

Procedures 2

PROC MEANS

PROC MEANS *opciones estadísticas*;

Algunas opciones:

DATA=*data-set* Data set que se quiere utilizar. Si se omite utiliza el último data set leído

NOPRINT no imprime al output

Algunos estadísticos:

ALPHA nivel de significancia para intervalo de confianza

CLM intervalo de confianza (nivel de confianza = 95% por default, para cambiarla usamos ALPHA =)

N número de valores no perdidos

NMISS número de valores perdidos

MIN valor mínimo

MAX valor máximo

MEAN media

MEDIAN mediana

PROBT p-value para hipótesis $H_0: \mu=0$

T t-Student para hipótesis $H_0: \mu=0$

STDDEV/STD desviación estándar

VARDEF=DF | N divide la suma de cuadrado por n-1 o n.

Q1 cuartilo 1

Q3 cuartilo 3

Algunos statements opcionales:

BY *var-list*; – hace un análisis para cada nivel de las variables en BY. Los datos tienen que estar ordenados por estas variables.

CLASS *var-list*; – hace un análisis para cada nivel de las variables en BY pero el output es más compacto y no hay que ordenarlos primero.

VAR *var-list*; – especifica qué variables numéricas se usarán en el análisis. Si se omite se usan todas.

OUTPUT OUT=*data-set* estadístico (*var-list*)=*name-list*; *data-set* es el archivo donde se va a guardar los estadísticos de las *var-list*. *Name-list* es el nombre que van a tener los estadísticos en el archivo *data-set*.

Ejemplo:

```
PROC MEANS DATA=estudiantes NOPRINT;  
VAR ex1-ex4;  
OUTPUT OUT=promedios MEAN(ex1-ex4)=prom1-prom4 STD(ex1-ex4)=st1-st4  
MEDIAN(ex1-ex4)=med1-med4;  
RUN;
```

PROC UNIVARIATE

PROC UNIVARIATE *opciones*;

Produce estadísticos para una variable

Calcula intervalos de confianza

Prueba de hipótesis para $H_0: \mu = \mu_0$

Prueba de Normalidad

Gráficas: tallo-hoja, box plot, histogramas, normal probability plot

Algunas opciones:

DATA= *data-set* Data set que se quiere utilizar. Si se omite utiliza el último data set leído

ALL – todos los estadísticos y gráficas

CIBASIC – intervalo de confianza para μ

ALPHA=valor – nivel de significancia para el intervalo de confianza

FREQ *var* – los valores de var representan la frecuencia de las observaciones

MU0= μ_0 – valor de μ en la hipótesis nula

PLOTS – histograma, Box-Plot y Normal probability plot

NORMAL – prueba normalidad de los datos

Algunos statements opcionales: además de los mencionados en PROC MEANS

HISTOGRAM *var-list* </opciones>; histogramas usando gráficas de alta resolución. Las *var-list* tienen que estar en VAR también. Algunas opciones:

CAXIS=*color* – color de los ejes

CBARLINE=*color* – color del borde de las barras

CFILL=*color* – color dentro de las barras

VAXIS= *lista de números* – escala en eje vertical

MIDPOINTS=*valores* – puntos medios (ejemplo de valores: 0 TO 100 BY 10)

PROBPLOT *var-list* </opciones>; Normal probability plot en alta resolución

VAR *var-list* si se omite utiliza todas las variables en el data set.

BY – hace un análisis para cada nivel de las variables en BY. Los datos tienen que estar ordenados por estas variables.

PROC FREQ

PROC FREQ *opciones*;

Construye tabla de contingencia

Algunas opciones:

DATA=*data-set* Data set que se quiere utilizar. Si se omite utiliza el último data set leído

NOPRINT no imprime al output. Se usa cuando solo queremos imprimir a un data set con la instrucción OUTPUT o la opción OUT= en la instrucción TABLES.

Algunos statements opcionales:

BY <DESCENDING> *var-list* <NOTSORTED>; – hace un análisis para cada nivel de las variables en BY. Los datos tienen que estar ordenados por estas variables si no se usa la opción <NOTSORTED>.

TABLES *var-combinations* </opciones_tables>;

var-combinations – var1*var2 var3*var5 ...

Opciones_tables:

MISSING incluye los valores perdidos en las estadísticas de frecuencia.

NOCOL suprime los por cientos de columnas

NOROW suprime los por cientos de filas

NOPERCENT suprime los por cientos a base del total

CHISQ pruebas de ji-cuadrado

FISHER prueba exacta de Fisher para tablas mayores de 2X2

OUT=*data-set* escribe las frecuencias al archivo *data-set*

```
DATA formato;
INPUT nombre $17. sexo edad;
CARDS;
Alberto Gómez 1 19
Ana García 0 21
Rolando Pérez 1 25
Manuel Hernández 1 18
Irma Rivera 0 17
María Álvarez 0 27
Marta Santiago 0 26
RUN;
PROC FORMAT;
VALUE genero 1='hombre'
0='mujer';
VALUE age 10-19='teen'
20-29='adulto';
PROC FREQ;
TABLE edad*sexo / norow nocol;
FORMAT edad age. sexo genero.;
RUN;
```

PROC CORR

PROC CORR <opciones>;

Algunas opciones:

DATA=*data-set* Data set con que vamos a trabajar. Si se omite utiliza el último data set leído

NOPRINT no imprime al output.

SPEARMAN

PEARSON – default

HOEFFDING

KENDALL

Algunos statements opcionales:

BY <DESCENDING> *var-list* <NOTSORTED>; – calcula coeficientes de correlación para cada nivel de las variables en BY. Los datos tienen que estar ordenados por estas variables si no se usa la opción <NOTSORTED>.

VAR *var-list*; – especifica qué variables numéricas se usarán en el análisis. Si se omite se usan todas.

WITH *var-list* – especifica con qué variables se van a correlacionar las variables enumeradas en VAR. Si se omite hace todas las variables en VAR contra todas ellas mismas.

Gráficas

Algunos statements opcionales:

GOPTIONS – define temporariamente opciones para muchas gráficas.

Algunas opciones:

BORDER | NOBORDER
CBACK=*background-color*
COLORS=<(colors-list | NONE)>
CPATTERN=*pattern-color*
CSYMBOL=*symbol-color*
CTEXT=*text-color*
CTITLE=*title-color*
HSIZE=*horizontal-size* <IN | CM>
ROTATE=LANDSCAPE | PORTRAIT
ROTATE | NOROTATE
VSIZE=*vertical-size* <IN | CM>

SYMBOL<1 ... 99> (se pueden crear 99 símbolos: symbol1, symbol2,...)
<COLOR=*color-del-símbolo*>
<MODE=EXCLUDE | INCLUDE> incluye o no los valores fuera del range de los ejes. Default es EXCLUDE.
<opciones>;

Se puede usar en GPLOT pero no en GCHART

Algunas opciones de SYMBOL:

BWIDTH=*ancho-caja* – se usa con INTERPOL=BOX

CI=*line-color*

CO=*outline-color*

CV=*plot-symbol-color*

FONT=*font-name*. Para dejar de usar un font hay que definir otro o ponerlo nulo: FONT=,

VALUE=*special-symbol* | *text-string* | NONE

special-symbol – nombre del símbolo (ver tabla al final). Para usar un símbolo especial no se puede usar la opción de FONT= o hay que poner FONT=, (nulo)

text-string – se usa con FONT=

INTERPOL= BOX | JOIN | R

BOX – box-whisker plot

JOIN – conecta los puntos con líneas rectas

R <TYPE><0><CLM|CLI><50 ... 99>

TYPE: L – regresión lineal

Q – regresión cuadrática

C – regresión cúbica

0 – fuerza intercepto a 0

CLM | CLI – intervalo de confianza para la media | para la predicción

50 ... 99 – nivel de confianza

Ejemplo INTERPOL=RLCLM95 – regresión lineal simple e intervalo de confianza para la predicción a un nivel de confianza de 95%.

TITLE or FOOTNOTE or NOTE <1...10> <'text'> <text-options>; - TITLE y FOOTNOTE son statements globales, NOTE es LOCAL (está en efecto solo dentro del PROC que se usa).

Text-options:

COLOR=*color* – afecta todo texto, líneas y rectángulos subsiguientes

FONT=*font*

HEIGHT=*text-height* <unidades> – Default es 1. El TITLE1 tiene el doble de tamaño que los otros por default

BOX=1...4 – dibuja un rectángulo. Los números representan el grueso de las líneas (1 fina, 4 gruesa)

LSPACE=*line-space* <unidades>

MOVE=(*x,y*) <unidades>

ANGLE=*grados* – rota la oración completa

DRAW=(*x, y, ..., x_n, y_n*) <unidades> – traza líneas

ROTATE=*grados* – rota cada letra

Unidades pueden ser: CELLS (caracteres), CM (centímetros), IN (pulgadas), PT (puntos), PCT (por ciento del área de output de la gráfica)

PROC GCHART

PROC GCHART<DATA=*input-data-set*><GOUT= *libref.output-catalog* >;

GOUT – catálogo dónde se guarda la gráfica

Algunos statements opcionales:

BY *var-list*; – hace un análisis para cada nivel de las variables en BY. Los datos tienen que estar ordenados por estas variables.

HBAR | VBAR | HBAR3D | VBAR3D *chart-variable(s)* </ *option(s)*>; gráfica de barras verticales u horizontales

PIE *chart-variable(s)* </ *option(s)*>; - pie chart

chart-variables – define las categorías. Cada variable hace una gráfica diferente.

Debe incluirse por lo menos una de estos últimos dos statements.

También se pueden usar los statements: FORMAT, LABEL, WHERE, TITLE

Algunas opciones:

CAXIS=*color* color de los ejes

CFRAME=*color* color del background.

CFREQ – presenta la frecuencia acumulada en la tabla de estadísticos.

CLM=*nivel de confianza* – intervalo de confianza. Necesita las siguientes opciones:

Para intervalo de confianza para la media de *var*: TYPE=MEAN

SUMVAR=*var*

Para intervalo de confianza para el por ciento de observaciones de cada punto medio: TYPE=PCT

COUTLINE=*color* | SAME – color de los bordes de las barras. SAME – usa el mismo color que adentro de las barras

CTEXT=*color* – color de todos los textos

DISCRETE – considere las variables numéricas como discretas.

FREQ – presenta la frecuencia en la tabla de estadísticos. Cuando especificamos algún estadístico solo sale este.

GROUP=*group-var* – hace un set de barras separadas para cada *group-var*.

GSPACE=*group-spacing* – separación entre grupos

INSIDE=*statistic* – (solo en VBAR o VBAR3D) coloca el valor del estadístico dentro de la barra. *statistic*=FREQ, CFREQ, CPERCENT, MEAN, PERCENT, SUM.

LEVELS=*n* – *n*=número de puntos medios para una variable numérica.

MEAN – presenta la media de la variable en SUMVAR en la tabla de estadísticos.

MIDPOINTS=*value-list* – *value-list*=*n1 n2 ... nk* ó *n1 TO n2* <BY incremento>

NOSTATS – en HBAR o HBAR3D – omite la tabla de estadísticos.

OUTSIDE=*statistic* – (solo en VBAR o VBAR3D) coloca el valor del estadístico sobre la barra.

PATTERNID=BY | GROUP | MIDPOINT | SUBGROUP

BY – cambia el patrón (relleno) de las barras cada vez que cambia el valor de la variable en el BY statement.
GROUP – cambia el patrón (relleno) de las barras cada vez que cambia el valor de la variable en el GROUP statement.
MIDPOINT – cambia el patrón (relleno) de las barras cada vez que cambia el punto medio (un patrón para cada barra).
SUBGROUP – cambia el patrón (relleno) de las barras cada vez que cambia el valor de la variable en la opción SUBGROUP=.

SHAPE=3D-bar-shape – (HBAR y HBAR3D) forma de las barras. Puede ser: BLOCK, CYLINDER, HEXAGON, PRISM, STAR.

SPACE=*ancho-barras*

SUM – presenta la suma de la variable en SUMVAR en la tabla de estadísticos.

SUMVAR=*summary-variable* – variable de la gráfica

SUBGROUP=*variable-subgrupo* – divide las barras en partes. Una parte para cada valor diferente de la variable subgrupo. (Como si fuera un histograma sobre otro).

TYPE=*statistic* – estadístico de la gráfica:

Si usamos SUMVAR entonces statistic puede ser: MEAN, SUM
Si no usamos SUMVAR entonces statistic puede ser: FREQ, CFREQ,
PERCENT, CPERCENT

WOUTLINE=*ancho-borde-barras*

PROC GPLOT

```
PROC GPLOT <DATA=data set> < UNIFORM><GOUT= libref.output-catalog>;
```

UNIFORM – pone misma escala en todos los ejes de todas las gráficas

GOUT – catálogo dónde se guarda la gráfica

Algunos statements opcionales:

PLOT *plots* </opciones>; – variable independiente vs variable dependiente

plots: x-var*y-var
(x-var-list)*(y-var-list)

(gráfica de X vs Y)
(genera una gráfica diferente para cada combinación de las variables en x-var-list con las variables en y-var-list)

x-var*y-var=var3

(genera una gráfica en los mismos ejes para cada valor de var3)

BUBBLE *plots* </opciones>; gráfica donde en el eje de x hay una variable, otra en el eje de y, y hay una tercera variable representada por círculos. El valor de esta variable determina el tamaño del círculo.

plots: x-var*y-var=c-var

Debe incluirse por lo menos uno de estos dos statements (PLOT o BUBBLE).

También se pueden usar los statements: FORMAT, LABEL, WHERE, TITLE, SYMBOL

Algunas opciones:

BCOLOR=color-burbuja (opción solo para BUBBLE)

CAXIS=*color* color de los ejes

CFRAME=*color* color del background.

CTEXT=*color* – color de todos los textos

FRAME/NOFRAME

GRID

NOAXIS – no pone los ejes

OVERLAY – produce todas las gráficas en los mismos ejes (opción solo para PLOT).

REGQN – traza ecuación de regresión especificada en la opción de INTERPOL del SYMBOL statement.

PROC BOXPLOT

PROC BOXPLOT < DATA=*data-set*><GOUT= *libref.output-catalog* >;

GOUT=*libref.output-catalog* – catálogo donde se guardan las gráficas.

Algunos statements:

PLOT (*vars-análisis*)**var-grupo* </opciones> – debe haber por lo menos un PLOT statement.

vars-análisis – variables que van a analizarse. Hace una gráfica para cada una.

vars-grupo – variable que identifica los grupos.

Algunas opciones de PLOT:

IDCOLOR= especifica color del símbolo para outliers
IDSYMBOL= especifica símbolo de outliers.

BY *vars*; Boxplot separado para cada variable en *vars*.

PROC GREPLAY

PROC GREPLAY opciones;

- Se usa para manejar la presentación de gráficas.
- Crea templates para presentar las gráficas
- Opera en ambiente de window (interactivo) y no window

Algunas opciones:

GOUT=*libref.output-catalog* – catálogo donde se guardan las gráficas.

IGOUT= *libref.input-catalog* – catálogo de dónde se obtienen las gráficas.

NOFS – usa line-mode en lugar de operar en ambiente de windows

TC=*libref.template-catalog* – catálogo donde están los templates. Catálogo de templates provisto por SAS: SASHELP.TEMPLT

TEMPLATE=*nombre-template* – nombre del template que se va a usar. Los templates dividen el área en sectores.

Algunos statements:

GOUT *libref.output-catalog*; – permite cambiar el catálogo donde se guardan las gráficas.
El default es WORK.GSEG

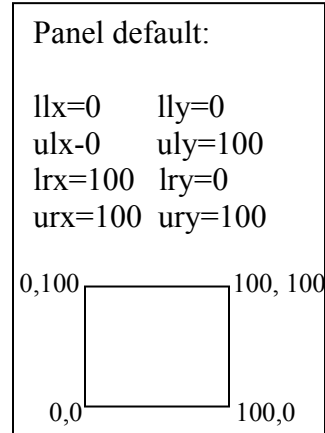
IGOUT *libref.input-catalog*; – permite cambiar el catálogo de dónde se obtienen las gráficas.

TC *libref.template-catalog*; – permite cambiar el catálogo donde están los templates.
Catálogo de templates provisto por SAS: SASHELP.TEMPLT

TDEF *nombre-template número-del-panel* / <opciones> <DES='descripción del panel pero no se imprime'>; – define o modifica templates en el catálogo de templates actual.

Algunas opciones de TDEF:

- COLOR=*border-color*
- DEF – especifica el panel default
- LLX=*x* – lower left x coordinate
- LLY=*y* – lower left y coordinate
- LRX=*x* – lower right x coordinate
- LRY=*y* – lower right y coordinate
- ROTATE=*degrees*
- SCALEX=*factor*
- SCALEY=*factor*
- ULX=*x* – upper left x coordinate
- ULY=*y* – upper left y coordinate
- URX=*x* – upper right x coordinate
- URY=*y* – upper right y coordinate



TEMPLATE *nombre-plantilla*; – permite cambiar el nombre del template que se va a usar.

TREPLAY *número-del-panel* : *nombre-de-la-gráfica*;

VALUE=	Plot Symbol	VALUE=	Plot Symbol	VALUE=	Plot Symbol	VALUE=	Plot Symbol
PLUS	+	% (percent)	♣	PAW	⋯	. (period)	⌘
X	×	& (ampersand)	♣	POINT	.	< (less than)	ℏ
STAR	*	' (single quote)	♣	DOT	●	, (comma)	♁
SQUARE	□	= (equals)	☆	CIRCLE	○	/ (slash)	ψ
DIAMOND	◇	- (hyphen)	⊙	_ (underscore)	◻	? (question mark)	ℙ
TRIANGLE	△	@ (at)	♀	" (double quote)	♠	((left parenthesis)	⌌
HASH	#	* (asterisk)	♀	# (pound sign)	♥) (right parenthesis)	♂
Y	Y	+ (plus)	⊕	\$ (dollar sign)	◇	: (colon)	*
Z	Z	> (greater than)	♂				

Note: The words or special characters in the VALUE= column are entered exactly as shown.