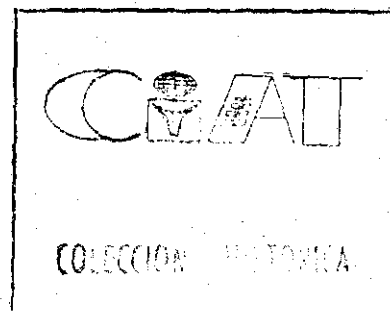


CIAT
QA
76
.25
M4
1978



CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL

(C.I.A.T.)

/ GUIA SIMPLIFICADA PARA UTILIZAR SAS

Unidad de Biometría

5652

* Gastón Mendoza
Jorge Augusto Porras

Octubre de 1978

CIAT
BIBLIOTECA

31 OCT. 1978

44358

INDICE

1.	Lenguaje SAS	1
1.1.	Introducción	1
1.2.	Tarjetas de control	3
1.3.	Creando conjuntos de datos SAS	6
1.4.	Instrucción SET	16
1.5.	Instrucción MERGE	25
1.6.	Instrucción UPDATE	30
1.7.	Procedimientos SORT y PRINT	31
1.8.	Sentencias de Encabezado	33
1.9.	Sentencias de Comentario	33
1.10	Sentencias para transformar y crear variables	33
	Tabla de Funciones más utilizadas en SAS	34
	Tabla de Operadores usados en SAS	36
1.11	Instrucciones de Programa	37
2.	Conceptos Generales de los Procedimientos SAS	39
2.1.	Introducción	39
2.2.	Instrucciones BY	39
3.	Procedimientos SAS más utilizados	41
	Procedimiento MEANS	42
	Procedimiento FREQ	46
	Procedimiento CORR	50
	Procedimiento PLOT	53
	Procedimiento ANOVA	55
	Procedimiento GLM	59
	Procedimiento DUNCAN	66
	Procedimiento TTEST	70
	Procedimiento STEPWISE	72
	Procedimiento NLIN	76
Apéndice A:	Salida impresa del job conteniendo algunos de los ejemplos del Capítulo 1.	
Apéndice B:	Instrucciones de Programa más útiles.	

GUIA SIMPLIFICADA PARA UTILIZAR SAS

1. Lenguaje SAS:

1.1. Introducción:

En esta guía simplificada se pretende ilustrar cómo utilizar SAS (Statistical Analysis System) para resolver los problemas de análisis estadísticos más comunes. Cabe señalar que sólo hemos considerado las opciones más útiles. Para situaciones más complicadas consultar el manual completo "A User's Guide to SAS 76" o solicitar ayuda al personal de la Unidad de Biometría.

SAS es un sistema integrado para el manejo de datos y análisis estadísticos desarrollado en North Carolina State University.

Un conjunto de datos SAS (o archivo SAS) es una colección de observaciones. Cada observación consiste de los valores de una o más variables correspondientes a un mismo ente o "individuo". Si una variable asume valores numéricos es llamada variable numérica; y si asume valores alfanuméricos es llamada variable alfanumérica.

En SAS el usuario elige los nombres de sus variables y de sus conjuntos de datos. Estos nombres están constituidos hasta por 8 caracteres consecutivos, el primero de los cuales debe ser una letra o un "under score". Ejemplos: VARIEDAD
REND_77.

Sentencias SAS (definidas como un conjunto de comandos y/o instrucciones) están delimitadas por punto y coma (;) y pueden comenzar y terminar en cualquier posición de una tarjeta. Más de una sentencia SAS puede aparecer en una tarjeta. Una sentencia SAS puede ocupar más de una tarjeta. Los elementos de una sentencia deben estar separados por uno o más blancos excepto cuando se les conecta con caracteres especiales.

Ejemplos válidos: X=Y; A = B; X = Y+Z; PROC GLM;

Ejemplos no válidos: X = Y X = Y; PROCGLM;

Un programa SAS (grupo de instrucciones SAS a leerse conjuntamente con ciertas instrucciones de control) tiene dos tipos de sentencias:

Sentencias que crean
conjuntos de datos SAS



Grupos de sentencias de este
tipo comienzan con DATA.

Sentencias que analizan
conjuntos de datos SAS



Grupos de sentencias de este
tipo comienzan con PROC.

1.2. Tarjetas de control necesarias para un job SAS:

Tarj. 1 //CIATppØuØJOBØ(CIAT,USEX),CIAT9999nombre,
MSGLEVEL=(2,Ø),

Tarj. 2 //ØCLASS=clase,

Tarj. 3 //ØTIME=tiempo,

Tarj. 4 //ØEXEC SAS76

} Programa SAS

Tarj. 5 /*

Tarj. 6 //

Nota: Ø significa un espacio en blanco

Ø denota el número 0

Tarjeta 1, columnas 5-6: Las letras y nombres en minúsculas significan que deben reemplazarse por un número o nombre adecuado. pp son dos letras que identifican el programa o proyecto al que pertenece el usuario, según la tabla siguiente:

Códigos a usar en las columnas 5-6, tarjeta de control No. 1:

AD	Adiestramiento
AL	Agronomía de Leguminosas
AZ	Arroz
BL	Biblioteca
BI	Biometría
EF	Economía Fríjol
EG	Economía Ganado de Carne
EY	Economía Yuca
FR	Fríjol
GC	Ganado de Carne
HA	Hatos I
IA	Ingeniería Agrícola
IB	IBRN
MT	Metereología
MZ	Maíz
OP	Operaciones de Campo
PF	Pastos y Forrajes
RH	Colección de Rhizobium
SA	Salud Animal
SB	Sistema de Información de Becarios
SI	SIFRI
UV	Universidad del Valle (PRIMOPS)

VI	IBYAN
YU	Yuca
PR	Porcinos

Tarjeta 1, columna 8: Letra o número de libre elección por el usuario para diferenciar sus jobs.

Tarjeta 1, columna 36: Nombre del usuario consistente de la inicial del nombre seguido de punto y a continuación el apellido.

Tarjeta 2.: El siguiente cuadro da una idea para la adecuada asignación del parámetro "clase" cuando no se usan cintas, ni archivos en cinta:

Poca memoria requerida ($\leq$ 192 K): G

Mucha memoria requerida ($>$ 192 K): L

Tarjeta 3.: El parámetro "tiempo" indica el tiempo máximo en minutos de procesamiento. Para trabajos pequeños usar 1, para trabajos normales usar 2 y para trabajos grandes usar 3.

La tarjeta 4 llama los programas de que consiste SAS.

Las tarjetas 5 y 6 indican fin del job.

Los archivos SAS pueden crearse de tres maneras:

- A partir de datos crudos
- A partir de archivos SAS previamente creados
- A partir de archivos creados por paquetes estadísticos tales como SAS72, SPSS, OSIRIS, BMDP y DATATEX.

A continuación se presentan diversos ejemplos de la primera modalidad.

1.3. Creando conjuntos o archivos de datos SAS:

Ejemplo 1: ENTRADA DE DATOS CON FORMATO LIBRE (UNA OBSERVACION POR TARJETA)

```
DATA; INPUT  X  Y; CARDS;
```

```
2    3
```

```
3    5
```

```
4    6
```

El resultado de las instrucciones y datos anteriores es la creación de un conjunto de datos que SAS automáticamente lo nombra DATA1, el cual consiste de 3 observaciones (cada una de las tres últimas líneas) y 2 variables (denominadas X e Y). Si posteriormente aparece otra instrucción del tipo DATA; el nuevo conjunto de datos (cuya estructura no nos interesa por el momento) sería automáticamente denominado DATA2 y así sucesivamente.

Ejemplo 2: ENTRADA DE DATOS CON FORMATO LIBRE (VARIAS OBSERVACIONES POR TARJETA)

```
DATA A2; INPUT  X  Y  @@; CARDS;
```

```
2    3    3    5    4    6
```


El resultado de las instrucciones y datos anteriores es la creación de un conjunto de datos denominado (por el usuario) A2 el cual tiene exactamente la misma estructura que el creado en el ejemplo anterior. Note que el uso de dos símbolos @ permite escribir las observaciones en la misma línea. La regla básica es escribir los valores dejando al menos un blanco de por medio.

Ejemplo 3: ENTRADA DE DATOS ALFANUMERICOS

```
DATA A3; INPUT VAR $ REND @@; CARDS;  
P720 580.3 P642 328.8 P005 .
```

El resultado de las instrucciones y datos anteriores es la creación del conjunto de datos A3 el cual consiste de 3 observaciones y dos variables. Note que el símbolo \$ después de VAR indica que la variable VAR es alfanumérica. Así mismo, note que el valor de la variable REND en la tercera observación es faltante y se le representa con un punto.

Ejemplo 4: ENTRADA DE DATOS CON FORMATO FIJO (ESPECIFICANDO LAS COLUMNAS)

```
DATA A4; INPUT X 1-2 Y 4; CARDS;  
2 3  
3 5  
4 6
```

En este ejemplo usamos un formato fijo que indica que los valores de la variable numérica X vienen perforados en las columnas 1 y 2 (con ajuste por la derecha) y los valores de la variable numérica Y vienen perforados en la columna 4 (está implícito que los valores posibles de Y son enteros no negativos de un dígito). El conjunto de datos A4 es exactamente igual al obtenido en el ejemplo 1.

Ejemplo 5: ENTRADA DE DATOS CON FORMATO FIJO

```
DATA A5; INPUT (NAME SCORE1 SCORE2)
($35. @50 2*5); CARDS;
```

Col. 1	Col. 35	Col. 50	Col. 55
↓		↓	↓
JUAN PEREZ		98.5	43.2
JOHN DOE		-31.2	-5.
	↓		
CENTRO INT. DE AGRICULTURA TROPICAL		45.8	0

El resultado es un conjunto de datos llamados A5 con 3 observaciones y 3 variables. Los valores de la variable alfanumérica NAME se leen en las columnas 1-35; los de la variable numérica SCORE1 se leen en las columnas 50-54; y los de la variable numérica SCORE2 se leen en las columnas 55-59. Note que el valor de SCORE2 en la tercera observación es 0. Si tal valor fuera faltante bastaría dejar en blanco las columnas 55-59.

Ejemplo 6: ENTRADA DE DATOS CON IDENTIFICADORES
COMUNES

Supongamos que deseamos crear un conjunto de datos consistente de la respuesta de una determinada variedad en diversas repeticiones. La forma más convencional sería perforar grupos de 3 valores correspondientes a VAR, REP y RESP. Sin embargo, SAS permite no sólo la lectura de datos, sino también la generación de valores de variables, como se ilustra a continuación:

```
DATA A6; INPUT  VAR  $  NREP @ ; REP=0 ; DROP NREP;
L1 : REP+1; INPUT RESP @ ; OUTPUT;
IF REP < NREP THEN GO TO L1; DELETE; CARDS;
P304    4  15.3  16.0  14.8  15.9
P058    3  14.1  13.8  14.1
```

El resultado es el conjunto de datos siguientes:

P304	1	15.3
P304	2	16.0
P304	3	14.8
P304	4	15.9
P058	1	14.1
P058	2	13.8
P058	3	14.1

Conjunto de datos A6 con 7
observaciones y 3 variables:
VAR REP RESP

El ejemplo 6 ilustra como usar varias instrucciones SAS no discutidas anteriormente. En el apéndice B se presentan algunas de las instrucciones SAS de mayor uso.

A estas alturas conviene presentar un ejemplo completo de un job SAS.

Ejemplo de un Job SAS

```
//CIATAD01 JOB (CIAT,USEX),CIAT9999G:MENDOZA,MSGLEVEL=(2,0),
```

```
// CLASS=G,
```

```
// TIME=1
```

```
// EXEC SAS76,V=765
```

```
DATA ALPE;
```

```
INPUT NOMBRE $ 1-10 SEXO $ 12 EDAD 14-15 ALTURA  
17-18 PESO 20-22;
```

```
CARDS;
```

datos →

```
PROC PRINT; TITLE EJEMPLO DE UN JOB SAS;
```

```
PROC PLOT;
```

```
PLOT ALTURA*PESO = SEXO;
```

```
/*
```

```
//
```

EJEMPLO 7: ENTRADA DE SOLO ALGUNAS VARIABLES

Supongamos que las tarjetas de datos son tales que cada tarjeta corresponde a una observación y que la estructura es la siguiente:

<u>Variable</u>	<u>Tipo</u>	<u>Columnas</u>	<u>Comentario</u>
NOMBRE	Alfanumérica	1-10	
SEXO	Alfanumérica	12	'M' y 'F'
EDAD	Numérica	14-15	En años
ALTURA	Numérica	17-18	En pulgadas
PESO	Numérica	20-22	En libras

Si por ejemplo, en la aplicación presente, sólo estamos interesados en las variables NOMBRE y PESO, bastarían las instrucciones:

```
DATA A7; INPUT NOMBRE $ 1-10 PESO 20-22;  
CARDS;
```

EJEMPLO 8: GRABACION DE UN ARCHIVO OS EN DISCO.

Supongamos que a partir de las tarjetas de datos descritos en el ejemplo anterior deseamos grabarlas en el disco DANE04 bajo el nombre CIAT.EJEM8. Para lograr esto es necesario que después de la tarjeta de EXEC SAS76, se incluyan sentencias JCL que definan al sistema operacional donde grabar el archivo:

```
//D04 DD DSN=CIAT.EJEM8,UNIT=DISK,VOL=SER=DANE04,  
// SPACE=(TRK,(3,1)),DISP=(NEW,KEEP),DCB=(LRECL=18,RECFM=FB,  
// BLKSIZE=12960)  
DATA A8;  
INPUT NOMBRE $ SEXO $ EDAD ALTURA PESO;  
FILE D04;  
PUT NOMBRE $10. SEXO $1. EDAD 2. ALTURA 2. PESO 3.;  
datos CARDS;  
/*  
//
```

EJEMPLO 9: GRABACION DE UN ARCHIVO SAS EN DISCO

Supongamos la misma situación planteada en el ejemplo 8 excepto de que no sólo nos interesa grabar los datos sino la información total del archivo SAS: nombre del archivo y nombre y tipo de las variables y datos.

```
//D04 DD DSN=CIAT.EJEM9,UNIT=DISK,VOL=SER=DANE04,  
// SPACE=(TRK,(3,1)),DISP=(,KEEP)  
DATA D04.A9;  
INPUT NOMBRE $ SEXO $ EDAD ALTURA PESO;  
CARDS;
```

DATOS →

```
/*  
//
```

La diferencia principal entre el archivo OS (de nombre CIAT.EJEM8) y el archivo SAS D04.A9 es que el primero puede servir de entrada a cualquier lenguaje o paquete, mientras que el segundo sólo puede ser utilizado por SAS aunque con la ventaja de que sólo es necesario recordar los nombres y tipo de las variables.

EJEMPLO 10: ENTRADA DE DATOS GRABADOS EN DISCO.

Si deseáramos crear el archivo SAS A10 utilizando los datos grabados en el ejemplo 8, pero sólo las variables SEXO y PESO, bastarían las siguientes instrucciones después del EXEC SAS76:

```
//D04 DD DSN=CIAT.EJEM8,UNIT=DISK,VOL=SER=DANE04,
```

```
// DISP=OLD
```

```
DATA A10; INFILE D04;
```

```
INPUT SEXO $ 11 PESO 17-18;
```

```
* Aquí vendrían los procedimientos de manipulación y análisis  
que se desearan;
```

```
/*
```

```
//
```


EJEMPLO 11: USO DE UN ARCHIVO SAS GRABADO EN DISCO.

Si en una corrida futura deseáramos usar el archivo SAS D04.A9 creado en el ejemplo 9 pero sólo reteniendo las variables SEXO y PESO, bastarían las siguientes instrucciones después del EXEC SAS76:

```
//D04 DD DSN=CIAT.EJEM9,UNIT=DISK,VOL=SER=DANE04,
```

```
// DISP=OLD
```

```
DATA All; SET D04.A9; KEEP SEXO PESO;
```

```
* Aquí vendrían los procedimientos de manipulación y  
análisis que se desearán;
```

```
/*
```

```
//
```

Veamos ahora instrucciones que nos permiten crear archivos SAS a partir de archivos SAS previamente creados.

1.4. Instrucción SET:

Se utiliza para crear un nuevo conjunto de datos SAS a partir de uno o más conjunto SAS previamente creados.

Ejemplo 1: FORMACION DE SUBCONJUNTOS DE DATOS

```
DATA B1; SET A4; IF X < 4;
```

El resultado es el subconjunto de A4 (ver ejemplo 4 de creación de conjuntos de datos) consistente de sus dos primeras observaciones en las cuales la condición $X < 4$ es satisfecha:

2	3
3	5

B1

Ejemplo 2: CONCATENACION DE CONJUNTOS DE DATOS

```
DATA B2; SET A4 B1;
```

El resultado es el conjunto de datos B2 cuyas 3 primeras observaciones son:

2	3
3	5
4	6

y las dos siguientes:

2 3

3 5

Ejemplo 3: INTERCALACION DE CONJUNTOS DE DATOS

DATA B3; SET A4 B1; BY X;

El conjunto de datos B3 contiene 5 observaciones
(3 de A4 y 2 de B1) ordenadas según los valores de X:

2	3
2	3
3	5
3	5
4	6

(Obs. proveniente de A1)

(Obs. proveniente de B1)

(Obs. proveniente de A1)

(Obs. proveniente de B1)

(Obs. proveniente de A1)

Ejemplo 4: Subarchivo de Observaciones

DATA CLASE; INPUT NOMBRE \$ SEXO \$ EDAD ALTURA PESO; CARDS;

Obs	NOMBRE	SEXO	EDAD	ALTURA	PESO
1	ALFREDO	M	14	1.65	50
2	ALICIA	F	13	1.60	40
3	BARBARA	F	13	1.40	35
4	CAROLA	F	14	1.35	32
5	HENRY	M	14	1.40	40
6	JAIME	M	12	1.45	40
7	JAVIER	M	12	1.20	30
8	JEREMIAS	M	15	1.35	35
9	JUAN	M	13	1.30	30
10	JULIO	M	12	1.20	30
11	JULIA	F	11	1.20	30
12	JULIETA	F	14	1.30	30
13	LUISA	F	15	1.40	35
14	MARIA	F	15	1.45	40
15	PEDRO	M	16	1.60	38

* Sólo las niñas;

DATA NINAS; SET CLASE;
IF SEXO='M' THEN DELETE;

* Archivo resultante;

Obs	NOMBRE	SEXO	EDAD	ALTURA	PESO
1	ALICIA	F	13	1.60	40
2	BARBARA	F	13	1.40	35
3	CAROLA	F	14	1.35	32
4	JULIA	F	11	1.20	30
5	JULIETA	F	14	1.30	30
6	LUISA	F	15	1.40	35
7	MARIA	F	15	1.45	40

Ejemplo 5: Subarchivo de Variables

* Sólo NOMBRE SEXO EDAD;

DATA SUB1; SET CLASE;

DROP ALTURA PESO;

* Archivo Resultante:

Obs	NOMBRE	SEXO	EDAD
1	ALFREDO	M	14
2	ALICIA	F	13
3	BARBARA	F	13
4	CAROLA	F	14
5	HENRY	M	14
6	JAIME	M	12
7	JAVIER	M	12
8	JEREMIAS	M	15
9	JUAN	M	13
10	JULIO	M	12
11	JULIA	F	11
12	JULIETA	F	14
13	LUISA	F	15
14	MARIA	F	15
15	PEDRO	M	16

Ejemplo 6: Adición de nueva variable

* CUAL ES EL AÑO DE NACIMIENTO;

DATA AR2; SET CLASE;

ANO_NAC=1978-EDAD;

* Archivo Resultante;

Obs	NOMBRE	SEXO	EDAD	ALTURA	PESO	ANO_NAC
1	ALFREDO	M	14	1.65	50	1964
2	ALICIA	F	13	1.60	40	1965
3	BARBARA	F	13	1.40	35	1965
4	CAROLA	F	14	1.35	32	1964
5	HENRY	M	14	1.40	40	1964
6	JAIME	M	12	1.45	40	1966
7	JAVIER	M	12	1.20	30	1966
8	JEREMIAS	M	15	1.35	35	1963
9	JUAN	M	13	1.30	30	1965
10	JULIO	M	12	1.20	30	1966
11	JULIA	F	11	1.20	30	1967
12	JULIETA	F	14	1.30	30	1964
13	LUISA	F	15	1.40	35	1963
14	MARIA	F	15	1.45	40	1963
15	PEDRO	M	16	1.60	38	1962

Ejemplo 7: Combinando las tres opciones anteriores

* PARA LAS NIÑAS SE DESEA TENER SOLO NOMBRE SEXO EDAD Y AÑO
NACIMIENTO;

DATA SUB2; SET CLASE;

IF SEXO='M' THEN DELETE;

DROP ALTURA PESO;

ANO_NAC = 1978-EDAD;

* Archivos Resultantes;

Obs	NOMBRE	SEXO	EDAD	ANO_NAC
1	ALICIA	F	13	1965
2	BARBARA	F	13	1965
3	CAROLA	F	14	1964
4	JULIA	F	11	1967
5	JULIETA	F	14	1964
6	LUISA	F	15	1963
7	MARIA	F	15	1963

Ejemplo 8: Subdivisión de archivos

* Sólo los nombres pero separados para niñas y niños;

DATA NINAS NINOS;

SET CLASE;

IF SEXO='F' THEN OUTPUT NINAS;

IF SEXO='M' THEN OUTPUT NINOS;

KEEP NOMBRE;

* Archivos Resultantes;

Obs	NOMBRE	Obs	NOMBRE
1	ALICIA	1	ALFREDO
2	BARBARA	2	HENRY
3	CAROLA	3	JAIME
4	JULIA	4	JAVIER
5	JULIETA	5	JEREMIAS
6	LUISA	6	JUAN
7	MARIA	7	JULIO
		8	PEDRO

Ejemplo 9: Concatenación

DATA JUNTOS;

SET NINAS (IN=INF)

NINOS (IN=INM);

IF INF THEN SEXO='F';

IF INM THEN SEXO='M';

* Archivo Resultante;

Obs	NOMBRE	SEXO
1	ALICIA	F
2	BARBARA	F
3	CAROLA	F
4	JULIA	F
5	JULIETA	F
6	LUISA	F
7	MARIA	F
8	ALFREDO	M
9	HENRY	M
10	JALME	M
11	JAVIER	M
12	JEREMIAS	M
13	JUAN	M
14	JULIO	M
15	PEDRO	M

Ejemplo 10: Intercalación

DATA JUNTOSXN;

SET NINAS (IN=INF)

NINOS (IN=INM);

BY NAME;

IF INF THEN SEXO='F';

IF INM THEN SEXO='M';

* Archivo Resultante;

OBS	NOMBRE	SEXO
1	ALFREDO	M
2	ALICIA	F
3	BARBARA	F
4	CAROLA	F
5	HENRY	M
6	JAIME	M
7	JAVIER	M
8	JEREMIAS	M
9	JUAN	M
10	JULIO	M
11	JULIA	F
12	JULIETA	F
13	LUISA	F
14	MARIA	F
15	PEDRO	M

1.5. Instrucción MERGE:

Se utiliza para crear nuevos conjuntos de datos combinando observaciones de dos o más conjuntos de datos.

Ejemplo 1: MERGE SIN BY

```
DATA M1; INPUT VAR RESP1 @@; CARDS;  
4 4.1. 3 3.8 5 5.6  
  
DATA M2; INPUT VAR RESP2 @@; CARDS;  
4 4.3 3 3.6 5 5.  
  
DATA M12; MERGE M1 M2;
```

El resultado final es el conjunto de datos M12 conteniendo 3 observaciones, 3 variables VAR RESP1 RESP2 y que puede representarse de la siguiente manera:

4	4.1	4.3
3	3.8	3.6
5	5.6	5.0

M12

Note que para que el conjunto M12 tenga un significado claro, es necesario que las observaciones de los conjuntos M1 y M2 se correspondan, (en este caso observaciones correspondientes tienen el mismo valor de VAR).

Ejemplo 2: MERGE USANDO BY

```
DATA M1S; SET M1;  
PROC SORT; BY VAR;
```

Resultado:

3	3.8
4	4.1
5	5.6



M1S

```
DATA M2S; SET M2;  
PROC SORT; BY VAR;
```

Resultado:

3	3.6
4	4.3
5	5.0



M2S

```
DATA M12S; MERGE M1S M2S; BY VAR;
```

Resultado:

3	3.8	3.6
4	4.1	4.3
5	5.6	5.0



M12S

Note que el MERGE usando BY requiere que cada uno de los conjuntos de datos cuyas observaciones van a combinarse, deben estar ordenados en las variables del BY. Este ordenamiento se logra con el procedimiento SORT explicado con mayor detalle posteriormente.

Ejemplo 3: Unión

Sean los Conjuntos:

A1					A2		
Obs	NAME	SEXO	EDAD	ALTURA	Obs	NAME	PESO
1	ALFREDO	M	14	1.65	1	ALFREDO	50
2	ALICIA	F	13	1.60	2	ALICIA	40
3	BARBARA	F	13	1.40	3	HENRY	40
4	HENRY	M	14	1.40	4	JAIME	40

* Incorporar el dato de peso;

DATA TOTAL;

MERGE A1 A2 ;

BY NAME;

* Archivo Resultante;

Obs	NAME	SEXO	EDAD	ALTURA	PESO
1	ALFREDO	M	14	1.65	50
2	ALICIA	F	13	1.60	40
3	BARBARA	F	13	1.40	.
4	HENRY	M	14	1.40	40
5	JAIME	.	.	.	40

Ejemplo 4: Intersección

Sean los Conjuntos:

INGLES			MATEMÁTICAS		
Obs	NAME	NOTAI	Obs	NAME	NOTAM
1	ALFREDO	4	1	ALICIA	4
2	ALICIA	5	2	BARBARA	5
3	BARBARA	4	3	CAROLA	5
4	JAIME	4	4	JAIME	5
5	JULIA	5	5	PEDRO	4

* Quienes vieron INGLES Y MATEMATICAS;

```
DATA RESP1; MERGE  INGLES  (IN=A)
                MATEMAT (IN=B);

BY NAME; KEEP NAME;

IF A AND B;
```

* Archivo Resultante;

Obs	NAME
1	ALICIA
2	BARBARA
3	JAIME

Ejemplo 5: Complementación

Considere los conjuntos INGLES y MATEMAT del ejemplo anterior.

* Quienes vieron inglés pero no vieron matemáticas;

```
DATA RESP2 ;  
    MERGE INGLES (IN=A)  
          MATEMAT (IN=B);  
BY NAME; KEEP NAME;  
IF A AND NOT B;
```

* Archivo Resultante;

Obs	NAME
1	ALFREDO
2	JULIA

1.6. Instrucción UPDATE:

Se utiliza para actualizar un conjunto de datos SAS
previamente creado.

Ejemplo 1: Actualización

MAESTRO				ACTUAL			
Obs	NAME	NOTAI	NOTAM	Obs	NAME	NOTAI	NOTAM
1	ALFREDO	4	.	1	ALFREDO	.	4
2	ALICIA	5	4	2	ALICIA	4	.
3	BARBARA	4	5	3	BARBARA	5	4
4	CAROLA	.	5	4	JULIA	4	
5	JAIME	4	5	5	RAUL	5	5
6	JULIA	5	.				
7	PEDRO	.	4				

* Actualize archivo maestro con nuevas notas;

DATA NUEVO;

UPDATE MAESTRO

ACTUAL

; BY NAME;

* Archivo Resultante;

Obs	NAME	NOTAI	NOTAM
1	ALFREDO	4	4
2	ALICIA	4	4
3	BARBARA	5	4
4	CAROLA	.	5
5	JAIME	4	5
6	JULIA	4	.
7	PEDRO	.	4
8	RAUL	5	5

1.7. Procedimientos SORT y PRINT:

Procedimiento SORT:

Se utiliza para reordenar las observaciones con un conjunto de datos previamente creado, o para crear un nuevo conjunto cuyas observaciones sean las de un conjunto de datos previamente creado, pero ordenada de una manera diferente. El procedimiento SORT no produce salida impresa. Su forma general es:

```
PROC SORT DATA = conjunto_de_datos_1  
              OUT = Conjunto_de_datos_2; BY lista_de_  
              variables;
```

El parámetro "conjunto_de_datos_1" da el nombre del conjunto de datos a ser ordenado. Si la sentencia DATA=conjunto_de_datos_1 se omite, se sobreentiende que el conjunto de datos a usar es el último que se creó.

El parámetro "conjunto_de_datos_2" da el nombre del conjunto de datos creado por el procedimiento SORT. Si la sentencia OUT = conjunto_de_datos_2 se omite, el procedimiento sólo reordena las observaciones en el conjunto de datos de entrada.

La sentencia BY indica al procedimiento SORT como reordenar las observaciones. SORT reordena las observaciones

primero de acuerdo a los valores ascendentes de la primera variable en la lista del BY. Si dos o más observaciones tienen el mismo valor de la primera variable, SORT las reordena de acuerdo a los valores ascendentes de la segunda variable en la lista del BY, y así sucesivamente.

Nota: Si en algún JOB se incluye algún procedimiento SORT, la tarjeta // EXEC SAS76 debe reemplazarse por
// EXEC SAS76,SORT=

Procedimiento PRINT:

Se utiliza para imprimir un listado de los valores de algunas o todas las variables de un conjunto de datos. Su forma más útil es:

```
PROC PRINT  N  DATA = nombre_del_conjunto_de_datos;  
              BY lista_de_variables;
```

La opción N pide imprimir el número de observaciones en el conjunto de datos. La opción DATA = nombre_del_conjunto_de_datos indica al procedimiento PRINT el nombre del conjunto de datos a ser impreso; si se le omite, el procedimiento PRINT se aplica al último conjunto de datos que se creó. La sentencia BY indica que se desea listados separados para cada subconjunto formado por el BY; para poder usar la sentencia BY se requiere que los datos estén ordenados de la manera que los ordenaría un SORT con la misma sentencia BY.

1.8. Sentencias de Encabezados:

Se utilizan para imprimir hasta 10 líneas de título en las páginas impresas por procedimientos SAS. Su forma general es:

TITLEn título;

El valor de n puede ser cualquier número entero entre 1 y 10, TITLE y TITLE1 son equivalentes. Hasta 132 caracteres pueden usarse en cada TITLE. Cuando SAS encuentra una sentencia TITLEn, SAS cancela todas las sentencias TITLEn con valores mayores que n. Sentencias de TITLE asociadas con un procedimiento deben escribirse después de la sentencia PROC.

1.9. Sentencias de Comentario:

Son de la forma *comentario;

El comentario puede tener cualquier longitud y contener cualquier carácter excepto punto y coma.

1.10. Sentencias para transformar y crear nuevas Variables:

Ejemplo: DATA C1;

INPUT A B @@;

Sentencias de
programa : X=A+B; A=B-A; R=SQRT (X**2+A**2);

CARDS;

Tarjeta de datos : 1 3 2 4 3 5

Resultado: El conjunto de datos C1 consiste de 4 variables A B X R cuya estructura es: A B X R

2	3	4	$\sqrt{20}$
2	4	6	$\sqrt{40}$
2	5	8	$\sqrt{68}$

Nota: Las expresiones aritméticas SAS siguen reglas similares a las de la aritmética FORTRAN. Los operadores son:

+ (adición), -(substracción), *(multiplicación), /(división), ** (exponenciación). En el apéndice 1 del manual completo de SAS se describen todas las funciones aritméticas SAS. En la siguiente tabla se indican las funciones aritméticas SAS más usadas:

FUNCIONES SAS MAS UTILIZADAS

<u>Nombre SAS</u>	<u>Significado</u>	<u>Forma</u>
ABS	Valor absoluto	ABS (argumento)
SQRT	Raíz cuadrada	SQRT (argumento no negativo)
ARSIN	Arco seno	ARSIN (argumento entre -1 y 1)
SIN	Seno trigonométrico	SIN (argumento en radianes)

<u>Nombre SAS</u>	<u>Significado</u>	<u>Forma</u>
EXP	Función exponencial en base e	EXP (argumento)
LOG	Logaritmo natural	LOG (argumento positivo)
LOG10	Logaritmo decimal	LOG10 (argumento positivo)

En el apéndice A se presenta la salida del job que contiene algunos de los ejemplos de este capítulo. La primera parte reproduce las tarjetas de control y un resumen de los recursos de computador utilizados en el job. A continuación se reproduce el listado del programa SAS con mensajes informativos intercalados. Finalmente, aparecen los resultados solicitados por el programa SAS.

Para referencia, se tabulan a continuación los operadores utilizados en SAS para formar expresiones.

Operadores Aritméticos

<u>Símbolo</u>	<u>Operación</u>	<u>Ejemplo</u>
**	Exponenciación	Y=X**2 ;
*	Multipliación	A= B*C ;
/	División	G= H/I ;
+	Adición	R= S+T ;
-	Substracción	U= V-X ;

Operadores de Comparación

<u>Símbolo</u>	<u>Abreviación</u>	<u>Comparación</u>
<	LT	Menor que
< =	LE	Menor que o igual a
>	GT	Mayor que
> =	GE	Mayor que o igual a
=	EQ	Igual a
≠	NE	Diferente de

Operadores Lógicos

<u>Símbolo</u>	<u>Abreviación</u>	<u>Significado</u>
&	AND	Conjunción
	OR	Disyunción

1.11. Sentencias de Programa

Son sentencias SAS que van después de la sentencia INPUT y antes de la sentencia CARDS. Sirven muchos propósitos como crear nuevas variables, omitir observaciones, proveer a SAS información extra acerca del archivo que se está creando, cambiar el orden en que SAS ejecuta los comandos.

Para entender como usar las sentencias de programas, es necesario saber como SAS crea un archivo de datos.

Cuando SAS crea un conjunto de datos, para cada observación se pasa por los siguientes pasos:

1. SAS usa la descripción en la sentencia de INPUT para leer la observación.
2. SAS usa los valores leídos en esta observación para ejecutar cualquier sentencia de programa que se presente.
3. SAS añade la observación al archivo que se está creando.

EJEMPLO: Con referencia al ejemplo 7 de la sección 1.3.

DATA;

INPUT NOMBRE \$ SEXO \$ EDAD ALTURA PESO;

PESOKG = PESO *.45; * PESOKG, PESO EN KGS;

ALTURA = ALTURA /12; *ALTURA, ALTURA EN PIES;

IF EDAD LT 12 OR EDAD GE 14 THEN LIST;

* LISTA LAS OBSERVACIONES PARA LAS QUE EDAD ES MENOR QUE 12
Y AQUELLAS PARA LAS QUE EDAD ES MAYOR O IGUAL A 14;

IF SEXO='M' AND EDAD GT 15 THEN DELETE;

* NO ENTRAN AL ARCHIVO LAS OBSERVACIONES DE VARONES MAYORES
DE 15 AÑOS;

CARDS;

datos →

* Aquí vendrían los procedimientos de manipulación de datos
y de análisis que se desearan;

2. Conceptos generales de los procedimientos SAS

2.1. Introducción:

Después de haber creado un conjunto de datos (ó archivo) SAS, es posible usar los procedimientos SAS para imprimir, ordenar, graficar y analizar estadísticamente los datos.

Los procedimientos SAS se llaman mediante instrucciones de la forma: PROC nombre_del_procedimiento;

Asociados con cada instrucción o sentencia de procedimiento SAS existen opciones y parámetros para especificar información adicional sobre lo que se desea. Para una información detallada de las opciones y parámetros de cada procedimiento, ver el manual completo. En esta guía simplificada sólo incluiremos los procedimientos más utilizados y de ellos las opciones y parámetros más usuales.

Por otro lado, una instrucción PROC es a menudo seguida por una o más sentencias que proveen información adicional. Exceptuando a la sentencia BY, la presencia de otras sentencias, tales como VARIABLES, CLASSES, ID, WEIGHT y OUTPUT, depende del procedimiento bajo consideración.

2.2. Sentencias o Instrucciones BY:

Cuando se usa una sentencia BY con un procedimiento SAS, SAS divide el conjunto de datos en subconjuntos de acuerdo

a los valores de las variables en la sentencia BY y a continuación aplica el procedimiento a cada subconjunto por separado. Por esta razón cuando se usa una sentencia BY, SAS espera que el conjunto de datos esté ordenado según las variables del BY.

3. Procedimientos SAS más utilizados:

En el capítulo 1 se incluyeron los procedimientos de utilidad SORT y PRINT. A continuación se presentan versiones resumidas de los procedimientos estadísticos SAS más utilizados, ilustrándolos con ejemplos que presenten sus características más saltantes. La relación completa de los procedimientos disponibles en SAS76 pueden obtenerla de los manuales "A User's Guide to SAS 76" y "SAS Supplementary User's Guide".

Procedimiento MEANS:

Este procedimiento produce estadísticos descriptivos univariados para variables numéricas. Su forma general más usual es:

```
PROC MEANS {opciones_y_parámetros};  
{VARIABLES lista_de_variables;} {BY lista_de_variables;}  
{OUTPUT OUT=nombre_del_conjunto_de_salida  
          opción_1=lista_de_variables_1  
          :  
          opción_n=lista_de_variables_n;} (*)
```

Las llaves encierran expresiones opcionales. Y aquellas con (*) pueden utilizarse más de una vez dentro del mismo PROC MEANS. Algunas de las opciones y parámetros que pueden incluirse en la instrucción PROC MEANS son:

DATA=nombre_del_conjunto_de_entrada, informa a SAS el nombre del conjunto de datos a ser usado; si se le omite, SAS usa el conjunto de datos creado más recientemente,

NOPRINT Solicita no imprimir ninguna salida, cuando el único objetivo es crear un archivo de salida.

La instrucción "BY lista_de_variables" se usa si se desea hacer el análisis para cada subconjunto definido por las combinaciones diferentes de valores de las variables del BY.

La instrucción "VARIABLES etc." solicita que sólo se procesen las variables numéricas listadas en tal instrucción.

La instrucción "OUTPUT etc." solicita la creación de un archivo de salida conteniendo los valores de los estadísticos solicitados en las opciones 1 a n. Algunas de estas opciones son N (número de observaciones utilizadas en el cálculo de los estadísticos), MEAN (media), STD (desviación estándar), MIN (el mínimo valor), MAX (el máximo valor), STDERR (error estándar de la media), SUM (suma), VAR (varianza), y CV (coeficiente de variación).

Ejemplo 1: Procesando todas las variables numéricas del último conjunto de datos.

PROC MEANS ;

Esta instrucción pide procesar todas las observaciones y todas las variables numéricas del conjunto de datos creados más recientemente, e imprime, para cada variable numérica, el número de observaciones utilizadas en el cálculo de los estadísticos (N), la media (MEAN), la desviación estándar de las observaciones individuales (STD), el valor mínimo (MIN), el valor máximo (MAX), el error estándar de la media (STDERR), la suma (SUM), la varianza (VAR) y el coeficiente de variación (CV).

Ejemplo 2: Obteniendo algunos estadísticos de algunas variables de un conjunto de datos que no es el creado más recientemente.

Supongamos que no se desea procesar el último conjunto de datos que se creó sino uno anterior llamado FIRST; además supongamos que sólo deseamos el número de observaciones, la media, la varianza y el coeficiente de variación de las variables RENDXHA y RENDXPAR. En este caso las instrucciones SAS requeridas son:

```
PROC MEANS DATA=FIRST N MEAN VAR CV ;  
VARIABLES RENDXHA RENDXPAR ;
```

Ejemplo 3: Obteniendo un análisis para cada combinación (AÑO, LUGAR).

Supongamos ahora que el conjunto de datos FIRST contiene datos de 3 años en 5 lugares y que deseamos las mismas estadísticas del ejemplo anterior, pero para cada una de las 15 combinaciones (AÑO, LUGAR). Esto se lograría con las siguientes instrucciones:

```
DATA; SET FIRST;  
PROC SORT; BY AÑO LUGAR;
```

* EL SORT ANTERIOR NOS PERMITE USAR LA SENTENCIA
BY AÑO LUGAR EN EL SIGUIENTE PROC MEANS;

PROC MEANS N MEAN VAR CV;

VARIABLES RENDXHA RENDXPAR; BY AÑO LUGAR;

Ejemplo 4: Creando un archivo de estadísticos.

PROC MEANS DATA=FIRST NOPRINT;

VARIABLES RENDXHA RENDXPAR ; BY AÑO LUGAR;

OUTPUT OUT=ESTDES MEAN=MRENDXHA MRENDXPA

STD =SRENDXHA;

* EL ARCHIVO ESTDES CONTIENE 15 OBSERVACIONES,
CADA UNA CONTENIENDO LAS VARIABLES MRENDXHA
MRENDXPA SRENDXHA AÑO Y LUGAR ;

Procedimiento FREQ

Este procedimiento permite obtener desde tablas de una entrada hasta tablas cruzadas de dos o más entradas tanto para variables numéricas como para variables alfanuméricas.

Su forma general más usual es:

```
PROC FREQ {DATA=nombre_del_conjunto_de_datos};
```

```
TABLES tabla_deseada_1
```

```
.
```

```
.
```

```
tabla_deseada_n {/opciones_y_parámetros};(*)
```

Las expresiones entre llaves son opcionales y aquellas con (*) pueden utilizarse más de una vez dentro del mismo PROC FREQ.

La expresión "DATA=nombre_del_conjunto_de_datos" informa a SAS el nombre del conjunto de datos a ser usado; si se le omite, SAS usa el conjunto de datos creado más recientemente.

Cada instrucción TABLES solicita una o más tablas. Cada "tabla_deseada" es de la forma nombre_de_variable_1 * nombre_de_variable_2 *... dependiendo del número de entradas que se desee. Las opciones y parámetros más útiles de la instrucción TABLES son:

EXPECTED Solicita la impresión de la frecuencia esperada en cada celda bajo la hipótesis de independencia.

CHISQ Solicita la impresión de los resultados de la prueba de chi-cuadrado de independencia de las tablas de doble entrada pedidas en la instrucción TABLES.

NOFREQ Solicita no imprimir las frecuencias de las celdas
NOPERCENT Solicita no imprimir los porcentajes de frecuencia
 total de las celdas

NOROW (NOCOL) Solicita no imprimir los porcentajes de frecuencia de fila (columna) de las celdas.

OUT=nombre_del_conjunto_de_datos_de_salida, solicita la creación de un nuevo conjunto de datos conteniendo los valores de las variables y conteos de frecuencia de la última "tabla_deseada_n" de la instrucción TABLES. El nuevo conjunto de datos contiene una observación por cada combinación de valores de las variables. Cada observación contiene estas variables mas una variable llamada COUNT que da la frecuencia para aquella combinación.

NOPRINT Solicita no imprimir ninguna salida, cuando el único objetivo es crear un archivo de salida.

Ejemplo 1: Tabla de una entrada o tabla de distribución de frecuencias.

PROC FREQ;

TABLES ESTCIV; TABLES SEXO;

El resultado lo constituyen dos tablas obtenidas procesando el último archivo creado. La primera

tabla muestra la distribución de frecuencias de la variable ESTCIV incluyendo, para cada nivel de ESTCIV, la correspondiente frecuencia absoluta, la frecuencia relativa en %, la frecuencia acumulada absoluta y la frecuencia acumulada relativa en %. La segunda tabla muestra la distribución de frecuencias de la variable SEXO y tiene la misma estructura de la anterior.

Ejemplo 2: Tablas de dos entradas

Supongamos que deseamos saber cuantas variedades del Banco de Promisorios de Fríjol existen en cada combinación de color primario de semilla (COLPSEMI) y hábito de crecimiento (HABITO).

```
PROC FREQ; TABLES COLPSEMI * HABITO;
```

Ejemplo 3: Tabla de contingencia

Supongamos que deseamos investigar si en las variedades del Banco de Promisorios de Fríjol hay una relación de dependencia entre las características color primario de la semilla (COLPSMI) y resistencia a mosaico común (MOSAICOM). La hipótesis nula es de que no hay tal dependencia de modo que la decisión de rechazarla se tomará sólo

si la evidencia es muy grande (es decir, si el valor observado de X^2 es superior al tabulado).

```
PROC FREQ;
```

```
TABLES MOSAICOM * COLPSEMI /EXPECTED CHISQ ;
```

Ejemplo 4: Detección de Celdas Incompletas.

Existen procedimientos SAS que requieren datos balanceados como por ejemplo ANOVA y LATTICE. Supongamos que deseamos determinar las celdas incompletas del IBYAN (International Bean Yield and Adaptation Nursery) para decidir que experimentos no pueden procesarse usando el procedimiento LATTICE.

```
PROC FREQ DATA=IBYAN;
```

```
TABLES EXPNO*GROUP*TREATMNT / OUT=CHK NOPRINT;
```

```
DATA CHK; SET CHK; IF COUNT NE 1 THEN OUTPUT;
```

```
PROC PRINT; TITLE1 CELDAS INCOMPLETAS;
```

Procedimiento CORR:

Este procedimiento produce estadísticas descriptivas univariadas y coeficientes de correlación entre variables. Su forma general más usual es:

```
PROC CORR {opciones y parámetros};  
{VARIABLES lista_de_variables_numéricas};  
{WITH lista_de_variables_numéricas};
```

Las llaves encierran expresiones opcionales. Algunas de las opciones y parámetros que pueden aparecer en la instrucción PROC CORR son:

PEARSON	Solicita coeficientes de correlación de Pearson.
SPEARMAN	Solicita coeficientes de correlación de "rankings" de Spearman.
KENDALL	Solicita coeficientes de correlación de Kendall
NOSIMPLE	Solicita no imprimir estadísticas descriptivas univariadas

OUTP=conjunto_de_datos_2, solicita la creación de un nuevo conjunto de datos denominado conjunto_de_datos_2 que incluye medias, desviaciones estándares, números de valores y coeficientes de correlación de Pearson. Si en lugar de P de la instrucción "OUTP= etc." se coloca S o K, se obtiene un conjunto de datos similar al descrito anteriormente, pero en lugar de coeficientes de correlación de Pearson, se obtienen coeficientes de correlación de Spearman

o Kendall, respectivamente.

DATA=nombre_del_conjunto_de_datos, indica el archivo a usar.

La instrucción "VARIABLES etc." indica que no considere todas las variables numéricas del archivo sino tan sólo las incluidas en la lista.

La instrucción "WITH etc." sólo tiene sentido usarla si se usa la instrucción "VARIABLES etc.", en cuyo caso el procedimiento calculará coeficientes de correlación entre cada variable en la instrucción "VARIABLES etc." y cada variable en la instrucción "WITH etc.".

Ejemplo 1: Procesando todas las variables numéricas del último conjunto de datos:

PROC CORR;

Esta instrucción pide obtener las estadísticas de cada variable numérica y el coeficiente de correlación de PEARSON para cada par de variables numéricas del último conjunto de datos que se creó.

Ejemplo 2: Obteniendo sólo los coeficientes de correlación de algunas variables de un conjunto de datos que no es el creado más recientemente:

```
PROC CORR DATA=FIRST SPEARMAN KENDALL NOSIMPLE;  
VARIABLES RENDXHA RENDXPAR DENSIDAD;
```

Estas instrucciones piden obtener sólo los coeficientes de correlación de PEARSON ("default"), SPEARMAN y KENDALL de cada par de las variables numéricas RENDXHA RENDXPAR DENSIDAD.

Ejemplo 3: Uso de la sentencia WITH

```
PROC CORR DATA=FIRST;  
VARIABLES RENDXHA RENDXPAR;  
WITH N P K;
```

Estas instrucciones piden obtener las estadísticas descriptivas de las variables RENDXHA y RENDXPAR y los 6 coeficientes de correlación de PEARSON correspondientes a las combinaciones (RENDXHA, N), (RENDXHA, P), (RENDXHA, K), (RENDXPAR, N), (RENDXPAR, P) y (RENDXPAR, K).

Procedimiento PLOT

Este procedimiento permite obtener gráficas bidimensionales.

Su forma general más usual es:

```
PROC PLOT {opciones_y_parámetros};{BY lista_de_variables;}  
  PLOT gráfico_deseado_1  
      .  
      .  
      gráfico_deseado_n{/opciones_y_parámetros};
```

Cada "gráfico_deseado" es de la forma

```
variable_vertical * variable_horizontal{='caracter' c  
variable};
```

Las llaves encierran expresiones opcionales.

Ejemplo 1: Obteniendo gráficas separadas de PESO (eje horizontal) y ALTURA (eje vertical) para cada sexo.

```
PROC PLOT DATA=D04.A9 ; BY SEXO;  
  PLOT ALTURA*PESO ;
```

Dado que no se especificó el caracter para representar puntos en el gráfico, SAS usa 'A' para representar puntos únicos, 'B' para representar puntos dobles, etc.

Ejemplo 2: Obteniendo gráficos simultáneos de EDAD (eje horizontal) y PESO (eje vertical) para ambos sexos:

```
PROC PLOT DATA=D04.A9 ;  
  PLOT PESO*EDAD=SEXO ;
```

Ejemplo 3: Obteniendo gráficos simultáneos de EDAD VS PESO y de EDAD VS ALTURA para ambos sexos combinados

```
PROC PLOT DATA=D04.A9 ;  
    PLOT PESO*EDAD='P'  
    ALTURA*EDAD='A'/OVERLAY ;
```

Ejemplo 4: Graficando media (MREND) VS Desviación Estándar (SREND) para verificar homogeneidad de la varianza.

```
PROC PLOT DATA=AVSD ;  
    PLOT SREND*MREND ;
```

Ejemplo 5: Graficando el rendimiento promedio (MREND) como función del nivel de fósforo(P):

Supongamos que el archivo AVSD fué originalmente creado por el procedimiento MEANS aplicado BY ANO al conjunto FIRST que contiene las variables ANO, P y REND. Entonces AVSD sólo contendría las variables ANO MREND y SREND pero no contendría P. Para incluir P en AVSD se podrían haber usado las instrucciones:

```
DATA AVSD ; MERGE FIRST AVSD ; BY ANO ;
```

Ahora estamos listos para solicitar el gráfico:

```
PROC PLOT DATA=AVSD; BY ANO;  
    PLOT P*MREND;
```


Procedimiento ANOVA

Este procedimiento sirve para analizar diseños balanceados en una forma más eficiente en la utilización de memoria y tiempo de computador que el procedimiento GLM. Aunque este procedimiento funciona para diseños no balanceados, su salida es de validez cuestionable.

Su forma general es:

```
PROC ANOVA {DATA=nombre_del_conjunto_de_datos};  
  {BY lista_de_variables ;}  
  {CLASSES lista_de_variables ;}  
  MODEL lista_de_variables_dependientes=variables_o_efectos_  
    independientes;  
  {TEST H=efectos_bajo_hipótesis E=efectos_de_error;}(*)  
  {MANOVA H=efectos_bajo_hipótesis E=efecto_de_error{/opciones_y_  
    parámetros;}}(*)  
  {MEANS lista_de_efectos{/opciones_y_parámetros;}}(*)
```

Las expresiones entre llaves son opcionales y aquellas con (*) pueden utilizarse más de una vez dentro del mismo PROC ANOVA.

La expresión "DATA=nombre_del_conjunto_de_datos" informa a SAS el nombre del conjunto de datos a ser usado; si se le omite, SAS usa el conjunto de datos creado más recientemente.

La instrucción "BY lista_de_variables" se usa si se desea hacer el análisis para cada subconjunto definido por las combinaciones diferentes de valores de las variables del BY. La instrucción "CLASSES lista_de_variables" se utiliza para indicar a SAS que las variables categóricas de la lista definen factores

en el modelo. La instrucción "TEST H=etc." solicita en forma simbólica hacer la prueba de una hipótesis "adicional" (en el sentido de que su denominador no es la línea de error del modelo completo) que se desea probar.

La instrucción "MODEL etc." define en forma simbólica el modelo a utilizar. La instrucción "MANOVA etc." solicita un análisis de varianza multivariado donde el parámetro H especifica las matrices de hipótesis y el parámetro E especifica la matriz de error. Las matrices de hipótesis de cada efecto se solicitan usando la opción PRINTH y la matriz de error usando la opción PRINTE. La instrucción MEANS permite obtener las medias de todas las variables continuas para cada nivel de cualquier efecto que involucre variables de clase y que esté o no incluido en el modelo. Sus opciones y parámetros más útiles son:

DUNCAN pide realizar la prueba de rango múltiple de Duncan para las medias de cada efecto principal listado en la instrucción MEANS.

ALPHA=.nn pide utilizar el nivel de significancia .nn en la prueba de Duncan. El "default" es .05 y las otras posibilidades son .01 y .1

E=efecto especifica el cuadrado medio de error a utilizar en la prueba de Duncan; si se omite este parámetro, SAS usa el cuadrado medio del residuo.

Ejemplo 1: Modelo ANOVA conteniendo sólo efectos principales A, B y C.

```
PROC ANOVA; CLASSES A B C ; MODEL Y=A B C ;
```

Ejemplo 2: Modelo ANOVA conteniendo efectos principales A, B y C y todas sus interacciones.

```
PROC ANOVA; CLASS A B C ;
```

* NOTE QUE LA PALABRA CLAVE CLASSES PUEDE TAMBIEN
ESCRIBIRSE CLASS;

```
MODEL Y=A B C A*B A*C B*C A*B*C ;
```

Ejemplo 3: Modelo ANOVA conteniendo efectos anidados.

```
PROC ANOVA; CLASSES A B C ;
```

```
MODEL Y=A B(A) C(A B);
```

* LA NOTACION B(A) INDICA QUE B ESTA ANIDADO
DENTRO DE A;

Ejemplo 4: Diseño Balanceado en Parcelas Divididas (Split plot design)

```
PROC ANOVA; CLASSES REP A B ;
```

```
MODEL Y = REP A REP*A B A*B;
```

* LAS HIPOTESIS DE A Y A*B NO ES NECESARIO SOLI-
CITARLAS EXPLICITAMENTE MEDIANTE INSTRUCCIONES

TEST PUES SON HECHAS CONTRA EL ERROR DEL MODELO (ERROR (B)). POR OTRO LADO LA PRUEBA DEL FACTOR A NO ES CONTRA EL ERROR (B) SINO CONTRA EL ERROR (A)=REP*A. ESTA PRUEBA SE PIDE DE LA MANERA SIGUIENTE;

TEST H = A E = REP*A ;

Ejemplo 5: Diseño Balanceado en Parcelas Sub-divididas (Split-split plot design).

```
PROC ANOVA; CLASSES REP A B C ;
MODEL Y=REP A REP*A B A*B REP*B(A) C A*C
B*C A*B*C ;
TEST H=A E=REP*A ;
TEST H=B A*B E=REP*B(A) ;
```

Ejemplo 6: Diseño Balanceado en Franjas Divididas (Split-block design)

```
PROC ANOVA;CLASSES REP A B ;
MODEL Y=REP A REP*A B REP*B A*B ;
TEST H=A E=REP*A ;
TEST H=B E=REP*B ;
```

Procedimiento GLM: (General Linear Model)

Este procedimiento usa el principio de mínimos cuadrados para ajustar un modelo de efectos fijos a prácticamente cualquier tipo de datos. Permite análisis univariados y multivariados, incluyendo regresión lineal simple y múltiple, análisis de varianza y análisis de covarianza.

Las instrucciones asociadas a GLM mayormente usadas son:

```
PROC GLM {DATA=nombre_del_conjunto_de_datos};
{BY lista_de_variables;}
{CLASSES lista_de_variables;}
MODEL lista_de_variables_dependientes = variables_o_
      efectos_independientes (/ lista_de_opciones);
{TEST H = efectos_bajo_hipótesis E = efecto_de_error /
      opciones;}
{MANOVA H = efectos_bajo_hipótesis E = efecto_de_error
      (/opciones_y_parámetros;}}(*)
{MEANS lista_de_efectos / opciones_y_parámetros;}}(*)
{OUTPUT OUT=nombre_del_conjunto_de_salida PREDICTED=
      lista_de_predichos
      RESIDUAL=lista_
      de_residuos;}
```

Las expresiones entre paréntesis no son indispensables y aquellas con (*) pueden utilizarse más de una vez dentro del mismo PROC GLM. La expresión "DATA=nombre_del_conjunto_de_datos" informa a SAS el nombre del conjunto de datos a ser

usado; si se le omite, SAS usa el conjunto de datos creado más recientemente. La instrucción "BY lista_de_variables" se usa si se desea hacer el análisis para cada subconjunto definido por las combinaciones diferentes de valores de las variables del BY. La instrucción "CLASSES lista_de_variables" se utiliza para indicar a SAS que las variables presentes en la lista definen posibles factores en el modelo. La instrucción "TEST H=etc." solicita en forma simbólica hacer la prueba de una hipótesis "adicional" (en el sentido de que su denominador no es la línea de error del modelo completo) que se desea probar; después de "/" debe incluirse uno o ambos de los siguientes parámetros:

HTYPE=n, especifica alguna de las sumas de cuadrados que se utilizó en el modelo (en consecuencia n debe ser 1, 2, 3, ó 4)

ETYPE=n, idem a lo anterior.

La instrucción "MODEL etc." define en forma simbólica el modelo lineal a utilizar. Las opciones más útiles de la instrucción MODEL son:

NOINT, indica que se desea un modelo sin intercepto.

P , pide imprimir los valores observados, predichos y residuales de las variables dependientes y el estadístico de Durbin-Watson.

SSL , pide utilizar la suma de cuadrados tipo I ("apropiada" para modelos balanceados, puramente anidados y regre-

sión polinómica).

SS3 , pide utilizar suma de cuadrados tipo III (para diseños no balanceados con celdas vacías, esta suma de cuadrados coincide con la que se obtendría utilizando el modelo de Harvey con efectos fijos). Si no se especifica el tipo de suma de cuadrados, GLM imprime las sumas de cuadrados tipo I y tipo IV (esta última es en muchos casos igual a la suma de cuadrados tipo III).

La instrucción "MANOVA etc." solicita un análisis de varianza multivariado donde el parámetro H especifica las matrices de hipótesis y el parámetro E especifica la matriz de error. Las matrices de hipótesis de cada efecto se solicitan usando la opción PRINTH y la matriz de error usando la opción PRINTE. La instrucción MEANS permite obtener las medias de todas las variables continuas para cada nivel de cualquier efecto que involucre variables de clase y que esté o no incluido en el modelo. Sus opciones y parámetros más útiles son:

DUNCAN pide analizar la prueba de rango múltiple de Duncan para las medias de cada efecto principal listado en la instrucción MEANS.

ALPHA=.nn pide utilizar el nivel de significancia .nn en la prueba de Duncan. El "default" es .05 y las otras posibilidades son .01 y .1

E=efecto especifica el cuadrado medio de error a utilizar en la prueba de Duncan; si se omite este parámetro,

SAS usa el cuadrado medio del residuo

La instrucción "OUTPUT etc." solicita la creación de un nuevo conjunto de datos que contiene todas las variables en el conjunto original más aquellas variables conteniendo valores predichos y residuos. El parámetro OUT da el nombre del nuevo conjunto de salida; si se le omite, SAS lo denomina siguiendo la convención DATAn.

Ejemplo 1: Regresión Lineal Simple de Y en X con Intercepto.

```
PROC GLM; MODEL Y=X;
```

Ejemplo 2: Regresión Lineal Simple de Y en X sin Intercepto.

```
PROC GLM; MODEL Y=X/NOINT;
```

Ejemplo 3: Regresión Lineal Múltiple de Y en X1 y X2.

```
PROC GLM; MODEL Y=X1 X2/P;
```

* NOTE QUE SE ESTA PIDIENDO IMPRIMIR VALORES OBSERVADOS, PREDICHOS Y RESIDUOS DE LA VARIABLE
Y;

Ejemplo 4: Modelo ANOVA conteniendo sólo efectos principales A, B y C.

```
PROC GLM; CLASSES A B C; MODEL Y=A B C ;
```


Ejemplo 5: Modelo ANOVA conteniendo efectos principales A, B y C y todas sus interacciones.

```
PROC GLM; CLASS A B C ;  
* NOTE QUE LA PALABRA CLAVE CLASSES PUEDE TAMBIEN  
  ESCRIBIRSE CLASS;  
MODEL Y=A B C A*B A*C B*C A*B*C ;
```

Ejemplo 6: Modelo ANOVA conteniendo efectos anidados.

```
PROC GLM; CLASSES A B C ;  
MODEL Y=A B(A) C(A B) ;  
* LA NOTACION B(A) INDICA QUE B ESTA ANIDADO  
  DENTRO DE A;
```

Ejemplo 7: Modelos ANCOVA (Análisis de Covarianza)

```
PROC GLM; CLASS A; MODEL Y=A X;  
* MODELO CON UNA PENDIENTE COMUN EN TODOS LOS  
  NIVELES DE A ;  
PROC GLM; CLASS A ; MODEL Y=A X(A);  
* MODELO CON PENDIENTES SEPARADAS PARA CADA NIVEL  
  DE A;
```

Ejemplo 8: Diseño Balanceado en Parcelas Divididas (Split plot design).

```
PROC GLM; CLASSES REP A B ;  
MODEL Y = REP A REP*A B A*B ;
```

* LAS HIPOTESIS DE A Y A*B NO ES NECESARIO SOLICITARLAS EXPLICITAMENTE MEDIANTE INSTRUCCIONES TEST PUES SON HECHAS CONTRA EL ERROR DEL MODELO (ERROR (B)). POR OTRO LADO LA PRUEBA DEL FACTOR A NO ES CONTRA EL ERROR (B) SINO CONTRA EL ERROR (A)=REP*A. ESTA PRUEBA SE PIDE DE LA MANERA SIGUIENTE;

```
TEST H = A E = REP*A/HTYPE=1 ETYPE=1;
```

Ejemplo 9: Diseño Balanceado en Parcelas Sub-divididas (Split-split plot design).

```
PROC GLM; CLASSES REP A B C ;  
MODEL Y=REP A REP*A B A*B REP*B(A) C A*C  
B*C A*B*C ;  
TEST H=A E=REP*A/HTYPE=1 ETYPE=1 ;  
TEST H=B A*B E=REP*B(A)/HTYPE=1 ETYPE=1;
```

Ejemplo 10: Diseño Balanceado en Franjas Divididas (Split-block design).

```
PROC GLM; CLASSES REP A B ;  
MODEL Y=REP A REP*A B REP*B A*B ;  
TEST H=A E=REP*A/HTYPE=1 ETYPE=1 ;  
TEST H=B E=REP*B/HTYPE=1 ETYPE=1 ;
```

Ejemplo Completo: Estimación del Rendimiento de Fríjol en Kg/Ha. en función de la cantidad de P utilizado (Municipio de

Quixadá, Estado de Ceará, Brasil).

DATA QUIXADA; AÑO=1968; DROP CONTADOR;

L1: AÑO+1; IF AÑO>1974 THEN DELETE; CONTADOR=0 ;

L2: CONTADOR+1; IF CONTADOR>27 THEN GO TO L1 ;

INPUT REND P @@; OUTPUT; GO TO L2; CARDS;

521	0	548	0	546	0	505	60	455	60	460	60
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
625	0	819	0	450	0	1100	60	1388	60	1575	60
866	120	1366	120	850	120						

PROC PRINT N; BY AÑO ;

TITLE DATOS DE RENDIMIENTO DE FRIJOL, QUIXADA, CEARA,
BRASIL;

PROC GLM; BY AÑO; MODEL REND = P P*P ;

TITLE2 REGRESION DE REND EN P Y P**2 PARA CADA AÑO ;

Procedimiento DUNCAN:

Este procedimiento permite realizar la prueba de Duncan de rango múltiple. El usuario especifica la variable de clasificación a utilizar para separar las observaciones en grupos en base a los valores de esta variable de clasificación. Su forma general más usual es:

```
PROC DUNCAN {DATA=nombre_del_conjunto_de_datos} {ALPHA=.--}  
            {MULTIPLE};  
  
CLASSES variable_o_variables_de_clasificación;  
  
VARIABLES lista_de_variables_numéricas;  
  
DF lista_de_grados_de_libertad_de_los_términos_de_error;  
  
MS lista_de_cuadrados_medios_de_las_variables_numéricas;
```

Ejemplo 1: Duncan a $\alpha = .05$ con una sola variable de clasificación y aplicado a una variable numérica del último conjunto de datos.

Supongamos que deseamos aplicar el procedimiento DUNCAN a $\alpha = .05$ para la comparación de las medias de rendimiento (REND) de 5 variedades (VAR) en un experimento con 6 bloques completos al azar. Supongamos que en una corrida anterior obtuvimos que $DF=20, MS=11898.5$. Las instrucciones necesarias son:

```
PROC DUNCAN;  
CLASSES VAR;  
VARIABLES REND;  
DF 20;  
MS 11898.5;
```

El resultado impreso de las instrucciones anteriores da las medias de las cinco variedades en orden de rendimiento. Al lado de cada media sale impreso el nombre o código de la variedad correspondiente. Además, a la izquierda de las medias ordenadas aparece una serie vertical de letras ordenadas. Si dos medias están conectadas por al menos una serie de la misma letra, entonces ellas no son significativamente diferentes. Si no existe ninguna línea que conecte dos medias, entonces tales medias son significativamente diferentes.

Note que cuando el α a utilizar es .05 no es necesaria la opción ALPHA=.--. Los otros valores posibles para α son .01 y .10.

Ejemplo 2: Duncan a α = .01 con dos variables de clasificación, aplicado a tres variables numéricas de un conjunto de datos que no es el creado últimamente.
Supongamos que deseamos aplicar el procedimiento

DUNCAN a $\alpha = .01$ para la comparación de las medias en las celdas (a_i, b_j) de las variables X, Y, Z del conjunto de datos EXP76_1 de un experimento para el cual una corrida anterior nos dió la siguiente información: $DF_X=DF_Z=24, DF_Y=23$, $MS_X=6.656, MS_Y=5.834, MS_Z=9.123$. Las instrucciones requeridas son:

```
PROC DUNCAN DATA=EXP76_1 ALPHA=.01 MULTIPLE;
CLASSES A B;
VARIABLES X Y Z;
DF 24 23 24;
MS 6.656 5.834 9.123;
```

Note que debido a que hay más de una variable en la instrucción CLASSES es necesario usar la opción MULTIPLE. Además note que tanto en la instrucción DF como en la MS, el primer valor está asociado con la primera variable; X en la lista de la instrucción VARIABLES; en forma similar, el segundo valor corresponde a la variable Y y el tercer valor a la variable Z.

Nota: En la versión SAS 76.5 los procedimientos ANOVA y GLM tienen incorporada dentro de la instrucción MEANS la opción DUNCAN que solicita la prueba de Duncan de rango múltiple para todos los efectos principales dados

en la instrucción MEANS. De este modo ya no es necesario hacer dos corridas para hacer la prueba de Duncan, sino solicitarla en el mismo procedimiento de análisis de varianza. Sin embargo, se corre el riesgo de hacer una prueba de Duncan aunque la prueba de F del efecto bajo consideración no sea significativa. Por supuesto, tal prueba carecería de valor estadístico.

Procedimiento TTEST

Este procedimiento calcula estadísticos t para pruebas relacionadas a la hipótesis de que las medias de dos grupos de datos son iguales. Produce dos tipos de estadísticos t: asumiendo que las varianzas de ambos grupos son iguales y asumiendo que no lo son. Para cada caso, da los grados de libertad y niveles de significancia. Además da el estadístico F para la prueba de homogeneidad de ambas varianzas. Su forma general es:

```
PROC TTEST {DATA=nombre_del_conjunto_de_datos};  
  CLASS variable_de_agrupación ;  
  {VARIABLES lista_de_variables;}
```

Las expresiones entre llaves son opcionales. La expresión "DATA=etc." indica el nombre del archivo a procesarse; si se le omite, SAS usa el archivo más reciente. La instrucción "CLASS etc." indica la variable que diferencia los dos grupos. La instrucción "VARIABLES etc." pide no procesar todas las variables numéricas sino tan sólo aquellas en la lista.

Ejemplo:

```
DATA A8; INPUT NOMBRE $ SEXO $ EDAD ALTURA PESO;  
CARDS;
```

datos →

PROC TTEST DATA=A8; CLASS SEXO;

VARIABLES ALTURA PESO;

* DE NO INCLUIR LA INSTRUCCION "VARIABLES ETC." SE HARIA UN
ANALISIS SEPARADO PARA EDAD ALTURA Y PESO;

Procedimiento STEPWISE:

Dados un conjunto de variables independientes y una variable dependiente, este procedimiento puede encontrar cuales de las variables independientes deberían incluirse en el modelo de regresión de acuerdo a uno de los cinco criterios de "optimalidad" descritos más abajo.

Su forma general es:

```
PROC STEPWISE {DATA=nombre_del_conjunto_de_datos};  
MODEL lista_de_variables_dependientes=lista_de_variables_  
independientes [/opciones_y_parámetros]; (*)
```

Las expresiones entre llaves son opcionales y aquellas con (*) pueden utilizarse más de una vez dentro del mismo PROC STEPWISE. La expresión "DATA=nombre_del_conjunto_de_datos" informa a SAS el nombre del conjunto de datos a ser usado; si se le omite, SAS usa el conjunto de datos creados más recientemente. La instrucción "MODEL etc." identifica a la colección de variables independientes y a la variable o variables dependientes. STEPWISE hará calculos separados para cada variable dependiente dentro de cada instrucción MODEL. Las opciones y parámetros más útiles de la instrucción MODEL son:

NOINT solicita modelo sin intercepto

FORWARD solicita aplicar el criterio de "optimalidad" de selección hacia adelante: encuentra primero el

modelo de una sola variable independiente con máximo R^2 . Luego, selecciona entre las otras variables independientes, la de mayor F, para incluirla como segunda variable independiente en el modelo si tal F es significativo (al nivel deseado para entrar al modelo). El proceso continúa añadiendo al modelo variables una a una hasta que ninguna produzca un F significativo.

BACKWARD solicita aplicar el criterio de "optimalidad" de eliminación hacia atrás: empieza con el modelo que incluye todas las variables independientes. Luego, se elimina la variable independiente con la menor contribución al modelo. El proceso continúa eliminando del modelo variables una a una hasta que todas las variables restantes produzcan valores "parciales" de F significativos (al nivel deseado para permanecer en el modelo).

STEPWISE pide aplicar como criterio de "optimalidad" una variante de la técnica de selección hacia adelante en la cual, después de que una variable se incluye en el modelo, se realiza un chequeo de cada una de las variables ya incluidas y si no produce un valor "parcial" de F significativo es eliminada del modelo. El proceso continúa añadiendo variables una a

una hasta que ninguna produzca un F significativo. En cierto modo, este criterio es una mezcla de los dos criterios anteriores y es la opción "default".

MAXR indica que el criterio de "optimalidad" a usar es el denominado máximo incremento en R^2 , el cual primero busca el modelo de una variable con máximo R^2 , luego busca el modelo de dos variables con máximo R^2 , y así sucesivamente.

MINR indica que el criterio de "optimalidad" a usar es el de mínimo incremento en R^2 , el cual es una variante del anterior y usualmente resulta en el mismo "mejor" modelo.

SEENTRY=.nn indica el nivel de significancia para entrar al modelo. El valor "default" es .50.

SLSTAY=.nn indica el nivel de significancia para permanecer en el modelo. El valor "default" es .10

INCLUDE=i indica que en el proceso de búsqueda del modelo "óptimo" se incluyen de todas maneras las primeras i variables independientes en la instrucción MODEL.

Los siguientes dos parámetros sólo aplican a las técnicas

MAXR y MINR:

START=m

STOP=n

indican que el proceso de búsqueda del modelo "óptimo" se inicia con las primeras m variables en la

instrucción MODEL y termina cuando se haya llegado al modelo "óptimo" conteniendo n variables.

Procedimiento NLIN:

Este procedimiento usa el principio de mínimos cuadrados (simples o ponderados) para obtener estimativos de los coeficientes de un modelo no lineal. Dado que los estimativos iniciales se obtienen por una búsqueda exhaustiva en una malla y luego son mejorados por métodos iterativos, este procedimiento es relativamente costoso y en general sólo debería usarse para ajustar modelos a datos provenientes de experimentos confirmatorios de gran precisión.

La forma general más usual de NLIN es:

```
PROC NLIN {DATA=nombre_del_conjunto_de_datos}
          {BEST=n}
          ;
{BY lista_de_variables;}
PARAMETERS nombre_del_parámetro_1=rango
          .
          .
          ;
{BOUNDS lista_de_expresiones ;}
{instrucciones SAS para crear variables temporales}
MODEL variable_dependiente=expresión;
DER.nombre_del_parámetro_1=expresión;...;
{{OUTPUT OUT=nuevo_conjunto_de_datos{/parámetros}};
```

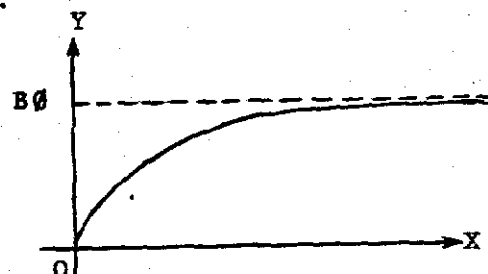
Las expresiones entre llaves son opcionales. La expresión "DATA=nombre_del_conjunto_de_datos" informa a SAS el nombre

del conjunto de datos a ser usado; si se le omite, SAS usa el conjunto de datos creado más recientemente. La opción BEST=n solicita que sólo se impriman las sumas de cuadrados de residuos de las n combinaciones iniciales de los parámetros que tengan las menores S.C. de residuos. La instrucción "BY lista_de_variables" se usa si se desea usar el procedimiento para cada subconjunto definido por las combinaciones diferentes de valores de las variables del BY. La instrucción "PARAMETERS etc." identifica los parámetros a ser estimados y define sus valores iniciales a través de la especificación "rango". Cada expresión de la instrucción "BOUNDS etc." restringe el estimativo del parámetro involucrado al rango especificado en la expresión. La instrucción "MODEL etc." define el modelo a ajustar. Cada instrucción "DER.etc." contiene la expresión algebraica de la derivada parcial de la expresión en la instrucción MODEL con respecto al parámetro cuyo nombre aparece inmediatamente después de los caracteres "DER.". La instrucción "OUTPUT etc." solicita crear el conjunto de datos SAS denominado nuevo_conjunto_de_datos, el cual contiene todas las variables en el conjunto de datos al que se aplica NLIN y además las variables cuyos nombres se dan en los parámetros utilizados. Los parámetros más útiles son:

PREDICTED = nombre_de_la_variable_de_valores_predichos
 RESIDUAL = nombre_de_la_variable_de_residuos

Ejemplo 1: Ajuste de una exponencial creciente asintótica
 que pase por el origen.

DATA NLIN1;
 INPUT X Y ;
 CARDS;
 datos →

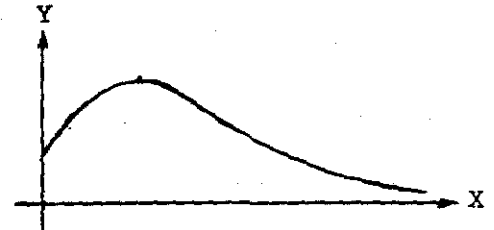


```
TITLE1 AJUSTE DEL MODELO  Y=B0*(1-EXP(-B1*X));
PROC NLIN  BEST=10;
PARAMETERS  B0=0  TO  10
              B1= ./01 TO .09 BY .005;
MODEL  Y=B0*(1-EXP(-B1*X));
DER.B0=1-EXP(-B1*X);
DER.B1=B0*X*EXP(-B1*X);
OUTPUT OUT=PRERE PREDICTED=YHAT RESIDUAL=YRES;
TITLE2  OBTENIENDO ESTIMATIVOS DE LOS PARAMETROS,
PREDICHOS Y RESIDUOS;
PROC PLOT DATA=PRERE;
TITLE2  EXAMEN DE PREDICHOS Y RESIDUOS;
PLOT  YRES * X  /VREF=0;
PLOT  Y*X='O'  YHAT*X='P'/OVERLAY;
```


Ejemplo 2: Ajuste de una exponencial con intercepto positivo, un máximo absoluto y asintótica al eje horizontal.

DATA NLIN2;
INPUT X Y;
CARDS;

datos →



TITLE1 AJUSTE DEL MODELO $Y = \exp(B_0 + B_1 \cdot X + B_2 \cdot X^2 + B_3 \cdot X^3)$;

PROC NLIN DATA=NLIN2 BEST=10;

PARAMETERS $B_0 = 0, 5, 10$

$B_1 = 0$ TO 3

$B_2 = -.5$ TO $-.1$ BY .1

$B_3 = -.09$ TO $-.01$ BY .01 ;

* PARA AHORRAR TIEMPO DE PROCESAMIENTO DEFINAMOS
LA SIGUIENTE VARIABLE TEMPORAL;

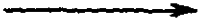
$E = \exp(B_0 + B_1 \cdot X - B_2 \cdot X^2 + B_3 \cdot X^3)$;

MODEL $Y = E$; DER. $B_0 = E$; DER. $B_1 = E \cdot X$;

DER. $B_2 = E \cdot X^2$; DER. $B_3 = E \cdot X^3$;

Ejemplo 3: Linealización de un Modelo No Lineal

El ejemplo anterior es muy probable que demande una cantidad considerable de tiempo y memoria del computador. Un procedimiento alternativo de ajustar un modelo muy similar es ajustar un polinomio cúbico al logaritmo de Y.

```
DATA NLIN3; INPUT X Y; LY=LOG(Y); CARDS;  
datos   
TITLE REGRESION LOG CUBICA DE Y EN X;  
PROC GLM DATA=NLIN3;  
MODEL LY=X X*X X*X*X;  
OUTPUT OUT=PRERES PREDICTED=LYP RESIDUAL=LYR;  
DATA NLIN3; SET PRERES; YEST=EXP(LYP);  
PROC PLOT DATA=NLIN3;  
    PLOT Y * X='O' YEST * X ='E'/OVERLAY;  
    PLOT LY* X='O' LYP * X ='P'/OVERLAY;  
    PLOT LYR*X='R'/HREF=0;  
TITLE GRAFICOS DE OBSERVADOS, ESTIMADOS, PREDICHOS  
Y RESIDUALES DEL MODELO LOG CUBICO;
```

APENDICE A

Salida Impresa del Job Conteniendo
algunos de los Ejemplos del Capítulo 1

GUIA SIMPLIFICADA DE SAS

TARJETAS DE CONTROL Y RESUMEN DE
RECURSOS UTILIZADOS

//CIATERC1 JOB (CIAT,USEX).CIAT0250J.SANDERS.MSGLEVEL=(2.0).

// CLASS=G.

// TIME=2

LOG IFF403J CIATERC1 STARTED TIME=17.40.23 USER=CIAT

LOG IFF281J 16B NOW OFF-LINE

LOG CIATERC1 SAS76 30.43 SEC. 192K. 186 IN. 91 OUT. 000 CP

LOG IFF404J CIATERC1 ENDED TIME=17.51.16

// EXEC SAS76.SORT=

//ET12F001 DD DUMMY

//SYSIN DD * GENERATED STMT

IEF1421 - STEP WAS EXECUTED - COND CODE 0000

IEF373J STEP /SAS76 / START 77188.1740

IEF374J STEP /SAS76 / STOP 77188.1751 CPU 0MIN 30.43SEC STOR VIRT 192K

JOBNAME	STEPNAME	CPU TIME	USED	COMP CODE
CIATERC1	SAS76	30.43 SEC.	192K.	000 CP

IEF298J CIATERC1 SYSOUT=A.

IEF375J JOB /CIATERC1/ START 77188.1740

IEF376J JOB /CIATERC1/ STOP 77188.1751 CPU 0MIN 30.43SEC

*** DEPRD - SENCO *** SISTEMA/370 MODELO 145, OS/VSI RELEASE 04.0

A-1

PAGINA 1 DEL LISTADO DEL
PROGRAMA SAS

STATISTICAL ANALYSIS SYSTEM

1 DATA INPUT X Y CARDS

NOTE THE DATA STATEMENT USED 0.81 SECONDS AND 92K.

```
5      PROC PRINT N DATA=DATA1
```

6 TITLE GUIA SAS SIMPLIFICADA TITLE2 EJEMPLO 1

7 TITLE3 ENTRADA DE DATOS CON FORMATO LIBRE (UNA OBSERVACION POR TARJETA)

NOTE THE PROCEDURE PRINT USED 1.00 SECONDS AND 132K AND PRINTED PAGE 1.

DATA A2 INPUT X Y @@ CARDS

NOTE SAS WENT TO A NEW LINE WHEN INPUT STATEMENT REACHED PAST THE END OF A LINE.

NOTE DATA SET TRABAJO.A2, HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.41 SECONDS AND 94K.

```
10      PROC PRINT N DATA=A2
```

11 TITLE2 EJEMPLO 2

12 TITLE3 ENTRADA DE DATOS CON FORMATO LIBRE (VARIAS OBSERVACIONES POR TARJETA)

NOTE THE PROCEDURE PRINT USED 1.00 SECONDS AND 96K AND PRINTED PAGE 2.

GUIA SIMPLIFICADA DE SAS

PAGINA 1 (Cont.) DEL LISTADO DEL PROGRAMA SAS

13 DATA A3 INPUT VAR 5 READ 20 CARDS

NOTE SAS WENT TO A NEW LINE WHEN INPUT STATEMENT
REACHED PAST THE END OF A LINE.

NOTE DATA SET TRABAJO.A3 HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.42 SECONDS AND 94K.

15 PROC PRINT

16 TITLE2 EJEMPLO 3-ENTRADA DE DATOS ALFANUMERICOS

NOTE THE PROCEDURE PRINT USED 0.99 SECONDS AND 96K AND PRINTED PAGE 3.

17 DATA A4 INPUT X 1-2 Y 4 CARDS

NOTE DATA SET TRABAJO.A4 HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.42 SECONDS AND 94K.

21 PROC PRINT

22 TITLE2 EJEMPLO 4

23 TITLE3 ENTRADA DE DATOS CON FORMATO FIJO (ESPECIFICANDO LAS COLUMNAS)

NOTE THE PROCEDURE PRINT USED 0.98 SECONDS AND 94K AND PRINTED PAGE 4.

24 DATA A5 INPUT (NAME SCORE1 SCORE2) (5 38. 8 30 285.) CARDS

NOTE DATA SET TRABAJO.A5 HAS 3 OBSERVATIONS AND 3 VARIABLES.

NOTE THE DATA STATEMENT USED 0.45 SECONDS AND 94K.

28 PROC PRINT

29 TITLE2 EJEMPLO 5-ENTRADA DE DATOS CON FORMATO FIJO

NOTE THE PROCEDURE PRINT USED 1.00 SECONDS AND 96K AND PRINTED PAGE 5.

30 DATA A6 INPUT VAR 5 READ 20 CARDS

GUIA SIMPLIFICADA DE SAS

PAGINA 2 DEL LISTADO DEL PROGRAMA SAS

2 STATISTICAL ANALYSIS SYSTEM

31 L1 REP+1 INPUT RESP @ OUTPUT
32 IF REPNREP THEN GO TO L1 DELETE CARDS

NOTE DATA SET TRABAJO.A6 HAS 7 OBSERVATIONS AND 3 VARIABLES.
NOTE THE DATA STATEMENT USED 0.51 SECONDS AND 94K.

35 PROC PRINT
36 TITLE2 EJEMPLO 6
37 TITLE3 ENTRADA DE DATOS CON IDENTIFICADORES COMUNES

NOTE THE PROCEDURE PRINT USED 1.08 SECONDS AND 96K AND PRINTED PAGE 6.

38 DATA B1 SET A4 IF X#4

NOTE DATA SET TRABAJO.B1 HAS 2 OBSERVATIONS AND 2 VARIABLES.
NOTE THE DATA STATEMENT USED 0.46 SECONDS AND 94K.

39 PROC PRINT
40 TITLE2 INSTRUCCION SET
41 TITLE3 EJEMPLO 1 - FORMACION DE SUBCONJUNTOS DE DATOS

NOTE THE PROCEDURE PRINT USED 1.01 SECONDS AND 96K AND PRINTED PAGE 7.

42 DATA B2 SET A4 B1

NOTE DATA SET TRABAJO.B2 HAS 5 OBSERVATIONS AND 2 VARIABLES.
NOTE THE DATA STATEMENT USED 0.54 SECONDS AND 94K.

GUIA SIMPLIFICADA DE SAS

PAGINA 2 (Cont.) DEL LISTADO DEL PROGRAMA SAS

43 PROC PRINT
44 TITLE3 EJEMPLO 2 - CONCATENACION DE CONJUNTOS DE DATOS

NOTE THE PROCEDURE PRINT USED 1.04 SECONDS AND 96K AND PRINTED PAGE 8.

45 DATA B3 SET A4 B1 BY X

NOTE DATA SET TRABAJO.B3 HAS 5 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.58 SECONDS AND 98K.

46 PROC PRINT

47 TITLE3 EJEMPLO 3 - INTERCALACION DE CONJUNTOS DE DATOS

NOTE THE PROCEDURE PRINT USED 1.06 SECONDS AND 96K AND PRINTED PAGE 9.

48 DATA M1 INPUT VAR RESP1 @@ CARDS

NOTE SAS WENT TO A NEW LINE WHEN INPUT STATEMENT
REACHED PAST THE END OF A LINE.

NOTE DATA SET TRABAJO.M1 HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.44 SECONDS AND 94K.

50 PROC PRINT

51 TITLE2 INSTRUCCION MERGE

52 TITLE3 EJEMPLO 1 - MERGE SIN BY - PRINT DEL CONJUNTO M1

NOTE THE PROCEDURE PRINT USED 1.08 SECONDS AND 96K AND PRINTED PAGE 10.

53 DATA M2 INPUT VAR RESP2 @@ CARDS

NOTE SAS WENT TO A NEW LINE WHEN INPUT STATEMENT
REACHED PAST THE END OF A LINE.

NOTE DATA SET TRABAJO.M2 HAS 3 OBSERVATIONS AND 2 VARIABLES.

GUIA SIMPLIFICADA DE SAS

PAGINA 3 DEL LISTADO DEL
PROGRAMA SAS

3 STATISTICAL ANALYSIS SYSTEM

NOTE THE DATA STATEMENT USED 0.42 SECONDS AND 94K.

55 TITLE3 EJEMPLO 1 - MERGE SIN BY - PRINT DEL CONJUNTO M2
56 PROC PRINT

NOTE THE PROCEDURE PRINT USED 1.02 SECONDS AND 96K AND PRINTED PAGE 11.

57 DATA M12 MERGE M1 M2

NOTE DATA SET TRABAJO.M12 HAS 3 OBSERVATIONS AND 3 VARIABLES.

NOTE THE DATA STATEMENT USED 0.53 SECONDS AND 98K.

58 PROC PRINT

59 TITLE3 EJEMPLO 1 - MERGE SIN BY - PRINT DEL RESULTADO DEL MERGE

NOTE THE PROCEDURE PRINT USED 1.04 SECONDS AND 96K AND PRINTED PAGE 12.

60 DATA M15 SET M1

NOTE DATA SET TRABAJO.M15 HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.44 SECONDS AND 94K.

60 PROC SORT BY VAR

NOTE DATA SET TRABAJO.M15 HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE PROCEDURE SORT USED 3.23 SECONDS AND 102K.

GUIA SIMPLIFICADA DE SAS

PAGINA 3 (cont.) DEL LISTADO DEL PROGRAMA DE SAS

61 PROC PRINT

62 TITLE3 EJEMPLO 2 - MERGE USANDO BY - PRINT DEL CONJUNTO M1 ORDENADO

NOTE THE PROCEDURE PRINT USED 0.89 SECONDS AND 124K AND PRINTED PAGE 13.

63 DATA M2S SET M2

NOTE DATA SET TRABAJO.M2S HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE DATA STATEMENT USED 0.45 SECONDS AND 124K.

63 PROC SORT BY VAR

NOTE DATA SET TRABAJO.M2S HAS 3 OBSERVATIONS AND 2 VARIABLES.

NOTE THE PROCEDURE SORT USED 3.09 SECONDS AND 130K.

64 PROC PRINT

65 TITLE3 EJEMPLO 2 - MERGE USANDO BY - PRINT DEL CONJUNTO M2 ORDENADO

NOTE THE PROCEDURE PRINT USED 0.92 SECONDS AND 124K AND PRINTED PAGE 14.

66 DATA M12S MERGE M1S M2S BY VAR

NOTE DATA SET TRABAJO.M12S HAS 3 OBSERVATIONS AND 3 VARIABLES.

NOTE THE DATA STATEMENT USED 0.59 SECONDS AND 124K.

67 PROC PRINT

68 TITLE3 EJEMPLO 2 - MERGE USANDO BY - PRINT DEL RESULTADO DEL MERGE

69 * ESTE ES UN EJEMPLO DE SENTENCIA DE COMENTARIO

NOTE THE PROCEDURE PRINT USED 1.01 SECONDS AND 124K AND PRINTED PAGE 15.

70 DATA C1 INPUT A B @@ X=A+B A=B-A R=SQRT(X**2+A**2) CARDS

NOTE SAS WENT TO A NEW LINE WHEN INPUT STATEMENT

A-7

GUIA SIMPLIFICADA DE SAS

PAGINA 4 DEL LISTADO DEL
PROGRAMA DE SAS

4

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

REACHED PAST THE END OF A LINE.

NOTE DATA SET TRABAJO.C1 HAS 3 OBSERVATIONS AND 4 VARIABLES.

NOTE THE DATA STATEMENT USED 0.54 SECONDS AND 124K.

72 PROC PRINT

73 TITLE2 EJEMPLO DE TRANSFORMACION Y CREACION DE NUEVAS VARIABLES

NOTE THE PROCEDURE PRINT USED 0.99 SECONDS AND 124K AND PRINTED PAGE 16.

NOTE SAS USED 132K MEMORY.

NOTE BARR, GOODNIGHT, SALL AND HELWIG

SAS INSTITUTE INC.

P.O. BOX 10066

RALEIGH, N.C. 27605

A-8

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977

EJEMPLO 1

ENTRADA DE DATOS CON FORMATO LIBRE (UNA OBSERVACION POR TARJETA)

OBS	X	Y
1	2	3
2	3	5
3	4	6

N=3

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977

EJEMPLO 2

ENTRADA DE DATOS CON FORMATO LIBRE (VARIAS OBSERVACIONES POR TARJETA)

OBS	X	Y
1	2	3
2	3	5
3	4	6

N=3

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977

EJEMPLO 3-ENTRADA DE DATOS ALFANUMERICOS

OBS	VAR	RENO
1	P720	580.3
2	P642	328.8
3	P005	+

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977

EJEMPLO 4

ENTRADA DE DATOS CON FORMATO FIJO (ESPECIFICANDO LAS COLUMNAS)

OBS	X	Y
1	2	3
2	3	5
3	4	6

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977

EJEMPLO 5-ENTRADA DE DATOS CON FORMATO FIJO

OBS	NAME	SCORE1	SCORE2
1	JUAN PEREZ	98.5	43.2
2	JOHN DOE	-31.2	-5.0
3	CENTRO INT. DE AGRICULTURA TROPICAL	45.8	0.0

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977

EJEMPLO 6

ENTRADA DE DATOS CON IDENTIFICADORES COMUNES

OBS	VAR	REP	RESP
1	P304	1	15.3
2	P304	2	16.0
3	P304	3	14.8
4	P304	4	15.9
5	P058	1	14.1
6	P058	2	13.8
7	P058	3	14.1

GUIA SAS SIMPLIFICADA

INSTRUCCION SET

8 36 TUESDAY, JULY 5, 1977

7

EJEMPLO 1 - FORMACION DE SUBCONJUNTOS DE DATOS

OBS	X	Y
1	2	3
2	3	5

GUIA SAS SIMPLIFICADA

INSTRUCCION SET

8 36 TUESDAY, JULY 5, 1977

8

EJEMPLO 2 - CONCATENACION DE CONJUNTOS DE DATOS

OBS	X	Y
1	2	3
2	3	5
3	4	6
4	2	3
5	3	5

GUIA SAS SIMPLIFICADA

INSTRUCCION SET

8 36 TUESDAY, JULY 5, 1977

9

EJEMPLO 3 - INTERCALACION DE CONJUNTOS DE DATOS

OBS	X	Y
1	2	3
2	2	3
3	3	5
4	3	5
5	4	6

GUÍA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977 10

INSTRUCCIÓN MERGE

EJEMPLO 1 - MERGE SIN BY - PRINT DEL CONJUNTO M1

OBS	VAR	RESP1
1	4	4.1
2	3	3.8
3	5	5.6

GUÍA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977 11

INSTRUCCIÓN MERGE

EJEMPLO 1 - MERGE SIN BY - PRINT DEL CONJUNTO M2

OBS	VAR	RESP2
1	4	4.3
2	3	3.6
3	5	5.0

GUÍA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977 12

INSTRUCCIÓN MERGE

EJEMPLO 1 - MERGE SIN BY - PRINT DEL RESULTADO DEL MERGE

OBS	VAR	RESP1	RESP2
1	4	4.1	4.3
2	3	3.8	3.6
3	5	5.6	5.0

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977 13

INSTRUCCION MERGE

EJEMPLO 2 - MERGE USANDO BY - PRINT DEL CONJUNTO M1 ORDENADO

ORS	VAR	RESP1
1	3	3.8
2	4	4.1
3	5	5.6

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977 14

INSTRUCCION MERGE

EJEMPLO 2 - MERGE USANDO BY - PRINT DEL CONJUNTO M2 ORDENADO

ORS	VAR	RESP2
1	3	3.6
2	4	4.3
3	5	5.0

GUIA SAS SIMPLIFICADA

8 36 TUESDAY, JULY 5, 1977 15

INSTRUCCION MERGE

EJEMPLO 2 - MERGE USANDO BY - PRINT DEL RESULTADO DEL MERGE

ORS	VAR	RESP1	RESP2
1	3	3.8	3.6
2	4	4.1	4.3
3	5	5.6	5.0

GUIA SAS SIMPLIFICADA

8 36 TUE SDAY, JULY 5, 1977

16

EJEMPLO DE TRANSFORMACION Y CREACION DE NUEVAS VARIABLES

OBS	A	B	X	R
1	2	3	4	4.47214
2	2	4	6	6.32456
3	2	5	8	8.24621

APENDICE B

Instrucciones de Programa más Útiles

I. Instrucciones de Control y Modificación de Variables

a. Instrucciones de Asignación

Forma: Nombre_de_variable=expresión_SAS ;

Uso: Crear nuevas variables o modificar variables existentes

Ejemplos: X=A+B ; A=A+B ; C=SQRT(A**2+B**2) ;

b. Instrucciones de Suma

Forma: Nombre_de_variable + expresión_SAS;

Uso : Acumular totales y subtotales, contadores

Ejemplos: REP+1; SUMX+X; SUMX2 + X*X;

c. Instrucciones IF-THEN

Forma: IF expresión THEN instrucción;

Uso : Cuando la acción indicada por "instrucción" se desea cumplir sólo si se cumple la condición indicada por "expresión"

Ejemplos: IF X=. THEN X=0.; IF N=1 THEN N=100;

d. Instrucciones GO TO

Forma: GO TO label_de_Instrucción;

Uso : Cambiar la secuencia de ejecución de las instrucciones de un programa

Ejemplos: IF REP<NREP THEN GO TO L1;

a. Label de Instrucciones

Forma: Label: Instrucción;

Uso : Rotular la instrucción para referencias tipo
GO TO.

Ejemplos: L1: REP+1;

II. Instrucciones de Selección de Observaciones y Variables

a. Instrucciones IF para crear subarchivos

Forma: IF expresión;

Uso : Crear un subarchivo con las observaciones para
las que "expresión" es verdadera

Ejemplo: IF SEXO='F' ;

b. Instrucciones DROP

Forma: DROP lista_de_variables;

Uso : Indicar que las variables en la lista no sean
incluidas en el archivo

Ejemplo: DROP NREP ;

c. Instrucciones KEEP

Forma: KEEP lista_de_variables;

Uso : Indicar que las variables en la lista son las
únicas a incluir en el archivo

Ejemplo: KEEP SEXO PESO;

d. Instrucciones DELETE

Forma: DELETE; ó IF-THEN DELETE;

Uso : Eliminar la observación en proceso

Ejemplo: IF SEXO='M' AND EDAD GT 20 THEN DELETE;

III. Instrucciones para describir los conjuntos de datos

a. Instrucciones LABEL

Forma: LABEL nombre_de_variable=etiqueta ... ;

Uso : Asignar etiquetas a variables

Ejemplo: LABEL N=NITROGENO

P=FOSFORO

K=POTASIO ;

b. Instrucciones LIST

Forma: LIST; ó IF-THEN LIST;

Uso : Listar la línea de entrada bajo proceso

Ejemplo: IF X=. OR Y<10 THEN LIST;

c. Instrucciones MACRO

Forma: MACRO nombre_de_la_macro información_de_la_macro %

Uso : Simplificar o volver modular un programa dando un nombre a un segmento del mismo que se repite en el job o que de job a job cambia ligeramente

te pero no en estructura o uso

Ejemplo: MACRO CHECK IF CCN=2644858 THEN DELETE; %
DATA SEMANAL ; SET NOMSEM ; CHECK
DATA MENSUAL ; SET NOMMEN ; CHECK