

## **4 GRAFICOS**

### **4.1 INTRODUCCION**

En este tema vamos a ver cómo se pueden mejorar la presentación de los trabajos realizados con el sistema SAS. Para ello es necesario tener contratado el software SAS/GRAPH. Los programas para este software son procedimientos SAS. Generalmente requieren pocas sentencias para producir resultados. Existe además de los procedimientos una herramienta denominada ANNOTATE que nos ayudara considerablemente en el diseño de todos los gráficos y que no es otra cosa que un conjunto de datos especial.

Los gráficos que produce SAS se almacenan en catálogos SAS. Los catálogos son miembros especiales de una librería que contiene información necesaria para reconstruir los gráficos. Se pueden enviar los gráficos a un catálogo específico utilizando la opción GOUT=nombre en la sentencia PROC de todos los procedimientos gráficos. Si no se utiliza dicha opción los gráficos por defecto se almacenan en el catálogo por defecto: WORK.GSEG. Lógicamente GSEG se borra al final de cada sesión SAS. Se pueden utilizar nombres con un nivel o dos niveles para identificar al catálogo, de la misma manera que cuando se creaban conjuntos de datos. En este último caso se guardarían de forma permanente (el primero determina la librería y el segundo el catálogo). Por defecto cualquier gráfico producido por un procedimiento se añade al final del catálogo correspondiente, a no ser que se especifique: goptions goutmode=replace; en cuyo caso cuando se ejecuta un procedimiento que produce gráficos, se rempazan todos los gráficos existentes en el catálogo especificado en la sentencia gout=.

Cuando la gráfica se envía a un catálogo, SAS automáticamente asigna un nombre (generalmente el nombre del procedimiento con el que se crea) y una descripción para identificar la gráfica en el catálogo. No obstante se pueden utilizar las opciones NAME= y DESCRIPTION= para que se puedan determinar nombres y descripciones propias.

Existe un procedimiento especial, cuya sintaxis es sencilla: proc goptions; run; que produce un listado en la ventana log con las opciones gráficas que se encuentran activas en el momento de ser ejecutado. Hay que tener cuidado y no confundir el procedimiento goptions con la sentencia goptions (la cual se puede utilizar en cualquier parte) que es precisamente la que permite alterar los valores por defecto de las opciones gráficas.

En el momento de definir colores para cualquier parte de un gráfico (áreas, líneas, texto,etc...), estos se pueden definir con sus nombres (cyan, green, red, yellow,...) o bien en escalas, así el código CXrrggbb, propone un color según escalas de rojo, verde y azul. Las cantidades se encuentran en valores hexadecimales (desde el 0 hasta la F). Los valores rr, gg, y bb representan el grado de rojos, verdes y azules respectivamente. Así el color rojo se indica como CXFF0000, el blanco como CXFFFFFF, verde como CX00FF00. Cualquier combinación es válida. Otro código viene representado en escala de grises; GRAYii, donde ii representa el brillo de gris y se obtiene en forma hexadecimal (de 0 a F). Blanco se define como GRAY00, y negro como GRAYFF. No obstante el color con el que aparecen los gráficos dependerá de las características del Hardware correspondiente y a veces tendrán que transformarse los colores para que puedan obtenerse en la salida correspondiente (plotter, impresora, pantalla, cañón, etc..).

La salida por defecto que posee SAS en la versión de Windows es la ventana GRAPH, en ella se pueden remodelar los gráficos obtenidos por los procedimientos, así como crear otros nuevos con el editor. Esta ventana dispone de una barra de menús con la que se puede operar fácilmente. Para entrar en ella hay que ejecutar en primer lugar un procedimiento gráfico (gplot, gchart, capability,...), y después se opera a través de GSEDIT, la forma de poder actuar en esta

ventana se comentará en un capítulo próximo.

## **4.2 RESALTAR LOS TEXTOS EN LOS GRÁFICOS.**

Son las sentencias conocidas TITLE, NOTE y FOOTNOTE. Estas se pueden utilizar como formas básicas de texto (rótulos, pancartas, etc.) a través del procedimiento GSLIDE o como etiquetas engarzadas en gráficos más complicados.

Se pueden utilizar las sentencias TITLE y FOOTNOTE en cualquier lugar de un programa. La sentencia NOTE solo se puede utilizar dentro de un procedimiento.

La sentencia TITLE indica líneas de texto que se representarán en la parte alta del gráfico. La sentencia FOOTNOTE especifica las líneas de texto que se generarán en la parte baja del gráfico. NOTE produce líneas de texto en el área que se encuentra tras los títulos y antes de las notas a pie de página.

Las opciones de estas tres sentencias son las mismas, siempre van delante del texto. Las principales son:

C = color      Color del título, nota o nota a pie de página (texto). Los principales colores de los que dispone el SAS son: *black, blue, brown, charcoal, cream, cyan, gold, gray, green, lilac, lime, magenta, maroon, olive, orange, pink, purple, red, rose, salmon, steel, tan, violet, white, yellow.*

F = tipo de fuente      Especifica el tipo de letra a usar. las principales son: *brush, centb, centbi, centx, centxi, complex, cscript, duplex, gitalic, italic, oldeng, script, simplex, swiss, swissb, swissbi, swissi, swissl, swissx, swissxb, titalic, triplex, zapf, zapfb, zapfbi, zapfi.* También se pueden indicar caracteres especiales, como *cartog, cgreek, cyrillic, german, electron, music, greek, weather, hebrew, math, marker, special*

H = n      Especifica la altura de los caracteres. Se puede indicar por CM (centímetros) PCT (porcentaje de pantalla) , IN (pulgadas) CELLS (celdas, valor por defecto) como todos los tamaños que se indiquen en el sas-graph. Un ejemplo:

*title c=red h=2 in 'Primer ejemplo';*

escribe el título Primer ejemplo, con un tamaño de 2 pulgadas, en color rojo.

J = (l, r, c)      Señala el lugar donde aparece el texto (a la izquierda, a la derecha, en el centro respectivamente).

LS = n      Controla el espacio por encima de las líneas de texto en las sentencias Title y Note y por debajo de las mismas en Footnote. Las unidades a utilizar son las mismas que en la opción H=

ANGLE = angulo (-90, +90)      Determina el ángulo con la horizontal en el que aparecerá el texto. Cuando se utiliza con un Title o note, si a= -90 el texto aparecerá en el lado derecho, si a=90 aparecer en el izquierdo. Con Footnote el resultado será el inverso.

LA = angulo      Indica el numero de grados desde la horizontal a los que el texto va a ser rotado. es parecido a Angle, salvo cuando el ángulo es de  $\pm 90^\circ$ , que el texto no se colocará en las esquina derecha e izquierda sino que utilizará la posición determinada por la opción j (por defecto central).

R = angulo      Señala el ángulo de rotación de los caracteres del texto. Se hace carácter a carácter o a grupo de caracteres.

BOX = 0,1,2,3,      Pide que se cree una caja que envuelva al texto. El número indica el grosor de las líneas de la caja. A mayor número mayor grosor.

BC = color      Especifica el color del fondo del texto, si este se incluye en una caja.  
 BS = n          Sirve para determinar la cantidad de espacio reservado para la caja. Donde n indica los centímetros (o pcn -porcentaje-, in -pulgadas-) de separación entre el texto y la caja, según se hayan determinado previamente las unidades de uso.  
 U =(0,1,2,3)    Manifiesta el grosor de subrayado del texto.

Algunos ejemplos

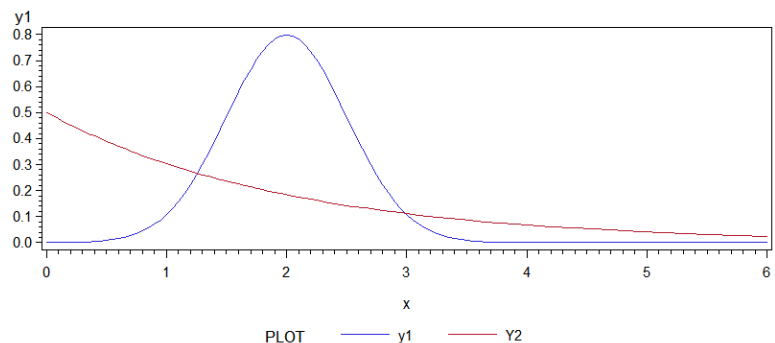
```
TITLE C=GREEN J=L BOX=3 U=2 H=3 'DISTRIBUCIONES NORMAL Y EXPONENCIAL DE MEDIAS 2' ;
```

**DISTRIBUCIONES NORMAL Y EXPONENCIAL DE MEDIAS 2**

```
TITLE BC=yellow c=blue  
LA=90 BOX=2 U=2 F=courier  
'DISTRIBUCION NORMAL' ;
```

DISTRIBUCION NORMAL

```
FOOTNOTE BC=yellow c=blue  
R=90 BOX=2 F=TITALIC J=C  
'VARIAS DISTRIBUCIONES' ;
```



V A R I A S                      D I S T R I B U C I O N E S

### 4.3. SENTENCIAS QUE RESALTAN EL DISEÑO GRAFICO

En este capítulo se verán las sentencias que más se utilizan para alterar la apariencia de las gráficas en su salida final.

### 4.3.1.SENTENCIA PATTERN.

Esta sentencia se utiliza para diseñar las áreas y barras de los gráficos, determinando los colores y tipo de las mismas. La sentencia tiene la siguiente sintaxis:

PATTERNn opciones;

La sentencia es aditiva, permanece su efecto hasta que no se retira por medio de otra sentencia PATTERN. Para volver a componer una opción anterior (resetear) se utiliza la palabra o letras clave de la opción seguida de un blanco. (se pueden o no utilizar comas entre opciones reseteadas).

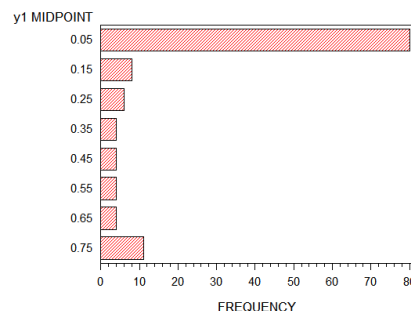
Las opciones que se utilizan son:

C = color especifica el color de una barra o de un área.

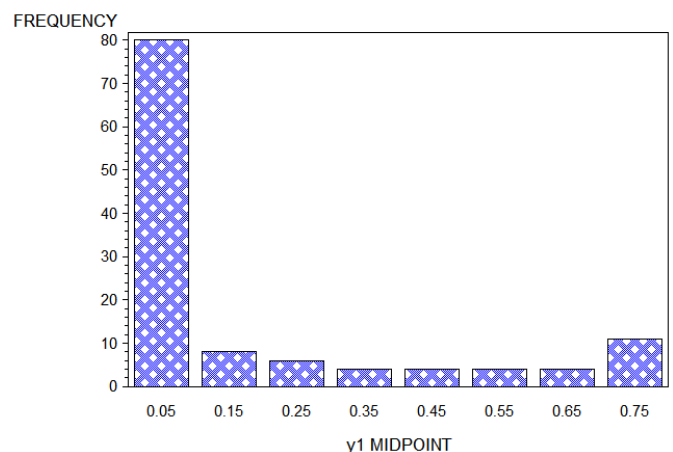
VALUE= valor mide la densidad interior de las barras y las líneas internas que se dibujan en las áreas: los valores a tomar son:

- ◆ E transparente.
- ◆ S la traza interior es solida (sin líneas).
- ◆ para incluir rayas hay diferentes estilos:rayas hacia la izquierda (L), derecha (R), cruzadas (X). La densidad indica el grosor de las líneas. 1 divide la barra en 3, de 2 hasta 5 se divide el area en mas partes y las trazas son más gruesas:Ejemplos: value=L3, value=R2, value=X5

```
proc gchart data=expon;  
PATTERN C=RED Value=r2;  
    hbar y1 ;  
run;
```



```
proc gchart data=expon;  
PATTERN C=blue Value=x5;  
    vbar y1 ;  
run;
```



### 4.3.2 SENTENCIA SYMBOL.

Esta sentencia se utiliza para señalar los símbolos que se utilizan en la representación gráfica de los valores observados de los datos, así como la creación de líneas de interpolación,

con lo que se construyen gráficas continuas.

La sintáxis es :

SYMBOLn opciones (n 1-255)

donde n es un identificador de la sentencia, muy utilizado si se pretenden dibujar más de una gráfica. Por defecto n=1.

Las opciones que se pueden utilizar son:

CV = color color para el símbolo representante de los valores observados.

CI = color color a utilizar para la línea de interpolación.

C = color indica el color de la gráfica. tiene el mismo efecto que cv y ci a la vez.

F = fuente indica la fuente de donde se extraerá el símbolo: son los caracteres especiales que se deberán haberse señalado previamente.

W =n Señala el grosor de la línea de interpolación.

H = n Manifiesta la altura de los caracteres.

L = n Proporciona el tipo de línea que se dibuja en la gráfica, los posibles valores de n son :

- ♦ 1 línea solida es la que aparece por defecto,
- ♦ 2-46 representa líneas disjuntas, la separación disminuye según aumente el número indicado (de 2 a 46).

MODE =(INCLUDE, EXCLUDE) Incluye o excluye, según se indique en la sentencia, a los valores observados fuera de los ejes en los cálculos de interpolación. El valor por defecto es EXCLUDE.

V = símbolo Especifica el simbolo a utilizar. Posibles valores son: plus, x, star, square, diamon, triangle, circle, dot .....). Por defecto no se utiliza ninguno (none).

I = tipo de interpolación. los posibles valores son:

I=none (no hay interpolacion).

I=join (línea continua entre observaciones).

I=needle (líneas verticales entre las observaciones y el eje horizontal).

I=Spline, es el método de interpolación más utilizado,

I=Lx (donde x puede valer 1, 3, 5) método de interpolación de Lagrange.

I=STDkxxx se usa cuando existen múltiples valores de la variable y con cada nivel de x; y se quiere ajustar la media con (+ -) 1, 2 3 desviaciones típicas para cada x. El valor de k puede ser 1, 2, 3 (según las desviaciones que se utilicen). Los valores de xxx pueden ser m, p, j, b, t, o combinaciones de esas letras (excepto b con t). Tienen que ver con la forma en la que se unen las líneas y en el método de determinación de la desviación típica muestral.

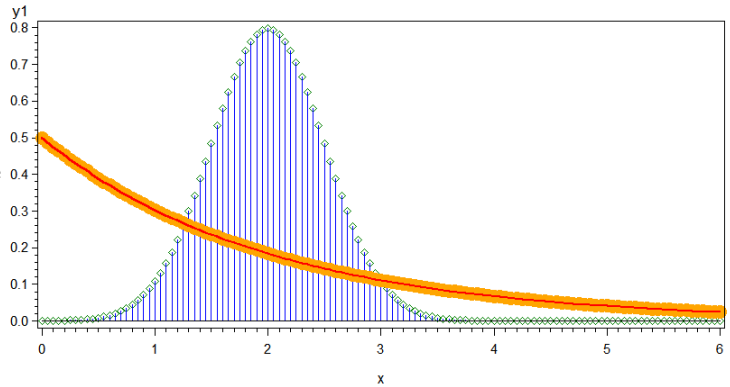
I=Rxxxxx, donde la interpolación ocurre a través de un ajuste por regresión. La primera x puede ser una L,Q,C según se quiera ajustar a una recta, una curva cuadrática o una cubica. Si le sigue un 0 desaparece el término independiente. Si se añade CLM o CLI calcula el intervalo de confianza para la media y las observaciones respectivamente. a continuación pueden aparecer el nivel de confianza (90, 95, 99).

```

proc gplot data=expon;
symbol i=needle v=diamond
cv=green ci=blue ;

symbol2 i=spline v=dot cv=orange
ci=red h=2 w=3;
plot (y1 y2)*x /overlay;
run;

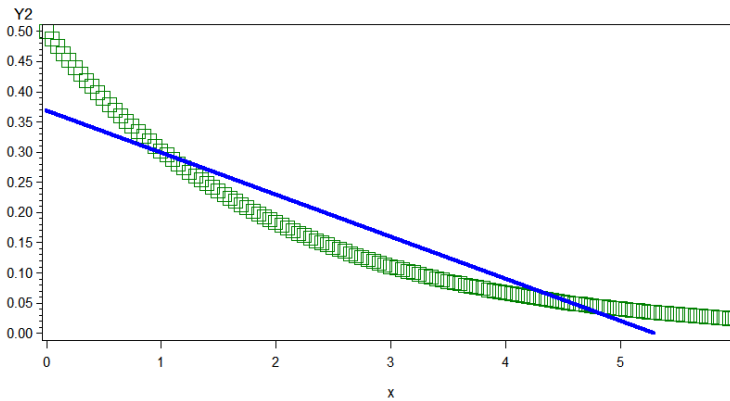
```



```

proc gplot data=expon;
symbol i=r h=1.5 v=square w=5;
plot y2*x;
run;

```



### 4.3.3 SENTENCIA AXIS.-

Se utiliza para el diseño de los ejes de una gráfica. La sintaxis es:

AXISn opciones (n 1-255 sirve para identificarla)

Las opciones más comunes son:

- ◆ COLOR= color Sirve para señalar el color del eje.
- ◆ LABEL = "none", "texto". Será el texto que aparezca para identificar la variable que se representa en el eje. Al igual que con las sentencias title, note, footnote se pueden incluir características del texto antes de escribir el mismo: ejemplo label=(H=2 c=blue 'camisas'). Si no se utiliza esta opción, por defecto se escribe el valor de la variable.
- ◆ LOGBASE = (n, pi, e) el eje tendrá escala logarítmica en la base indicada (n puede ser cualquier valor mayor que 1, e representa al logaritmo neperiano).
- ◆ LOGSTYLE = (power, expand) especifica los valores de los ejes cuando se usa la opción logbase. Si ponemos power los valores serán potencias de la base (por ejemplo 1,2,3,4...). Con expand aparecerán los valores elevados a la potencia, así si la base es 10 los valores que aparecerán en los ejes podrán ser (1,10,100,1000,...).
- ◆ MAJOR = (none, descripcion) parámetros de las marcas de los ejes. La descripción puede incluir color= height= number= (número de marcas en el eje) y width.
- ◆ ORDER = especificación es la forma de indicar los valores de las marcas de referencia en los ejes. Dos ejemplos:  
*order= 10 to 100 by 10* las marcas se pondrán en (10,20,30....80,90,100).  
*order = 14 15 16 18* (aparecerán justo esas marcas).
- ◆ ORIGIN = (x,y) marca las coordenadas del origen.
- ◆ STYLE = n especifica el estilo de la línea del eje (solida, 1, 2,..., etc). La n actúa igual de forma similar a como lo hacía en la sentencia symbol.

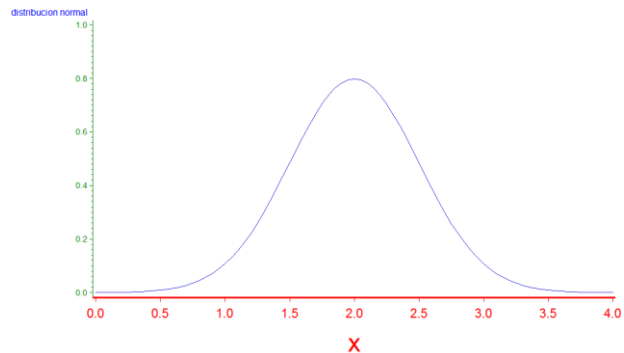
- ◆ VALUE = (NONE| descripción) permite describir los niveles del eje asociados con las marcas de referencias. None no escribe valores, la descripción comprende opciones como: Angle=, Color=, Font=, Height=, Justify=, Rotate=, Tick=. Un ejemplo:  
value=(c=red font=arial h=2)
- ◆ WIDTH = n especifica la anchura de la línea (n entero positivo).

Un ejemplo en el que se define axis1 para el eje horizontal de un gráfico generado por el procedimiento Gplot, y axis2 para el eje vertical.

```
axis1 c=red label=(h=3) order=0 to
4 by 0.5 value=(h=1.5) width=3
style=1 ;

axis2 c=green label=(c=blue
'distribucion normal') order=0 to 1
by 0.2 ;

proc gplot data=expon;
plot y1*x / vaxis=axis2 haxis=axis1
noframe;
run;
```



#### 4.4 PROCEDIMIENTO GCHART.

Este procedimiento realiza histogramas de variables continuas o diagramas de barra para las variables discretas (sean numéricas o carácter), verticales (sentencia VBAR), horizontales (sentencia HBAR), bloques (sentencia BLOCK), diagramas circulares denominados tartas (sentencia PIE) y estrellas (sentencia STAR). En general se trata de resumir gráficamente los datos. Sobre todo es muy utilizada para relacionar dos o más variables. En cualquiera de las gráficas la variable que se incluya en la sentencia será la que determine las etiquetas que contendrán a las barras o secciones. Las opciones de las sentencias controlan el tipo de estadístico que se representará (el total, las medias, porcentajes, ...) y si se realizan agrupaciones en función de otras variables.

La altura y la longitud de cada barra, segmento o línea es proporcional a su frecuencia, a la media u otro estadístico de alguna variable adicional. Asimismo el análisis se puede estratificar en niveles correspondientes a una tercera variable.

Si la variable es numérica el sistema determina el número apropiado de intervalos y grupos de datos. No obstante existen opciones para elegir los puntos medios de los intervalos y el número de los mismos.

En la sentencia **Proc GChart** se puede incluir las sentencias DATA= conjunto de datos de trabajo, ANNOTATE, GOUT= catalogo sas como en todos los procedimientos gráficos. Esta última sentencia indica donde puede ser guardado el gráfico que se obtenga en el procedimiento. El catalogo deberá estar formado por dos nombres, el primero de los cuales llamará a una librería de referencia previamente especificada con una sentencia del estilo libname. Caso de ser definida con un solo nombre, la gráfica se archivar en el catalogo WORK, que es temporal, sólo sobrevive a una sesión de trabajo.

La sintaxis específica del procedimiento es

```
PROC CHART DATA=nombre c.datos;
VBAR variables / opciones;
HBAR variables / opciones;
```

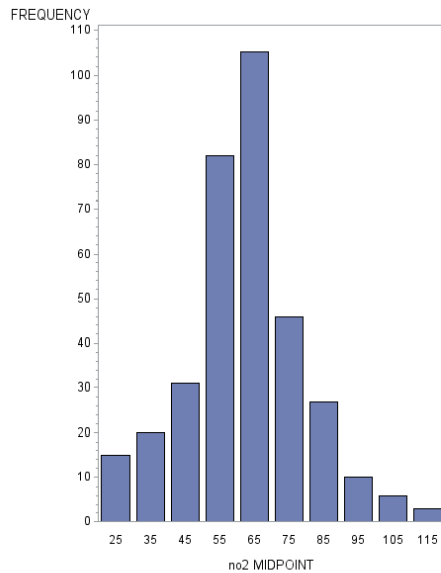
BLOCK variables / opciones;  
PIE variables /opciones;  
STAR variables / opciones;

Si se indican más de una variable en una sentencia, aparecerán tantas gráficas como variables. Las opciones (que siempre se deben encontrar tras el símbolo /) comunes a todas estas sentencias son:

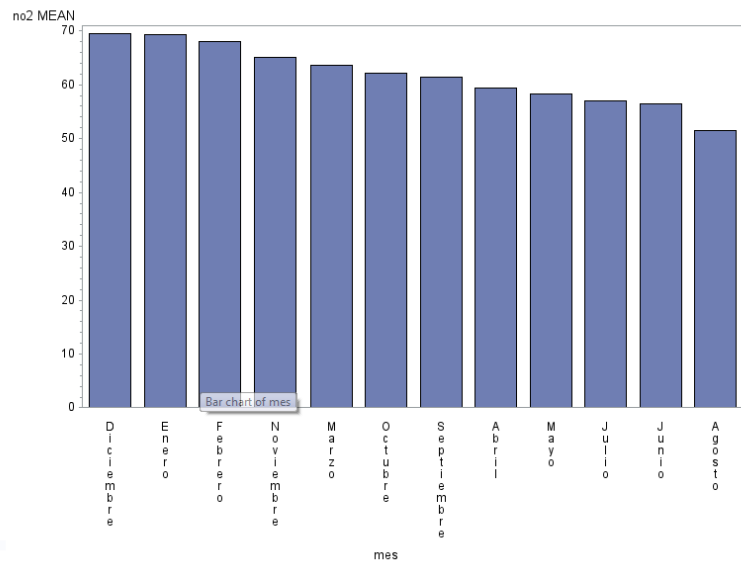
- SUMVAR= variable** Señala la variable que se quiere utilizar para determinar el contenido de cada gráfica. Así la sentencia: vbar x / sumvar=y. Extrae un diagrama de barras verticales, calculando el total de la variable y para cada valor de la variable x. Si incluimos la opción type=mean, lo que determinará cada barra será la media de la variable y para cada valor de la variable x.
- TYPE= [Freq|Cfreq| Pct|cpct| sum| mean ]** Representará el estadístico a describir por los diagramas de barras. Por defecto es Freq, excepto cuando se incluye sumvar, en cuyo caso es sum (totales).
- GROUP=variable** Sirve para indicarnos que aparecerá un diagrama (o subgráfica) para cada valor que tome la variable identificada en esta opción. Aparecen en la misma página para observar las diferencias. Los valores missing de una variable se tratan como un nivel de la misma.
- MIDDPOINTS= valores** Si la variable principal es continua, esta sentencia define los puntos medios de los intervalos para cada barra. La forma de definir los valores es similar a la de la sentencia DO del lenguaje básico.
- LEVELS = n** Muestra el número de barras cuando la variable indicada en la sentencia es continua. Por defecto las elige el sistema.
- DISCRETE** Le dice al sistema que la variable numérica que va a ser evaluada no es continua si no discreta, exigirá una barra para cada valor.
- MISSING** Considera que los valores missing de la variable que se representa en el eje de las abscisas, formen una barra o sector propio. Por defecto no se les considera.
- FREQ=variable** Señala a una variable que representa el contador de las observaciones que contiene realmente cada observación del conjunto de datos SAS (se utiliza si los datos ya vienen en frecuencias).



```
proc gchart data=contaminacion;
  vbar no2;
run;
```



```
proc gchart data=contaminacion;
  vbar mes /sumvar=no2 type=mean descending;
run;
```



Para las sentencias **HVAR**, **VBAR**, **BLOCK** se pueden incluir las siguientes opciones:

**SUBGROUP=variable** Cada barra mostrará diferentes colores y tamaños para cada valor de la variable que se indique en esta opción. Cada uno de los valores se representará de forma diferenciada en la gráfica.

**ANNOTATE =c.datos SAS** Especifica el conjunto de datos cuya información se pretende utilizar para diseñar el gráfico. Común a todas las sentencias gráficas.

**FRAME** Pide que se cree una cubierta que rodee al histograma.

**CFRAME = color** Color usado en el interior del cuadro generado por la opción frame (que es necesaria para que se ejecute esta).

**AUTOREF** Dibuja una línea horizontal (en vbar) o vertical (para el hvar) para cada valor de referencia del eje.

**CTEXT = color** Indica el color del texto del diagrama.

**DESCENDING** Ordena los bloques o barras en orden decreciente.

**ASCENDING** Ordena los bloques o barras en orden creciente.

**NAME=nombre** Definición del gráfico para su identificación en el catálogo. Debe contener como máximo 8 caracteres.

**DES = 'secuencia'** Descripción de hasta 40 caracteres que acompañará al gráfico en el catálogo al que se destine para su identificación.

**NOLEGEND** Elimina la leyenda que produce el sistema.

**PATTERNID = [group|subgroup|midpoint|by ]** Especifica la variable gráfica que controla el cambio en el patrón gráfico (pattern). Por defecto es la identificada en la opción subgroup (cada subgroup diferente color).

**AXIS = axisn** Indica que el eje tenga el diseño del n-esimo axis.

**CAXIS= color** Señala el color a utilizar por el eje (cuando no haya sentencia AXIS).

**RAXIS = axisn ó valores** Indica los valores que representan las marcas principales de los

GAXIS =AXISn

MAXIS =AXISn

NOAXIS

SPACE = n

WIDTH = n

G100

NOSTATS

REF=valor

BLOCKMAX = n

CAXIS = color

COURLINE = color

ejes bien por encontrarse definidos en el eje n-esimo, bien por ser especificado a través de valores (por ejemplo:10 to 100 by 10). Si los valores son mayores que el valor máximo, aquellos se truncan en este

Indica el número de la sentencia axis que caracteriza a la definición de cada grupo (el significado de cada gráfica) si se escribió la opción group.

Indica el número de la sentencia axis que caracteriza a los puntos intermedios de los intervalos (midpoints).

Suprime los ejes.

n determina el espacio entre barras en celdas.

Modifica la anchura de las barras a n celdas.

Fuerza a las barras a totalizar el 100% de cada grupo, se usa cuando se utilice la sentencia group.

Elimina los estadísticos en la salida de datos junto a las barras (sólo se utiliza en Hbar).

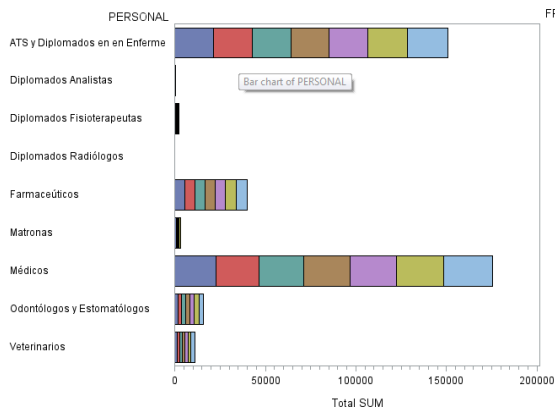
Aparece una línea de referencia en el eje para el valor indicado.

Escala los bloques, siendo el mayor el que toma el valor n.

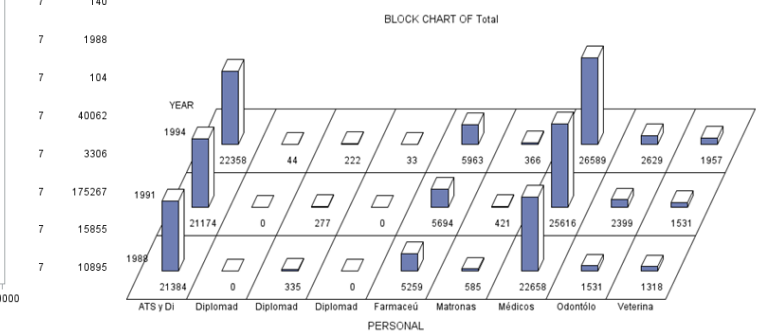
Especifica el color que se usara para dibujar los ejes.

Indica el color utilizado en la línea que se encuentra enfrente de los bloques. Estas tres últimas opciones se utilizan sólo en la sentencia block.

```
proc gchart data=salud;
  hbar personal / sumvar=total subgroup=year;
run;
```



```
proc gchart data=salud;
  where year in (1988, 1991, 1994);
  block personal / sumvar=total group=year;
run;
```



Existen posibilidades especiales para la sentencia pie (tartas), entre ellas se encuentra la posibilidad de diseñar un patrón gráfico (pattern). Este deberá ser concluido de la siguiente forma:

Pnxaaa

- ◆ n indica la densidad , los posibles valores de menor a mayor son (e-1-2-3-4-5-s).
- ◆ x (xindica que las líneas interiores sean cruzadas, si se omite son paralelas).
- ◆ aaa significa el ángulo con la horizontal (0-360).

Otras opciones:

ACROSS = n

Número de "pasteles" (graficas) incluidas en una página.

ANGLE = n Indica el ángulo con respecto a la horizontal de comienzo del primer sector (por defecto 0 ).

DOWN = n Número de graficas que aparecen verticalmente.

EXPLODE = lista Separa los sectores del circulo con los nombres indicados en la lista.

INVISIBLE = lista Indica los valores para los que no se quiere que aparezca sector.

MIDPOINTS = (valores) Se marcan los puntos intermedios.

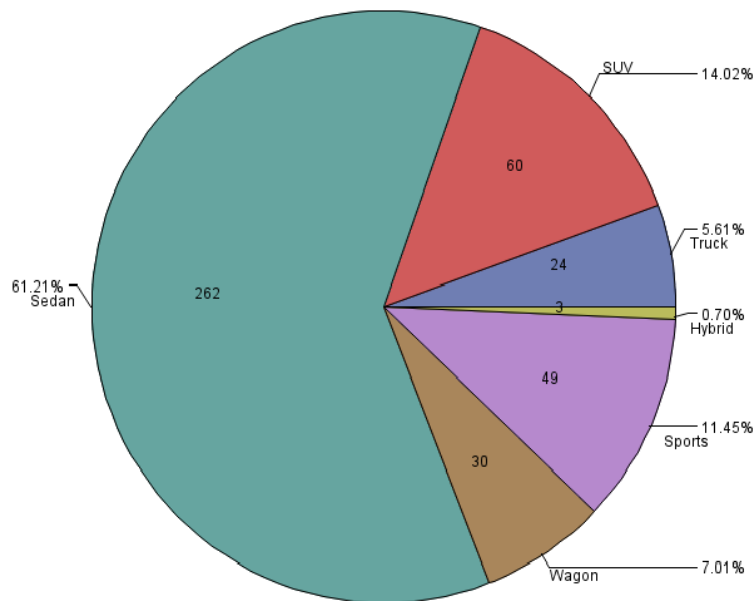
OTHERS = n Agrupa valores con menos de n por ciento en un sector (por defecto n=4).

SLICE = (arrow, inside, none, outside) Indica donde aparecen los puntos de marca con respecto al sector: con flechas, en el interior, en ningún sitio, y por fuera.

VALUE = (arrow, inside, none, outside) Realiza la misma acción que la opción anterior pero para el valor de la variable respuesta.

PERCENT = (arrow, inside, none, outside) Lo mismo pero con respecto a los porcentajes.

```
proc gchart data=sashelp.cars;
pie type / other=0 midpoints="Truck" "SUV" "Sedan" "Wagon" "Sports" "Hybrid"
value=inside percent=arrow slice=outside;
```



También hay opciones específicas de la sentencia star (estrella).-

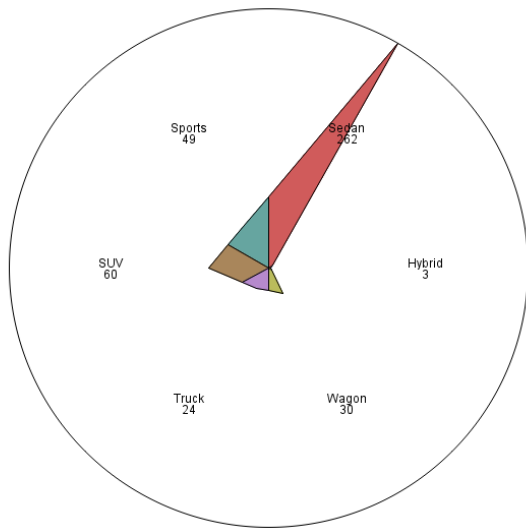
STARMIN = n radio minimo.

STARMAX = n radio maximo.

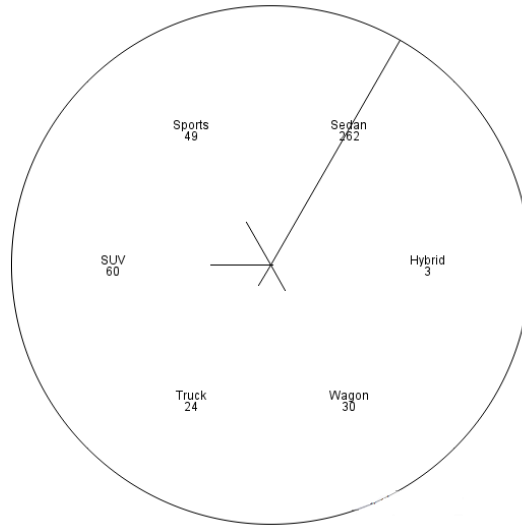
NOCONNECT indica que las líneas no se conectaran.

SLICE, VALUE, PERCENT la misma propiedad que para la sentencia pie.

```
proc gchart data=sashelp.cars;
star type / value=inside ;
run;
```



```
proc gchart data=sashelp.cars;
star type / value=inside noconnect ;
run;
```



## 4.5 PROCEDIMIENTO GPLOT.

Crea la gráfica de una variable frente a otra. La forma de crearla es indicando en cada punto de la gráfica una o más observaciones del conjunto de datos.

Las variables pueden ser numéricas o carácter. Las gráficas se escalan de forma automática aunque se pueden controlar mediante las sentencia AXIS.

Se pueden obtener varios tipos de gráficas:

- ◆ gráficos solapados en los mismos ejes (varias variables representadas a la vez). Opción Overlay
- ◆ Burbujas con círculos cuyo tamaño depende proporcionalmente del valor que tome otra tercera variable. Sentencia Buble
- ◆ Gráficas frente a dos ejes verticales . Sentencia PLOT2 o BUBLE2
- ◆ Gráficas que unen los puntos representativos de las observaciones a través de interpolación (controlada con la sentencia SYMBOL).
- ◆ Gráficas donde los ejes se encuentran en escala logarítmica (controlada mediante la sentencia AXIS).

La sintaxis del procedimiento es:

```
PROC GPLOT opciones;
BY variables ;
PLOT lo que se quiera dibujar / opciones ;
BUBLE lo que se quiera dibujar / opciones;
RUN;
```

Las opciones que se pueden incluir en la sentencia **PROC PLOT** son:

DATA= conjunto de datos; (por defecto el último creado).

UNIFORM. Ejes uniformes, idéntica escala en vertical y horizontal cuando se tienen más de una gráfica.

GOUT = librería Indica la carpeta donde se guardarán las gráficas.

Las sentencias que se utilizan en este procedimiento son:

SENTENCIA BY: Sirve para obtener diferentes gráficas según los valores de las variables aquí indicadas.

SENTENCIA PLOT

Sirve para realizar las gráficas. La forma de su utilización es:

$\text{PLOT (variable eje vertical * var eje horizontal) = una variable '}$

(si no se incluye la variable representa sin más el cruce de las dos variables, si se especifica generará tantas curvas como valores diferentes tome esa variable).

Se pueden disponer de forma distributiva. Así por ejemplo la sentencia

$\text{plot (y x)*(a b);}$

imprimirá 4 gráficas  $y^*a$   $y^*b$   $x^*a$   $x^*b$ .

Si se utiliza una variable auxiliar

$\text{Plot } y^*x=Z;$

produce una curva para cada valor de Z, donde el máximo número de valores permitidos a Z es 16.

Si la sentencia es:  $\text{PLOT } y^*x = n;$  nos indica que se utilice como formato de representación gráfica, la n-esima sentencia symbol.

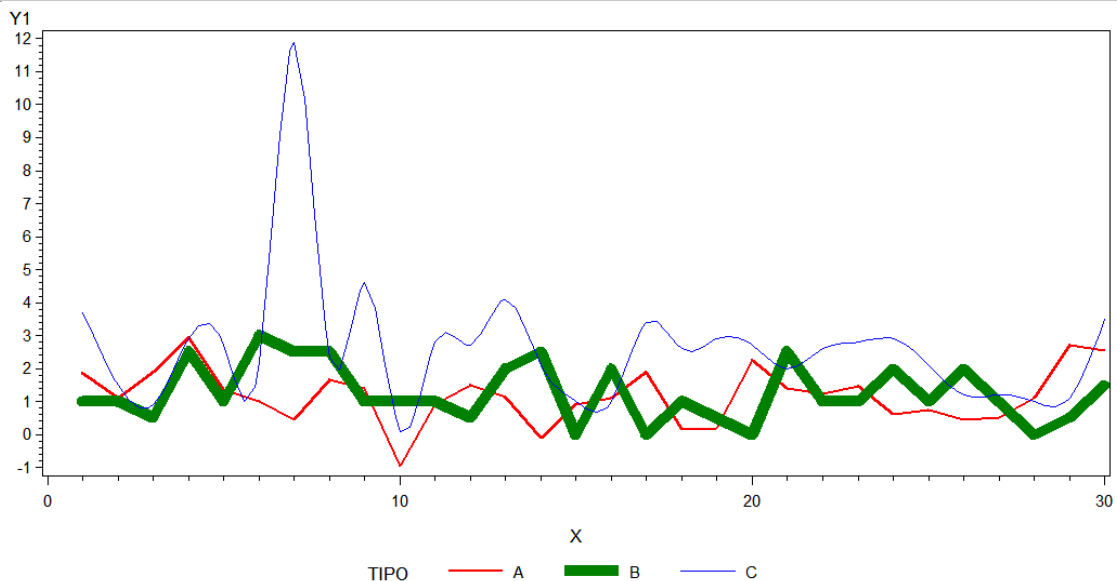
Ejemplos de esta sentencia son:

$\text{PLOT } y^*x = 3;$  (utilizará la sentencia symbol3 como referencia)

$\text{PLOT } y^*x = \text{sexo}$  (una gráfica para sexo masculino y otra para sexo femenino)

OTRO EJEMPLO:

```
PROC GPLOT;  
SYMBOL1 I=J CI=RED; SYMBOL2 I=J CI=GREEN ; SYMBOL3 I=SPLINE CI=BLUE;  
PLOT Y1*X=TIPO /LEGEND;  
RUN;
```



Dentro de las opciones de la sentencia PLOT se encuentran:

VAXIS Indicará los límites entre los que se encuentra el eje vertical. Por ejemplo:

*PLOT y\*x / vaxis= 10 to 100 by 5;*

señala que el eje vertical se extenderá desde el valor 10 hasta el 100, habiendo entre ellos marcas para los valores 10, 15, 20.....90, 95, 100. Esta opción admite el valor de un eje previamente creado: por ejemplo:

*vaxis=axis1*

Axis1 se ha debido crear previamente como se indico en el apartado 4.3.3

HAXIS Misma especificación de Vaxis pero para el eje horizontal.

VZERO El eje vertical comienza en el 0 (ignorando por tanto los valores negativos).

HZERO El eje horizontal comienza en el 0.

AUTOVREF Dibuja automáticamente líneas verticales.

AUTOHREF Dibuja automáticamente líneas horizontales.

HREF= valores Indica en qué lugar del eje horizontal se generan las líneas verticales. Se puede indicar un valor o un conjunto de la misma forma que como en la sentencia DO del lenguaje básico.

VREF = valores Genera líneas horizontales donde se indique (similar a HREF).

CH = color Marca el color de la línea vertical.

CV= color Indica el color de la línea horizontal.

LH = valor Señala el tipo de línea vertical que se utilizará en el eje horizontal, los posibles valores van desde 1 (solida) a 46 (muy discontinua).

LV = valor Ahora indica el Tipo de línea horizontal.

GRID Produce líneas horizontales y verticales que envuelven la totalidad de la grafica.

AREAS = n Especifica el área que va a ser coloreada, si es que se pretende colorear alguna. El area 1 es el espacio que se encuentra bajo la primera curva, el area 2 el que se encuentra sobre el área 1 y por debajo de la segunda curva, etc..

OVERLAY Permite varias curvas en una pantalla. (no debe especificarse cuando se utilice la forma  $y*x=z$ ). Por defecto cada curva o representación de observaciones se presenta en una gráfica diferente.

SKIPMISS Crea un salto en la curva, discontinuidad, debido a la existencia de un valor missing si la interpolación que se utiliza es join (unión a través de líneas rectas).

CAXIS = color Indica el color de los ejes así como el de sus marcas (tanto las mayores como las menores.

CTEXT = color Color de los textos en los ejes (nombres de variables y los valores de las marcas).

HMINOR = n Número de marcas pequeñas (de segundo orden) entre las marcas grandes de referencias(estas se especifican a traves de haxis y vaxis).

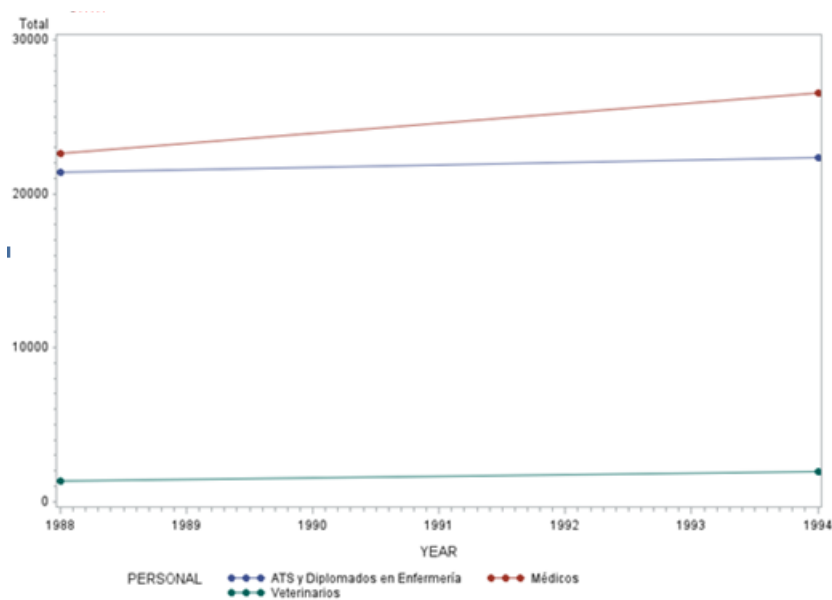
NOAXES Elimina los ejes en la grafica.

VREVERSE Intercambia el orden de los valores en el eje vertical.

HREVERSE Igual para el eje horizontal.

Ejemplo:

```
proc gplot data=salud;
  where personal in ('Médicos' 'ATS y Diplomados en Enfermería'
  'Veterinarios');
  plot total*year=personal /legend;
run;
```



## SENTENCIA BUBLE.-

La sintaxis de la sentencia es: `buble y*x = z / opciones` donde  $z$  es una variable.

Esta sentencia hace que se genere una gráfica con círculos en el plano  $xy$  cuyos tamaños dependen del valor que tome la variable  $z$ . Si  $z$  es positivo el círculo tiene traza continua, si es negativo traza discontinua.

Las opciones de la sentencia `plot` pueden utilizarse en esta sentencia. Además se tienen estas otras adicionales:

`BCOLOR=color` Color de los círculos.

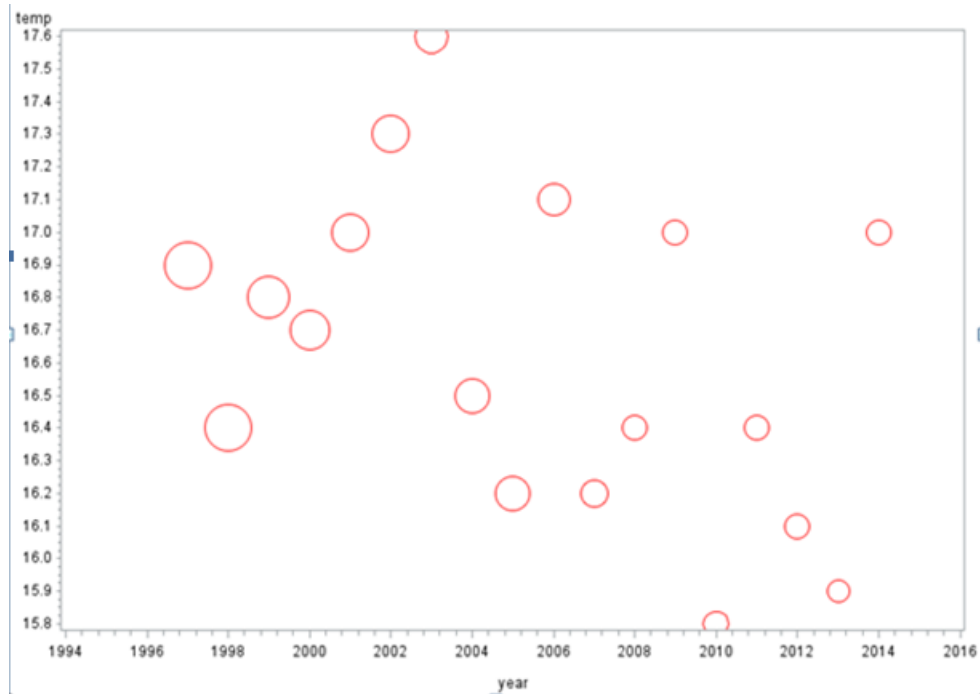
`BSCALE=area|radio` Sirve para determinar la escala que se utilizará para los círculos, es decir si serán proporcionales al área o al radio. Por defecto es el área.

`BSIZE=valor` Indica el tamaño del círculo mayor. Todos los demás se obtendrán multiplicando por este valor el cociente entre la observación correspondiente y el valor máximo.

`BLABEL` Etiqueta los círculos o burbujas con los valores formateados de la tercera variable. El color de este texto se especifica con la opción `CTEXT=`

`BFONT=fuente` Señala la fuente de letra a utilizar por las etiquetas. Por defecto `SIMPLEX`

```
proc gplot data=contaminacion;
  bubble temp*year=co / haxis=1994 to 2016 by 2
    bcolor=red bsize=20;
run;
```



### Sentencia symbol.-

Se utiliza según vimos con anterioridad, puede ser incluida en el interior del procedimiento o bien haberla generado previamente y llamarla desde el interior. Los datos que no se encuentren entre los límites de la gráfica no se incluyen para aproximar la curva, a no ser que se encuentre dentro de la sentencia symbol la opción:

MODE = include

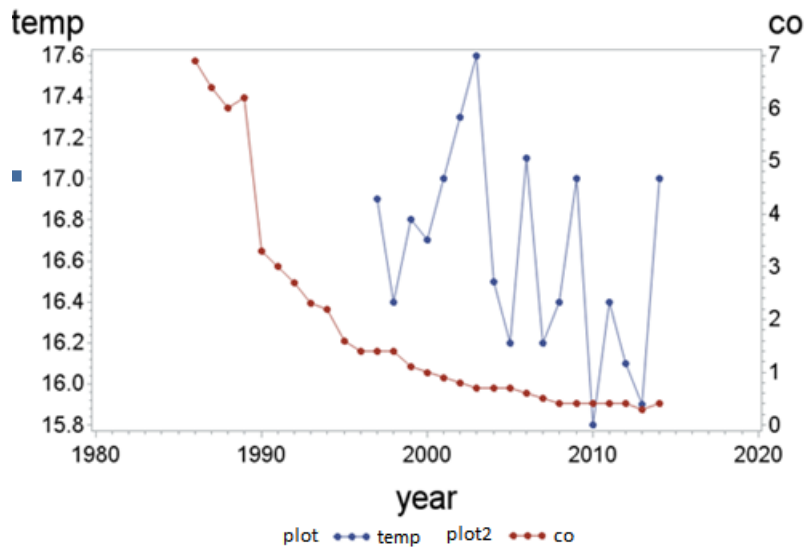
### SENTENCIAS PLOT2 Y BUBLE2.-

Tienen la particularidad de incluir un eje vertical en la parte derecha de la pagina, por lo demás tienen las mismas características y opciones que las sentencias PLOT y BUBLE.

Ejemplo:

```
proc gplot data=contaminacion;
  plot temp*year ;
  plot2 co*year;
run;
```





#### **4.6 PROCEDIMIENTO G3D**

El procedimiento G3D produce gráficos en el espacio tridimensional dependiendo de las variables especificadas en las sentencias PLOT o SCATTER. Una superficie tridimensional se obtiene con la sentencia PLOT. Mientras que un gráfico disperso de puntos se obtiene con la sentencia SCATTER. Se puede especificar el tamaño, forma y color a través de variables para la obtención del gráfico.

La sintaxis es:

PROC G3D [DATA=c.datos GOUT=grafico salida ANNOTATE=c.datos utilizado para la anotación y mejora de la gráfica. ] ;

PLOT yvariable\*xvariable=zvariable / opciones;

SCATTER yvariable\*xvariable=zvariable / opciones;

Es obligatoria la inclusión de una de estas dos sentencias.

Las opciones de la sentencia plot son:

ANNOTATE=c. datos annotate Indica el conjunto de datos utilizado para la anotación.

CAXIS=color Color usado para dibujar los ejes. Si no se indica, el segundo color de la lista colors=, será el que se utilice.

CBOTTOM=color Color utilizado para dibujar la superficie inferior de la gráfica. Si no se indica el cuarto color de la lista será el que se use.

CTEXT=color Color utilizado para el texto de la gráfica. Si no se indica se usa el primero de la lista de los colores (incluye las variables y las marcas de los ejes).

CTOP=color Color utilizado para especificar la superficie superior de la gráfica. Si no aparece esta opción se usará el tercer color de la lista colors=.

DESCRIPTION= 'texto' Especifica un texto de hasta 40 caracteres que aparecerá en la descripción del gráfico si se utiliza el procedimiento Greplay. Sirve para identificar la gráfica. Si no se escribe nada SAS/GRAPH indica descripciones por defecto que hacen referencia al procedimiento utilizado.

**GRID** Dibuja un enrejado de líneas coincidiendo con las marcas de los ejes.  
**SIDE** Produce una cara para una superficie del gráfico.  
**NAME='nombre'** Identificación a través del nombre que se le quiere dar al gráfico. Por defecto es el nombre del procedimiento con el que se ha obtenido. Muy útil si se pretende reproducir.  
**ROTATE=ángulo** o **ángulo1 ángulo2...o ROTATE=ángulo TO ángulo BY** incremento. Especifica uno o más ángulos con los que se rotará la grafica alrededor del eje z. Por defecto 70°. Se produzcan tantas gráficas como ángulos se indiquen.  
**TILT=ángulo** **TILT=ángulo1 ángulo2... TILT=ángulo TO ángulo BY** incremento. Especifica uno o más ángulos con los que se rotará la gráfica alrededor del eje y. Por defecto 70°. Se produzcan tantas gráficas como ángulos se indiquen.  
**NOAXIS** Requiere que los ejes y las etiquetas de los mismos no aparezcan.  
**NOLABEL** Pide que las etiquetas de los ejes (si es que hay) no se dibujen (pero si las marcas).  
**XTICKNUM= n YTICKNUM= n ZTICKNUM= n** Especifica el número de marcas que se localizarán en cada uno de los ejes.  $n \geq 2$ . Por defecto  $n=4$ .  
**XYTYPE= 1 ó 2 ó 3** Especifica el patrón a utilizar para dibujar la gráfica. Si vale 1, sólo se dibujan las líneas paralelas al eje X. Si vale 2 se dibujan las líneas paralelas al eje y, y si vale 3 se dibujan ambas líneas paralelas. Por defecto **XYTIPE=3**.  
**ZMAX= valor ZMIN= valor** Indican los valores máximo y mínimo de la variable Z que pertenecerán a la gráfica.

Todas las anteriores opciones pueden utilizarse en la sentencia SCATTER salvo CTOP, CBOTTOM, SIDE y XTYPE=. Además existen opciones exclusivas la sentencia SCATTER:

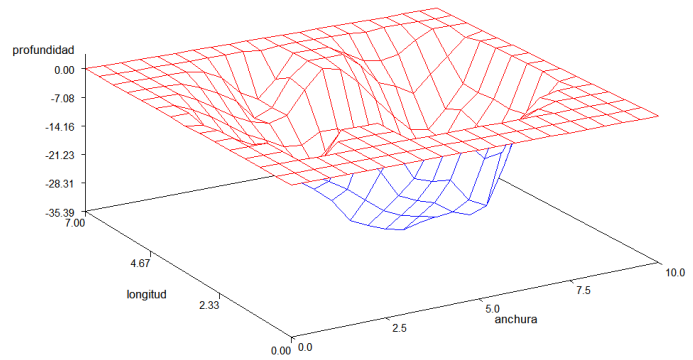
**NONEEDLE** Indica que la línea que conecta la base del plano con el punto tridimensional no se dibujará. Por defecto si lo hace.  
**SHAPE= 'nombre del símbolo'** Indica el nombre del símbolo entre caracteres (podría ser una variable carácter cuyos valores serían los símbolos). Se utilizarán para representar a los puntos tridimensionales. shape-variable. Los valores posible son: *BALLOON DIAMOND PRISM CLUB FLAG PYRAMID CROSS HEART SPADE CUBE PILLAR SQUARE CYLINDER POINT STAR*  
 Si la especificación de la opción SHAPE es con una variable cada observación vendrá representada por un símbolo diferente. El valor por defecto es PYRAMID.  
**SIZE= valor | variable** indicadora del tamaño. Indica una constante o una variable numérica que determina el tamaño del símbolo utilizado. Si se utiliza una variable, cada observación se representara por un símbolo de diferente tamaño. El valor por defecto es 1.

ejemplos:

```

proc g3d data=lago;
PLOT
longitud*anchura=profundidad /
cbottom=blue ctop=red rotate=30
      tilt=50 xticknum=5
zticknum=6 ;
run;

```



```

PROC G3D data=lago;
SCATTER
longitud*anchura=profundidad/
  SHAPE='CUBE' GRID CAXIS=RED
CTEXT=RED;
run;

```

