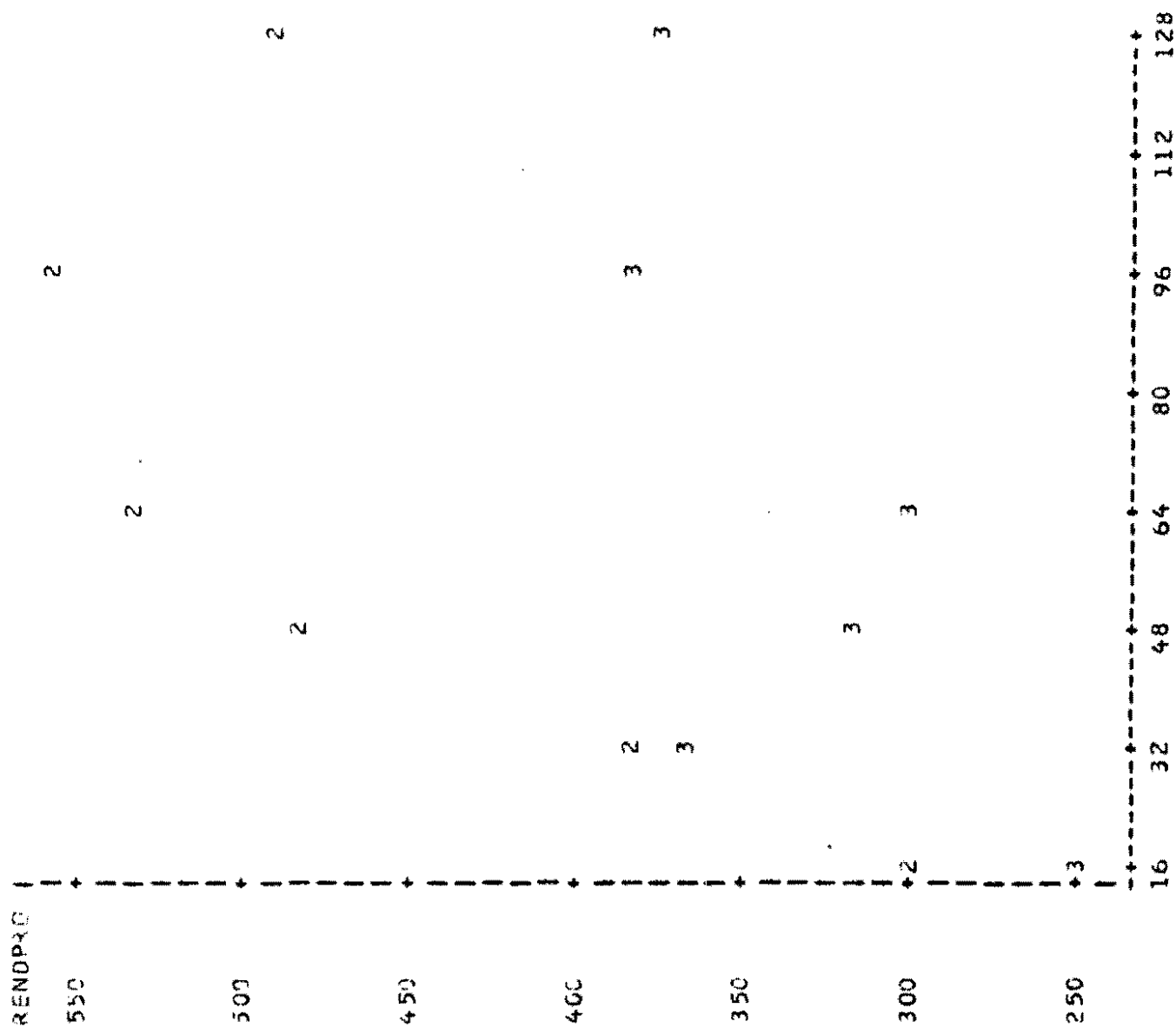
PROC REG

EJEMPLO NO. 1:

```
PROC SORT DATA = A ;
    BY VARIEDAD DENSIDAD ;
PROC MEANS NOPRINT DATA=A ;
    BY VARIEDAD DENSIDAD ;
    VAR REND ;
    OUTPUT OUT = B
        MEAN= RENDPRO ;
DATA B1 ;
    SET B ;
    DENC = (DENSIDAD<5)*16*DENSIDAD +
        (DENSIDAD=5)*96 +
        (DENSIDAD=6)*128 ;
    DEN2 = DENS**2 ;
PROC PLOT DATA = B1 ;
    PLOT RENDPRO*DENS=VARIEDAD /
        HPOS=50 VPOS=40 ;
    TITLE PROMEDIO DE RENDIMIENTO VS DENSIDAD ;
```

NUMEROS DE REPRODUCCIONES DE GENOTIPOS

PLOT OF REPRO-DENS SYMBOL IS VALUE OF VARIEDAD



GENE

PROC REG

EJEMPLO NO. 1 (CONT.)

```
PROC REG DATA=B1 ;
  BY VARIEDAD ;
  RECTA : MODEL RENDPRO = DENS / P CLM ;
           OUTPUT OUT=B2 PREDICTED=P_REND
                RESIDUAL =E_REND
                U95M      =S_REND
                L95M      =I_REND ;

  ID DENS ;
  TITLE REGRESION LINEAL SIMPLE ;
PROC PLOT DATA=B2 ;
  BY VARIEDAD ;
  PLOT RENDPRO*DENS = 'Q' P_REND*DENS='P' /
       OVERLAY HPOS=50 VPOS=40 ;
  PLOT E_REND*DENS='Q' / VREF=0 HPOS=50 VPOS=40;
  PLOT P_REND*DENS='Q' S_REND*DENS='S'
       I_REND*DENS='I' / OVERLAY HPOS=50 VPOS=40;
```

REGRESSION LINEAL SIMPLE
VARIEDAD=2

MODEL: RECTA
DEP VARIABLE: RENDPRO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	24699.675	24699.675	4.375	0.1046
ERRUR	4	22580.143	5645.036		
C TOTAL	5	47279.818			
ROOT MSE		75.133453	R-SQUARE	0.5224	
DEP MEAN		458.267	ADJ R-SQ	0.4030	
C.V.		16.39514			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T
INTERCEP	1	353.455	59.977718	5.843	0.0043
DENS	1	1.684559	0.805330	2.092	0.1046

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% MEAN	UPPER95% MEAN	RESIDUAL
1	16	302.625	377.408	49.347	240.401	514.415	-74.783
2	32	383.350	404.361	40.062	293.132	515.589	-21.011
3	48	480.575	431.314	33.270	338.944	523.684	49.261
4	64	536.200	458.267	30.673	373.106	543.428	77.933
5	96	557.925	512.173	40.052	400.944	623.401	45.752
6	128	488.925	566.078	59.978	399.556	732.601	-77.153

SUM OF RESIDUALS 7.38964E-13
SUM OF SQUARED RESIDUALS 22580.14

INTERPRETACION

- COEFICIENTE DE DETERMINACION

MEDIDA DE LA VARIABILIDAD EXPLICADA POR LA
ECUACION DE REGRESION

$$R^2 = \frac{SCR}{SCT}$$

- ESTIMADOR DE LA VARIANZA DEL ERROR

$$s^2 = \frac{SCE}{n-p-1}$$

INTERPRETACION

- DOXIMA DE HIPOTESIS DE LA ECUACION DE REGRESION:

. HIPOTESIS NULA

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

. HIPOTESIS ALTERNATIVA

H_1 : Algún β es no nulo

. ESTADISTICO

$$F_c = \frac{\frac{SCR}{p}}{\frac{SCE}{n-p-1}}$$

. DISTRIBUCION DEL ESTADISTICO BAJO H_0 CIERTA:

$$F_c \sim F_{(p, n-p-1)}$$

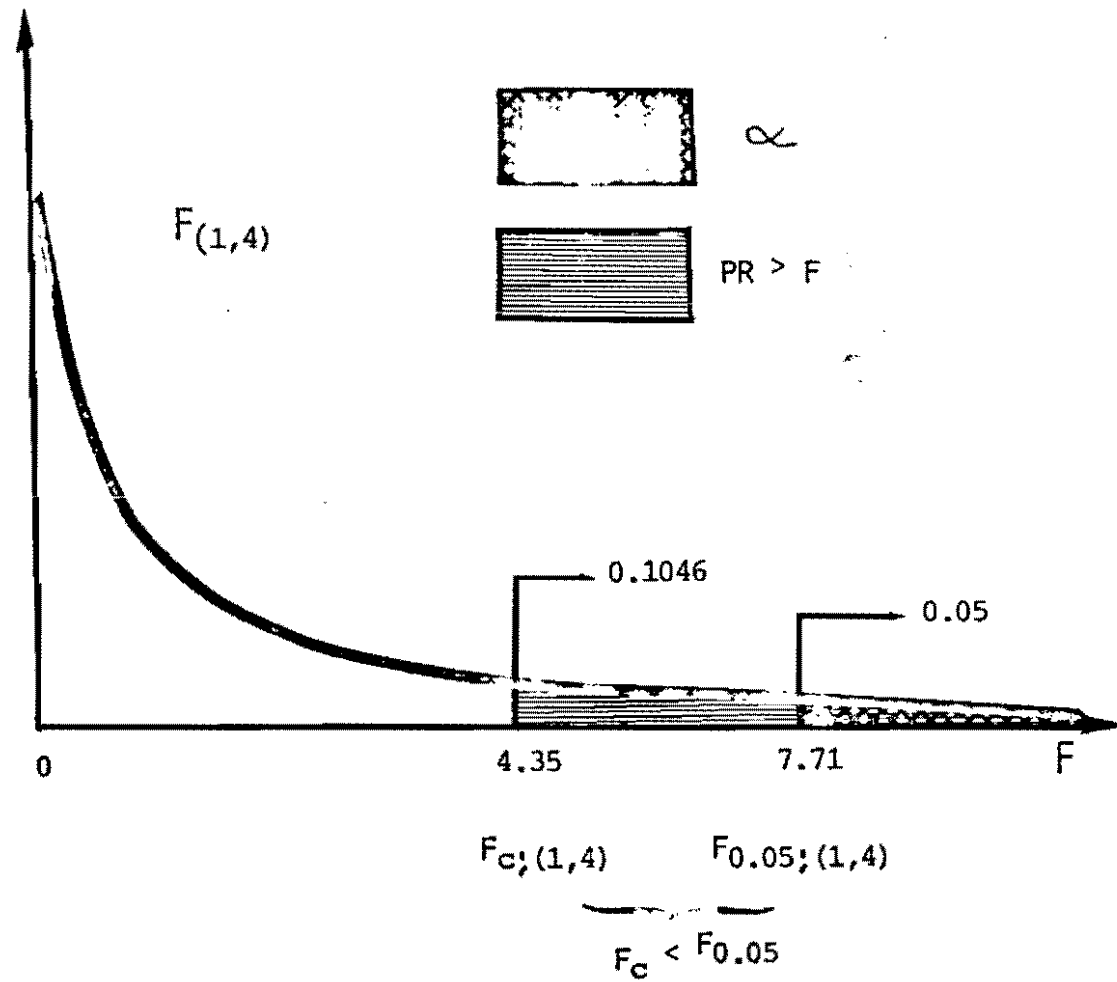
. DECISION AL NIVEL α : RECHAZAR H_0 SI $F_c > F_{\alpha}^2(p, n-p-1)$

MODEL: RECTA
DEP VARIABLE: RENDPRO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	24699.675	24699.675	4.375	0.1046
ERROR	4	22580.143	5645.036		
C TOTAL	5	47279.818			
ROOT MSE		75.133453	R-SQUARE	0.5224	
DEP MEAN		458.267	ADJ R-SQ	0.4030	

INTERPRETACION: PROBABILIDAD DE F

DECISION: NO RECHAZAR H_0



INTERPRETACION

- DOCIMAS DE HIPOTESIS ACERCA DE LOS PARAMETROS

• HIPOTESIS NULA

$$H_0: \beta_k = 0 \quad k=1,2,\dots,p$$

• HIPOTESIS ALTERNATIVA

$$H_1: \beta_k \neq 0$$

• ESTADISTICO

$$T_c = \frac{b_k}{s_b} = \frac{\text{Estimador de } \beta_k}{\text{Err. estand. del Estimador}}$$

• DISTRIBUCION DEL ESTADISTICO BAJO H_0 CIERTA:

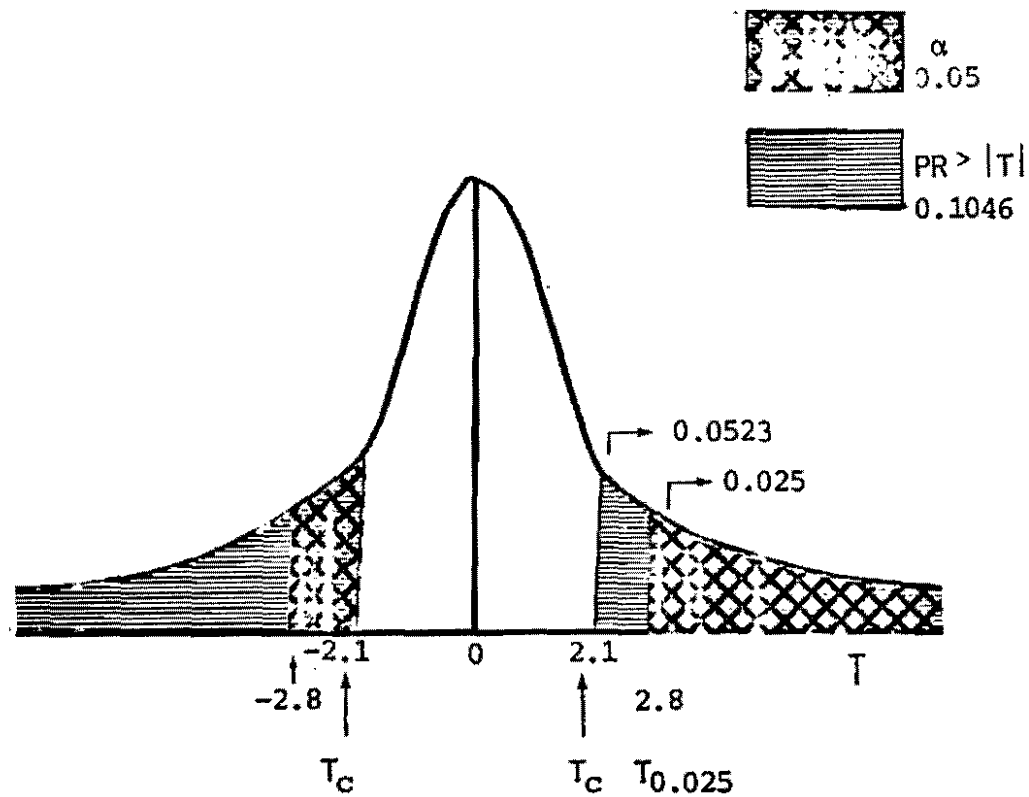
$$T_c \sim T(n-p-1)$$

• DECISION AL NIVEL α : RECHARZAR H_0 SI $|T_c| > T_{\alpha/2; (n-p-1)}$

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T
INTERCEP	1	350.455	59.977718	5.843	0.0043
DENS	1	1.684559	0.805330	2.092	0.1046

INTERPRETACION DE LA PROBABILIDAD DE T

DECISION: NO RECHAZAR H_0



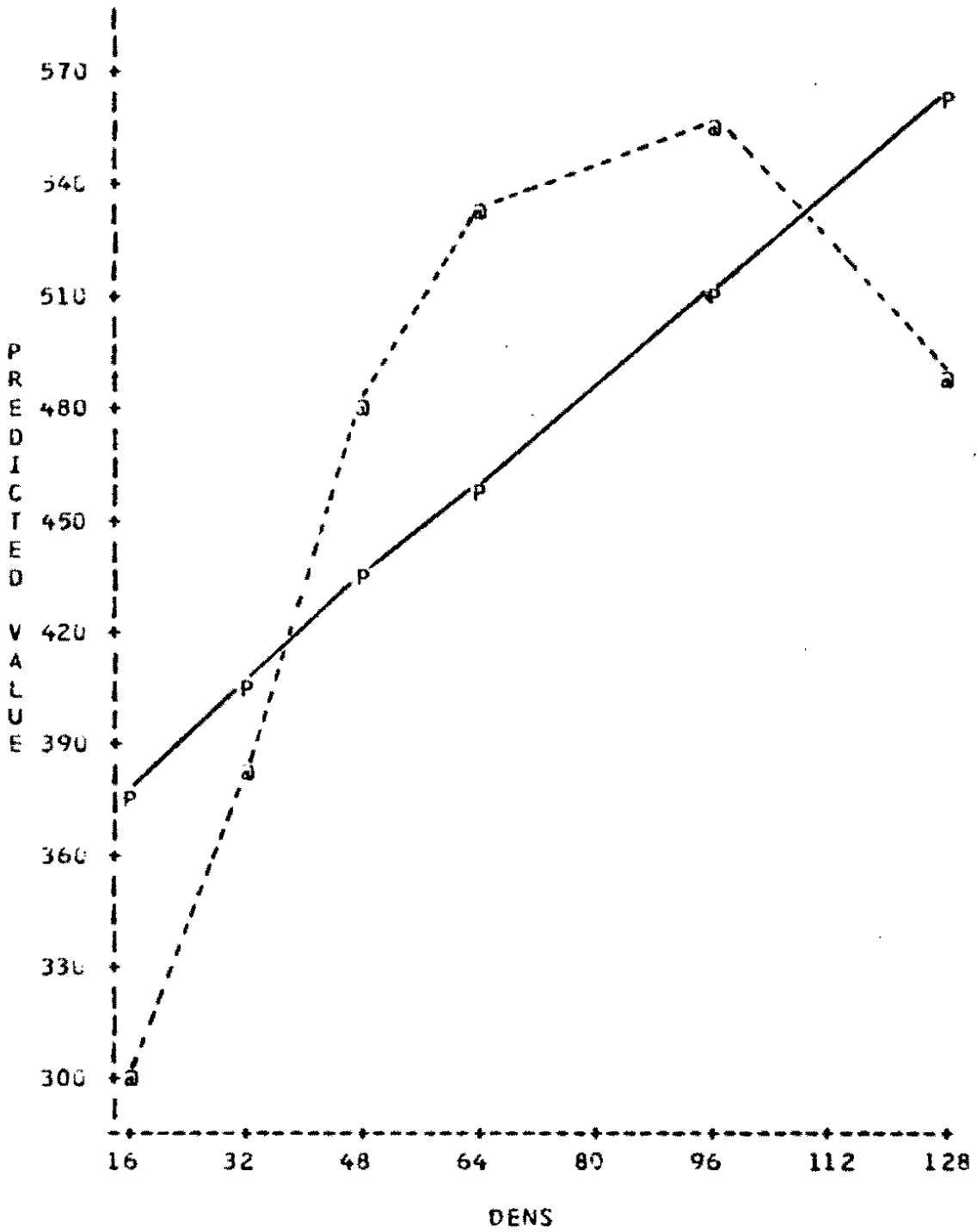
$$|T_c| < t_{0.025}$$

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% MEAN	UPPER95% MEAN	RESIDUAL
1	16	302.625	377.408	49.347	240.401	514.415	-74.783
2	32	383.350	404.361	40.062	293.132	515.589	-21.011
3	48	480.575	431.314	33.270	338.944	523.684	49.261
4	64	536.200	458.267	30.673	373.106	543.428	77.933
5	96	557.925	512.173	40.762	400.944	623.401	45.752
6	128	488.925	566.078	59.978	399.556	732.601	-77.153

REGRESION LINEAL SIMPLE
VARIEDAD=2

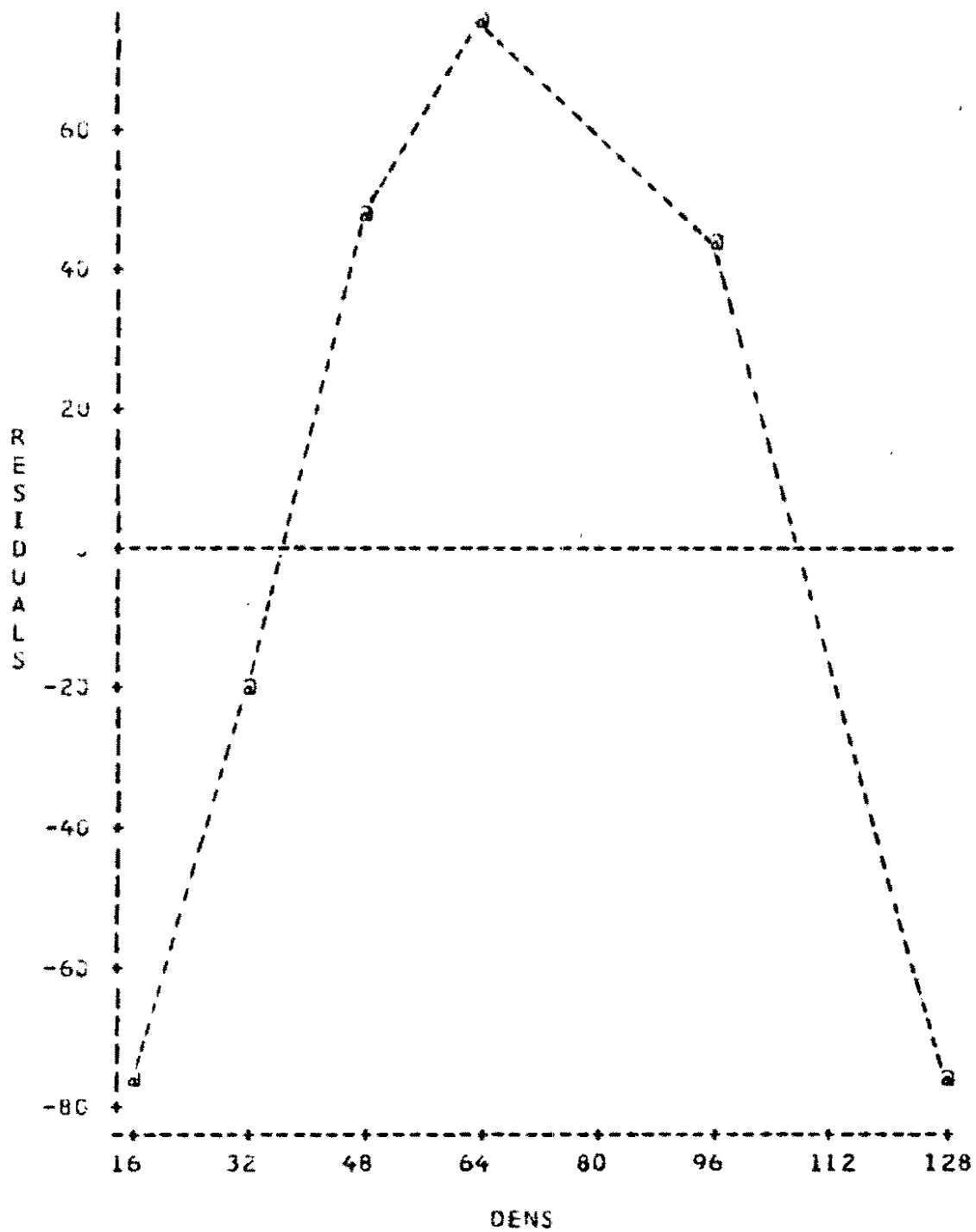
PLOT OF RENDPRO.DENS
PLOT OF P_REND^DENS

SYMBOL USED IS a
SYMBOL USED IS P



REGRESION LINEAL SIMPLE
VARIEDAD=2

PLOT OF E_REND*DENS SYMBOL USED IS a



ESTIMACION POR INTERVALO DE CONFIANZA $(1-\alpha)$ EN REGRESION LINEAL SIMPLE:

- PARA LA MEDIA DE LA POBLACION Y DADO X.

$$\bar{Y}_X \pm T_{\frac{\alpha}{2}} \cdot \sqrt{\text{CME} \cdot \left[\frac{1}{n} + \frac{(x - \bar{X})^2}{(n-1) \cdot S_x^2} \right]}$$

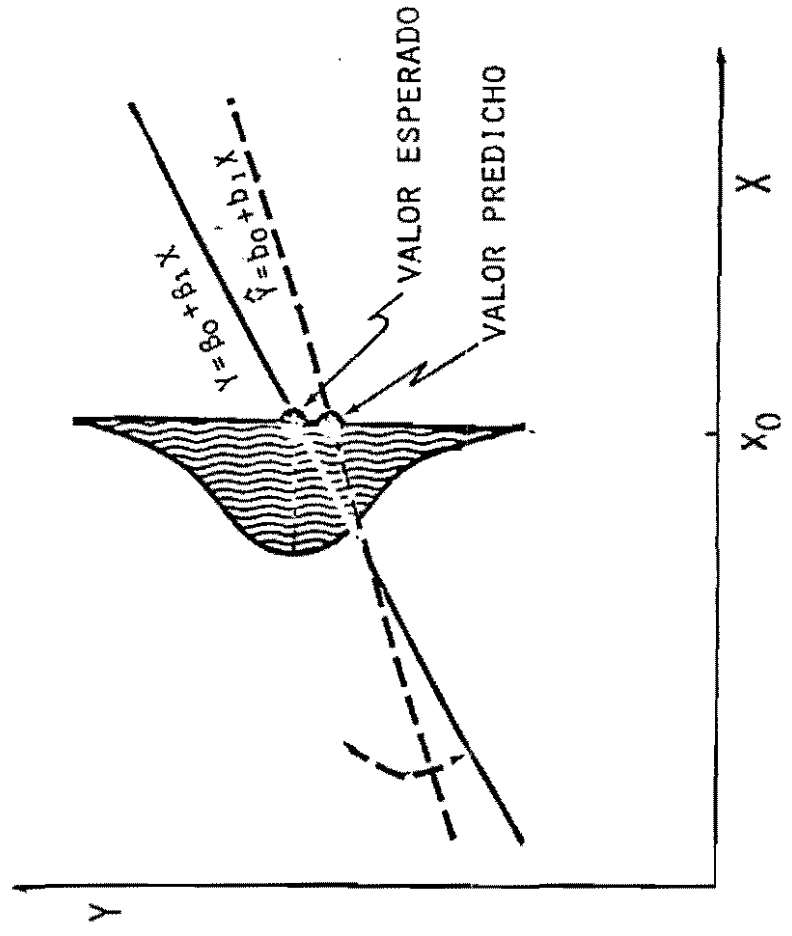
- PARA UN VALOR Y DADO X.

$$\bar{Y}_X \pm T_{\frac{\alpha}{2}} \cdot \sqrt{\text{CME} \cdot \left[1 + \frac{1}{n} + \frac{(x - \bar{X})^2}{(n-1) \cdot S_x^2} \right]}$$

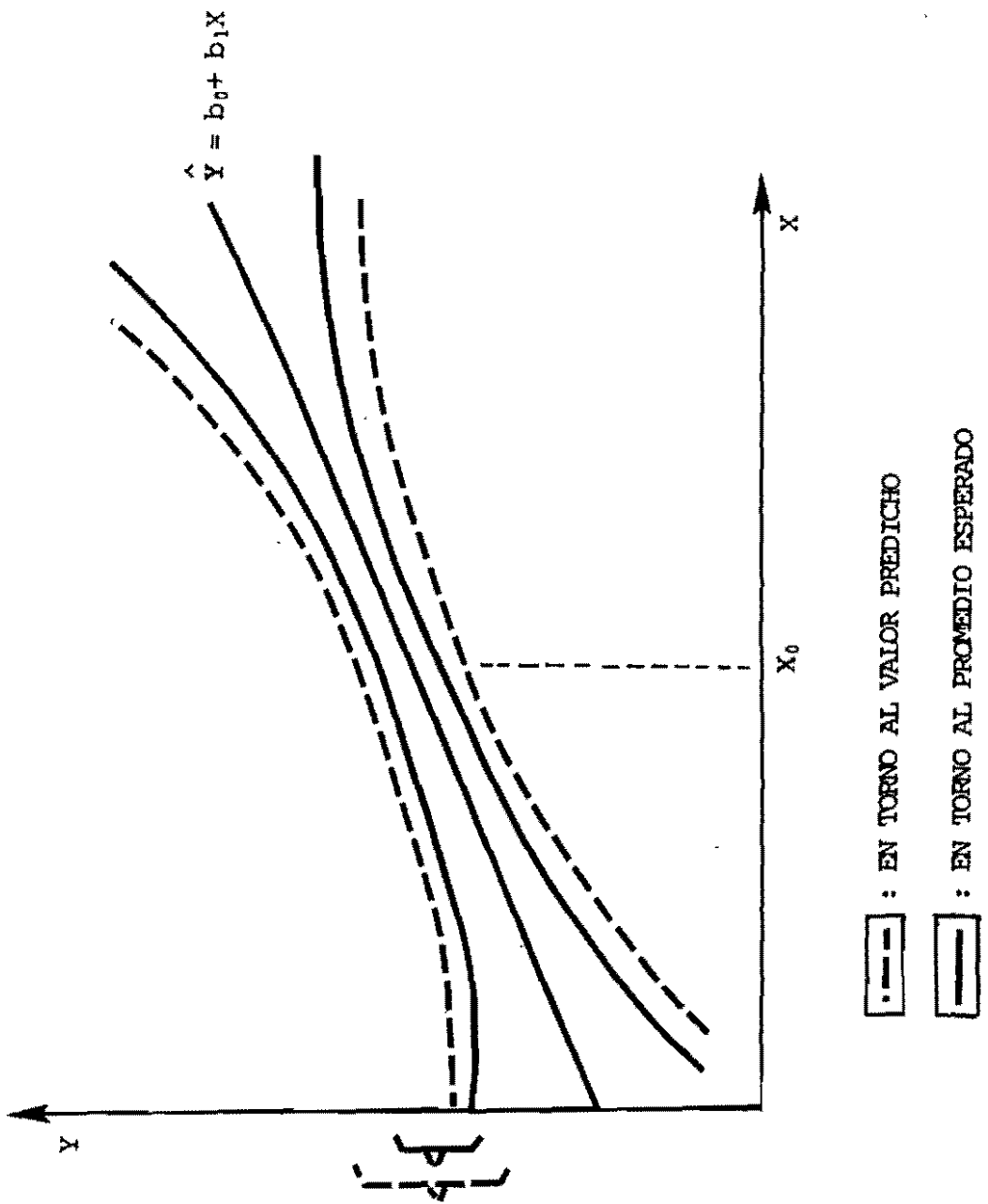
NOTA: - SUPUESTO ADICIONAL: PARA CADA VALOR X, FIJO, Y TIENE UNA DISTRIBUCION NORMAL.

- SE PUEDEN TAMBIEN CONSTRUIR INTERVALOS DE CONFIANZA PARA LA ESTIMACION DE CADA UNO DE LOS PARAMETROS Y DE LA VARIANZA DE Y CONDICIONADA POR X.

DISTRIBUCION DE Y PARA UN X FIJO

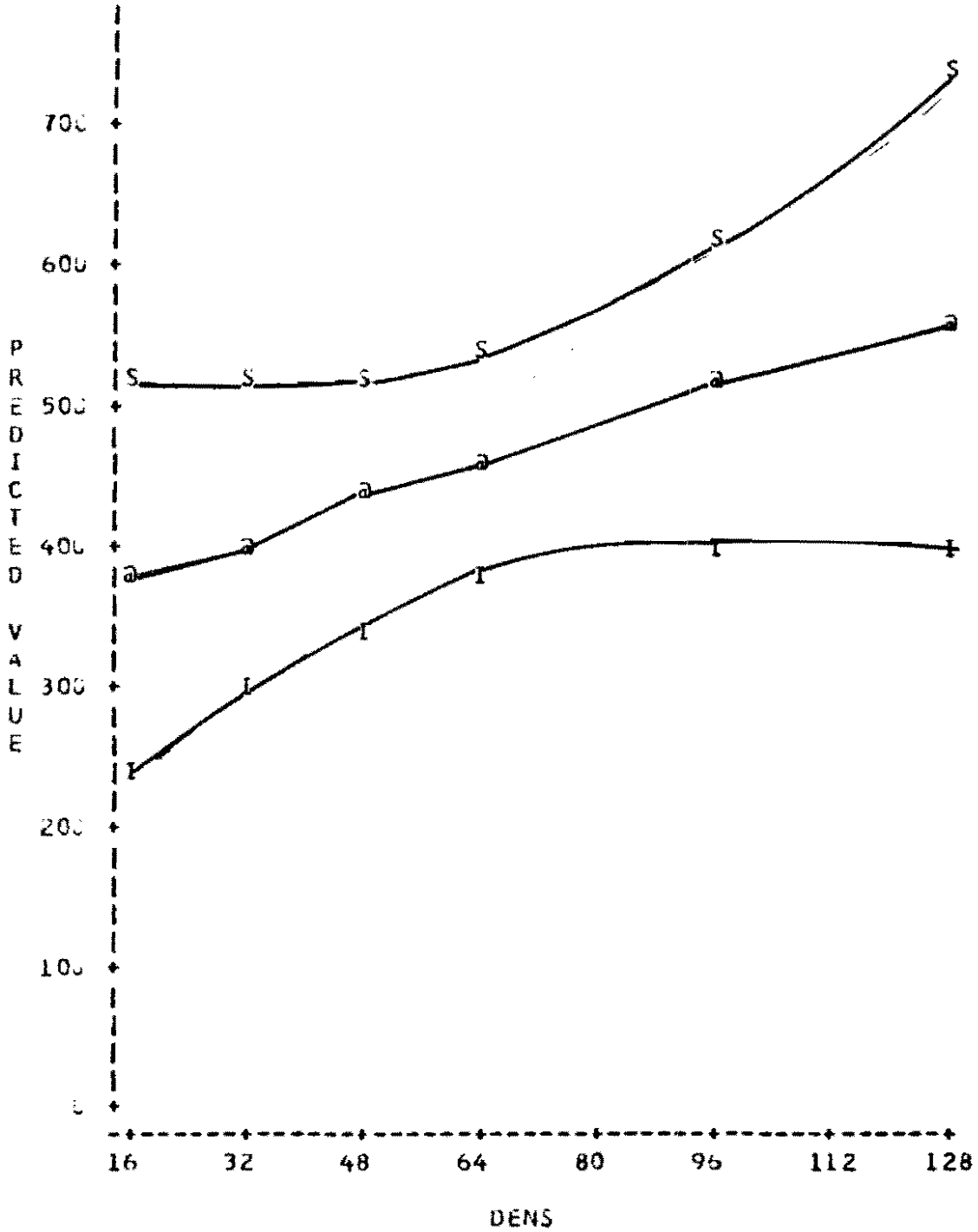


LOS DOS TIPOS DE INTERVALO DE CONFIANZA
QUE PRODUCE PROC REG



REGRESION LINEAL SIMPLE
 VARIEDAD=2

PLOT OF P_REND*DEN\$ SYMBOL USED IS @
 PLOT OF S_REND*DEN\$ SYMBOL USED IS S
 PLOT OF I_REND*DEN\$ SYMBOL USED IS I



PROC REG

EJEMPLO NO. 2:

```
PROC REG DATA=B1 ;
  BY VARIEDAD ;
  CUADR : MODEL RENDPRO = DENS DEN2 / P CLM;
          OUTPUT OUT=B3 P=P_REND      R=E_REND
                    U95M=S_REND      L95M=I_REND;

  ID DENS ;
  TITLE REGRESION CUADRATICA ;
PROC PLOT DATA=B3 ;
  BY VARIEDAD ;
  PLOT RENDPRO*DENS = 'Q' P_REND*DENS='P' /
        OVERLAY HPOS=50 VPOS=40 ;
  PLOT E_REND*DENS='Q' / VREF=0 HPOS=50 VPOS=40 ;
  PLOT P_REND*DENS='Q' S_REND*DENS='S'
        I_REND*DENS='I' / OVERLAY HPOS=50 VPOS=40;
```

REGRESION CUADRATICA
 VARIEDAD=2

MODEL: CUADR
 DEP VARIABLE: RENDPRO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	2	46931.652	23465.826	202.195	0.0006
ERROR	3	348.166	116.055		
C TOTAL	5	47279.818			
ROOT MSE		10.772902	R-SQUARE	0.9926	
DEP MEAN		458.267	ADJ R-SQ	0.9877	
C.V.		2.350793			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HC: PARAMETER=0	PROB > T
INTERCEP	1	166.351	15.839578	10.502	0.0018
DENS	1	8.839347	0.529630	16.688	0.0005
DEN2	1	-0.049363	0.003566553	-13.841	0.0008

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% MEAN	UPPER95% MEAN	RESIDUAL
1	16	302.625	295.143	9.241	265.735	324.552	7.482
2	32	383.350	398.662	5.759	380.334	416.990	-15.312
3	48	480.575	476.906	5.797	458.457	495.356	3.669
4	64	536.200	529.876	6.791	508.266	551.487	6.324
5	96	557.925	559.995	6.703	538.662	581.328	-2.070
6	128	488.925	489.017	10.245	456.413	521.622	-0.092338

SUM OF RESIDUALS 1.08002E-12
 SUM OF SQUARED RESIDUALS 348.1663

REGRESION CUADRATICA
 VARIEDAD=2

MODEL: CUADR
 DEP VARIABLE: RENDPRO

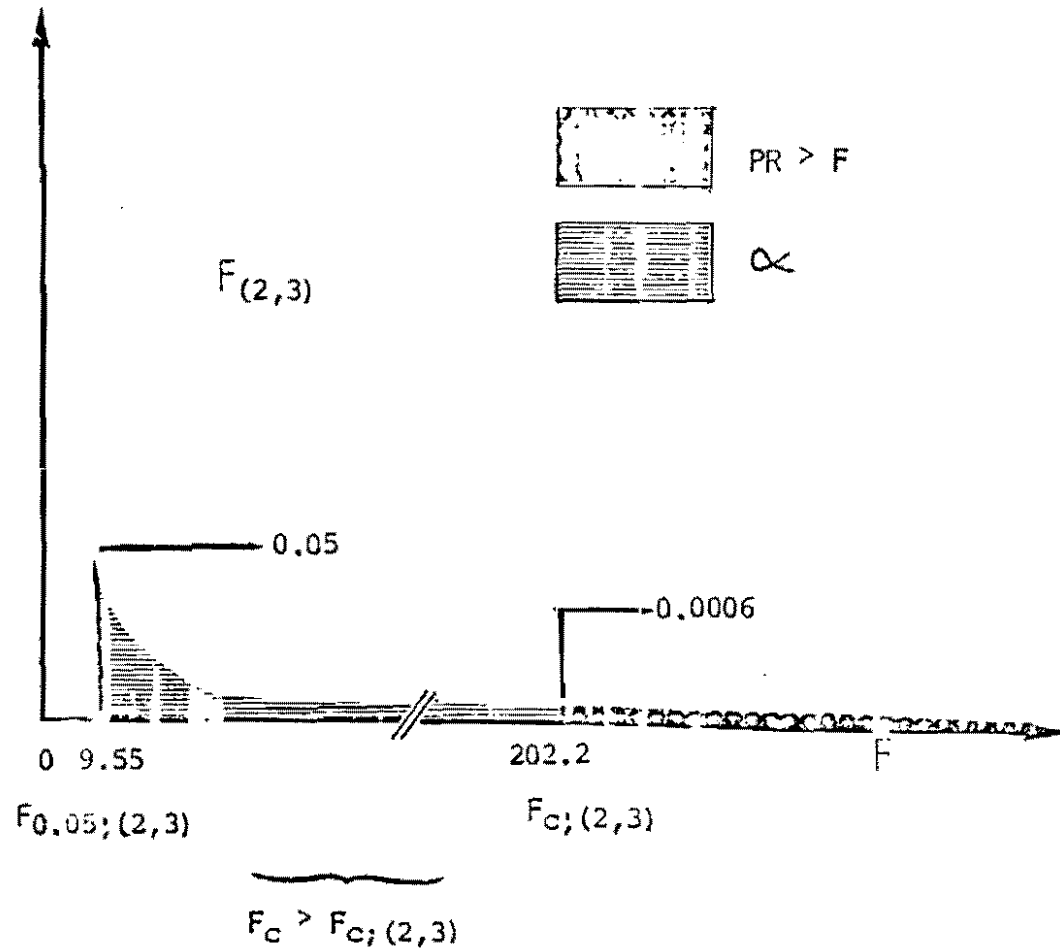
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	2	46931.652	23465.826	202.195	0.0006
ERROR	3	348.166	116.055		
C TOTAL	5	47279.818			
ROOT MSE		10.772902	R-SQUARE	0.9926	
DEP MEAN		458.267	ADJ R-SQ	0.9877	
C.V.		2.350793			

-243-

INTERPRETACION DE LA PROBABILIDAD F

DECISION: RECHAZAR H_0

-111-



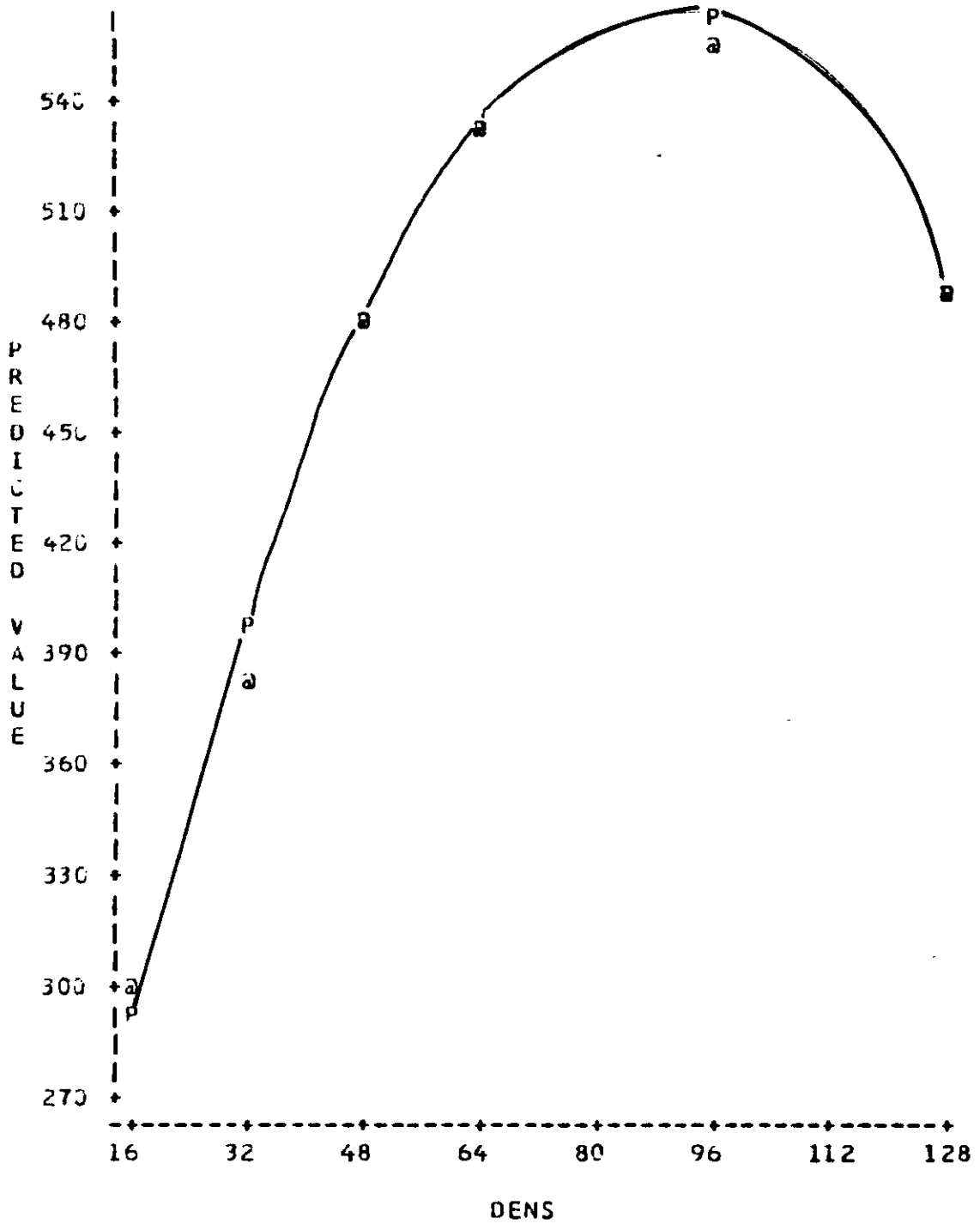
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=C	PROB > T
INTERCEP	1	166.351	15.839578	10.502	0.0018
DENS	1	8.839347	0.529680	16.688	0.0005
DENZ	1	-0.049363	0.003566553	-13.841	0.0008

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% MEAN	UPPER95% MEAN	RESIDUAL
1	16	302.625	295.143	9.241	265.735	324.552	7.482
2	32	383.350	398.662	5.759	380.334	416.990	-15.312
3	48	480.575	476.906	5.797	458.457	495.356	3.669
4	64	536.200	529.876	6.791	508.266	551.487	6.324
5	96	557.925	559.995	6.703	538.662	581.328	-2.070
6	128	488.925	489.017	10.245	456.413	521.622	-0.092338

REGRESION CUADRATICA
VARIEDAD=2

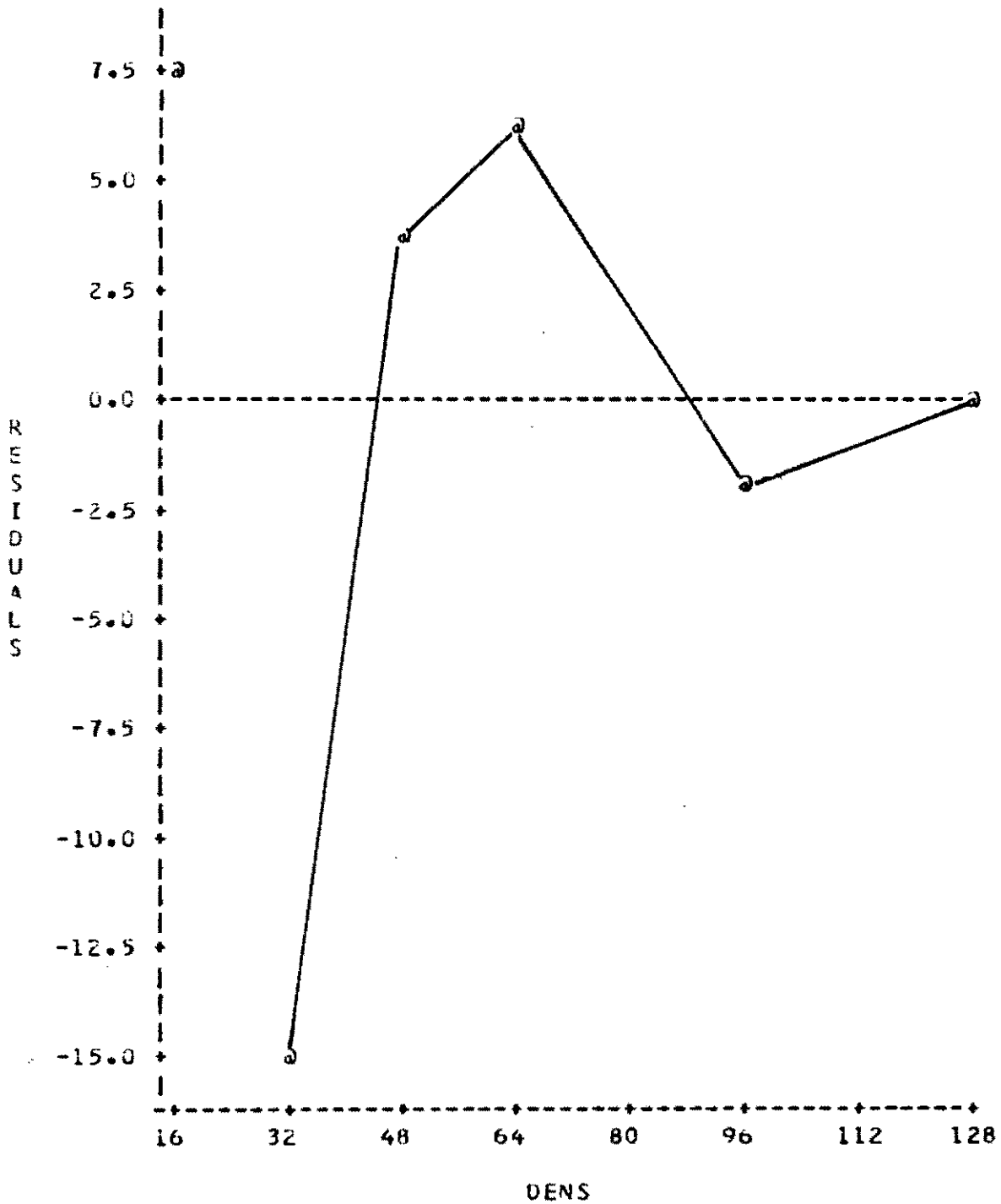
PLOT OF RENDPRO/DENS
PLOT OF P_REND/DENS

SYMBOL USED IS a
SYMBOL USED IS P



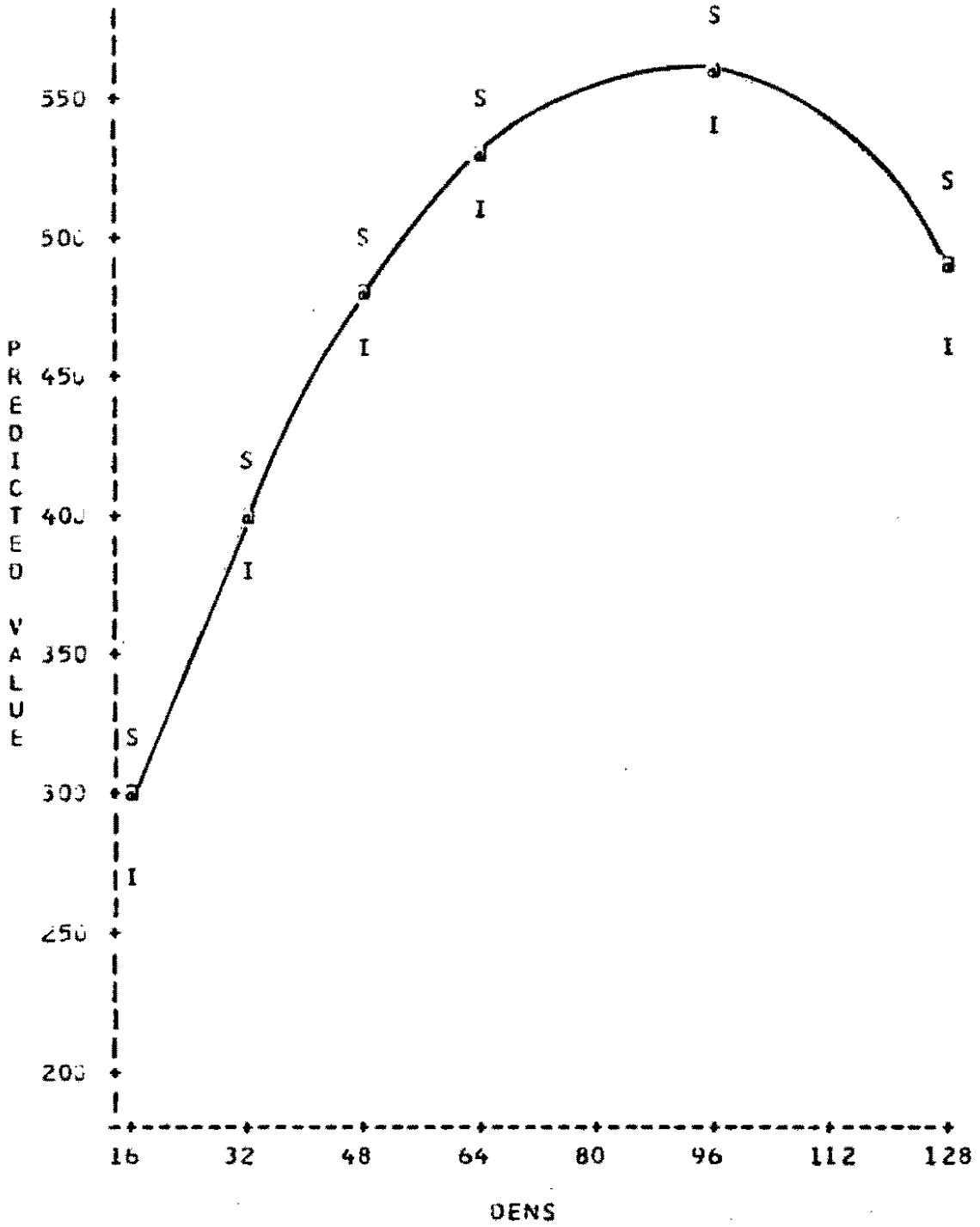
REGRESION CUADRATICA
VARIEDAD=2

PLOT OF E_REND=DENS SYMBOL USED IS @



REGRESION CUADRATICA
 VARIEDAD=2

PLOT OF P_REND#DENS SYMBOL USED IS @
 PLOT OF S_REND#DENS SYMBOL USED IS S
 PLOT OF I_REND#DENS SYMBOL USED IS I



PROC REG

EJEMPLO NO. 3:

```
DATA B ;
  SET A ;
  IF VARIEDAD=3 THEN DELETE ;
PROC PLOT DATA=B ;
  PLOT REND*DENS='a' / VPOS=40 HPOS=50;
  TITLE .RENDIMIENTO VS. DENSIDAD ;
PROC SORT DATA=B ;
  BY DENS ;
PROC MEANS NOPRINT DATA=B ;
  BY DENS ;
  VAR REND ;
  OUTPUT OUT=C VAR=S2 ;
DATA D ;
  MERGE B C ;
  BY DENS ;
  W=1/S2 ;
  DEN2=DENS*2 ;
```


PROC REG

EJEMPLO NO. 3 (CONT.)

```
PROC REG DATA=D ;
  M1 : MODEL RFND = DENS ;
        OUTPUT OUT=D1 P=P1_REND
                R=E1_REND ;
        WEIGHT W ;
  M2 : MODEL REND = DENS DEN2 ;
        OUTPUT OUT=D2 P=P2_REND
                R=E2_REND ;
        WEIGHT W ;
  TITLE ANALISIS DE REGRESION ;
PROC PLOT DATA=D1 ;
  PLOT RFND*DENS = 'Q' P1_REND*DENS='P' /
        OVERLAY HPOS=50 VPOS=40 ;
  PLOT E1_REND*DENS='*' /
        VREF=0 HPOS=50 VPOS=40;
PROC PLOT DATA=D2 ;
  PLOT REND*DENS = '*' P2_REND*DENS='P' /
        OVERLAY HPOS=50 VPOS=40;
  PLOT E2_REND*DENS='*' /
        VREF=0 HPOS=50 VPOS=40 ;
```

ANALISIS DE REGRESION

MODEL: M1
 DEP VARIABLE: KEND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	19.607336	19.607336	16.177	0.0006
ERROR	22	26.665343	1.212061		
C TOTAL	23	46.272679			

ROOT MSE	1.100936	R-SQUARE	0.4237
DEP MEAN	468.149	ADJ R-SQ	0.3975
C.V.	0.2351682		

-252-

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T
INTERCEPT	1	316.220	42.367043	7.464	0.0001
DENS	1	2.710141	0.673822	4.022	0.0006

ANALISIS DE REGRESION

MODEL: M2
 DEP VARIABLE: REND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	2	28.081335	14.040667	16.208	0.0001
ERROR	21	18.191344	0.866254		
C TOTAL	23	46.272679			

ROOT MSE	0.930728	R-SQUARE	0.6069
DEP MEAN	468.149	ADJ R-SQ	0.5694
C.V.	0.1968104		

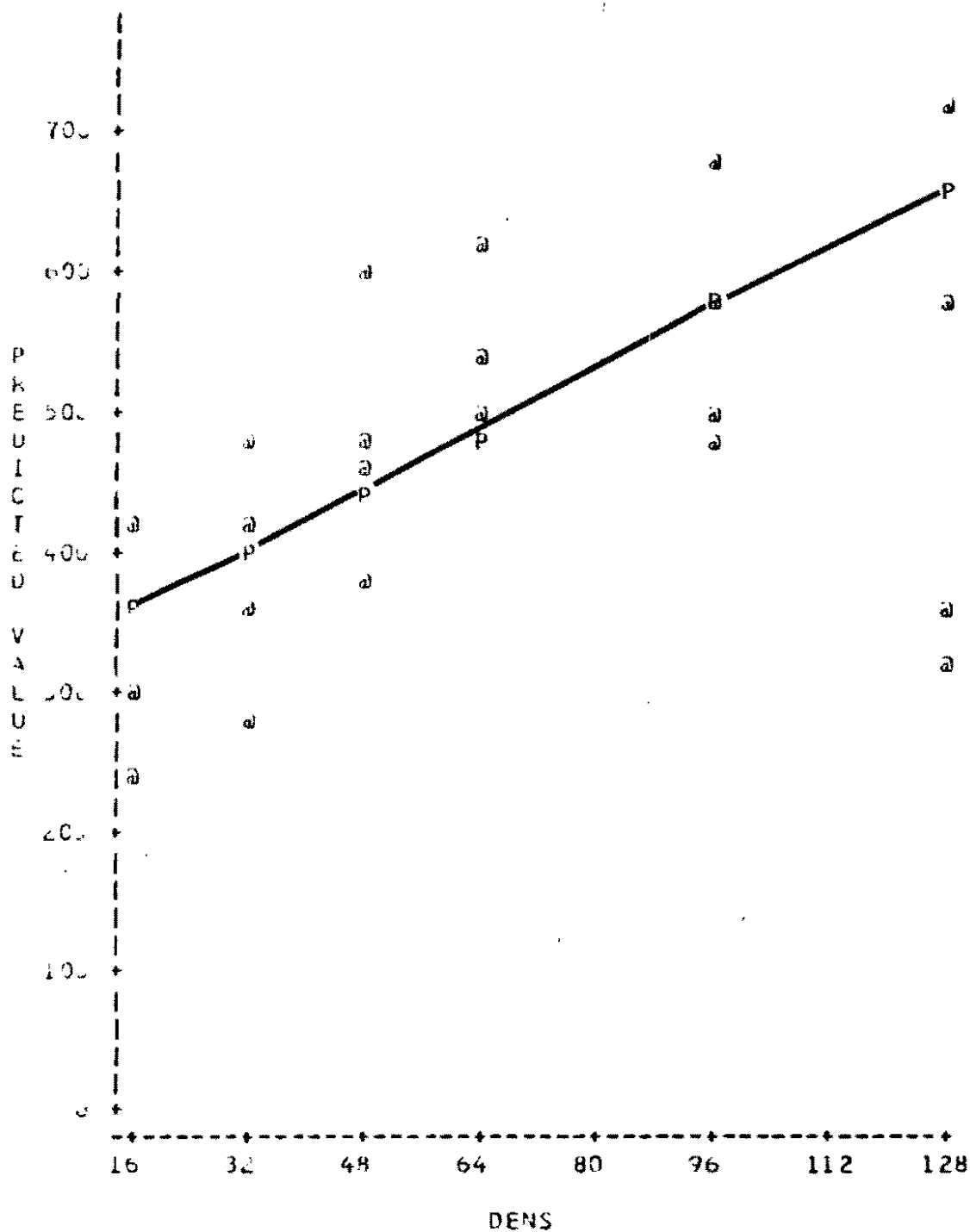
-253-

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HC: PARAMETER=0	PROB > T
INTERCEP	1	164.896	60.197163	2.739	0.0123
DENS	1	8.950024	2.074788	4.314	0.0003
DEN2	1	-0.050205	0.016052	-3.128	0.0051

GRAFICAS

PLOT OF REND+DENS
 PLOT OF P1_REND+DENS

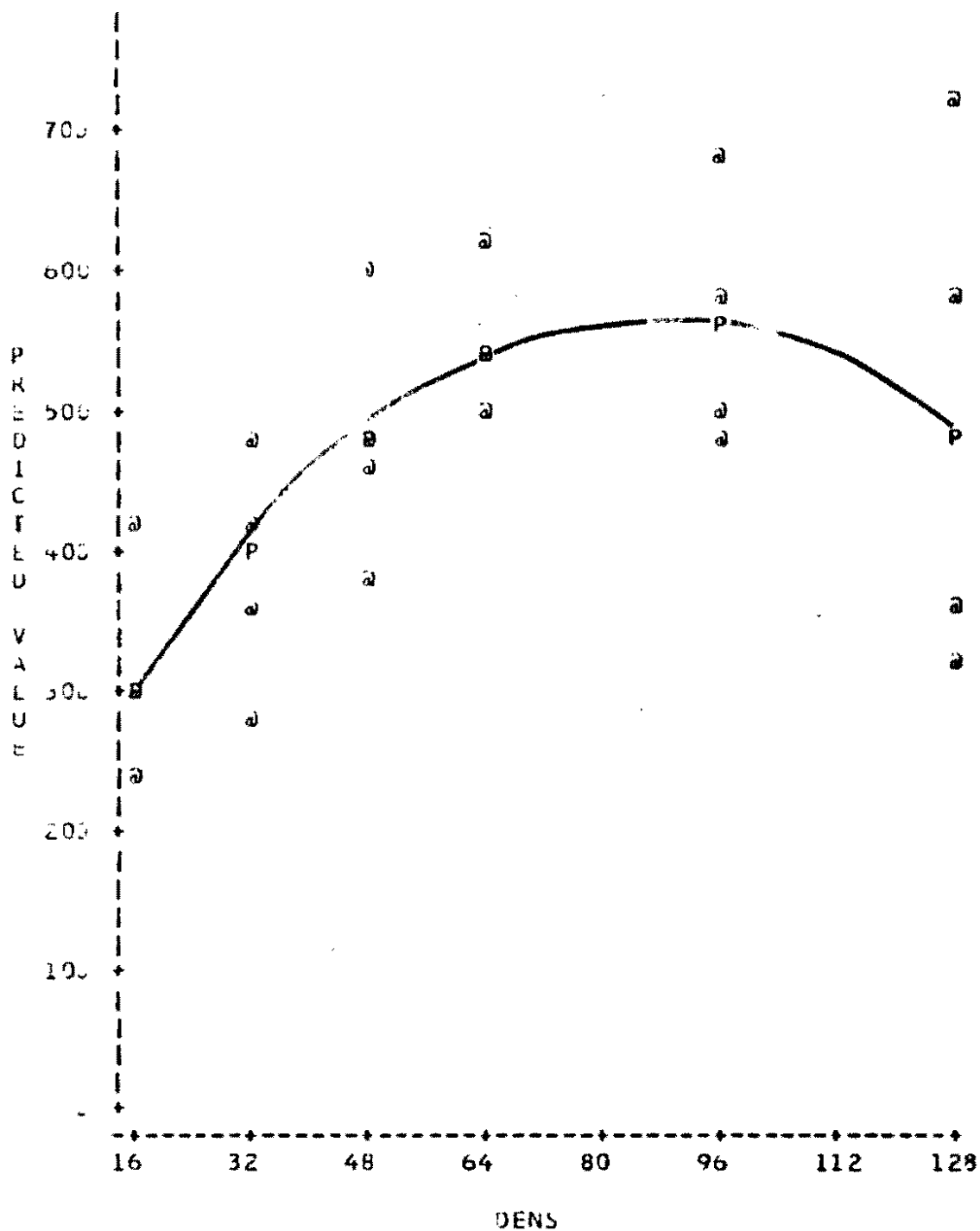
SYMBOL USED IS @
 SYMBOL USED IS P



NOTE: 20 OBS HIDDEN

GRAFICAS

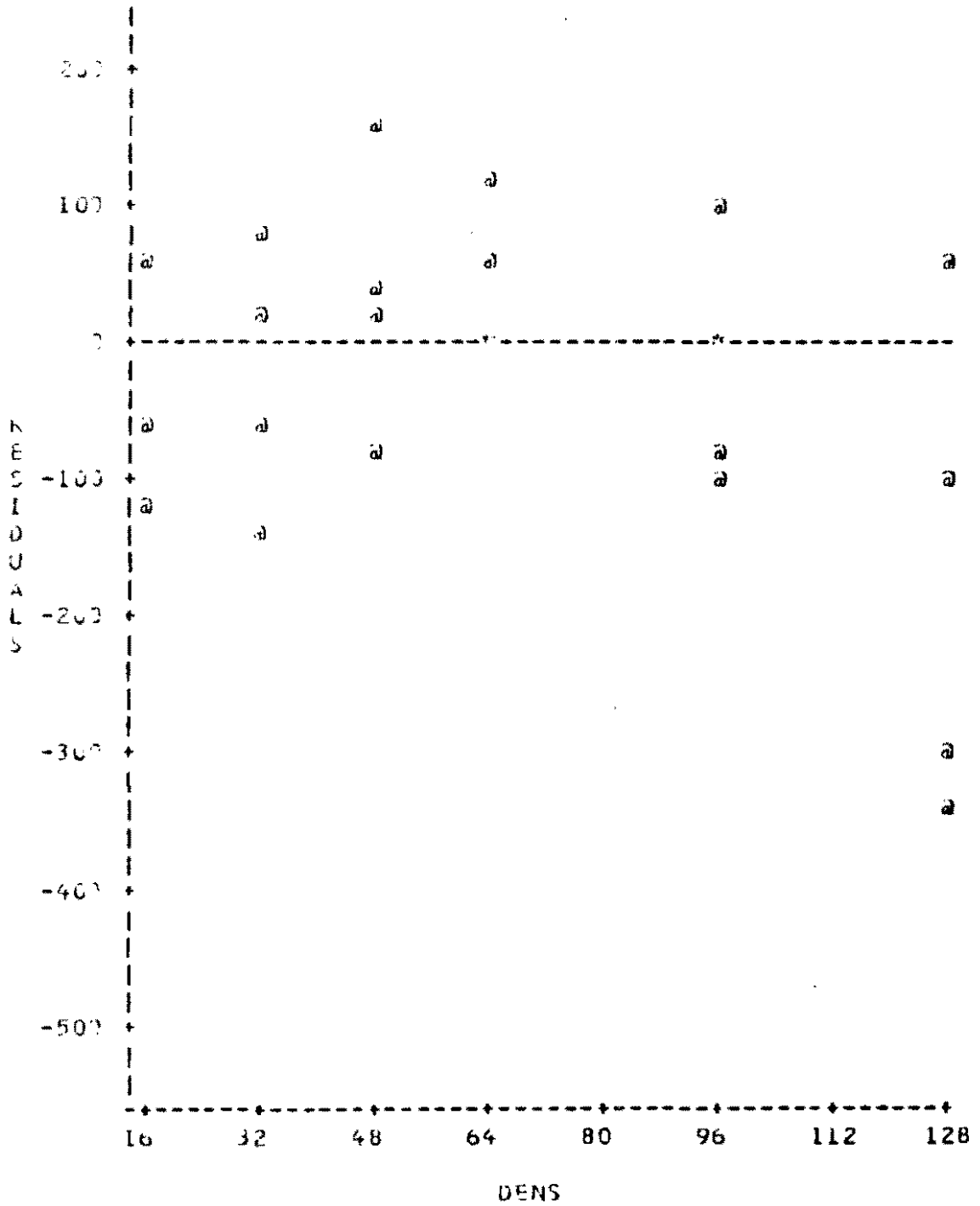
PLOT OF REND+DENS SYMBOL USED IS a
 PLOT OF %_REND+DENS SYMBOL USED IS P



NOTE: 20 OBS HIDDEN

GRAFICAS

PLOT OF EI_REVD#DENS SYMBOL USED IS @

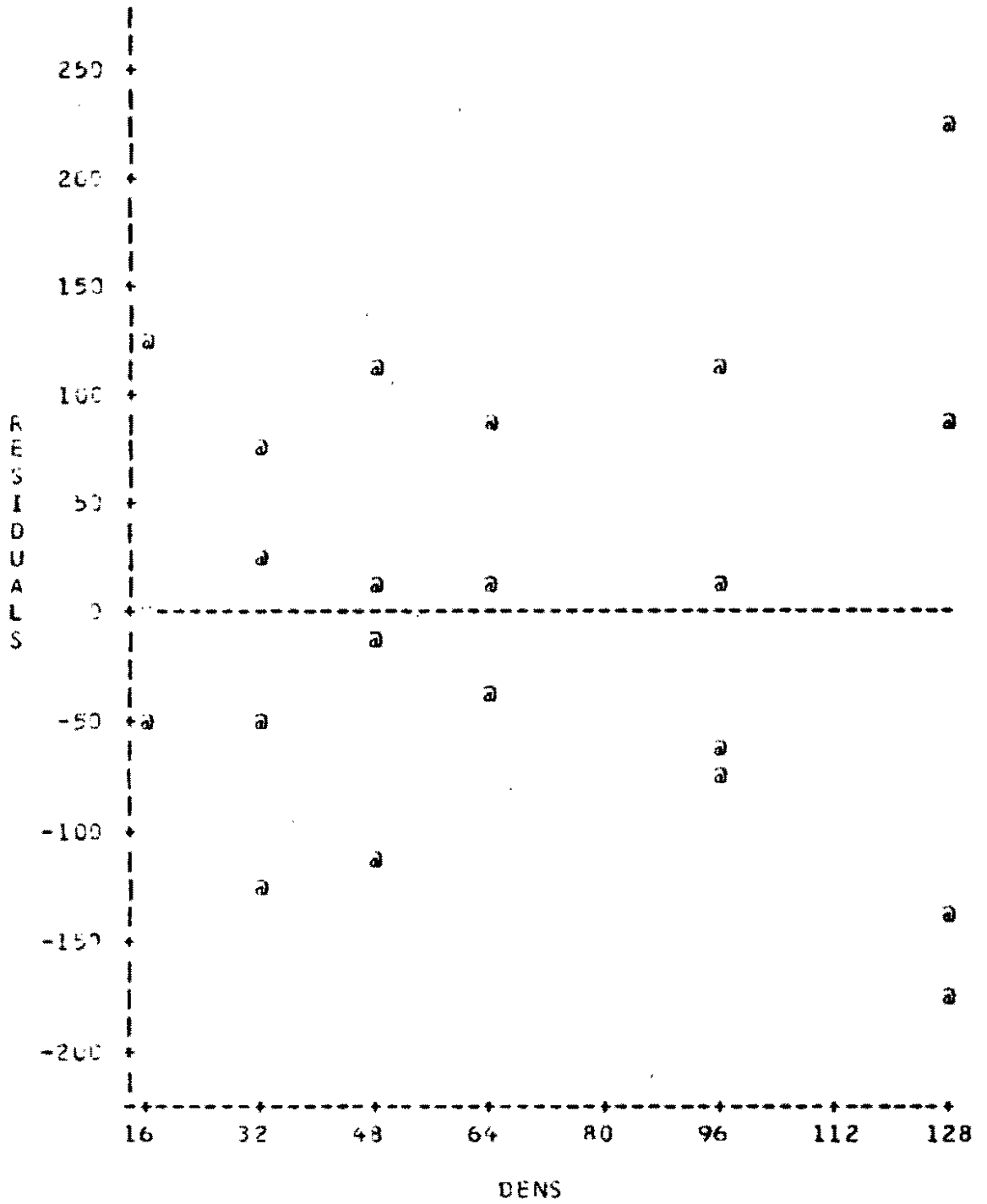


NOTE: 2 OBS HIDDEN

GRAFICAS

PLOT OF E2_REND DENS

SYMBOL USED IS a



NOTE: 2 OBS HIDDEN

PROC REG

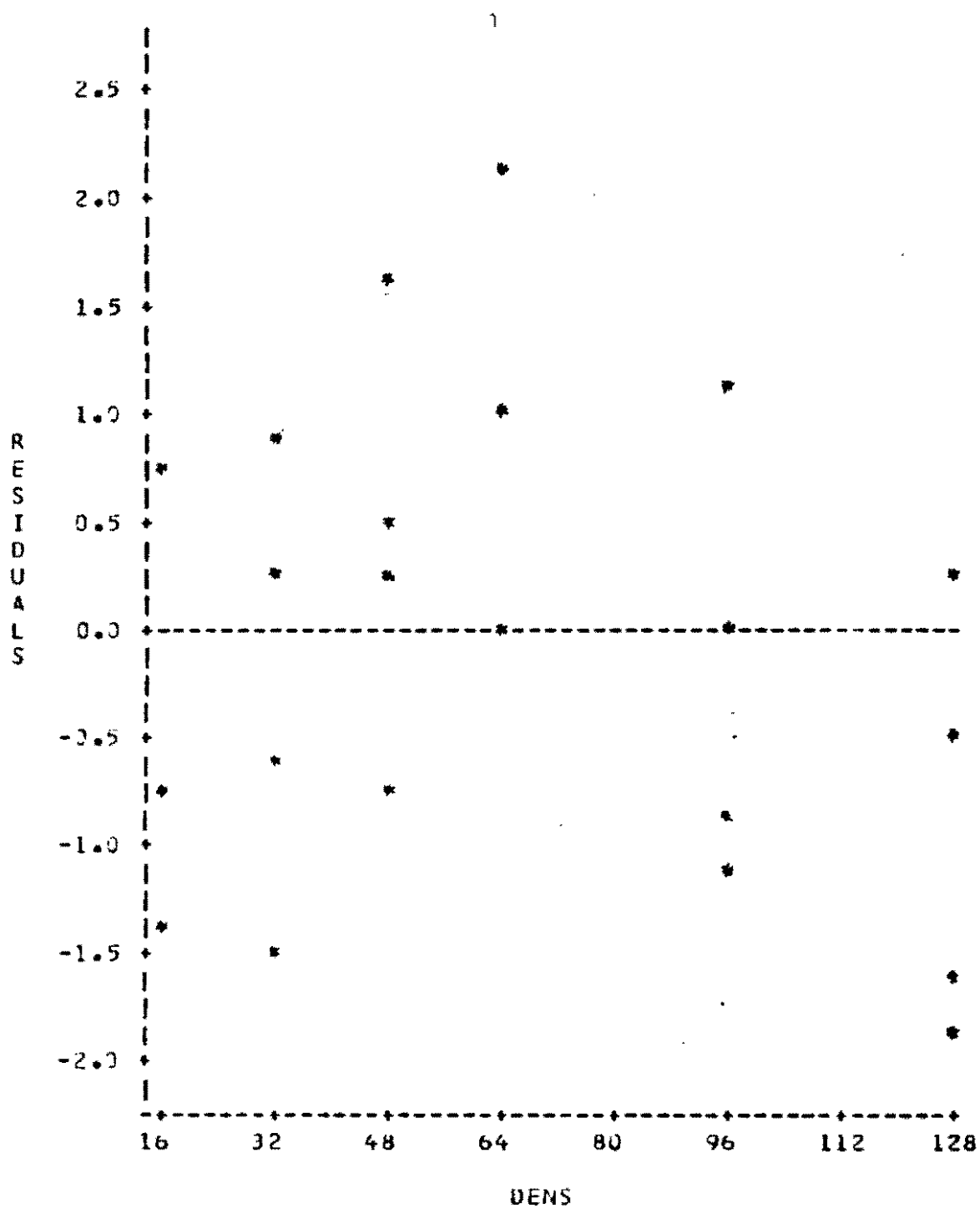
EJEMPLO NO. 3 (CONT.)

```
DATA D1;
    SET D1;
    E1_REND = E1_REND*SQRT(W);
DATA D2;
    SET D2;
    E2_REND = E2_REND*SQRT(W);
PROC PLOT DATA = ;
    PLOT E1_REND*DENS = '*' /
        VPOS=40 HPOS=50 ;
PROC PLOT DATA = ;
    PLOT E2_REND*DENS = '*' /
        VPOS=40 HPOS=50;
```


ANALISIS DE REGRESION

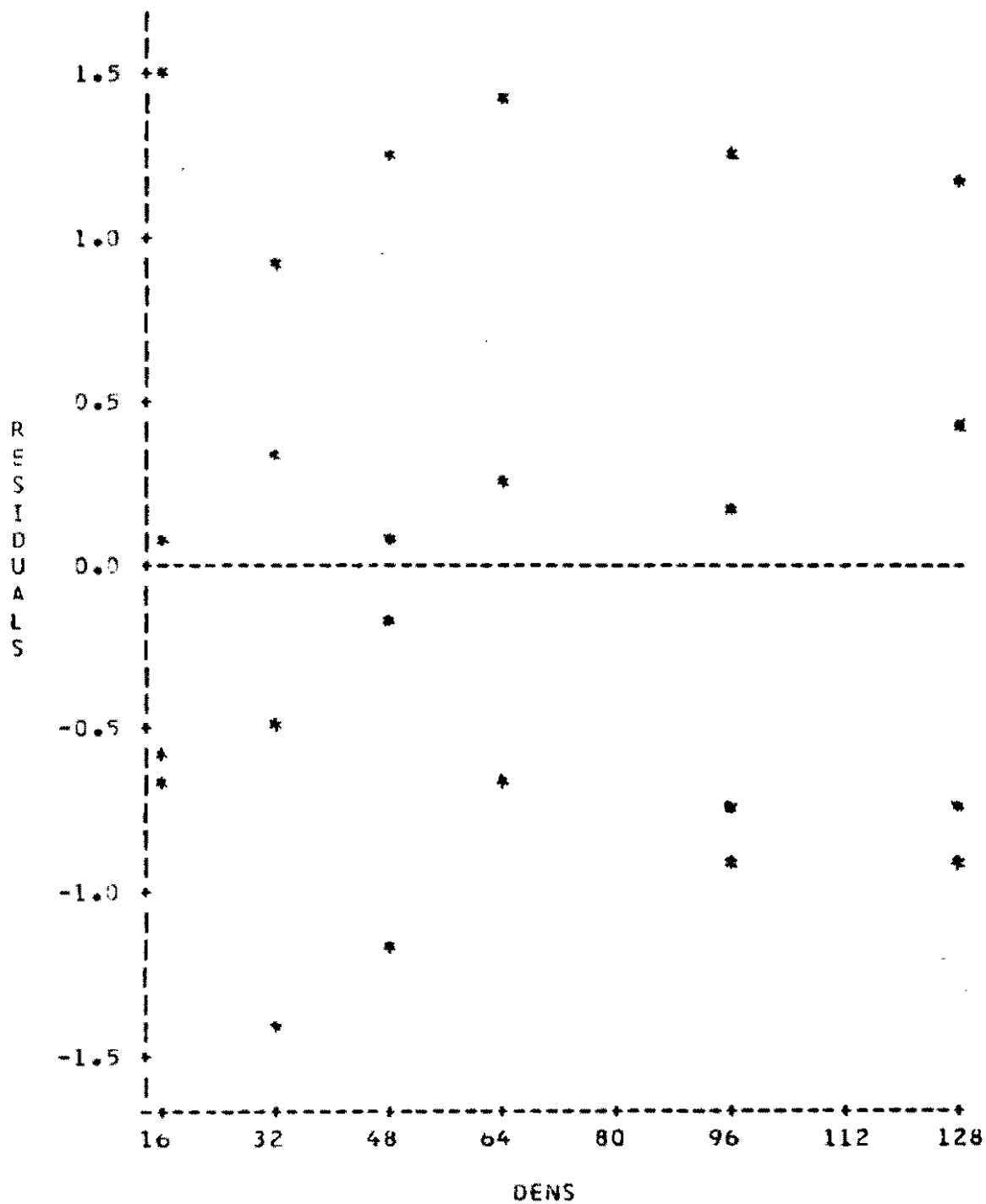
PLOT OF EI_REND vs DENS

SYMBOL USED IS *



ANALISIS DE REGRESIÓN

PLOT OF E2_REND.DENS SYMBOL USED IS *



PROC REG

EJEMPLO NO. 4:

```
DATA A ;
  SET A ;
  IF VARIEDAD=2;
PROC REG DATA=A OUTEST = AEST OUTSSCP =ASS SIMPLE ALL;
  M1 : MODEL REND = VAINASXP GRANOSXV P100SEM ;

  OUTPUT OUT=AFIN P=PREND P=RREND ;
  GV : TEST GRANOSXV VAINASXP ;
  GP : TEST GRANOSXV P100SEM ;
  VP : TEST VAINASXP P100SEM ;
  G : TEST GRANOSXV ;
  V : TEST VAINASXP ;
  P : TEST P100SEM ;
PROC PRINT DATA=AEST ;
  TITLE ESTIMADOS DE LOS PARAMETROS;
PROC PRINT DATA=ASS ;
  TITLE SUMA DE CUADRADOS ;
PROC PRINT DATA=AFIN ;
  TITLE ARCHIVO FINAL ;
```

DESCRIPTIVE STATISTICS

	SUM	MEAN	UNCORRECTED SS	VARIANCE	STD DEVIATION
REND	10998.39828491	458.26659520	5441189.58004171	17434.39330044	132.03936269
VAINASXP	331.00000000	13.79166667	5223.00000000	28.60688406	5.34954037
GRANOSXV	104.59999084	4.35833295	469.33992058	0.58514494	0.76494767
P100SEM	604.29985905	25.17916079	17337.46240223	92.24778625	9.60457111

-262-

SUMS OF SQUARES AND CROSSPRODUCTS

SSCP	REND	VAINASXP	GRANOSXV	P100SEM	INTERCEP
REND	5441190	152034	46892.01	274791.9	10998.4
VAINASXP	152034	5223	1419.9	7879.198	331
GRANOSXV	46892.01	1419.9	469.3399	2713.449	104.6
P100SEM	274791.9	7879.198	2713.449	17337.46	604.2999
INTERCEP	10998.4	331	104.6	604.2999	24

MODEL CROSSPRODUCTS X*X X*Y Y*Y

X*X	INTERCEP	VAINASXP	GRANOSXV	P100SEM	REND
INTERCEP	24	331	104.6	604.2999	10998.4
VAINASXP	331	5223	1419.9	7879.198	152034
GRANOSXV	104.6	1419.9	469.3399	2713.449	46892.01
P100SEM	604.2999	7879.198	2713.449	17337.46	274791.9
REND	10998.4	152034	46892.01	274791.9	5441190

SEQUENTIAL PARAMETER ESTIMATES

-263-

INTERCEP	458.267				
VAINASXP	450.978	0.528506			
GRANOSXV	844.095	-2.27805	-81.3179		
P100SEM	821.126	-1.12499	-92.5711	2.22851	

X*X INVERSE, B, SSE

INVERSE	INTERCEP	VAINASXP	GRANOSXV	P100SEM	REND
INTERCEP	2.246265	-0.0377197	-0.34634	-0.00694683	821.1257
VAINASXP	-0.0377197	0.001794269	0.0009620969	0.0003487257	-1.12499
GRANOSXV	-0.34634	0.0009620969	0.09608368	-0.00340336	-92.5711
P100SEM	-0.00694683	0.0003487257	-0.00340336	0.0006739811	2.228506
REND	821.1257	-1.12499	-92.5711	2.228506	309626.6

MODEL: M1
 DEP VARIABLE: REND

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	3	91364.399	30454.800	1.967	0.1514
ERROR	20	309627	15481.332		
C TOTAL	23	400991			
ROOT MSE		124.424	R-SQUARE	0.2278	
DEP MEAN		458.267	ADJ R-SQ	0.1120	
C.V.		27.15101			

-776-

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T	TYPE I SS	TYPE II SS	STANDARDIZED ESTIMATE	TOLERANCE
INTERCEP	1	321.126	186.481	4.403	0.0003	5040199	300164	0.000000	.
VAINASXP	1	-1.124992	5.270453	-0.213	0.8331	183.780	705.361	-0.045570	0.847060
GRANOSXV	1	-92.571068	38.568165	-2.400	0.0262	93812.107	89186.872	-0.536295	0.773320
PI00SEM	1	2.228506	3.230190	0.690	0.4982	7368.513	7368.513	0.162102	0.699308

TEST: GV	NUMERATOR:	87963.6	DF:	1	F VALUE:	5.6819
	DENOMINATOR:	15481.3	DF:	20	PROB >F :	0.0272
TEST: GP	NUMERATOR:	90735.9	DF:	1	F VALUE:	5.8610
	DENOMINATOR:	15481.3	DF:	20	PROB >F :	0.0251
TEST: VP	NUMERATOR:	384.667	DF:	1	F VALUE:	0.0248
	DENOMINATOR:	15481.3	DF:	20	PROB >F :	0.8763
TEST: G	NUMERATOR:	89186.9	DF:	1	F VALUE:	5.7609
	DENOMINATOR:	15481.3	DF:	20	PROB >F :	0.0262
TEST: V	NUMERATOR:	735.361	DF:	1	F VALUE:	0.0456
	DENOMINATOR:	15481.3	DF:	20	PROB >F :	0.8331
TEST: P	NUMERATOR:	7368.51	DF:	1	F VALUE:	0.4760
	DENOMINATOR:	15481.3	DF:	20	PROB >F :	0.4982

COVARIANCE OF ESTIMATES

COVB	INTERCEP	VAINASXP	GRANOSXV	P100SEM
INTERCEP	34775.17	-583.951	-5361.8	-107.546
VAINASXP	-583.951	27.77767	14.89454	5.398738
GRANOSXV	-5361.8	14.89454	1487.503	-52.6885
P100SEM	-107.546	5.398738	-52.6885	10.43413

CORRELATION OF ESTIMATES

CORRB	INTERCEP	VAINASXP	GRANOSXV	P100SEM
INTERCEP	1.0000	-0.5941	-0.7455	-0.1785
VAINASXP	-0.5941	1.0000	0.0733	0.3171
GRANOSXV	-0.7455	0.0733	1.0000	-0.4229
P100SEM	-0.1785	0.3171	-0.4229	1.0000

ESTIMADOS DE DATA=A

OBS	_TYPE_	_MODEL_	_SIGMA_	REND	VAINASXP	GRANOSXV	P10JSEM	INTERCEP
1	OLS	M1	124.424	-1	-1.125	-92.571	2.22851	821.126

SUMA DE CUADRADOS

OBS	_NAME_	INTERCEP	REND	VAINASXP	GRANOSXV	P10JSEM
1	REND	10998.4	5441190	152034	46892.0	274792
2	VAINASXP	331.0	152034	5223	1419.9	7879
3	GRANOSXV	104.6	46892	1420	469.3	2713
4	P10JSEM	604.3	274792	7879	2713.4	17337
5	INTERCEP	24.3	10998	331	104.6	604

-267-

ARCHIVO FINAL

UBS	VARIEDAD	DENSIDAD	REP	REND	VAINASXP	GRANDSXV	PLOGSEM	IDENT	PREND	RREND
1	2	1	1	299.9	22	4.3	21.9	211	447.125	-147.22
2	2	1	3	246.7	13	4.6	24.9	213	436.164	-189.46
3	2	1	2	423.7	23	5.0	22.5	212	382.537	41.16
4	2	1	4	240.2	13	5.1	22.1	214	383.638	-143.44
5	2	2	1	272.8	11	4.4	27.7	221	463.168	-190.37
6	2	2	3	429.5	19	3.7	24.0	223	510.722	-81.22
7	2	2	2	479.4	20	3.7	20.4	222	501.574	-22.17
8	2	2	4	351.7	20	5.1	23.4	224	378.601	-26.96
9	2	3	1	597.3	20	4.9	22.4	231	394.946	202.35
10	2	3	3	488.4	10	4.1	25.6	233	487.384	1.02
11	2	3	2	464.9	13	4.4	22.9	232	450.221	14.68
12	2	3	4	371.7	12	5.2	19.6	234	369.935	1.76
13	2	4	1	615.3	9	3.6	20.7	241	523.875	91.42
14	2	4	3	491.1	7	5.3	21.5	243	370.537	120.56
15	2	4	2	545.9	11	3.8	35.3	242	535.647	10.25
16	2	4	3	492.5	23	2.6	8.1	243	572.617	-80.12
17	2	5	1	675.5	15	4.0	22.2	251	483.439	192.00
18	2	5	3	498.2	11	4.4	26.7	253	460.939	37.26
19	2	5	2	577.0	11	4.7	27.3	252	434.505	142.49
20	2	5	4	481.0	8	4.0	23.1	254	493.320	-12.32
21	2	6	1	571.5	11	3.5	36.7	261	566.538	4.96
22	2	6	3	354.2	8	6.2	62.6	263	377.690	-23.49
23	2	6	2	714.4	16	3.6	23.1	262	521.349	193.05
24	2	6	4	315.6	5	4.4	19.6	264	451.867	-136.27

PROC STEPWISE

OBJETIVO:

SELECCIONAR EXPLORATORIAMENTE VARIABLES QUE MEJOR AJUSTAN ALGUNA ECUACION DE REGRESION LINEAL.

ESPECIFICACION:

```
PROC STEPWISE DATA = nombre de archivo;  
MODEL var.depend = v ar. indep/opciones;  
WEIGHT variable;  
BY variables;
```

PROPOSICION MODEL - OPCIONES -

NOINT	ELIMINA EL TERMINO INTERCEPTO DEL MODELO
BACKWARD	TECNICA DE ELIMINACION DE VARIABLES SI NO SON SIGNIFICATIVAS AL NIVEL SLSTAY
FORWARD	TECNICA DE INCLUSION DE VARIABLES QUE SON SIGNIFICATIVAS AL NIVEL SLENTRY.
STEPWISE	TECNICA DE INCLUSION DE VARIABLES QUE SON SIGNIFICATIVAS AL NIVEL SLENTRY Y LAS ELIMINA SI NO SON SIGNIFICATIVAS AL NIVEL SLSTAY.

OTRAS

MAXR	SLSTAY =	INCLUDE =
MINR	SLENTRY =	START =
		STOP =

PROC STEPWISE

EJEMPLO NO. 1:

```
PROC STEPWISE DATA=A;  
MODEL REND = GRANOSXV VAINASXV P100SEM /  
FORWARD BACKWARD STEPWISE ;  
TITLE ANALISIS DE REGRESION POR PASOS;
```

ANALISIS DE REGRESION POR PASOS

FORWARD SELECTION PROCEDURE FOR DEPENDENT VARIABLE REND

STEP 1 VARIABLE GRANOSXV eNTERED R SQUARE = 0.20145152 C(P) = 0.68367128

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	80780.26105593	80780.26105593	5.55	0.0278
ERROR	22	320210.81227741	14555.03692170		
TOTAL	23	400991.07333334			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	795.92496594				
GRANOSXV	-77.47417957	32.88599720	80780.26105593	5.55	0.0278

STEP 2 VARIABLE P100SEM REMOVED R SQUARE = 0.20145152 C(P) = 0.68367128

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	80780.26105593	80780.26105593	5.55	0.0278
ERROR	22	320210.81227741	14555.03692170		
TOTAL	23	400991.07333334			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	795.92496594				
GRANOSXV	-77.47417957	32.88599720	80780.26105593	5.55	0.0278

 ALL VARIABLES IN THE MODEL ARE SIGNIFICANT AT THE 0.1000 LEVEL.

ANALISIS DE REGRESION POR PASOS

BACKWARD ELIMINATION PROCEDURE FOR DEPENDENT VARIABLE REND

STEP 0 ALL VARIABLES ENTERED R SQUARE = 0.22784648 C(P) = 4.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	91364.40409722	30454.80136574	1.97	0.1514
ERROR	20	309626.66923611	15481.33346161		
TOTAL	23	400991.07333334			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	821.12592191				
GRANDSXV	-92.57106852	38.56816622	89186.87320378	5.76	0.0262
VAINASXP	-1.12499505	5.27045300	705.36465829	0.05	0.8331
PI00SEM	2.22850406	3.23018944	7368.50251945	0.48	0.4982

STEP 1 VARIABLE VAINASKP REMOVED R SQUARE = 0.22608743 C(P) = 2.04556227

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	90659.03943893	45329.51971947	3.07	0.0678
ERROR	21	310332.03389440	14777.71589973		
TOTAL	23	400991.07333334			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	797.47590022				
GRANDSXV	-91.96784024	37.58023067	88503.49175798	5.99	0.0233
P100SEM	2.44715288	2.99304424	9878.77838300	0.67	0.4228

STEP 2 VARIABLE PLOGSEM ENTERED R SQUARE = 0.22608743 C(P) = 2.04556227

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	90659.03943893	45329.51971947	3.07	0.0678
ERROR	21	310332.03389440	14777.71589973		
TOTAL	23	400991.07333334			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	797.47590022				
GRANOSXV	-91.90784024	37.58023067	88503.49175798	5.99	0.0233
PLOGSEM	2.44715288	2.99304424	9878.77838300	0.67	0.4228

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

ANALISIS DE REGRESION POR PASOS

STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE REND

STEP 1 VARIABLE GRANOSXV ENTERED R SQUARE = 0.20145152 C(P) = 0.68367128

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	80780.26105593	80780.26105593	5.55	0.0278
ERROR	22	320210.81227741	14555.03692170		
TOTAL	23	400991.07333334			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	795.92496594				
GRANOSXV	-77.47417957	32.88599720	80780.26105593	5.55	0.0278

NO OTHER VARIABLES MET THE 0.1500 SIGNIFICANCE LEVEL FOR ENTRY INTO THE MODEL.

