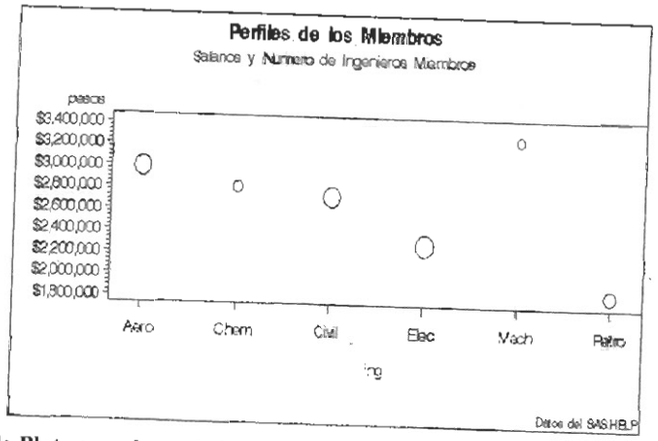
```
de los ejes de modo que las burbujas cercanas a los ejes no sean recortadas.*/
axis1 offset=(5,5);
/*Genera bubble plot. HAXIS= asigna la declaración AXIS1 al ej horizontal.
Los salarios promedios son asignados en formato dólares*/
proc gplot data=trabajos;
format pesos dollar11.;
bubble pesos*ing=num / haxis=axis1;
run;
quit;
```

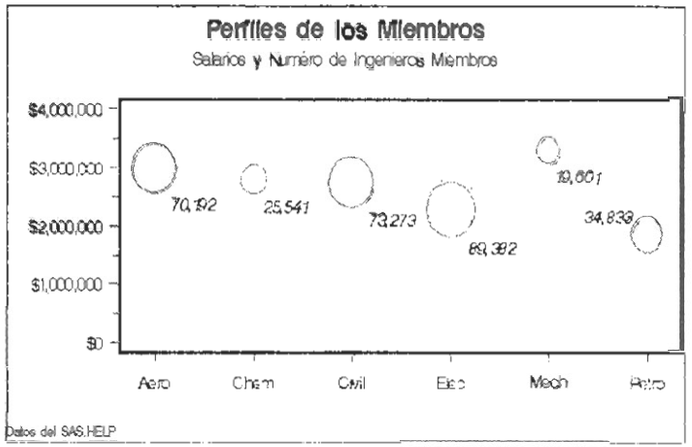


Ejemplo 3: Bubble Plot con valores de la variable bubble

```
/*Define el ambiente grafico*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=6 htext=4;
/*Define titulos y notas de pie.*/
title1 'Perfiles de los Miembros' ;
title2 h=4 'Salarios y Numero de Ingenieros Miembros' ;
footnote1 h=3 j=l 'Datos del SAS.HELP' ;
/*Define caracteristica de los ejes. AXIS1 suprime la etiqueta del eje horizontal y usa
OFFSET= para mover el primer y ultimo valor de marca principal del eje lejos de los ejes
verticales así las burbujas no son recortadas. AXIS2 usa ORDER= para ajustar los intervalos
de las marcas principales de eje vertical; aunque esto puede hacerse con VAXIS= en la
declaración BUBBLE, puede no suprimir  la etiqueta del eje y alterar otras caracteristicas*/
axis1 label=none offset=(5,5) width=3 value=(height=4);
axis2 order=(0 to 4000000 by 1000000) label=none major=(height=1.5) minor=(height=1)
```



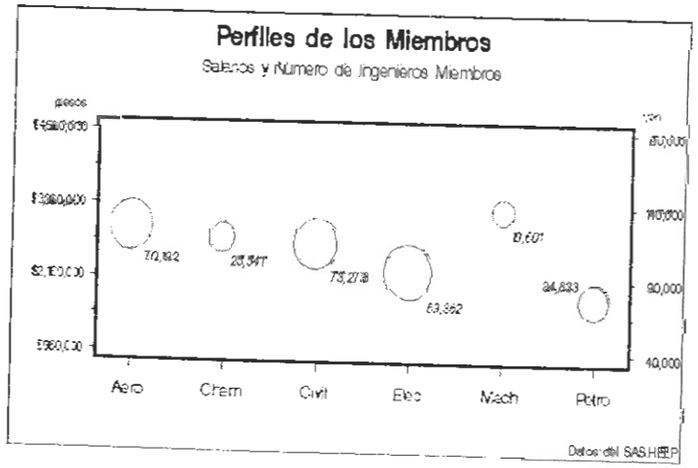
```
width=3 value=(height=4);
/*Genera bubble plot. VMINOR= especifica una marca de tick menor para el eje vertical.
BCOLOR= especifica los colores de las burbujas; BLABEL etiqueta las burbujas con los valores
de la variable num, y BFONT= especifica la fuente para el texto. BSIZE= incrementa el tamaño
de las burbujas aumentando el tamaño del factor de escala a 12. CAXIS= define el color de
las líneas de los ejes y de todas las marcas en estos*/
proc gplot data=trabajos;
format pesos dollar11. num comma7.0;
bubble pesos*ing=num / haxis=axis1
vaxis=axis2 vminor=1 bcolor=red
blabel bfont=swissi bsize=12
caxis=blue;
run;
quit;
```



Ejemplo 4: Bubble Plot con valores de la variable bubble y dos ejes verticales (escalas de valores diferentes)

```
/*Define el ambiente grafico*/
goptions reset=global gunit=pct border cback=white colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=6 htext=3;
/*Crea el conjunto de datos trabajos2 y calcula la variable YEN. El paso DATA usa una
declaración SET para leer el conjunto de datos en el archivo trabajos*/
data trabajos2;
set trabajos;
```

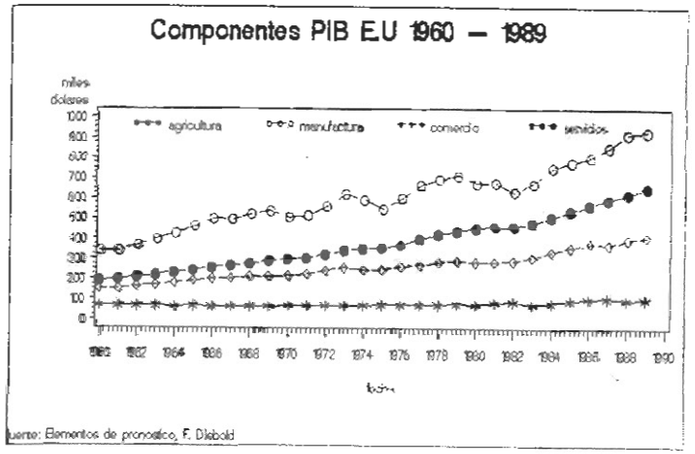
```
yen=pesos*125/3000;run;
/*Define títulos y notas de pie.*/
title1 'Perfiles de los Miembros' ;
title2 h=4 'Salarios y Numero de Ingenieros Miembros' ;
footnote j=r 'Datos del SAS.HELP' ;
/*Define características del eje horizontal.*/
axis1 offset=(5,5) label=none width=3 value=(h=4);
/*Genera el bubble plot con un segundo eje vertical. En la declaración BUBBLE, HAXIS=
especifica la definición AXIS1 y VAXIS= escala el eje izquierdo. En la declaración BUBBLE2,
VAXIS= escala el eje derecho. Ambos ejes representan el mismo rango de valores monetarios.
Las declaraciones BUBBLE y BUBBLE2 aseguran que las burbujas generadas por cada declaración
son identicas coordinando las especificaciones en BCOLOR=, que genera colores para las
burbujas; BSIZE=  incrementa el tamaño del factor de escala a 12 y CAXIS= especifica el
color de los ejes y de sus marcas. Las etiquetas de los ejes y los valores de sus marcas
usan el color por defecto, el cual es el primer color en la lista de colores*/
proc gplot data=trabajos2;
format pesos dollar11. num yen comma11.0;
bubble pesos*ing=num / haxis=axis1
vaxis=960000 to 4800000 by 1200000 hminor=0 vminor=1
blabel bfont=swissi bcolor=red bsize=12 caxis=blue;
bubble2 yen*ing=num / vaxis=40000 to 200000 by 50000
vminor=1 bcolor=red bsize=12 caxis=blue;
run;quit;
```



Ejemplo 5: Uso de la interpolación NEEDLE y dos ejes verticales

```
/*Define ambiente de graficación*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftext=swiss ftitle=swissb htitle=6 htext=3;
/* Crea el conjunto de datos MINNTEMP leyendo sólo las columnas de las variables mes (la
segunda variable que inicia en la décima columna en los datos) y de la variable f2 (que
inicia después de la columna 23)  */
data minntemp;
input @10 mes @23 f2;    /* Temperatura fahrenheit para Minneapolis */
/* Calcula temperatura en grados centigrados */
/* para Minneapolis                          */
c2=(f2-32)/1.8;
output;
datalines;
01JAN83  1   1   40.5  12.2  52.1
01FEB83  2   1   42.2  16.5  55.1
01MAR83  3   2   49.2  28.3  59.7
01APR83  4   2   59.5  45.1  67.7
01MAY83  5   2   67.4  57.1  76.3
01JUN83  6   3   74.4  66.9  84.6
01JUL83  7   3   77.5  71.9  91.2
01AUG83  8   3   76.5  70.2  89.1
01SEP83  9   4   70.6  60.0  83.8
01OCT83  10  4   60.2  50.0  72.2
01NOV83  11  4   50.0  32.4  59.8
01DEC83  12  1   41.2  18.6  52.5
;
 run;
/* Define titulos y notas de pies */
title1 'Temperatura promedio mensual para Minneapolis';
footnote1 j=l ' Fuente: 1984 American Express';
footnote2 j=l '          Appointment Book' ;
/* Define características de simbolos */
symbol1 interpol=needle ci=blue cv=red width=3 value=star height=3;
symbol2 interpol=none value=none;
/* define caracteristicas del  eje horizontal*/
```

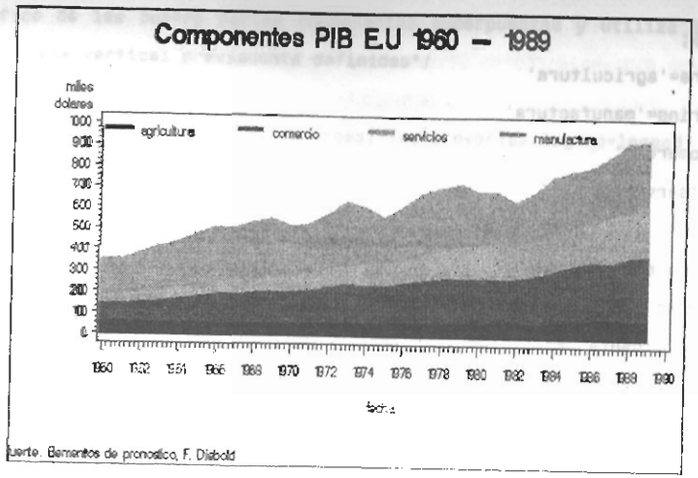
```
/*realiza grafico de las cuatro series componentes superpuestas y utiliza las leyendas de
cada serie y del eje vertical previamente definidas*/
proc gplot data=seis;
plot (agriculture manufacturing retail services)*fecha/overlay legend=legend1 vaxis=axis2;
run;
quit;
```




```
/*Define ambiente  de graficacion*/
goptions reset=global gunit=pct border cback=white
colors=(black, red, green, gray) ctext=black
ftitle=swissb ftext=swiss htitle=6 htext=3;
/*lectura de datos sobre series de componentes del PIB de E.U*/
data cinco;
infile'C:\Documents and Settings\Propietario\Mis documentos\clases\docum para
clases\estadística 3\diebold\FCST03\STYLEKP.dat' firstobs=2;
input AGRICULTURE MANUFACTURING RETAIL SERVICES;
run;

/*crea variable fecha y etiqueta variables originales*/
data seis;
set cinco;
fecha=intnx('year','31dec1959'd,_n_);
format fecha year7.;
t+1;
```



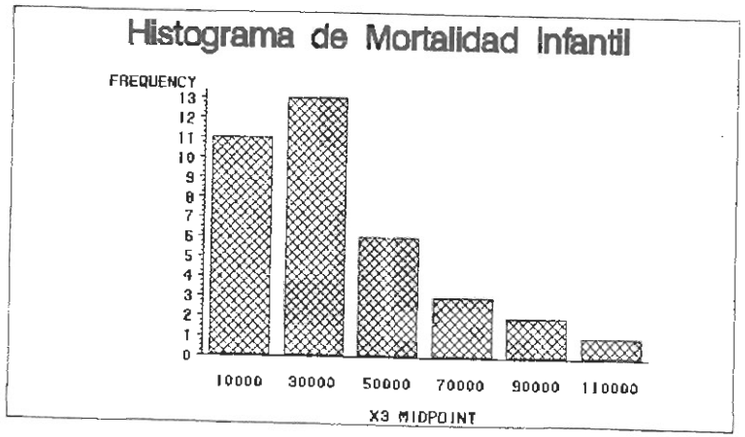
```
year=year(fecha);
label agriculture='agricultura'
      manufacturing='manufactura'
      retail='comercio'
      services='servicios';
run;
/*Define titulos y pie de notas*/
title1 'Componentes PIB E.U 1960 - 1989' ;
footnote1 h=3 j=l 'fuente: Elementos de pronostico, F. Diebold';
/*Define simbolos de graficacion*/
symbol1 i=join v=none c=black width=4 h=4;
symbol2 i=join v=none c=red width=4 h=4;
symbol3 i=join v=none c=green width=4 h=4;
symbol4 i=join v=none c=gray width=4 h=4;
/*define etiqueta del eje vertical*/
axis2 label=(h=3 'miles' justify=right 'dolares');
/*define leyenda del grafico*/
legend1 label=none
shape=symbol(4,2)
position=(top left inside)
mode=share;
/*Greafica las cuatro series mediante resalte de las areas entre ellas. Las series deben
nombrarse en la declaración plot en orden de menor a mayor
proc gplot data=seis;
plot (agriculture retail services manufacturing)*fecha/overlay legend=legend1 vaxis=axis2
areas=4;
run;
quit;
```



2.5.1.5 PROC GCHART

Este procedimiento permite realizar gráficas, que son útiles en la estadística descriptiva. Produce gráficos de barras verticales y horizontales, gráficos de bloques, gráficos pie y dunut charts, y gráficos de estrella para representar los valores de una variable, las relaciones entre dos o más variables o los valores de un estadístico calculado de una o más variables. Por ejemplo el siguiente programa muestra un histograma de la variable X3 (Mortalidad Infantil);

```
PROC GCHART DATA=misdatos;
  vbar x3;
  title 'Histograma de Mortalidad Infantil';
run;
```



La sintaxis de este PROC (sólo las declaraciones más usadas) es como sigue:

```
PROC GCHART <DATA=Base de datos>
<GOUT=<librería.>catálogo de salidas gráficas>
BLOCK variables </ opciones>;
HBAR | HBAR3D | VBAR | VBAR3D variables </ opciones>;
PIE | PIE3D | DONUT variables </ opciones>;
STAR variables </ opciones>;
Run;
Quit;
```

<GOUT=<librería.>catálogo de salidas gráficas>

Especifica el catálogo SAS en el cual se guardan las salidas gráficas producidas por el procedimiento GCHART.

BLOCK variables </ opciones>;

Crea gráficos de bloques en los cuales la altura de los bloques representa en valor del estadístico que se grafica para cada categoría de datos.

Las opciones pueden ser:

BLOCKMAX=Valor máximo. Especifica el valor del estadístico que se grafica para el bloque más alto y permite que se produzca una serie de gráficos de bloques usando la misma escala.

CAXIS=Color de rejilla. Especifica el color para rejilla de los puntos medios.

COUTLINE=Color de bordes de bloque. Es el mismo color usado para el interior.

CTEXT=Color del texto en el gráfico

DISCRETE Trata una variable numérica como una variable discreta en vez de continua. Un punto medio de separación es creado por cada valor formateado único de la variable. También crea una rejilla cuadrada separada y un bloque por cada valor único de la variable.

FREQ=Variable numérica. Especifica una variable en el conjunto de datos que represente un conteo o un peso por cada observación en el cálculo del estadístico que se grafica. Cada observación es contada el número de veces especificado por el valor de la variable numérica. FREQ= es válido con todos los estadísticos que se grafican.

G100 Calcula el porcentaje y el porcentaje acumulado separadamente para cada grupo. Los porcentajes individuales reflejan la contribución del punto medio al grupo y al 100% total para cada grupo.

GROUP=Variable de grupo. Organiza los datos de acuerdo a los valores de la variable de grupo. Esta variable puede ser de tipo carácter o numérica y siempre es tratada como una variable discreta.

LEGEND=LEGEND<1...99> Especifica las características de la leyenda para la leyenda de subgrupo. La declaración LEGEND es ignorada a menos que se use la opción SUBGROUP=

LEVELS=Número de puntos medios. Especifica el número de puntos medios para la variable numérica que se grafica.

MIDPOINTS=Lista de valores. Especifica los puntos medios para bloques. La forma de especificar la lista depende del tipo de variable. Para variables numéricas, la lista es o bien una lista explícita de valores, o un valor de inicio y un valor final con un incremento, o una combinación de ambas formas: n_1 <...n_k>; n_1 TO n_k <BY incremento>; n_1 <...n_k> TO n_l TO n_p <BY incremento>. Si una variable numérica tiene un formato asociado, los valores especificados deben ser los valores sin formato.

Para variables carácter, la lista de valores es una lista de valores caracteres únicos encerrados entre comillas y separados por espacios en blanco: : 'valor-l' <...'valor-n' >. Si una variable carácter tiene asociado un formato, los valores especificados deben ser los valores formateados.

NOHEADING Omite el encabezado automático que describe el tipo de estadístico graficado.

NOLEGEND Suprime la leyenda generada automáticamente por la opción SUBGROUP=

PATTERNID=BY | GROUP | MIDPOINT | SUBGROUP Especifica la forma en que se asigna el relleno de los patrones. Por defecto es PATTERNID=SUBGROUP. Los valores para PATTERNID= son: BY cambia los patrones cada vez que cambia el valor de la variable especificada por BY. Todos los bloques usan el mismo patrón si el procedimiento GCHART no incluye una declaración BY. GROUP cambia los patrones cada vez que el valor de la variable de grupo cambia. Todos los bloques en cada grupo (fila) usan el mismo patrón, pero se usa un patrón diferente para cada grupo. MIDPOINT cambia patrones cada vez que cambian los valores de los puntos medios. Si es usada la opción GROUP=, los respectivos patrones de puntos medios son repetidos por cada grupo. SUBGROUP cambia patrones cada vez que cambian los valores de la variable de subgrupo. Los bloques deben estar subdivididos por

SUBGROUP Cambia patrones cada vez que el valor de la variable subgrupo cambia. Los bloques deben ser subdivididos por la opción SUBGROUP= para el valor SUBGROUP que tiene un efecto. Sin SUBGROUP=, todas las caras de los bloques tienen el mismo patrón. Si se usa la opción SUBGROUP= y además se especifica un PATTERNID= con un valor diferente al de SUBGROUP, los segmentos de bloques usan el mismo patrón y resultan indistinguibles.

SUMVAR Especifica la variable a ser usada para calcular las sumas o medias. El procedimiento Gchart calcula la suma o la media de la variable a resumir para cada punto medio. Cuando se usa SUMVAR=, la opción TYPE= debe ser o SUM o MEAN. Con SUMVAR=, el ajuste por defecto es **TYPE=**SUM.

TYPE=estadístico . Especifica el estadístico a representar con el gráfico. Cuando no se usa la opción SUMVAR=, puede especificarse uno de los siguientes estadísticos:

FREQ Frecuencia (ajuste por defecto)

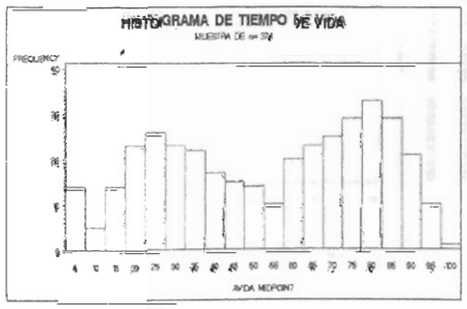
CFREQ Frecuencia acumulada

PERCENT o PCT porcentaje

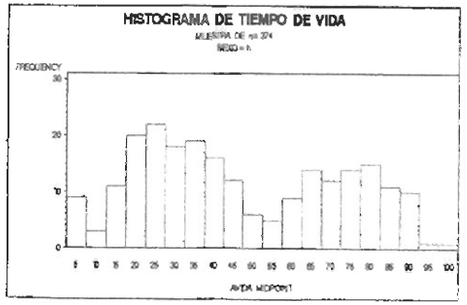
CPERCENT o CPCT porcentaje acumulado

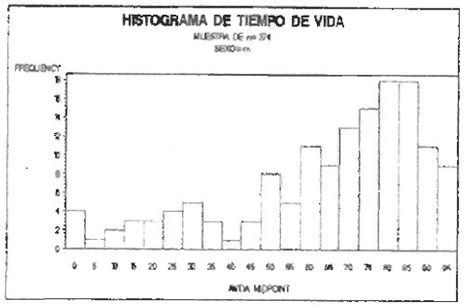
Existen muchas otras opciones que el usuario puede explorar. Veamos algunos de los gráficos obtenidos con el GCHART:

```
/*OPCIONES DE PRESENTACION DE SALIDAS*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=8 htext=3;
OPTIONS nodate nocenter nonumber ls=80 ps=60;
/*LECTURA DE DATOS Y DEFINICION DE VARIABLES*/
DATA VIDA;
INFILE 'C:\FECHAS.TXT';
INPUT SEXO $ DIAN MESN ANAC DIAM MESM AMUE;
AVIDA=AMUE+MESM/12+DIAM/360-ANAC-MESN/12-DIAN/360;
RUN;
/*HISTOGRAMA PARA LA VARIABLE AVIDA CON 20 BARRAS*/
PROC GCHART DATA=VIDA;
VBAR AVIDA/LEVELS=20 SPACE=0;
PATTERN1 VALUE=E C=RED;
TITLE1 'HISTOGRAMA DE TIEMPO DE VIDA';
TITLE2 'MUESTRA DE n=374';
RUN;
QUIT;
```

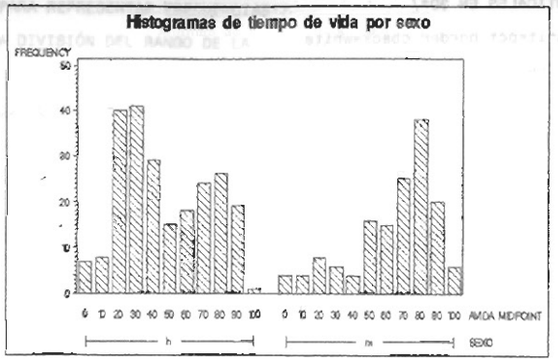


```
/*ORDENANDO LA BASE DE DATOS EN FORMA ASCENDENTE, POR LA VARIABLE SEXO*/
PROC SORT DATA=VIDA;
BY SEXO;
RUN;
/*OBTENIENDO HISTOGRAMAS SEPARADOS DE LA VARIABLE AVIDA, POR
CADA NIVEL DE LA VARIABLE SEXO*/
/*NOTA: EL USO DEL ARGUMENTO  BY...  EN UN PROC REQUIERE PRIMERO
ORDENAR LOS DATOS POR LOS NIVELES DE LA VARIABLE USADA PARA
LA DISCRIMINACION (EN EL EJEMPLO, ES SEXO)*/
PROC GCHART DATA=VIDA;
BY SEXO;
VBAR AVIDA/LEVELS=20 SPACE=0;
PATTERN1 VALUE=E;
RUN;
QUIT;
```

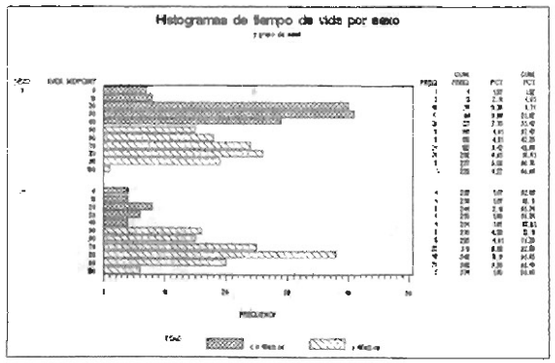




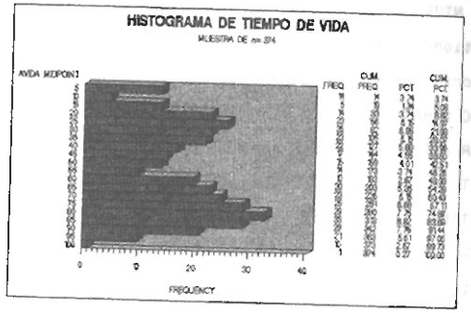

```
/*OBTENIENDO HISTOGRAMAS DE LA VARIABLE AVIDA, POR CADA NIVEL DE
LA VARIABLE SEXO, PERO UBICANDOLOS EN EL MISMO GRAFICO*/
PROC GCHART DATA=VIDA;
VBAR AVIDA/GROUP=SEXO;
PATTERN1 v=l1 C=BLUE;
title 'Histogramas de tiempo de vida por sexo';
RUN;
QUIT;
```

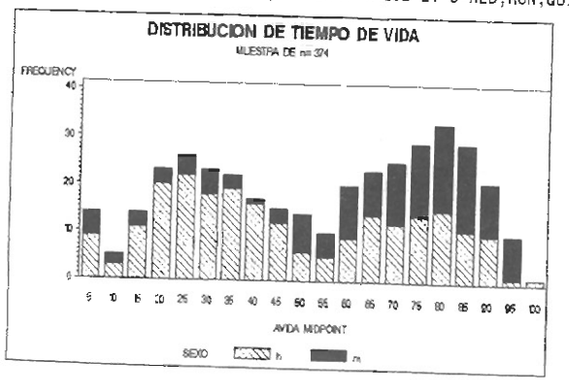

```
/*DEFINIENDO UNA NUEVA VARIABLE CATEGORICA A PARTIR DE CONDICIONES EVALUADAS SOBRE
UNA DE LAS VARIABLES EXISTENTES EN DATA VIDA*/
DATA VIDA2;
SET VIDA;
IF AVIDA<=45 THEN EDAD=" <=45anos ";IF AVIDA>45 THEN EDAD=">45años";run;
/*DIAGRAMA DE BARRAS HORIZONTALES DE LA VARIABLE VIDA DISCRIMINANDO POR
LOS NIVELES DE LAS VARIABLES SEXO Y EDAD*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=2;
PROC GCHART DATA=VIDA2;
HBAR AVIDA/GROUP=SEXO SUBGROUP=EDAD;
PATTERN1 C=BLUE VALUE=X2;
PATTERN2 C=RED V=L1;
TITLE1 'Histogramas de tiempo de vida por sexo';
TITLE2 'y grupo de edad';
RUN;
QUIT;
```



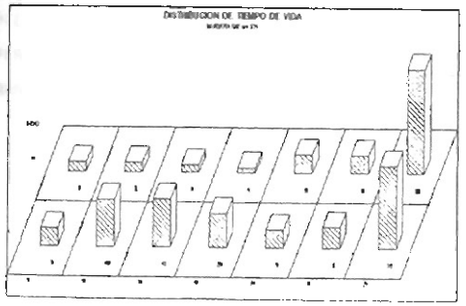
```
/*DIAGRAMA DE BARRAS VERTICALES EN 3D*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=5 htext=3;
TITLE1 'HISTOGRAMA DE TIEMPO DE VIDA';
TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
VBAR3D AVIDA/LEVELS=20 SPACE=0 COUTLINE=black;
PATTERN1 VALUE=s C=RED;
RUN;
QUIT;
/*DIAGRAMA DE BARRAS HORIZONTALES EN 3D*/
TITLE1 'HISTOGRAMA DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
HBAR3D AVIDA/LEVELS=20 SPACE=0 COUTLINE=black;
PATTERN1 VALUE=s C=RED;
RUN;QUIT;
```

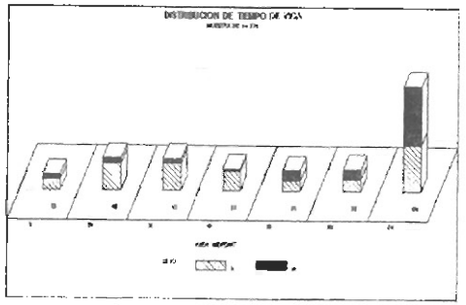



```
/*DIAGRAMAS DE BARRAS VERTICALES APILADOS*/
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
VBAR AVIDA/SUBGROUP =SEXO LEVELS=20 COUTLINE black;PATTERN1 VALUE=L1 C=RED;RUN;QUIT;
```




```
/*DIAGRAMA DE BLOQUES PARA REPRESENTAR FRECUENCIAS*/
/*NO PERMITE UNA AMPLIA DIVISIÓN DEL RANGO DE LA
VARIABLE DE INTERÉS*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=3 htext=2;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
BLOCK AVIDA/GROUP=SEXO MIDPOINTS=10 20 30 40 50 60 70 COUTLINE=black NOHEADING;
PATTERN1 VALUE=L1 C=RED;
RUN;QUIT;
/*OTRA VERSIÓN PERO APILANDO LOS BLOQUES SEGÚN SEXO*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=3 htext=2;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TTTLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
BLOCK AVIDA/SUBGROUP=SEXO MIDPOINTS=10 20 30 40 50 60 70 COUTLINE=black NOHEADING;
PATTERN1 VALUE=L1 C=RED;
RUN;QUIT;
```

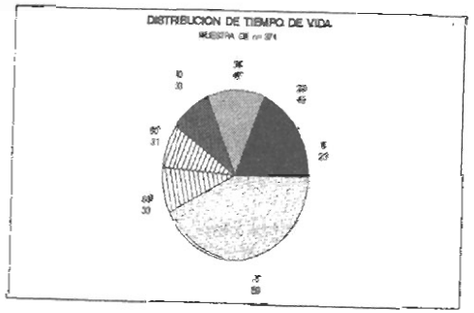




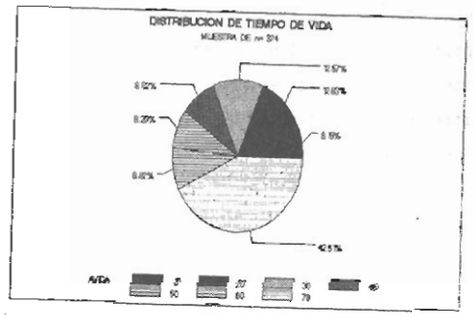
```
/*PIE CHARTS*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE AVIDA/  MIDPOINTS=10 TO 70 BY 10 COUTLINE=black NOHEADING; /*VER PIE 1*/
RUN;QUIT;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
```

```
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE AVIDA/  MIDPOINTS=10 TO 70 BY 10 LEGEND PERCENT=ARROW COUTLINE=black NOHEADING;/*VER PIE 2*/
RUN;QUIT;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE AVIDA/  EXPLODE=10 TO 70 BY 10 MIDPOINTS= 10 TO 70 BY 10 COUTLINE=black NOHEADING;/*VER PIE 3*/
RUN;QUIT;
```

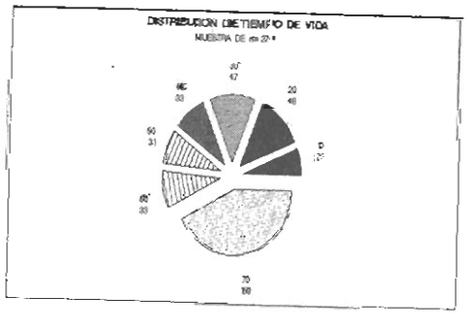
PIE 1



PIE 2



PIE 3

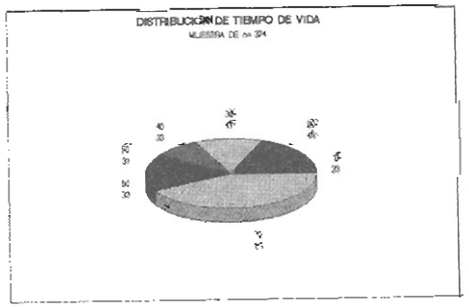



```
/*PIE CHARTS EN 3D*/
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE3D AVIDA/  MIDPOINTS=10 TO 70 BY 10 COUTLINE=black NOHEADING; /*VER PIE3D 1*/
RUN;QUIT;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
```

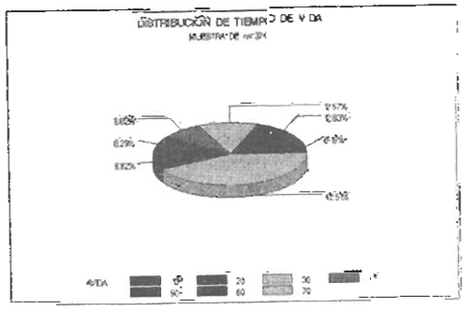


```
PROC GCHART DATA=VIDA;
PIE3D AVIDA/  MIDPOINTS=10 TO 70 BY 10 LEGEND PERCENT=ARROW COUTLINE=black NOHEADING;/*VER PIE3D 2*/
RUN;QUIT;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE3D AVIDA/  EXPLODE=10 TO 70 BY 10 MIDPOINTS= 10 TO 70 BY 10 COUTLINE=black NOHEADING;/*VER PIE3D 3*/
RUN;QUIT;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE3D AVIDA/  EXPLODE=10 TO 70 BY 10 MIDPOINTS= 10 TO 70 BY 10 COUTLINE=black NOHEADING legend
PERCENT=ARROW angle=30;/*VER PIE3D 4*/
RUN;QUIT;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red) ftitle=swissb ftext=swiss htitle=4 htext=3;
TITLE1 'DISTRIBUCION DE TIEMPO DE VIDA';TITLE2 'MUESTRA DE n=374';
PROC GCHART DATA=VIDA;
PIE3D AVIDA/  EXPLODE=10 TO 70 BY 10 MIDPOINTS= 10 TO 70 BY 10 COUTLINE=black NOHEADING legend
PERCENT=ARROW angle=60; /*VER PIE3D 5*/
RUN;QUIT;
```

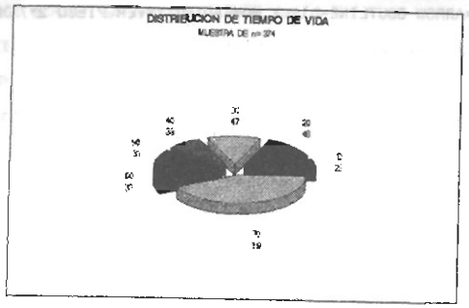
PIE3D 1



PIE3D 2

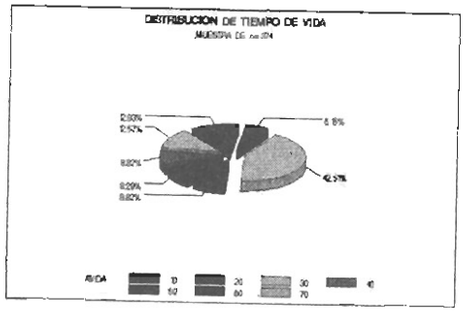


PIE3D 3



PIE3D 4 (angle=30)

PIE3D 5 (ANGLE=60)



2.5.1.6 El PROC BOXPLOT

Como su nombre lo indica, este procedimiento realiza gráficos box plot o de caja y bigote en alta resolución, esencialmente de una variable cuantitativa según niveles de una variable cualitativa o categórica, presentando los boxplots de cada nivel en el mismo plano, es decir, boxplots comparativos. Por tanto los boxplots de variables cuantitativas sin discriminar por subpoblaciones no son posibles con este procedimiento, a menos que se use una variable categórica artificial de un sólo nivel.

Antes de invocar este proc es necesario ordenar la base de datos en forma ascendente, según la variable categórica.

La sintaxis es como sigue:

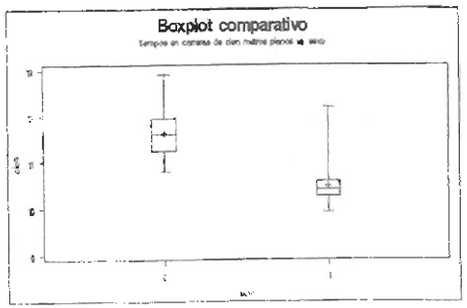
```
PROC BOXPLOT <opciones>;
PLOT variable-de- análisis*variable-de-grupo=variable-simbolo/<opciones>;
BY variables;
ID variables;
```



```
RUN;
QUIT;
```

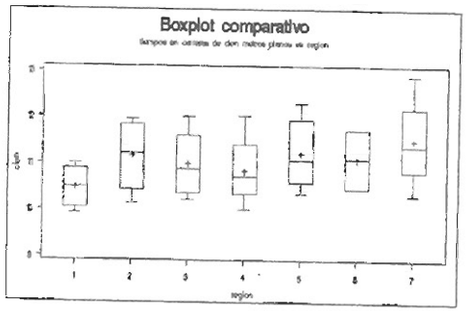
Ejemplo:

```
OPTIONS NODATE NOCENTER NONUMBER;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=6 htext=3;

/*lectura de datos y definición de variables*/
DATA UNO;
infile'c:\registros.txt' firstobs=2 expandtabs;
input cien dosc cuatroc ochoc milquin tresmil maraton pais $ region     cincomil     diezmil
sexo;
RUN;

/*creando un boxplot comparativo de la variable CIEN según niveles de la variable SEXO*/
proc sort data=uno;
by sexo;
run;
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. sexo';
proc boxplot data=uno;
plot cien*sexo;
run;
quit;
```

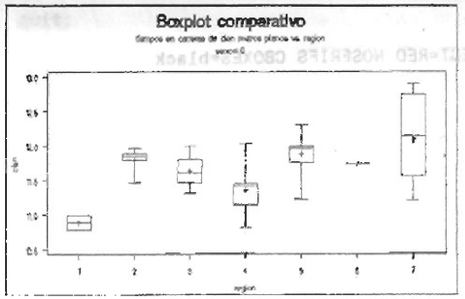


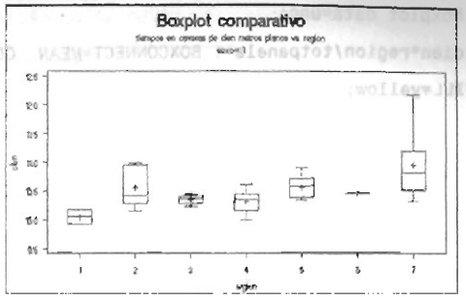

```
/*creando un boxplot comparativo de la variable CIEN según niveles de la variable REGION*/
```

```
/*para asegurar que todos los boxplots aparezcan en un sólo gráfico, se usa la opción */
/*totpanels=1. Esto es útil si la variable de grupo tiene muchos niveles*/
proc sort data=uno;
by region;
run;
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. region';
proc boxplot data=uno;
plot cien*region/totpanels=1;
run;
quit;
```

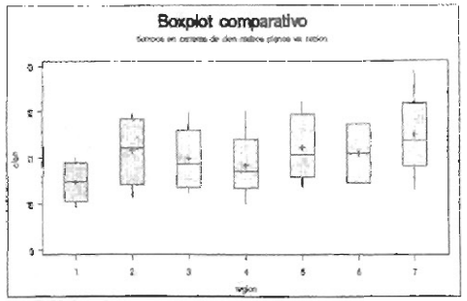



```
/*creando dos boxplots comparativos de CIEN vs. REGION para cada nivel de SEXO*/
proc sort data=uno;
by sexo region;
run;
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. region';
proc boxplot data=uno;
by sexo;
plot cien*region;
run;
quit;
```

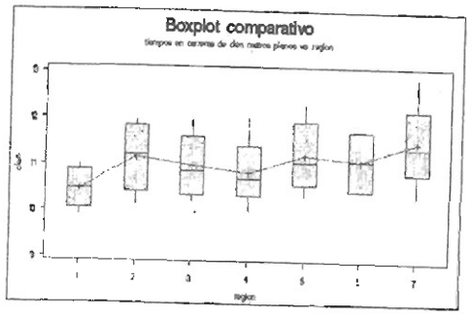



```
/*Boxplot comparativo por region eliminando las bases*/
de los extremos de los bigotes con la opción NSERIFS*/
/*Definiendo color de las lineas del boxplot con CBOXES*/
/*rellenando las cajas con color con la opción CBOXFILL*/
proc sort data=uno;
by region;
run;
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. region';
proc boxplot data=uno;
plot cien*region/totpanels=1 NOSERIFS CBOXES=black CBOXFILL=yellow;
run;
quit;
```

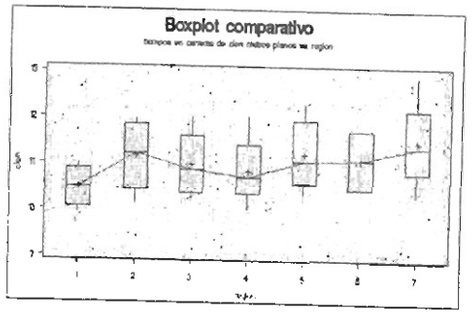



```
/*Además de las opciones NOSERIFS, CBOXES y CBOXFILL, se usa */
/*BOXCONNECT=MEAN (puede ser también igual a MEDIAN|MAX | MIN | Q1 | Q3 */
/*para conectar mediante lineas rectas el estadistico indicado.*/
/*CCONNECT especifica el color de la lineas anteriores*/
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. region';
```

```
proc boxplot data=uno;
plot cien*region/totpanels=1 BOXCONNECT=MEAN  CCONNECT=RED NOSERIFS CBOXES=black
CBOXFILL=yellow;
run;
quit;
```



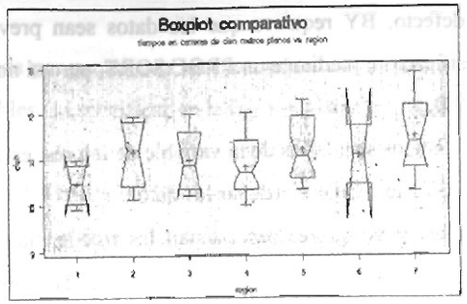

```
/*Adicional a las opciones del boxplot previo,*/
/*se especifica color para el fondo del gráfico con CFRAME*/
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. region';
proc boxplot data=uno;
plot cien*region/totpanels=1 BOXCONNECT=MEDIAN CCONNECT=RED NOSERIFS CBOXES=black
CBOXFILL=YELLOW CFRAME=CXF7E1C2;
run;
quit;
```




```
/*Boxplots con notches o intervalos de confianza para la mediana*/
TITLE1'Boxplot comparativo';
TITLE2'tiempos en carreras de cien metros planos vs. region';
proc boxplot data=uno;
plot cien*region/totpanels=1 notches CBOXES=black CBOXFILL=yellow CFRAME=CXF7E1C2;
run;
```



```
quit;
```



2.5.2 PROCEDIMIENTOS PARA ANÁLISIS CON UNA Y DOS MUESTRAS

2.5.2.1 PROC MEANS

Este procedimiento proporciona herramientas para resumir datos, calcular estadísticos descriptivos a través de todas las observaciones disponibles y dentro de grupos de observaciones definidos por los niveles de una variable categórica o de clasificación. Específicamente, permite calcular:

- Estadísticos descriptivos basados en los momentos.
- Cuantiles estimados, incluyendo la mediana
- Límites de confianza para la media
- valores extremos
- Prueba t

La sintaxis es como sigue:

```
PROC MEANS <opciones> <palabras claves_estadísticos)>;
 BY <DESCENDING> variable-1 <... <DESCENDING> variable-n><NOTSORTED>;
 CLASS variable(s) </ opciones>;
 FREQ variable;
 ID variable(s);
OUTPUT <OUT=conjunto-datos-SAS> <especificaciones de estadisticos de salida>
<id-especificaciones-grupo> <máximo-id-especificaciones>
<mínimo-id-especificaciones> </opciones> ;
 TYPES request(s);
VAR variable(s) < / WEIGHT=variable-peso>;
 WAYS lista;
 WEIGHT variable;
```

BY calcula estadísticos separados de la variable de interés, en cada grupo de la variable de clasificación especificada en BY. Por defecto, BY requiere que los datos sean previamente ordenados por los niveles de la variable de clasificación, mediante un PROC SORT, pero si no se ordenan necesariamente debe invocarse **NOTSORTED**.

CLASS también calcula estadisticos separados de la variable de interés, en cada grupo de la variable de clasificación; a diferencia de BY, no requiere ordenar los datos.

FREQ identifica una variable cuyos valores representan las frecuencias de cada observación de la variable de interés, cuando los datos son dados de manera resumida (como una tabla de frecuencias simples).

ID incluye variables de identificación adicional en el conjunto de salida de datos.

OUTPUT OUT= Crea un conjunto de datos de salida que contiene los estadísticos especificados y las variables de identificación.

TYPES identifica combinaciones específicas de variables de clase usadas para subdividir los datos.

VAR identifica las variables de análisis y su orden en los resultados.

WAYS Especifica el número de formas para hacer combinaciones únicas de variables de clase.

WEIGHT Identifica una variable cuyos valores ponderan cada observación en los cálculos estadísticos.

Las opciones en la línea del PROC MEANS son entre otras, las siguientes:

DATA el nombre del conjunto de datos SAS sobre el que se aplica el análisis.

ALPHA=valor Especifica el nivel de significancia para la prueba t y el nivel de confianza (1-alpha) para los intervalos de confianza para la media.

CLASSDATA= conjunto de datos SAS. Especifica un conjunto de datos SAS que contiene las combinaciones de los valores de las variables de clasificación que deben ser presentados en la salida y permite por tanto filtrar o suplementar al conjunto de datos de análisis. Cualesquier combinación de las variables de clasificación que ocurra en el conjunto de datos especificado en CLASSDATA= pero no dentro del conjunto de datos donde están las variables de análisis, aparecerán en la salida con una frecuencia de cero. El conjunto de datos en **CLASSDATA=** debe contener todas las variables de clasificación, sus tipos de datos y sus formatos deben coincidir con las variables de clase correspondientes en el conjunto de datos de análisis.

COMPLETETYPES crea todas las posibles combinaciones de las variables de clase aún si las combinaciones no ocurren dentro del conjunto de datos de análisis. La opción **PRELOADFMT** de la declaración **CLASS** asegura que el PROC MEANS produzca todos los rangos o valores de formato

definidos por el usuario para las combinaciones de las variables de clasificación, aún cuando la frecuencia sea cero.

NOPRINT Cancela la salida por pantalla de los resultados del PROC **MEANS**.

Entre las opciones disponibles, a especificar en la linea del proc means, se tienen entre otras:

PALABRAS CLAVES DE ESTADÍSTICOS

Especifican cuales estadísticos calcular y el orden en que aparecerán en los resultados por pantalla (si no se especifica noprint). Éstas son:

CLM para solicitar límites de confianza para las medias de las variables de análisis.

RANGE para solicitar el rango.

CSS Suma de cuadrados corregidas de las variables de análisis.

SKEWNESS ó **SKEW** para solicitar el coeficiente de asimetría muestral.

CV para solicitar el coeficiente de variación muestral.

STDDEV ó **STD** para la desviación estándar muestral.

KURTOSIS ó **KURT** para el coeficiente de curtosis muestral.

STDERR para solicitar el error estándar estimado de estimación de la media.

LCLM, UCLM para solicitar el límite inferior y superior, respectivamente, del intervalo de confianza bilateral para la media.

SUM para la suma de cada variable de análisis.

MIN, MAX para obtener respectivamente, el mínimo y el máximo muestral de cada variable de análisis.

MEAN, MEDIAN ó **P50**, para obtener respectivamente, la media y la mediana muestral.

USS para obtener la suma de cuadrados de cada variable de análisis.

VAR para la varianza muestral.

N para obtener el total de observaciones.

Q1 ó **P25**, **Q3** ó **P75** para obtener el primer y tercer cuartil muestral, respectivamente.

P5, **P10**, **P90**, **P95**, para obtener respectivamente los percentiles muestrales 5, 10, 90 y 95.

Los anteriores estadísticos también pueden ser guardados en un data creado con la declaración output out= invocando palabra-clave=nombre, por ejemplo mean=media.

A continuación se presentan algunos ejemplos:

```
OPTIONS NODATE NOCENTER NONUMBER;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=5 htext=3;

/*lectura de datos y definición de variables*/
DATA UNO;
infile 'c:\registros.txt' firstobs=2 expandtabs;
input cien dosc cuatroc ochoc milquin tresmil maraton pais $ region cincomil diezmil sexo;
RUN;

/*Obteniendo por defecto, la media, la desviaicón estándar y el valor máximo y mínimo*/
proc means data=uno;
var cien dosc;
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
cien	110	11.0448182	0.7033325	9.9300000	12.9000000
dosc	110	22.2910000	1.6304595	19.7200000	27.1000000

```
/*especificando los estadisticos de interés a calcular*/
proc means data=uno n mean std median q1 q3 min max range skewness kurtosis p5 p50 p95;
var cien dosc;
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Median	Lower Quartile	Upper Quartile
cien	110	11.0448182	0.7033325	10.9900000	10.4100000	11.6100000
dosc	110	22.2910000	1.6304595	21.9550000	20.8100000	23.5400000

Variable	Minimum	Maximum	Range	Skewness	Kurtosis	5th Pctl
cien	9.9300000	12.9000000	2.9700000	0.4036283	-0.8546393	10.1200000
dosc	19.7200000	27.1000000	7.3800000	0.4947259	-0.6321548	20.2200000

Variable	50th Pctl	95th Pctl
cien	10.9900000	12.2300000
dosc	21.9550000	25.0000000

```
/*omitindo la salida por pantalla, pero guardando estadisticos de intérés en un data llamado
estadisticos*/
proc means data=uno noprint;
var cien dosc;
output out=estadisticos(DROP=_FREQ_ _TYPE_) n=n100 n200 mean=media100 media200 std=desv100 desv200
median=mediana100 mediana200 q1=q1_100 q1_200 q3=q3_100 q3_200 min=min_100 min_200 max=max100 max200
range=rango100 rango200 skewness=asim100 asim200 kurtosis=kurtos100 kurtos200 p5=perc05_100 perc05_200
p50=perc50_100 perc50_200 p95=perc95_100 perc95_200;
run;

/*imprimiendo por pantalla el contenido del data creado en el proc means, estadisticos*/
proc print data=estadisticos;run;

Obs n100 n200 media100 media200 desv100 desv200 mediana100 mediana200 q1_100 q1_200
 1   110  110  11.0448  22.291  0.70333 1.63046    10.99      21.955     10.41  20.81
Obs q3_100  q3_200  min_100  min_200  max100  max200  rango100  rango200  asim100  asim200
 1   11.61   23.54    9.93    19.72    12.9    27.1     2.97      7.38    0.40363  0.49473
                              perc05_  perc05_  perc50_  perc50_  perc95_  perc95_
Obs kurtos100  kurtos200      100      200      100      200      100      200
 1   -0.85464   -0.63215     10.12    20.22    10.99    21.955   12.23     25

/*solicitando la media y la desviación estándar de las variables cien y dosc pero por nivel de la
variable sexo. Guardando media y desviaicón estándar en data estad2 */
proc means data=uno mean std;
class sexo;
var cien dosc;
output out=estad2 mean=media std=desv;
run;
```

The MEANS Procedure

sexo	N Obs	Variable	Mean	Std Dev
0	55	cien	11.6185455	0.4522103
		dosc	23.6416364	1.1110602

1	55	cien	10.4710909	0.3514292
		dosc	20.9403636	0.6446474

```
/*imprimiendo por pantalla el data estad2*/
proc print data=estad2;run;
```

Obs	sexo	_TYPE_	_FREQ_	media	desv
1	.	0	110	11.0448	0.70333
2	0	1	55	11.6185	0.45221
3	1	1	55	10.4711	0.35143

/*de nuevo solicitando la media y la desviación estándar de las variables cien y dosc mediante la declaración BY, observe que los datos no fueron ordenados previamente por lo que se usa la opción notsorted. Guardando algunos estadísticos en data estad3*/

```
proc means data=uno mean std;
by sexo notsorted;
var cien dosc;
output out=estad3 mean=media_100 media_200 std=desv100 desv_200;
run;
```

sexo=0

The MEANS Procedure

Variable	Mean	Std Dev
cien	11.6185455	0.4522103
dosc	23.6416364	1.1110602

sexo=1

Variable	Mean	Std Dev
cien	10.4710909	0.3514292
dosc	20.9403636	0.6446474

```
/*imprimendo por pantalla el data estad3*/
proc print data=estad3;run;
```

Obs	sexo	_TYPE_	_FREQ_	media_100	media_200	desv100	desv_200
1	0	0	55	11.6185	23.6416	0.45221	1.11106
2	1	0	55	10.4711	20.9404	0.35143	0.64465

/*usando dos criterios de clasificación (sexo y region) para el analisis de la variable cien y guardando estadisticos de interés en data estad4*/

```
proc means data=uno mean std;
class sexo region;
var cien;
output out=estad4 mean=media std=desv;
run;
```

The MEANS Procedure

Analysis Variable : cien

sexo	region	N Obs	Mean	Std Dev
0	1	2	10.8950000	0.1484924
	2	5	11.7880000	0.1938298
	3	4	11.6300000	0.2831960
	4	23	11.3530435	0.3045180
	5	13	11.8753846	0.3078045
	6	1	11.7300000	.
	7	7	12.0771429	0.6188892
1	1	2	10.0500000	0.1697056
	2	5	10.5520000	0.3856423
	3	4	10.3450000	0.0911043
	4	23	10.3195652	0.1763900
	5	13	10.5576923	0.1838094
	6	1	10.4600000	.
	7	7	10.9442857	0.6173291

```
/*imprimiendo por pantalla el data estad4*/
proc print data=estad4;run;
```

Obs	sexo	region	_TYPE_	_FREQ_	media	desv
1	.	.	0	110	11.0448	0.70333
2	.	1	1	4	10.4725	0.50493
3	.	2	1	10	11.1700	0.71215
4	.	3	1	8	10.9875	0.71394
5	.	4	1	46	10.8363	0.57749
6	.	5	1	26	11.2165	0.71633
7	.	6	1	2	11.0950	0.89803
8	.	7	1	14	11.5107	0.83558

9	0	.	2	55	11.6185	0.45221
10	1	.	2	55	10.4711	0.35143
11	0	1	3	2	10.8950	0.14849
12	0	2	3	5	11.7680	0.19383
13	0	3	3	4	11.6300	0.28320
14	0	4	3	23	11.3530	0.30452
15	0	5	3	13	11.8754	0.30780
16	0	6	3	1	11.7300	.
17	0	7	3	7	12.0771	0.61889
18	1	1	3	2	10.0500	0.16971
19	1	2	3	5	10.5520	0.38564
20	1	3	3	4	10.3450	0.09110
21	1	4	3	23	10.3196	0.17639
22	1	5	3	13	10.5577	0.18381
23	1	6	3	1	10.4600	.
24	1	7	3	7	10.9443	0.61733

```
/*usando dos criterios de clasificación (sexo y region) para el analisis de las variables cien y dosc,
y guardando estadísticos de interés en data estad5*/
proc means data=uno mean std;
class sexo region;
var cien dosc;
output out=estad5 mean=media100 media200 std=desv100 desv200;
run;
```

The MEANS Procedure

sexo	region	N Obs	Variable	Mean	Std Dev
0	1	2	cien	10.8950000	0.1484924
			dosc	22.0400000	0.2969848
	2	5	cien	11.7880000	0.1938298
			dosc	23.9720000	0.6513601
	3	4	cien	11.6300000	0.2831960
			dosc	23.6575000	0.7334110
	4	23	cien	11.3530435	0.3045180
			dosc	23.0326087	0.8012162
	5	13	cien	11.8753846	0.3078045
			dosc	24.2338462	0.7501060
	6	1	cien	11.7300000	.
			dosc	23.8800000	.
	7	7	cien	12.0771429	0.6188892
			dosc	24.7214286	1.6021280
1	1	2	cien	10.0500000	0.1697056
			dosc	19.9850000	0.3323402
	2	5	cien	10.5520000	0.3856423
			dosc	21.2500000	0.6242596
	3	4	cien	10.3450000	0.0911043
			dosc	20.7725000	0.2559134
	4	23	cien	10.3195652	0.1763900
			dosc	20.5982609	0.3556512
	5	13	cien	10.5576923	0.1838094
			dosc	21.2492308	0.3308943
	6	1	cien	10.4600000	.
			dosc	20.6600000	.
	7	7	cien	10.9442857	0.6173291
			dosc	21.6785714	1.0194513

```
/*imprimiendo data estad5*/
proc print data=estad5;run;
```

Obs	sexo	region	_TYPE_	_FREQ_	media100	media200	desv100	desv200
1	.	.	0	110	11.0448	22.2910	0.70333	1.83046
2	.	1	1	4	10.4725	21.0125	0.50493	1.21404
3	.	2	1	10	11.1700	22.6110	0.71215	1.55560
4	.	3	1	8	10.9875	22.2150	0.71394	1.62378
5	.	4	1	46	10.8363	21.8154	0.57749	1.37481
6	.	5	1	26	11.2165	22.7415	0.71633	1.62441
7	.	6	1	2	11.0950	22.2700	0.89803	2.27688
8	.	7	1	14	11.5107	23.2000	0.83558	2.03891
9	0	.	2	55	11.6185	23.6416	0.45221	1.11106
10	1	.	2	55	10.4711	20.9404	0.35143	0.64465
11	0	1	3	2	10.8950	22.0400	0.14849	0.29698
12	0	2	3	5	11.7880	23.9720	0.19383	0.65136
13	0	3	3	4	11.6300	23.6575	0.28320	0.73341
14	0	4	3	23	11.3530	23.0326	0.30452	0.80122
15	0	5	3	13	11.8754	24.2338	0.30780	0.75011
16	0	6	3	1	11.7300	23.8800	.	.
17	0	7	3	7	12.0771	24.7214	0.61889	1.60213
18	1	1	3	2	10.0500	19.9850	0.16971	0.33234
19	1	2	3	5	10.5520	21.2500	0.38564	0.62426
20	1	3	3	4	10.3450	20.7725	0.09110	0.25591
21	1	4	3	23	10.3196	20.5983	0.17639	0.35565
22	1	5	3	13	10.5577	21.2492	0.18381	0.33089
23	1	6	3	1	10.4600	20.6600	.	.
24	1	7	3	7	10.9443	21.6786	0.61733	1.01945

```
/*Un conjunto de ejemplos con 100 datos normales simulados*/
OPTIONS NODATE NOCENTER NONUMBER;
goptions reset=global gunit=pct border cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=6 htext=3;

/*GENERACIÓN DE 100 OBSERVACIONES DE UNA NORMAL MU=500 SIGMA=5*/
DATA DATOS (DROP=I);
DO I = 1 TO 100;
X=500+5*NORMAL(1500);
OUTPUT;
END;
RUN;

/*Solicitando un intervalo de confianza para la media del 99%*/
proc means data=datos alpha=0.01 clm;
var x;
run;

The MEANS Procedure
   Analysis Variable : X
   Lower 99%      Upper 99%
 CL for Mean    CL for Mean
 ___________________________
 499.1037791    501.6854105
 ___________________________

/*De nuevo, se piden los límites del intervalo de confianza del 99% pero invocando las palabras claves
para cada límite*/
proc means data=datos alpha=0.01 lclm uclm;
var x;
run;

The MEANS Procedure
   Analysis Variable : X
   Lower 99%      Upper 99%
 CL for Mean    CL for Mean
 ___________________________
 499.1037791    501.6854105
 ___________________________
```



```
/*solicitando el número de observaciones, el errore estándar estimado para la media muestral y el error
estándar muestral de los datos de X*/
proc means data=datos N STDERR STD;
var x;
run;

The MEANS Procedure
      Analysis Variable : X
  N       Std Error          Std Dev
 ___________________________________
 100      0.4914762         4.9147617
 ___________________________________
```

2.5.2.2 PROC UNIVARIATE

Está orientado al análisis de una sola variable. Este PROC es muy útil en el análisis exploratorio de datos. Tiene varias opciones que son interesantes de utilizar. Dentro de los argumentos en su sintaxis, se tienen, entre otros, los siguientes:

```
PROC UNIVARIATE <opciones>;
BY <DESCENDING> variable-1 <...<DESCENDING> variable-n>  <NOTSORTED>;
 CLASS variable-1<(variable-opciones)> <variable-2<(variable-opciones)>>
HISTOGRAM <variable(s)> </ opciones>;
 ID variable(s);
 INSET <palabras claves> </ opciones>;
 OUTPUT <OUT=SAS-data-set> Palabra clave - estadístico-1=nombre(s) <... Palabra clave -
estadístico -n=nombre(s)> <especificación de percentiles>;
 PROBPLOT <variable(s)> </ opciones>;
 QQPLOT <variable(s)> </ opciones>;
 VAR variable(s);
 WEIGHT variable;
RUN;
```

La declaración **BY** calcula estadísticos univariados separadamente para cada grupo o nivel de la variable especificada en **BY**. Se puede especificar más de una variable; si no se usa la opción **NOTSORTED**, las observaciones en los datos deben estar ordenadas por todas las variables que se especifiquen en esta declaración o deben ser indexadas apropiadamente. La opción **DESCENDING** especifica que los datos están ordenados en forma descendente. **NOTSORTED** especifica que las observaciones no están necesariamente ordenadas en forma alfabética o numérica. La declaración

CLASS también hace análisis de la variable especificada en **VAR**, por cada nivel o categoría de una variable categórica. A diferencia de **BY**, el uso del argumento **CLASS** no requiere ordenar previamente los datos según la variable de clasificación. Las opciones gráficas disponibles son **HISTOGRAM**, **PROBPLOT**, **QQPLOT**. A cada uno de estos gráficos podemos insertar los valores de algún estadístico con el argumento **INSET**, invocado inmediatamente después de indicar la gráfica a realizar sobre la cual se hará la inserción.

WEIGHT Especifica una variable numérica en el correspondiente data, cuyos valores ponderan los valores de las variables de análisis. Los valores de la variable de ponderación no pueden ser enteros. Si el valor de la variable de ponderación es cero, el procedimiento cuenta la observación en el número total de las observaciones; si la variable de ponderación es menor que cero, el procedimiento convierte los valores de ponderación a cero y cuenta la observación en el número total de observaciones; y si hay valores faltantes en la variable de peso, excluye la correspondiente observación de la variable de análisis. La variable de ponderación no cambia la forma en que el procedimiento determina el rango, la moda, valores extremos, observaciones extremas, o el número de valores faltantes. La prueba t de Student es la única prueba de localización que el proc univariate calcula cuando se ponderan las variables de análisis. Si se especifica **WEIGHT**, las opciones **CIPCTLDF**, **CIPCTLNORMAL**, **LOCCOUNT**, **NORMAL**, **ROBUSTSCALE**, **TRIMMED=**, y **WINSORIZED=** no estarán disponibles.

Entre las opciones que pueden especificarse en la línea del PROC UNIVARIATE, se tienen:

ALL

Solicita todos los estadísticos y tablas que generan las opciones **FREQ**, **MODES**, **NEXTRVAL=5**, **PLOT**, y **CIBASIC**. Si no hay ponderación de las observaciones, esta opción también solicita los estadísticos y tablas que generan las opciones **CIPCTLDF**, **CIPCTLNORMAL**, **LOCCOUNT**, **NORMAL**, **ROBUSTCALE**, **TRIMMED=0.25**, y **WINSORIZED=0.25**.

ALPHA=valor

Especifica el nivel de significancia (de confianza 1-ALPHA) para los intervalos de confianza. Por defecto ALPHA=0.05.

CIBASIC<(<TYPE=palabra clave> <ALPHA=valor>)>

Solicita intervalos de confianza para la media, la desviación estándar y la varianza basado en el supuesto de que los datos se distribuyen normales. Para muestras grandes, este supuesto no es requerido para la media debidos al teorema de límite central.**TYPE=Palabra clave**

Especifica el tipo de intervalo de confianza, mediante las palabras claves **LOWER**, **UPPER**, o **TWOSIDED**. Por defecto es **TWOSIDED**.

CIPCTLDF<(<TYPE=palabra clave> <ALPHA=valor>)>

Solicita intervalos de confianza para cuantiles usando un método libre de distribución, es decir, que no asume ninguna distribución paramétrica para la variable. **TYPE=palabra clave**, para esta opción, puede ser uno de los siguientes tipos de intervalos: **LOWER**, **UPPER**, **SYMMETRIC**, o **ASYMMETIC**. Por defecto es **SYMMETRIC**.

CIPCTLNORMAL<(<TYPE=palabra clave> <ALPHA=valor>)>

Solicita límites de confianza para cuantiles pero basado en el supuesto de que los datos son normales. **TYPE=palabra clave**, para esta opción, puede ser uno de los siguientes tipos de intervalos: **LOWER**, **UPPER** o **TWOSIDED**. Por defecto es **TWOSIDED**.

LOCCOUNT

Solicita una tabla que muestra el número de observaciones mayores, iguales y menores que el valor de **MU0=**. PROC UNIVARIATE usa estos valores para construir la prueba de signo y de rango con signo.

MODES

Solicita una tabla de todas las posibles modas. Por defecto, cuando los datos contienen más de una moda, el procedimiento muestra la moda más pequeña en la tabla de estadísticos básicos.

MU0=valor(es)

Especifica el valor de la media o parámetro de localización en la hipótesis nula para las pruebas de localización. Si se especifica un sólo valor, el procedimiento lo usa para las hipótesis nulas de todas las variables de análisis. Si se especifica múltiples valores, en VAR debe especificarse igual número de variables de análisis, en el orden correspondiente a las hipótesis nulas en mente.

NOPRINT

Suprime toda salida por pantalla del procedimiento UNIVARIATE.

NORMALTEST

Solicita pruebas de normalidad, entre ellas la prueba Shapiro Wilk y una serie de pruebas de bondad de ajuste basadas en la función de distribución empírica.

ROUND=unidad(es)

Especifica las unidades a usar para redondear las variables de análisis. Se puede asignar un valor diferente del número de redondeo a cada variable de análisis, especificando tantos números como variables hayan sido especificadas en el argumento **VAR**.

TRIMMED=valor(es)<(<TYPE=palabra clave> <ALPHA=valor>)>

Solicita una tabla de medias recortadas, donde los valores especificados indican el número o proporción de observaciones a recortar. Si el valor es una proporción p entre 0 y 0.5, el procedimiento recorta el número de observaciones que corresponde al menor entero mayor o igual a np, donde n es el número de observaciones. **TYPE=palabra clave**, especifica el intervalo de confianza para la media **LOWER**, **UPPER**, o **TWOSIDED**. Por defecto es **TWOSIDED** .

VARDEF=divisor

Especifica el divisor a usar para el cálculo de varianzas y desviaciones estándar. Por defecto es DF. Si se requiere calcular el error estándar de la media, los límites de confianza y la prueba t de Student, se usa el valor por defecto.

Veamos a través de los siguientes ejemplos el uso del PROC UNIVARIATE:

```
OPTIONS NODATE NOCENTER NONUMBER;
goptions reset=global gunit=pct bordar cback=white
colors=(black blue green red)
ftitle=swissb ftext=swiss htitle=6 htext=3;

/*GENERACIÓN DE 100 OBSERVACIONES DE UNA NORMAL MU=500 SIGMA=5*/
DATA DATOS (DROP=I);
DO I = 1 TO 100;
X=500+5*NORMAL(1500);
OUTPUT;
END;
RUN;
/*solicitando adicional a las medidas de resumen, pruebas de normalidad (NORMALTEST), histogramas
(HISTOGRAM) con densidad normal insertando en dicho histograma la media, la mediana, los cuartiles, el
coeficiente de asimetria y los valores máximos y mínimos de la variable X*/
/*con PROBPLOT Se pide también gráfico de probabilidad normal al cual se anade linea recta de ajuste
normal con media de 500 y desviación estándar de 5. Con QQPLOT se pide gráfico de cuantil cuantil
normal. A ambos gráficos se inserta valor del estadistico de la prueba Shapiro Wilk y el valor P
respectivo*/
/*en el argumento output out= se crea un data llamado univout donde se guardan la media, la mediana, la
desviación estándar y los percentiles 50, 95, 97.5 y 100 (ver opciones pctlpre= y pctlpts= )*/

proc univariate data=DATOS NORMALTEST;
var X;
histogram X/NORMAL MIDPERCENTS;
```



```
inset mean std median q1 q3 skewness min max;
pattern color='blue' v=l1;
PROBPLOT X/NORMAL(MU=500 SIGMA=5 COLOR=RED);
INSET NORMALTEST PROBN;
QQPLOT X/NORMAL(MU=500 SIGMA=5 COLOR=RED);
INSET NORMALTEST PROBN;
output out= univout mean= MEDIA median= MEDIANA STD=DESV_STD
pctlpre= PERCENTIL_  pctlpts= 50, 95 to 100 by 2.5;

/*imprimiendo por pantalla el contenido del data univout creado con el output out= univout en el proc
univariate*/
proc print data= univout; RUN;
run;quit;
```

Las salidas de la anterior programación se muestran a continuación. La tabla 1 muestra diferentes momentos muestrales de localización, dispersión, asimetría, curtosis, entre otros. La tabla 2 presenta medidas de localización y dispersión, así se duplica parte de la información presentada por la tabla 1; la tabla 3 presenta resultados para la prueba, en este caso bilateral sobre la media, con hipótesis nula de Mu=0, por la prueba t, y las pruebas no paramétrica de signo y de rango con signo. La tabla 4 presenta los resultados para evaluar el test de normalidad por cuatro métodos o pruebas disponibles. La tabla 5 ofrece estimaciones de cuantiles. La tabla 6 presenta los cuatro valores más pequeños y los más grandes en la muestra:

Tabla 1

The UNIVARIATE Procedure

Variable: X

Moments

N	100	Sum Weights	100
Mean	500.394595	Sum Observations	50039.4595
Std Deviation	4.9147617	Variance	24.1548826
Skewness	-0.2363992	Kurtosis	0.55617509
Uncorrected SS	25041866.4	Corrected SS	2391.33338
Coeff Variation	0.98217722	Std Error Mean	0.49147617

Tabla2

Basic Statistical Measures

Location		Variability	
Mean	500.3946	Std Deviation	4.91476
Median	500.8701	Variance	24.15488