

## 2. Descripción detallada y tabulación de variables.

### **Análisis de caso: Las ventas de 3 equipos durante un periodo de 3 meses.**

En nuestra empresa hay tres equipos de ventas que operan en 3 zonas aproximadamente equivalentes (con la misma cantidad y densidad de clientes y con las mismas oportunidades de clientes potenciales aproximadamente). Los vendedores trabajan a tiempo parcial. Hemos registrado los datos de las horas trabajadas (llamada horash), las ventas y beneficios (llamada benef) para cada equipo durante cada uno de los tres primeros meses del año. La primera variable, equipo, (llamada eq), está codificada con 1 para equipo 1, 2 para el equipo 2 y 3 para el equipo 3. La segunda variable es el mes y está codificada 1 para Enero, 2 para Febrero y 3 para Marzo. Vamos a describir estos datos de modo que tengamos una visión global, por equipos y por meses.

#### 1. Descripción detallada mediante estadísticos descriptivos y tabulación de variables.

El programa 2D es la elección cuando se desea una descripción estadística pormenorizada, ver la tabulación de la variable y representarla mediante histogramas (en realidad, para ser más exactos, en diagramas de barras en forma de columnas).

#### **INSTRUCCIONES. PROGRAMA 2D.**

```
/input var=5. format=free. case=9.
/var names=eq, mes, horash, ventas, benef. use = horash, ventas y benef.
/end
1 1 480 5000 2500
1 2 490 4950 2500
1 3 560 5010 2600
2 1 510 5000 2100
2 2 560 5210 2125
2 3 495 4990 2000
3 1 470 5100 2700
3 2 480 5300 2850
3 3 505 5700 3000
/end
```

#### **OUTPUT SELECCIONADO**

El Output seleccionado siguiente se presenta para la primera variable con los resultados muy espaciados para permitir extensos cuadros de notas que expliquen los principales estadísticos utilizados.

#### BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES

CASE NO.	Número de variable				
	1 eq	2 mes	3 horash	4 ventas	5 benef
1	1.00	1.00	480.00	5000.00	2500.00
2	1.00	2.00	490.00	4950.00	2500.00
3	1.00	3.00	560.00	5010.00	2600.00
4	2.00	1.00	510.00	5000.00	2100.00
5	2.00	2.00	560.00	5210.00	2125.00
6	2.00	3.00	495.00	4990.00	2000.00
7	3.00	1.00	470.00	5100.00	2700.00
8	3.00	2.00	480.00	5300.00	2850.00
9	3.00	3.00	505.00	5700.00	3000.00

Por defecto el programa muestra el listado de los 10 primeros casos en todas las variables. Si se Añade /PRINT CASE=0. antes de /END en las instrucciones no se lista ningún caso. Con esta instrucción se puede controlar el número de casos que se desea ver. (Este comando no tendrá efecto trabajando en modo

NUMBER OF CASES READ. . . . . 9

Número de casos que el programa ha leído. Importante para verificar que ha leído todos los casos sin errores.

```
*****
* horash *
*****
```

Variable analizada

	VALUE	ZSCORE	CASE #
MAX	560.0000000	1.635	3
MIN	470.0000000	-1.068	7

VARIABLE NUMBER . . . . . 3  
NUMBER OF DISTINCT VALUES . . . . . 7

NUMBER OF VALUES COUNTED . . . . .	9
NUMBER OF VALUES NOT COUNTED . . . . .	0

Número de valores contados para esta variable (en este caso 9) y número de valores no contados para esta variable. Puede haber valores no contados por ejemplo debido a valores faltantes o fuera de rango si se definen mínimos y máximos. En esta variable no hay ninguno.

	VALUE	VALUE/S.E.
SKEWNESS	0.72	0.883
KURTOSIS	-1.18	-0.725

Asimetría (SKEWNESS) y asimetría estandarizada (es decir, dividida por su error típico, S.E.).  
Curtosis y curtosis estandarizada (curtosis dividida por su error típico S.E.)

Para obtener la asimetría BMDP utiliza el Coeficiente de asimetría de Fisher. En una distribución normal el valor de asimetría sería 0. Bajo el supuesto de normalidad el error típico de la asimetría es raíz cuadrada de 6/N. La asimetría estandarizada (dividida por su error típico) puede leerse aproximadamente como una puntuación típica, de modo que valores absolutos mayores que 3 son extremados y mayores que 2 inusuales, indicando la presencia de distribuciones considerablemente asimétricas. Más allá de 2 en valores absolutos, si el signo es positivo la curva está sesgada a la derecha (acumulación de frecuencias en los valores bajos) y si es negativo la curva está sesgada a la izquierda (las frecuencias mayores se acumulan en los valores más altos).

Para obtener la curtosis BMDP utiliza el estadístico de curtosis más utilizado, basado en el cuarto momento. En una distribución normal el valor de Curtosis sería 0. El error típico de la curtosis es raíz cuadrada de 24/N. La curtosis estandarizada también puede leerse aproximadamente como una puntuación típica. Una curtosis estandarizada de -2 o menor indica una curva platicúrtica, más aplastada que la normal. Una curtosis estandarizada de +2 o más indica una curva leptocúrtica, más anudada que la curva normal.

	ESTIMATE	ST. ERROR	ESTIMATE
MEAN	505.5555420	11.1006889	ST. DEV.
MEDIAN	495.0000000	23.0940228	VARIANCE
MODE	NOT UNIQUE		RANGE
			(Q3-Q1) / 2

Estadísticos de tendencia central:  
Media, mediana y moda (esta última solo en caso de que sea única)

Error típico de la media y error típico de la Mediana.

Estadísticos de dispersión:  
Desviación típica, varianza, rango y rango semi-intercuartil.  
Para la varianza y la desviación típica el programa calcula con denominador N-1, es decir, ofrece la estimación insesgada del parámetro de estos estadísticos en la población a la que pertenece la muestra analizada

El error típico de la media es la desviación típica dividida por la raíz cuadrada de N.

Para el error típico de la mediana se utiliza una fórmula propuesta por Tukey que se basa en estadísticos de rango (para su fórmula ver BMDP Vol. I, pag. 143). El error típico de la mediana es un estadístico raramente

Intervalo confidencial de la media al 95% de confianza. (Es la media más(ó menos) t (al percentil 0'795 y 1 g.l.) por el error típico de la media. Con el 95% de confianza podemos afirmar que la media de la variable en la población está entre estos dos valores.

LOWER 95% C.L. OF MEAN	479.9588010
UPPER 95% C.L. OF MEAN	531.1522830

Q1	480.0000000
Q3	535.0000000

Cuartil 1 y Cuartil 3

S-	472.2534790
S+	538.8576050

Media más (y menos) una desviación típica

Tabulación de la variable

Valor de la variable	Frecuencia	Porcentaje	Porcentaje acumulado
470.	1	11.1	11.1
480.	2	22.2	33.3
490.	1	11.1	44.4
495.	1	11.1	55.6

Valor de la variable	Frecuencia	Porcentaje	Porcentaje acumulado
505.	1	11.1	66.7
510.	1	11.1	77.8
560.	2	22.2	100.0

```

*****
* ventas *
*****

          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 4      MAX    5700.0000000    2.334    9
          MIN    4950.0000000   -0.792    2
NUMBER OF DISTINCT VALUES . 8
NUMBER OF VALUES COUNTED . 9
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS    1.30    1.598
          KURTOSIS    0.42    0.256

          ESTIMATE      ST.ERROR      ESTIMATE
MEAN      5140.0000000    79.9652710    ST.DEV.    239.8958130
MEDIAN    5010.0000000    89.4893341    VARIANCE   57550.0000000
MODE      5000.0000000
          RANGE      750.0000000
          (Q3-Q1)/2    130.0000000

          LOWER 95% C.L. OF MEAN    4955.6103500
          UPPER 95% C.L. OF MEAN    5324.3896500
          Q1    4995.0000000
          Q3    5255.0000000
          S-    4900.1040000
          S+    5379.8960000

          PERCENTS
          VALUE  COUNT  CELL  CUM
          4950.  1  11.1  11.1
          4990.  1  11.1  22.2
          5000.  2  22.2  44.4
          5010.  1  11.1  55.6

          PERCENTS
          VALUE  COUNT  CELL  CUM
          5100.  1  11.1  66.7
          5210.  1  11.1  77.8
          5300.  1  11.1  88.9
          5700.  1  11.1 100.0
    
```

```

*****
* benef *
*****

          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 5      MAX    3000.0000000    1.475    9
          MIN    2000.0000000   -1.395    6
NUMBER OF DISTINCT VALUES . 8
NUMBER OF VALUES COUNTED . 9
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS    -0.06   -0.078
          KURTOSIS    -1.61   -0.985

Plot 9: (S)kip, (Q)uit Run, or ENTER to display ...

          ESTIMATE      ST.ERROR      ESTIMATE
MEAN      2486.1110800    116.1529620    ST.DEV.    348.4588930
MEDIAN    2500.0000000    216.5064540    VARIANCE   121423.6020000
MODE      2500.0000000
          RANGE      1000.0000000
          (Q3-Q1)/2    331.2500000

          LOWER 95% C.L. OF MEAN    2218.2775900
          UPPER 95% C.L. OF MEAN    2753.9445800
          Q1    2112.5000000
          Q3    2775.0000000
          S-    2137.6521000
          S+    2834.5700700

          PERCENTS
          VALUE  COUNT  CELL  CUM
          2000.  1  11.1  11.1
          2100.  1  11.1  22.2
          2125.  1  11.1  33.3
          2500.  2  22.2  55.6

          PERCENTS
          VALUE  COUNT  CELL  CUM
          2600.  1  11.1  66.7
          2700.  1  11.1  77.8
          2850.  1  11.1  88.9
          3000.  1  11.1 100.0
    
```

/end

2. Análisis descriptivo detallado y tabulación por grupos  
(análisis por equipos de la variable ventas).

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES

```
/input var=5. format=free. case=9.
/var names=eq, mes, horash, ventas, benef. use=eq, horash, ventas, benef.
/group var=eq. codes(eq)=1, 2, 3. names(eq)= equipo1, equipo2, equipo3.
/end
```

CASE NO.	1	3	4	5
NO. eq	horash	ventas	benef	
1	equipo1	480.00	5000.00	2500.00
2	equipo1	490.00	4950.00	2500.00
3	equipo1	560.00	5010.00	2600.00
4	equipo2	510.00	5000.00	2100.00
5	equipo2	560.00	5210.00	2125.00
6	equipo2	495.00	4990.00	2000.00
7	equipo3	470.00	5100.00	2700.00
8	equipo3	480.00	5300.00	2850.00
9	equipo3	505.00	5700.00	3000.00

Aquí se solicita que se utilice la variable eq (que codifica el equipo) para realizar grupos. A continuación se describen los códigos de la variable y, por el mismo orden, las etiquetas de cada uno de los códigos. En el listado de datos el programa utiliza esas etiquetas en la variable

NUMBER OF CASES READ. . . . . 9

GROUPING VARIABLE. . . eq

CATEGORY	FREQUENCY
equipo1	3
equipo2	3
equipo3	3

Se describe el número de casos registrado para cada grupo (en este caso hay tres datos por variable para cada equipo), y, a continuación comienzan los análisis para cada grupo de cada variable citada en USE

\*\*\*\*\*  
\* ANALYSIS OF horash FOR GROUP equipo1 \*  
\*\*\*\*\*

VARIABLE NUMBER	3	MAX	560.0000000	ZSCORE	1.147	CASE #	3
NUMBER OF DISTINCT VALUES	3	MIN	480.0000000		-0.688		1
NUMBER OF VALUES COUNTED	3			VALUE	VALUE/S.E.		
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS	0.36		0.256		
		KURTOSIS	-2.33		-0.825		

	ESTIMATE	ST.ERROR	ESTIMATE
MEAN	510.0000000	25.1661148	ST.DEV. 43.5889893
MEDIAN	490.0000000	0.0000000	VARIANCE 1900.0000000
MODE	NOT UNIQUE		RANGE 80.0000000
			(Q3-Q1)/2 40.0000000

LOWER 95% C.L. OF MEAN	401.7189030
UPPER 95% C.L. OF MEAN	618.2811280
Q1	480.0000000
Q3	560.0000000
S-	466.4110110
S+	553.5889890

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
480.	1	33.3	33.3	560.	1	33.3	100.0
490.	1	33.3	66.7				

\*\*\*\*\*  
\* ANALYSIS OF ventas FOR GROUP equipo1 \*  
\*\*\*\*\*

VARIABLE NUMBER	4	MAX	5010.0000000	ZSCORE	0.726	CASE #	3
NUMBER OF DISTINCT VALUES	3	MIN	4950.0000000		-1.141		2
NUMBER OF VALUES COUNTED	3			VALUE	VALUE/S.E.		
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS	-0.34		-0.243		
		KURTOSIS	-2.33		-0.825		

	ESTIMATE	ST.ERROR	ESTIMATE
MEAN	4986.6665000	18.5592155	ST.DEV. 32.1455040
MEDIAN	5000.0000000	0.0000000	VARIANCE 1033.3333700
MODE	NOT UNIQUE		RANGE 60.0000000
			(Q3-Q1)/2 30.0000000

LOWER 95% C.L. OF MEAN	4906.8125000
UPPER 95% C.L. OF MEAN	5066.5205100
Q1	4950.0000000
Q3	5010.0000000

```

S- 4954.5210000
S+ 5018.8120100

      VALUE      COUNT      PERCENTS
      VALUE      COUNT      PERCENTS
      CELL      CUM          CELL      CUM
4950.      1      33.3      33.3      5010.      1      33.3      100.0
5000.      1      33.3      66.7

-----

*****
* ANALYSIS OF benef      FOR GROUP equipo1 *
*****

      VALUE      ZSCORE      CASE #
      MAX      2600.0000000      1.155      3
      MIN      2500.0000000      -0.577      1
VARIABLE NUMBER . . . . .      5
NUMBER OF DISTINCT VALUES .      2
NUMBER OF VALUES COUNTED .      3
NUMBER OF VALUES NOT COUNTED      0
      SKEWNESS      0.38      VALUE/S.E.      0.272
      KURTOSIS      -2.33      VALUE/S.E.      -0.825

      ESTIMATE      ST.ERROR      ESTIMATE
MEAN      2533.3332500      33.3333321      ST.DEV.      57.7350273
MEDIAN      2500.0000000      0.0000000      VARIANCE      3333.3335000
MODE      2500.0000000
      RANGE      100.0000000
      (Q3-Q1)/2      50.0000000

      LOWER 95% C.L. OF MEAN      2389.9113800
      UPPER 95% C.L. OF MEAN      2676.7551300
      Q1      2500.0000000
      Q3      2600.0000000
      S-      2475.5981400
      S+      2591.0683600
    
```

```

      VALUE      COUNT      PERCENTS
      VALUE      COUNT      PERCENTS
      CELL      CUM          CELL      CUM
2500.      2      66.7      66.7      2600.      1      33.3      100.0

-----

*****
* ANALYSIS OF horash      FOR GROUP equipo2 *
*****

      VALUE      ZSCORE      CASE #
      MAX      560.0000000      1.126      3
      MIN      495.0000000      -0.784      1
VARIABLE NUMBER . . . . .      3
NUMBER OF DISTINCT VALUES .      3
NUMBER OF VALUES COUNTED .      3
NUMBER OF VALUES NOT COUNTED      0
      SKEWNESS      0.30      VALUE/S.E.      0.214
      KURTOSIS      -2.33      VALUE/S.E.      -0.825

      ESTIMATE      ST.ERROR      ESTIMATE
MEAN      521.6666870      19.6497116      ST.DEV.      34.0342979
MEDIAN      510.0000000      0.0000000      VARIANCE      1158.3335000
MODE      NOT UNIQUE
      RANGE      65.0000000
      (Q3-Q1)/2      32.5000000

      LOWER 95% C.L. OF MEAN      437.1207580
      UPPER 95% C.L. OF MEAN      606.2125850
      Q1      495.0000000
      Q3      560.0000000
      S-      487.6323850
      S+      555.7009890
    
```

```

      VALUE      COUNT      PERCENTS
      VALUE      COUNT      PERCENTS
      CELL      CUM          CELL      CUM
495.      1      33.3      33.3      560.      1      33.3      100.0
510.      1      33.3      66.7

-----

*****
* ANALYSIS OF ventas      FOR GROUP equipo2 *
*****

      VALUE      ZSCORE      CASE #
      MAX      5210.0000000      1.154      5
      MIN      4990.0000000      -0.617      2
VARIABLE NUMBER . . . . .      4
NUMBER OF DISTINCT VALUES .      3
NUMBER OF VALUES COUNTED .      3
NUMBER OF VALUES NOT COUNTED      0
      SKEWNESS      0.38      VALUE/S.E.      0.270
      KURTOSIS      -2.33      VALUE/S.E.      -0.825
    
```

```

MEAN          5066.6665000      71.7247772      ST.DEV.      124.2309650
MEDIAN        5000.0000000      0.0000000      VARIANCE    15433.3330000
MODE          NOT UNIQUE                               RANGE       220.0000000
                                                    (Q3-Q1)/2   110.0000000
    
```

```

                LOWER 95% C.L. OF MEAN  4758.0595700
                UPPER 95% C.L. OF MEAN  5375.2734400
                Q1      4990.0000000
                Q3      5210.0000000
                S-      4942.4355500
                S+      5190.8974600
    
```

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
4990.	1	33.3	33.3	5210.	1	33.3	100.0
5000.	1	33.3	66.7				

\*\*\*\*\*  
 \* ANALYSIS OF benef FOR GROUP equipo2 \*  
 \*\*\*\*\*

```

                VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 5      MAX  2125.0000000  0.756  3
                MIN  2000.0000000 -1.134  6
NUMBER OF DISTINCT VALUES . 3
NUMBER OF VALUES COUNTED . 3      VALUE      VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  -0.32    -0.229
                KURTOSIS  -2.33    -0.825
    
```

```

                ESTIMATE      ST.ERROR      ESTIMATE
MEAN          2075.0000000  38.1881294  ST.DEV.      66.1437836
MEDIAN        2100.0000000  0.0000000  VARIANCE    4375.0000000
MODE          NOT UNIQUE                               RANGE       125.0000000
                                                    (Q3-Q1)/2   62.5000000
    
```

```

                LOWER 95% C.L. OF MEAN  1910.6897000
                UPPER 95% C.L. OF MEAN  2239.3103000
                Q1      2000.0000000
                Q3      2125.0000000
                S-      2008.8562000
                S+      2141.1438000
    
```

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
2000.	1	33.3	33.3	2125.	1	33.3	100.0
2100.	1	33.3	66.7				

\*\*\*\*\*  
 \* ANALYSIS OF horash FOR GROUP equipo3 \*  
 \*\*\*\*\*

```

                VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 3      MAX  505.0000000  1.109  3
                MIN  470.0000000 -0.832  7
NUMBER OF DISTINCT VALUES . 3
NUMBER OF VALUES COUNTED . 3      VALUE      VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  0.26    0.181
                KURTOSIS  -2.33    -0.825
    
```

```

                ESTIMATE      ST.ERROR      ESTIMATE
MEAN          485.0000000  10.4083300  ST.DEV.      18.0277557
MEDIAN        480.0000000  0.0000000  VARIANCE    324.9999690
MODE          NOT UNIQUE                               RANGE       35.0000000
                                                    (Q3-Q1)/2   17.5000000
    
```

```

                LOWER 95% C.L. OF MEAN  440.2165530
                UPPER 95% C.L. OF MEAN  529.7834470
                Q1      470.0000000
                Q3      505.0000000
                S-      466.9722290
                S+      503.0277710
    
```

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
470.	1	33.3	33.3	505.	1	33.3	100.0
480.	1	33.3	66.7				

\*\*\*\*\*  
 \* ANALYSIS OF ventas FOR GROUP equipo3 \*  
 \*\*\*\*\*

```
*****
                                VALUE  ZSCORE  CASE #
                                MAX    5700.000000  1.091   9
VARIABLE NUMBER . . . . . 4    MIN    5100.000000  -0.873   2
NUMBER OF DISTINCT VALUES . 3
NUMBER OF VALUES COUNTED . 3                                VALUE  VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0    SKEWNESS  0.21    0.147
                                KURTOSIS  -2.33   -0.825

ESTIMATE  ST. ERROR  ESTIMATE
MEAN      5366.6665000  176.3834080  ST.DEV.    305.5050350
MEDIAN    5300.0000000  0.0000000   VARIANCE  93333.3281000
MODE      NOT UNIQUE                                RANGE     600.0000000
                                                (Q3-Q1)/2  300.0000000

                                LOWER 95% C.L. OF MEAN  4607.7495100
                                UPPER 95% C.L. OF MEAN  6125.5835000
                                                Q1     5100.0000000
                                                Q3     5700.0000000
                                                S-     5061.1616200
                                                S+     5672.1713900

                                PERCENTS
                                VALUE  COUNT  CELL  CUM
                                5100.   1    33.3  33.3
                                5300.   1    33.3  66.7
                                5700.   1    33.3  100.0

-----
```

```
*****
* ANALYSIS OF benef FOR GROUP equipo3 *
*****
```

```
                                VALUE  ZSCORE  CASE #
                                MAX    3000.0000000  1.000   9
VARIABLE NUMBER . . . . . 5    MIN    2700.0000000  -1.000   6
NUMBER OF DISTINCT VALUES . 3
NUMBER OF VALUES COUNTED . 3                                VALUE  VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0    SKEWNESS  0.00    0.000
                                KURTOSIS  -2.33   -0.825

ESTIMATE  ST. ERROR  ESTIMATE
MEAN      2850.0000000  86.6025391  ST.DEV.    150.0000000
MEDIAN    2850.0000000  0.0000000   VARIANCE  22500.0000000
MODE      NOT UNIQUE                                RANGE     300.0000000
                                                (Q3-Q1)/2  150.0000000

                                LOWER 95% C.L. OF MEAN  2477.3791500
                                UPPER 95% C.L. OF MEAN  3222.6208500
                                                Q1     2700.0000000
                                                Q3     3000.0000000
                                                S-     2700.0000000
                                                S+     3000.0000000

                                PERCENTS
                                VALUE  COUNT  CELL  CUM
                                2700.   1    33.3  33.3
                                2850.   1    33.3  66.7
                                3000.   1    33.3  100.0

-----
```



## Representaciones gráficas de variables.

### Cómo obtener gráficos de alta resolución (Plots) con BMDP.

La representación gráfica de la información estadística juega un papel importante en la comprensión de los datos. BMDP ofrece gráficos de datos estadísticos en dos formas:

- Como texto
- En alta resolución.

Los gráficos como texto forman parte del output normal de los programas, están formados utilizando caracteres (tales como "X" o "H" para formar histogramas) y pueden manejarse desde el editor de BMDP, desde el editor de DOS o desde cualquier procesador de texto.

Por ejemplo, el programa 2D puede producir histogramas utilizando caracteres dentro del output de texto.

Los gráficos en alta resolución pueden verse y editarse desde PLOT. Cuando se ejecuta un programa BMDP los gráficos que este produce (sea por defecto o solicitados por el usuario) pueden no aparecer en pantalla en alta resolución, y, en general, no pueden editarse en alta resolución desde el programa BMDP y no serán guardados en un archivo de gráficos si no se solicita explícitamente.

BMDP produce gráficos de alta resolución para todos los plots disponibles en los programas.

Para garantizar que podemos ver, editar, e imprimir estos gráficos en alta resolución es necesario:

- 1) guardarlos desde el programa BMDP, y,
- 2) después, acabado el programa, verlos desde PLOT

Se accede a PLOT desde el menú principal del programa, desde el mismo menú en el que se solicita RUN, EDIT o VIEW.

#### Guardar los PLOTS de alta resolución que hacen los programas BMDP:

Dentro de las instrucciones, generalmente como anterior al párrafo /END, introducir:

**/PLOT file='nombre.plt'.**

Todos los plots que produzca el programa serán guardados en alta resolución en el archivo llamado 'nombre.plt' en el directorio en que tengamos configurado guardar los plots. En ese archivo se guardarán todos los plots que produzca el programa, a modo de una lista de plots.

En la opción SINGLE CONFIGURE del menú principal es posible predeterminar el directorio donde se guardan los archivos tipo '.plt', así como los de cualquier otro tipo (\*.INP, \*.OUT, etc.) que utiliza el BMDP.

En algunos programas se pueden introducir múltiples problemas sucesivamente, en ese caso basta con incluir la instrucción de guardado de gráficos solo en el primer párrafo plot que se escriba.

Si se vuelve a escribir la instrucción para guardar gráficos otra vez el archivo se reescribirá sin previa advertencia y se perderán los gráficos guardados la primera vez con ese mismo nombre.

Observación: Los Plots en alta resolución NO se ven ni se pueden guardar en el archivo de output, que es un archivo de texto, aunque allí a veces queda la referencia de que se confeccionó el gráfico mientras se ejecutaba el programa, incluso aunque no se haya guardado el Plot. Tampoco es posible ver los plots guardados mediante VIEW ni con el EDITOR de BMDP. Para ver, editar e imprimir los gráficos después de ejecutado un programa hay que entrar en PLOT, y solo será posible si se han guardado mediante la mencionada instrucción situada en el archivo de instrucciones.

#### Acceder a los Plots.

Después de ejecutado el programa en el que hemos guardado plots como 'nombre.plt' salimos de ese programa y desde el menú principal escogemos la opción PLOT y, dentro de ella, abrir (open) la opción 'BMDP Plot File' que nos permitirá especificar el disco y directorio en el que buscar o seleccionar directamente el archivo tipo '.plt'.

Por ejemplo:

c:\GRAFICOS\\*.plt

nos mostrará (y permitirá abrir seleccionando) todos los archivos '.plt' que aparezcan en el directorio gráficos. (Si éste es el directorio donde queremos guardar los archivos tipo '.plt' conviene que lo especifiquemos -antes de ejecutar los programas que realizan los gráficos- en SINGLE CONFIGURE).

Una vez que PLOT haya abierto el archivo 'nombre.plt' nos presentará una lista de todos los gráficos que están allí almacenados. Para ver un gráfico simplemente lo seleccionados y lo presentamos con F6 (Plot).

Observación: Si en el procedimiento de instalación de BMDP no se selecciona adecuadamente la pantalla entonces no será posible visualizar gráficos de alta resolución con BMDP. Una vez instalado BMDP puede corregirse este problema entrando en PLOT desde el menú general y reconfigurando la elección de la pantalla

hasta encontrar una que funcione adecuadamente con el propio ordenador (seguir las instrucciones paso a paso que va dando el programa).

#### Editar un gráfico.

Es posible cambiar toda una serie de elementos en un gráfico, tales como título, rango de los ejes, etc. Puede hacerse una vez seleccionado un gráfico en la lista (y antes de presentarlo con F6) mediante el menú EDIT que aparece en pantalla asociado a la tecla F3.

#### Presentar dos o más Plots a la vez.

En algunos casos es muy interesante poder trazar dos plots a la vez (por ejemplo para comparar dos histogramas por grupos realizados por 3D). En ese caso, desde dentro de PLOT seleccionamos los plots uno a uno resaltándolos y utilizando F4 y después trazamos con F6.

Cuando se seleccionan dos o más plots antes de ejecutar F6, el programa pregunta cuantos plots coloca por fila y cuantos por columna.

Por ejemplo, si estamos interesados en imprimir juntos los histogramas de una variable para 3 grupos diferentes, obtenidos mediante el programa 2D haciendo análisis por grupos y guardando gráficos, acabado el análisis accedemos a PLOT, seleccionamos los gráficos de interés, cambiamos ejes y títulos mediante el menú EDIT y después seleccionamos los 3 gráficos con F4 antes de visualizarlos con F6 (respondiendo 1 gráfico por fila y 3 por columna a las preguntas que hace el programa al presionar F6 después de seleccionar los 3 gráficos con F4).

Otro ejemplo, si estamos interesados en ver los histogramas por grupos producidos por 3D uno encima de otro, para poder comparar visualmente las dos distribuciones, entonces seleccionamos los dos histograma mediante F4, ejecutamos F6 y contestamos que 1 gráfico por fila (primera pregunta) y 2 gráficos por columna (segunda pregunta).

Los plots se trazan en el orden en que se seleccionan. Al trazar dos o más plots en la misma pantalla, las elecciones para el primero (ejes, colores y otros atributos) se aplican a todos. Por ello para que tenga sentido trazar varios plots juntos estos tienen que tener un eje Y común y estar escalados en las mismas unidades.

#### Imprimir un gráfico.

Cuando tenemos el gráfico en pantalla, teniendo una impresora conectada al ordenador, basta con pulsar la letra **P** (y seguir instrucciones).

La impresora ha de estar debidamente configurada e instalada para BMDP (mediante el procedimiento de configuración en el que se selecciona pantalla, ploter -si hay- e impresora).

Observación: La impresora también se configura seleccionándola al instalar el programa, y puede ser reconfigurada a posteriori del mismo modo que la pantalla o el plotter en el caso de que disponga de uno conectado al ordenador.

3. Análisis gráfico de los datos: Obtención de histogramas.

Vamos a reanalizar la variable ventas con el programa 2D solicitando los gráficos que este programa puede hacer.

El tipo de gráfico que 2D puede hacer está especificado por defecto, es decir, no hay que decir explícitamente que se desea tales gráficos, simplemente se especifica donde se desea guardarlos en alta resolución.

**INSTRUCCIONES. PROGRAMA 2D.**

```



```

Se restringe el análisis a la variable ventas.

/plot file='2D001.plt'.

Se solicita que se guarden los gráficos en formato de alta resolución en el archivo 2d001.plt. El programa guardara en el directorio que tengamos establecido por defecto (modificable en SINGLE COMMANDS).

**OUTPUT SELECCIONADO**

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES

```



```

Los programas BMDP siempre repiten las instrucciones en OUTPUT para facilitar que después sepamos a que correspondían los análisis y para ofrecer un diagnóstico de errores en las

CASE	4
NO.	ventas
1	5000.00
2	4950.00
3	5010.00
4	5000.00
5	5210.00
6	4990.00
7	5100.00
8	5300.00
9	5700.00

NUMBER OF CASES READ. . . . . 9

```

*****
* ventas *
*****
VARIABLE NUMBER . . . . . 4
NUMBER OF DISTINCT VALUES . 8
NUMBER OF VALUES COUNTED . 9
NUMBER OF VALUES NOT COUNTED 0

```

	VALUE	ZSCORE	CASE #
MAX	5700.0000000	2.334	9
MIN	4950.0000000	-0.792	2

	VALUE	VALUE/S.E.
SKEWNESS	1.30	1.598
KURTOSIS	0.42	0.256

EACH 'H' REPRESENTS 1 RESPONSE.

```

      H
      H
      H
    HH H H H      H
    .....+.....+.....+.....
          5100      5700
                5400

```

Este es el histograma de la variable ventas en forma de gráfico como texto, como especifica el output cada H significa en esta ocasión un caso.

Aquí se anuncia que este gráfico se ha realizado también como plot en alta resolución. Este plot puede verse, después de editado e impreso con PLOT, en la página siguiente.

--- Plot 1 has been added to PLOTFILE.

	ESTIMATE	ST.ERROR	ESTIMATE	
MEAN	5140.0000000	79.9652710	ST.DEV.	239.8958130
MEDIAN	5010.0000000	89.4893341	VARIANCE	57550.0000000
MODE	5000.0000000		RANGE	750.0000000
			(Q3-Q1)/2	130.0000000

LOWER 95% C.L. OF MEAN	4955.6103500
UPPER 95% C.L. OF MEAN	5324.3896500
Q1	4995.0000000
Q3	5255.0000000
S-	4900.1040000
S+	5379.8960000

Este es un gráfico de síntesis, también en modo texto, donde puede apreciarse la posición de estadísticos conocidos: MIN, MAX, Q1, Q3, S-, S+, MODA, MEDIANA y MEDIA. Los valores numéricos de esos estadísticos están en el output.

```

S      Q      M      Q      S
-      MM     M      3      +
.      M      OE     E      .      M
.      I      DD     A      .      A
.      N      EI     N      .      X
+.....+.....+.....+.....+.....+.....+.....+.....+.....+.....+.....+.....+
4900      5100      5300      5500      5700
          5000      5200      5400      5600

```

--- Plot 2 has been added to PLOTFILE.

Este gráfico también lo guarda como PLOT, pero en este caso no lo adjuntamos porque para este gráfico concreto el modo texto es mucho más

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
4950.	1	11.1	11.1	5100.	1	11.1	66.7
4990.	1	11.1	22.2	5210.	1	11.1	77.8
5000.	2	22.2	44.4	5300.	1	11.1	88.9
5010.	1	11.1	55.6	5700.	1	11.1	100.0

```

-----
NUMBER OF INTEGER WORDS USED IN PRECEDING PROBLEM      894
Copying Plots from workfile to C:\E4\2D001.PLT ...
/end
PROGRAM TERMINATED

```

Por último, aquí se confirma que se han guardado los plots en este archivo.

4. Gráficos por grupos.

INSTRUCCIONES. PROGRAMA 2D.

```

/input
title is 'Análisis por grupos de ventas con graficos'.
           #poner un titulo al analisis es siempre opcional
var=5.
format=free.
case=9. #declarar el numero de casos es tambien opcional, pero recomendable.
/var
names=eq, mes, horash, ventas, benef.
use=mes, ventas. #restringe los analisis solo a la variable ventas.
           #la variable agrupadora (aquí mes) siempre tiene que
           #mencionarse en la instruccion USE
/group
var=mes. #solicita analisis por grupos
codes(mes)=1,2,3.
Names(mes)=Enero, Febrero, Marzo.
/print #las siguientes instrucciones del parrafo plot no funcionan en modo interactivo
no ezsc. #evita que aparezcan en el output las tipicas maximas y minimas
no ecas. #excluye del output los numeros de caso de max y min
no extr. #excluye del output el maximo y minimo.
no sk. #excluye del output la asimetria y curtosis.
case=0. #excluye del output el listado de los 10 primeros casos.
/plot file='2D002.plt'. #indica donde guardar los plots graficos.
/end
1 1 480 5000 2500
1 2 490 4950 2500
1 3 560 5010 2600
2 1 510 5000 2100
2 2 560 5210 2125
2 3 495 4990 2000
3 1 470 5100 2700
3 2 480 5300 2850
3 3 505 5700 3000
/end

```

Aquí se especifica que el análisis es por grupos.

OUTPUT SELECCIONADO

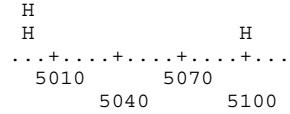
NUMBER OF CASES READ. . . . . 9

GROUPING VARIABLE. . . mes

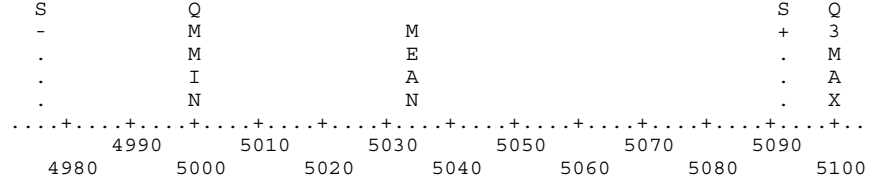
CATEGORY	FREQUENCY
Enero	3
Febrero	3
Marzo	3

\*\*\*\*\*  
\* ANALYSIS OF ventas FOR GROUP Enero \*  
\*\*\*\*\*

EACH 'H' REPRESENTS 1 RESPONSE.



	ESTIMATE	ST. ERROR		ESTIMATE
MEAN	5033.3335000	33.3333321	ST.DEV.	57.7350273
MEDIAN	5000.0000000	0.0000000	VARIANCE	3333.3335000
MODE	5000.0000000		RANGE	100.0000000



```

                PERCENTS
      VALUE   COUNT  CELL  CUM      VALUE   COUNT  CELL  CUM
      5000.    2    66.7  66.7      5100.    1    33.3  100.0
-----
*****
* ANALYSIS OF ventas  FOR GROUP Febrero *
*****

EACH 'H' REPRESENTS 1 RESPONSE.

      H          H  H
.....+.....+.....
      5000
          5200

      ESTIMATE      ST.ERROR      ESTIMATE
MEAN      5153.3335000    104.9338460    ST.DEV.      181.7507480
MEDIAN    5210.0000000      0.0000000    VARIANCE    33033.3359000
MODE      NOT UNIQUE                                RANGE      350.0000000

      Q  S
      1  -
      M  .
      I  .
      N  .
.....+.....+.....+.....+.....+.....
      4950      5000      5050      5100      5150      5200      5250      5300

                PERCENTS
      VALUE   COUNT  CELL  CUM      VALUE   COUNT  CELL  CUM
      4950.    1    33.3  33.3      5300.    1    33.3  100.0
      5210.    1    33.3  66.7
-----

```

```

*****
* ANALYSIS OF ventas  FOR GROUP Marzo *
*****

EACH 'H' REPRESENTS 1 RESPONSE.

      HH          H
.....+.....+.....+.....
      5100      5700
          5400

      ESTIMATE      ST.ERROR      ESTIMATE
MEAN      5233.3335000    233.4047550    ST.DEV.      404.2688900
MEDIAN    5010.0000000      0.0000000    VARIANCE    163433.3280000
MODE      NOT UNIQUE                                RANGE      710.0000000

      S          Q
      -          1 M
      .          M E
      .          I D
      .          N I
.....+.....+.....+.....+.....+.....+.....
      4900      5000      5100      5200      5300      5400      5500      5600      5700

                PERCENTS
      VALUE   COUNT  CELL  CUM      VALUE   COUNT  CELL  CUM
      4990.    1    33.3  33.3      5700.    1    33.3  100.0
      5010.    1    33.3  66.7
-----

```

Copying Plots from workfile to C:\E4\2D002.PLT ...  
PROGRAM TERMINATED

En la página siguiente pueden verse los 3 histogramas impresos juntos en PLOT después de editar sus ejes para unificar la escala.

5. Representación gráfica conjunta de dos variables (representación de la evolución de horas de trabajo, ventas y beneficios para cada equipo).

El programa 6D permite trazar diagramas de dispersión (scatterplots), lo que permite ver gráficamente la relación de una variable con otra. Aquí vamos a utilizar el programa 6D para situar en abcisas el mes y en ordenadas, sucesivamente, las horas trabajadas, las ventas y los beneficios. Lo haremos de forma que podamos distinguir dentro de cada uno de estos gráficos los datos correspondientes a cada uno de los tres equipos, de ese modo podremos ver su evolución comparativamente.

**INSTRUCCIONES. PROGRAMA 6D.**

```



```

En el párrafo /GROUP se establecen los grupos en la variable equipo (llamada eq) y se establecen nombres de conveniencia para los equipos (A, B y C) para poder distinguirlos fácilmente en los gráficos.

En el párrafo /PLOT se describe el tipo de graficos deseados. Se establece que la variable en el eje de X (abcisas) es el mes. En el eje Y colocaremos tres variables sucesivamente (horash, ventas y benef). La instrucción CROSS significa que debe hacer un gráfico para cada variable mencionada en el eje de X por cada variable mencionada en el eje de Y (en este caso 1 x 3 = 3 gráficos). La instrucción file señala donde debe guardar

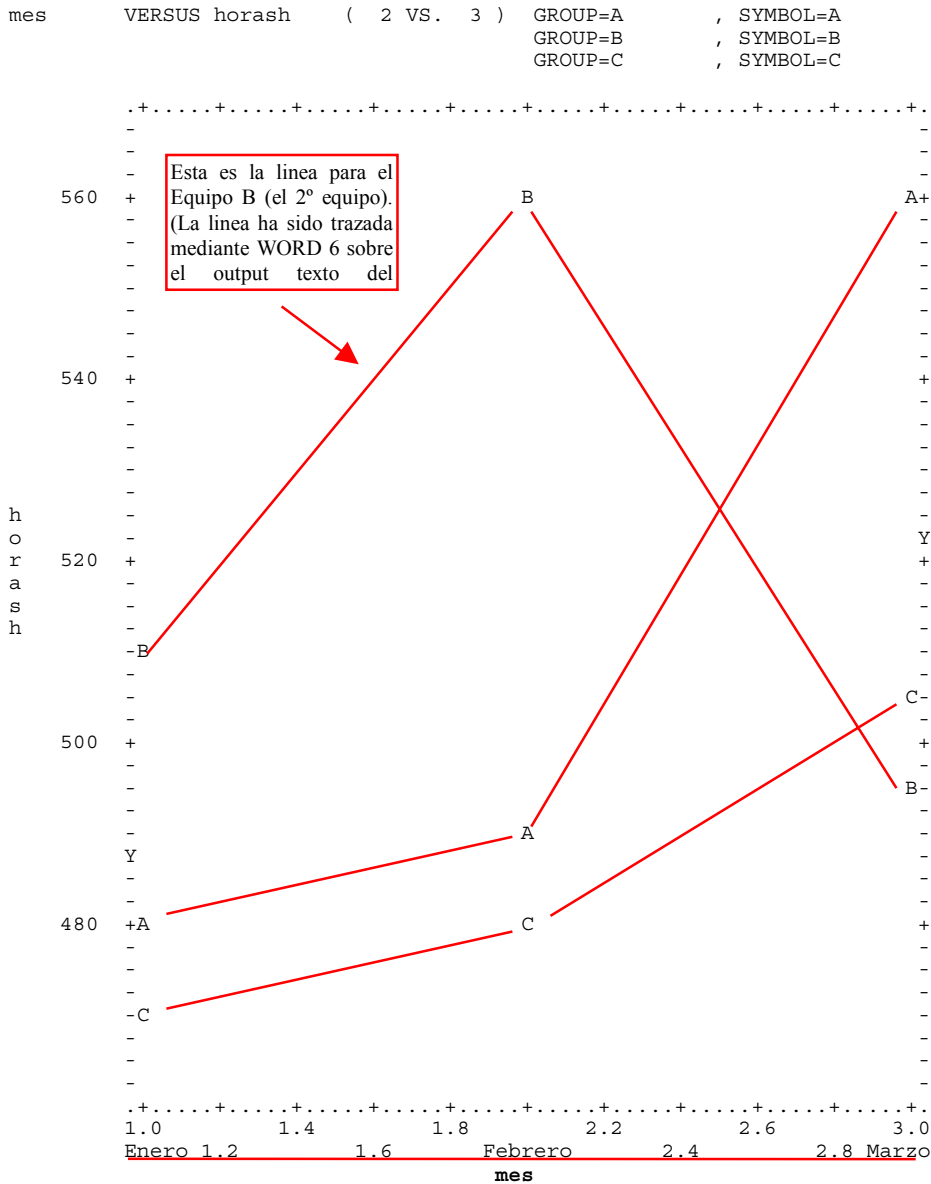
**OUTPUT SELECCIONADO**

```

BMDP6D - BIVARIATE (SCATTER) PLOTS

GROUPING VARIABLE . . . . . 1 eq
NUMBER OF CASES READ. . . . . 9
    
```

En el output se presentan los gráficos en modo texto, a continuación de cada uno de ellos se ha intercalado el gráfico en modo alta resolución correspondiente obtenido con PLOT.



N = 9  
R = .433  
P = .244

--REGRESSION LINE--  
Y= 472.22 +16.667\*X

-RES.MS-  
1029.4

	MEAN	S.D.
X	2.0000	.86603
Y	505.56	33.302

El programa informa de la recta de regresión de Y sobre X para el conjunto de los grupos. Esta ecuación permite pronosticar (Y) horash a partir del (X) mes.

Media cuadrática residual obtenida con la ecuación de regresión lineal anterior. Es la suma de los errores cuadráticos de estimación dividida por N-2.

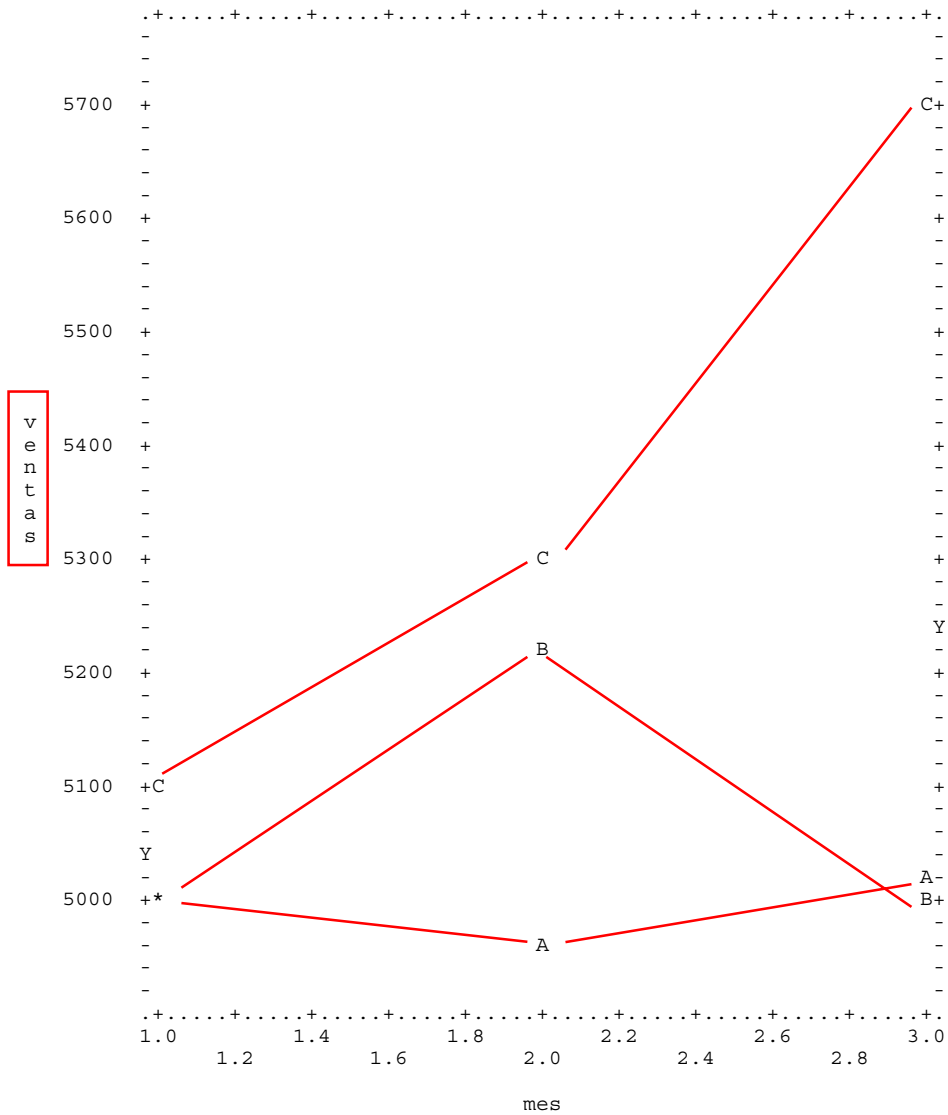
N es el número de casos del gráfico.  
R es el coeficiente de correlación de Pearson.  
p es la probabilidad de obtener por azar en estas circunstancias un coeficiente de correlación de ese tamaño o mayor. Si es menor o igual que 0.05 se dice que el coeficiente de correlación difiere significativamente de 0. Es decir, en ese caso creemos que en la población hay alguna correlación entre las variables X e Y. En este caso el coeficiente de correlación de Pearson no es de interés dada la naturaleza de las variables en

El programa presenta también la media y la desviación típica tanto de X (mes en este caso) como de Y (horash en ese caso).

--- Plot 1 has been added to PLOTFILE.



```
mes      VERSUS ventas    ( 2 VS. 4 )  GROUP=A      , SYMBOL=A
                                GROUP=B      , SYMBOL=B
                                GROUP=C      , SYMBOL=C
```



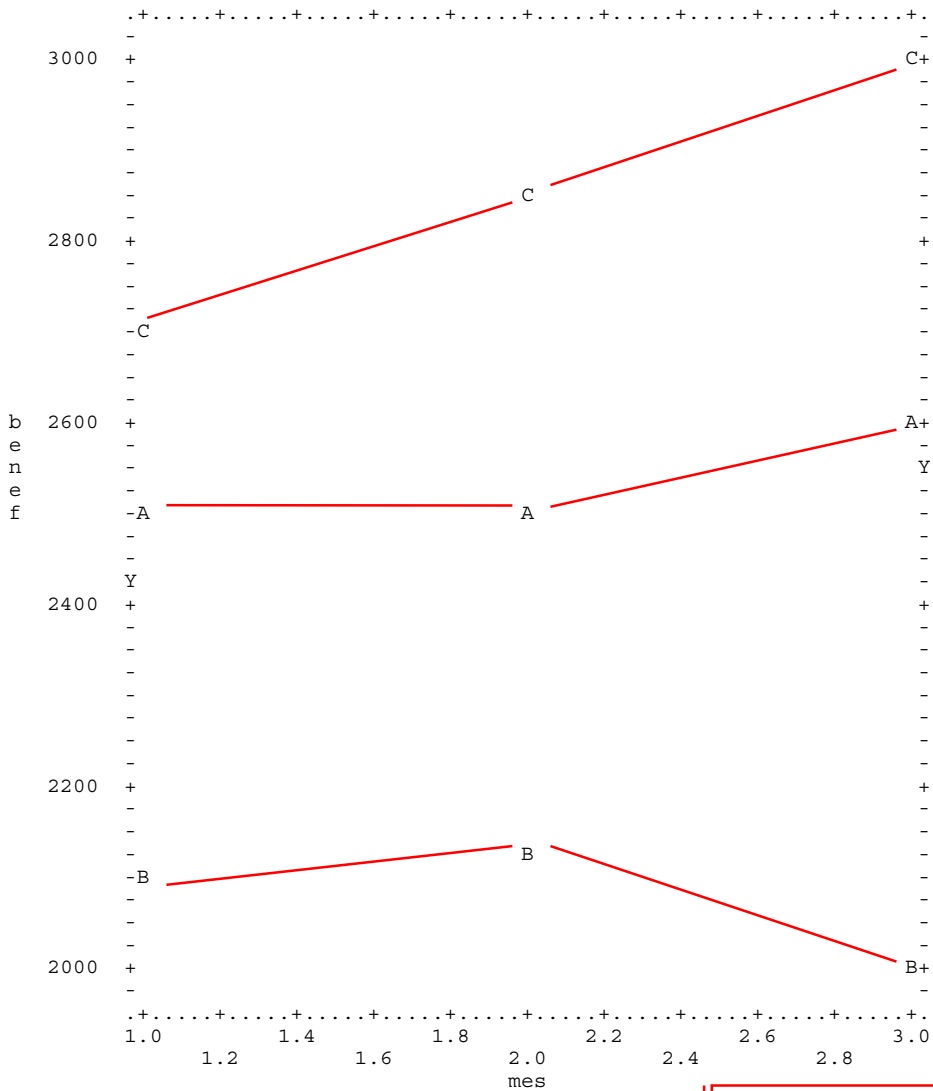
N = 9  
R = .361  
P = .340

```
--REGRESSION LINE--
Y= 4940.0 +100.00*X
-RES.MS-      X  MEAN  S.D.
              Y  2.0000 .86603
              Y  5140.0 239.90
```

Ecuación de regresión lineal para el conjunto de los datos. Dado el desigual comportamiento de los equipos, no sería adecuada para efectuar un buen pronóstico del comportamiento futuro de ninguno de ellos.

--- Plot 2 has been added to PLOTFILE.

```
mes      VERSUS benef      ( 2 VS. 5 )
GROUP=A      , SYMBOL=A
GROUP=B      , SYMBOL=B
GROUP=C      , SYMBOL=C
```



Estos gráficos por grupos (equipos), mediante 6D, colocando la variable meses en el eje de abscisas, han permitido ver con toda claridad que la evolución de los 3 equipos es muy dispar, algo que no sería tan patente si solo hubiéramos obtenido estadísticos

N = 9  
R = .124  
P = .750

```
--REGRESSION LINE--
Y= 2386.1 +50.000*X
RES.MS- 136627
MEAN X 2.0000 S.D. .86603
Y 2486.1 348.46
```

Dado el heterogeneo comportamiento de los tres equipos a lo largo de los tres meses, esta ecuación que pronostica la variable Y (beneficios) a partir de X (mes) teniendo en cuenta a los tres equipos a la vez, no permitiría un buen pronóstico. Sería mejor una ecuación particular para cada equipo

--- Plot 3 has been added to PLOTFILE.  
Copying Plots from workfile to C:\E4\6D001.PLT ...

NOTA:  
Se intercalan los mismos gráficos obtenidos con PLOT.  
Los gráficos en alta resolución incluyen una recta que es la línea de regresión que ha calculado el programa para el conjunto de los grupos.  
En gráfico en alta resolución para horash se han añadido las líneas para cada grupo.  
Las líneas de grupos en los gráficos en modo texto se han añadido en WORD 6.0 para mayor claridad en la lectura de los mismos.

**Obtención de puntuaciones diferenciales y puntuaciones típicas.**

6. Primer paso: Obtener la media y la desviación típica de las variables que deseamos transformar a diferenciales o a típicas.

Nuestro propósito con este análisis es obtener la media y la desviación típica de las variables, en particular de la variable W para obtener mediante transformación en un análisis posterior las puntuaciones diferenciales y típicas de esta variable. Vamos a utilizar para ello el programa 2D, pero también habiéramos podido utilizar para este fin el 1D. Las instrucciones siguientes servirían tanto para 1D como para 2D sin ninguna modificación.

**INSTRUCCIONES. PROGRAMA 2D.**

```

/input
var are 4. case=10. format is free.
/var
names are X, Y, Z, W. Freq is Z.
/end
2 2 4 10
3 1 9 7
4 2 15 8
5 1 26 9
6 2 36 5
7 3 49 1
4 2 15 8
5 2 25 7
5 2 26 7
5 3 15 8
/end
    
```

**OUTPUT SELECCIONADO**

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES

```

/input
var are 4. case=10. format is free.
/var
names are X, Y, Z, W. Freq is Z.
/end
    
```

CASE NO.	3 Z	1 X	2 Y	4 W
1	4.00	2.00	2.00	10.00
2	9.00	3.00	1.00	7.00
3	15.00	4.00	2.00	8.00
4	26.00	5.00	1.00	9.00
5	36.00	6.00	2.00	5.00
6	49.00	7.00	3.00	1.00
7	15.00	4.00	2.00	8.00
8	25.00	5.00	2.00	7.00
9	26.00	5.00	2.00	7.00
10	15.00	5.00	3.00	8.00

Obsérvese que al definir Z como frecuencias mediante la instrucción Freq is Z, el listado reordena las variables para presentar primero la columna de

Aquí comienza la descripción de las variables según el modelo usual del programa 2d.

NUMBER OF CASES READ. . . . . 10

```

*****
* X *
*****
    
```

```

                                VALUE   ZSCORE  CASE #
VARIABLE NUMBER . . . . . 1    MAX     7.0000000  1.400    6
                                MIN     2.0000000 -2.808    1
NUMBER OF DISTINCT VALUES . . . . . 6
NUMBER OF VALUES COUNTED. . . . . 220
NUMBER OF VALUES NOT COUNTED 0
                                VALUE   VALUE/S.E.
                                SKWNESS -0.28  -1.679
                                KURTOSIS -0.16  -0.497
'H' REPRESENTS 12 RESPONSES, '*' REPRESENTS LESS THAN 12 RESPONSES.
    
```

```

*
H
H
H *
H H
* H H H
H H H H
* * H H H H
...+...+...+...+...+...
2      6
4
    
```

Histograma en formato texto. Cada H representa 12 casos y cada \* menos de 12 casos, como expresa el output antes del

ESTIMATE                      ST. ERROR                      ESTIMATE

MEAN	5.3363638	0.0800982	ST.DEV.	1.1880476
MEDIAN	5.0000000	0.0000000	VARIANCE	1.4114572
MODE	5.0000000		RANGE	5.0000000
			(Q3-Q1)/2	0.5000000

LOWER 95% C.L. OF MEAN	5.1785021
UPPER 95% C.L. OF MEAN	5.4942255
Q1	5.0000000
Q3	6.0000000
S-	4.1483164
S+	6.5244112

```

          S      Q      Q      S
          -      M      M      3      +
M         .      E      E      .      .      M
I         .      D      A      .      .      A
N         .      I      N      .      .      X
    +-----+-----+-----+-----+-----+-----+

```

2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	
PERCENTS				PERCENTS							
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM				
2.	4	1.8	1.8	5.	92	41.8	61.4				
3.	9	4.1	5.9	6.	36	16.4	77.7				
4.	30	13.6	19.5	7.	49	22.3	100.0				

\*\*\*\*\*  
 \* Y \*  
 \*\*\*\*\*

VARIABLE NUMBER . . . . .	2	MAX	3.0000000	ZSCORE	1.317	CASE #	6
NUMBER OF DISTINCT VALUES .	3	MIN	1.0000000		-1.717		2
NUMBER OF VALUES COUNTED . .	220			VALUE	VALUE/S.E.		
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS	-0.15		-0.878		
		KURTOSIS	-0.74		-2.246		

'H' REPRESENTS 15 RESPONSES, '\*' REPRESENTS LESS THAN 15 RESPONSES.

```

          *
          H
          H
          H
          H      *
          H      H
*         H      H
H         H      H
H         H      H
    +-----+-----+-----+-----+

```

1.4	2.1	2.8	ESTIMATE	ST.ERROR	ESTIMATE
MEAN	2.1318183	0.0444461	ST.DEV.	0.6592416	
MEDIAN	2.0000000	0.0000000	VARIANCE	0.4345994	
MODE	2.0000000		RANGE	2.0000000	
			(Q3-Q1)/2	0.5000000	

LOWER 95% C.L. OF MEAN	2.0442216
UPPER 95% C.L. OF MEAN	2.2194149
Q1	2.0000000
Q3	3.0000000
S-	1.4725767
S+	2.7910600

```

          S      Q      S      Q
          -      M      M      +      3
M         .      E      E      .      M
I         .      D      A      .      A
N         .      I      N      .      X
    +-----+-----+-----+-----+

```

1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	
PERCENTS				PERCENTS							
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM				
1.	35	15.9	15.9	3.	64	29.1	100.0				
2.	121	55.0	70.9								

```

*****
* W *
*****

                Valor      En Tipicas      caso
VARIABLE NUMBER . . . . . 4      MAX      10.0000000    1.457      1
                MIN      1.0000000    -1.689      6
NUMBER OF DISTINCT VALUES . 6
NUMBER OF VALUES COUNTED . . 220
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  -0.75      -4.557
                KURTOSIS  -0.85      -2.582
    
```

'H' REPRESENTS 7 RESPONSES, '\*' REPRESENTS LESS THAN 7 RESPONSES.

```

*
H
H *
H * H H
H H H H *
H H H H H
H H H H H
H H H H H*
+ . . . . + . . . . + . . . .
0. 4. 8.
    
```

La curtosis estandarizada es menor que -2. Por tanto tenemos una distribución que podemos considerar platocúrtica, es decir, más aplanada que una distribución normal.

La asimetría estandarizada es negativa y mayor de 2 en valores absolutos: Tenemos por tanto una distribución asimétrica sesgada a la izquierda (es decir, los casos se acumulan en los valores más altos de la derecha y la cola de la izquierda es, relativamente, más delgada).

Media y desviación típica necesarias para obtener después diferenciales y típicas de la variable W

```

                ESTIMATE      ST. ERROR      ST. DEV.      ESTIMATE
MEAN      5.8318181      0.1928989      2.8611531
MEDIAN      7.0000000      0.0000000      VARIANCE      8.1861973
MODE      7.0000000      RANGE      9.0000000
                (Q3-Q1)/2      1.5000000
    
```

La posición de la mediana a la derecha de la media también indica que la curva concentra sus frecuencias en la parte derecha, y que por tanto se dice que “esta sesgada a la izquierda” (se menciona la cola más delgada en esta

```

LOWER 95% C.L. OF MEAN      5.4516420
UPPER 95% C.L. OF MEAN      6.2119942
Q1      5.0000000
Q3      8.0000000
S-      2.9706650
S+      8.6929712
    
```

```

                S      Q      M      M      Q      S
M      .      .      E      E      .      .      M
I      .      .      A      D      .      .      A
N      .      .      N      I      .      .      X
+ . . . . + . . . . + . . . . + . . . . + . . . . + . . . . + . . . .
1. 2. 3. 4. 5. 6. 7. 8. 9. 10
    
```

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	49	22.3	22.3	8.	45	20.5	86.4
5.	36	16.4	38.6	9.	26	11.8	98.2
7.	60	27.3	65.9	10.	4	1.8	100.0

/end

7. Transformación de una variable a puntuaciones diferenciales y a puntuaciones típicas conocida su media y su desviación típica.

El ejemplo introduce la transformación de una variable a diferenciales, mediante la creación de una variable transformada en la que a cada dato inicial se le sustrae la media aritmética, previamente conocida por un análisis anterior. A continuación, mediante una segunda nueva variable transformada, se transforma la misma variable a puntuaciones típicas dividiendo la puntuación diferencial por la desviación típica también previamente conocida.

Podemos crear variables transformadas mediante el párrafo /TRANS en cualquier conjunto de instrucciones para cualquier programa BMDP. Por ejemplo, si no necesitáramos una descripción detallada y tabulación de las nuevas variables transformadas podríamos haber utilizado en este análisis el 1D. Hemos preferido hacer todos los pasos con el 2D porque así pueda apreciarse en los histogramas y en la tabulación como las puntuaciones diferenciales de W (representadas por la nueva variable Wdif), y las típicas (representadas por la nueva variable que hemos llamado Wtip), conservan exactamente la misma distribución de las puntuaciones directas originales de la variable W, de las que solo son una transformación lineal.

Además el ejemplo introduce el formato stream, un tipo de formato libre menos utilizado que permite escribir los casos uno a continuación de otro, con varios casos por línea.

**INSTRUCCIONES. PROGRAMA 2D.**

```

/input
var are 4. case=10. format is stream. #stream permite varios casos por linea
/var names are X, Y, Z, W. Freq is Z.
/trans
Wdif = W - 5.8318181. #Crea una nueva variable con las diferenciales de W
Wtip = Wdif/2.8611531. #Crea una nueva variable con las típicas de W
#Los datos del valor de la media y de la desviación típica
#los hemos obtenido en el analisis anterior.
/end
2 2 4 10 3 1 9 7 4 2 15 8 5 1 26 9 6 2 36 5 7 3 49 1 4 2 15 8 5 2 25 7
5 2 26 7 5 3 15 8
/end
    
```

**OUTPUT SELECCIONADO**

```

NUMBER OF CASES READ. . . . . 10

*****
* Wdif *
*****
                Descripción de las puntuaciones diferenciales
                de la variable W (llamadas aquí Wdif)

VARIABLE NUMBER . . . . . 5          MAX      4.1681819    ZSCORE   CASE #
                . . . . . 6          MIN      -4.8318181   -1.689    6
NUMBER OF DISTINCT VALUES . . . . . 6
NUMBER OF VALUES COUNTED. . . . . 220          VALUE     VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0          SKEWNESS  -0.75     -4.557
                KURTOSIS   -0.85     -2.582
'H' REPRESENTS 7 RESPONSES, '*' REPRESENTS LESS THAN 7 RESPONSES.

      *
      H
H      H *
H      * H H
H      H H H
H      H H H *
H      H H H H
H      H H H H
H      H H H H*
.....+.....+.....+.....
-4          0          4

ESTIMATE      ST. ERROR      ESTIMATE
MEAN          0.000000    0.1928989    ST. DEV.     2.8611531
MEDIAN        1.1681819    0.0000000    VARIANCE     8.1861973
MODE          1.1681819                                RANGE        9.0000000
                                                    (Q3-Q1)/2    1.5000000

                LOWER 95% C.L. OF MEAN    -0.3801759
                UPPER 95% C.L. OF MEAN    0.3801759
                Q1                          -0.8318181
                Q3                          2.1681819
                S-                          -2.8611531
                S+                          2.8611531

                S          Q          Q          S
                -          1          M          M          3          +
    
```

2. Descripción detallada y tabulación de variables J. L. Meliá (1997) Análisis de datos con BMDP [www.uv.es/psicometria](http://www.uv.es/psicometria)

La distribución es idéntica a la de W, pero ahora la media es 0. Todos los valores están desplazados hacia abajo "el valor de una media"

```

M      .      .      E      E      .      .      M
I      .      .      A      D      .      .      A
N      .      .      N      I      .      .      X
.....+.....+.....+.....+.....+.....+.....+.....+.....+.....
      -4      -3      -2      -1      0      1      2      3      4
    
```

```

"
      VALUE      COUNT      PERCENTS      PERCENTS
      VALUE      COUNT      CELL      CUM      VALUE      COUNT      CELL      CUM
      -4.8318     49      22.3     22.3      2.1682     45      20.5     86.4
      -0.8318     36      16.4     38.6      3.1682     26      11.8     98.2
      1.1682      60      27.3     65.9      4.1682      4       1.8    100.0
    
```

```

*****
* Wtip *
*****
    
```

Descripción de las puntuaciones típicas

```

VARIABLE NUMBER . . . . . 6      MAX      1.4568189      ZSCORE      CASE #
NUMBER OF DISTINCT VALUES . 6      MIN      -1.6887660     -1.689      6
NUMBER OF VALUES COUNTED . . 220     VALUE      VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS   -0.75      -4.557
                                          KURTOSIS   -0.85      -2.582
'H' REPRESENTS      7 RESPONSES, '*' REPRESENTS LESS THAN 7 RESPONSES.
    
```

```

      *
      H
H      H *
H      * H H
H      H H H
H      H H H *
H      H H H H
H      H H H H
H      H H H H*
.....+.....+.....+.....+.....
      -1      0      1
-2
    
```

La distribución es igual que en las diferenciales o en las directas. Ahora a media vale 0 y la desviación típica y la varianza valen 1.

```

MEAN      ESTIMATE      ST.ERROR      ST.DEV.      ESTIMATE
MEDIAN    0.0000000      0.0674200      1.0000001
MODE      0.4082906      0.0000000      VARIANCE     1.0000002
                                          RANGE        3.1455851
                                          (Q3-Q1)/2   0.5242642
                                          LOWER 95% C.L. OF MEAN -0.1328750
                                          UPPER 95% C.L. OF MEAN 0.1328751
                                          Q1          -0.2907283
                                          Q3          0.7578000
                                          S-         -1.0000001
                                          S+          1.0000001
    
```

```

      S      Q      M      M      Q      S
      -      1      E      E      3      +
M      .      .      A      A      .      .      M
I      .      .      N      D      .      .      A
N      .      .      N      I      .      .      X
.....+.....+.....+.....+.....+.....+.....+.....+.....+.....
      -1.5     -1.2     -0.90    -0.30     0.0     0.30     0.60     0.90     1.2     1.5
    
```

```

      VALUE      COUNT      PERCENTS      PERCENTS
      VALUE      COUNT      CELL      CUM      VALUE      COUNT      CELL      CUM
      -1.6888     49      22.3     22.3      0.7578     45      20.5     86.4
      -0.2907     36      16.4     38.6      1.1073     26      11.8     98.2
      0.4083      60      27.3     65.9      1.4568      4       1.8    100.0
    
```

/end

8. Obtener un listado de las puntuaciones diferenciales y típicas obtenidas mediante transformación utilizando el programa 1D.

En el análisis anterior hemos obtenido por transformación las diferenciales y las típicas de W, y hemos analizado su distribución mediante 2D. Ahora nos proponemos obtener un listado de los datos con las puntuaciones diferenciales y típicas de cada caso utilizando el programa 1D.

**INSTRUCCIONES. PROGRAMA 1D.**

```
/input
var are 4. case=10. format is stream. #stream permite varios casos por linea
/var names are X, Y, Z, W. Freq is Z.
```

```
/trans
Wdif=W-5.8318181. #Crea una nueva variable con las diferenciales de W
Wtip= Wdif/2.8611531. #Crea una nueva variable con las típicas de W
```

```
/print
data. #para ver las tipicas de W necesitamos que nos muestre los datos el 1D
/end
2 2 4 10 3 1 9 7 4 2 15 8 5 1 26 9 6 2 36 5 7 3 49 1 4 2 15 8 5 2 25 7
5 2 26 7 5 3 15 8
/end
```

Esta es la instrucción que provoca la aparición en el output del listado de datos.

Hay que mantener las instrucciones de transformación siempre que deseemos hacer uso de esas transformadas. (Salvo cuando se guardan los datos en formato de archivo BMDP las transformadas NO son añadidas a los datos, por eso hay que mantener las instrucciones de transformación cada vez que deseemos hacer uso de estas variables transformadas.

**OUTPUT SELECCIONADO**

BMDP1D - SIMPLE DATA DESCRIPTION  
 NUMBER OF CASES READ. . . . . 10

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO. NAME	TOTAL FREQ.	MEAN	STANDARD DEV.	ST.ERR OF MEAN	COEFF OF VAR	SMALLEST VALUE	Z-SCR	LARGEST VALUE	Z-SCR
1 X	220	5.3364	1.1880	.08010	.22263	2.0000	-2.81	7.0000	1.40
2 Y	220	2.1318	.65924	.04445	.30924	1.0000	-1.72	3.0000	1.32
4 W	220	5.8318	2.8612	.19290	.49061	1.0000	-1.69	10.000	1.46
5 Wdif	220	0.0000	2.8612	.19290	4542E4	-4.8318	-1.69	4.1682	1.46
6 Wtip	220	0.0000	1.0000	.06742	2740E4	-1.6888	-1.69	1.4568	1.46

CASE NO.	3 Z	1 X	2 Y	4 W	5 Wdif	6 Wtip
1	4.00	2.00	2.00	10.00	4.17	1.46
2	9.00	3.00	1.00	7.00	1.17	0.41
3	15.00	4.00	2.00	8.00	2.17	0.76
4	26.00	5.00	1.00	9.00	3.17	1.11
5	36.00	6.00	2.00	5.00	-0.83	-0.29
6	49.00	7.00	3.00	1.00	-4.83	-1.69
7	15.00	4.00	2.00	8.00	2.17	0.76
8	25.00	5.00	2.00	7.00	1.17	0.41
9	26.00	5.00	2.00	7.00	1.17	0.41
10	15.00	5.00	3.00	8.00	2.17	0.76

Aquí podemos observar las puntuaciones típicas y diferenciales que son fruto de la transformación de la variable W



**Medias Ponderadas.**

9. Obtención de la media ponderada utilizando freq.

Podemos obtener la media ponderada de una variable X introduciendo los valores de esa variable y, en otra variable P las ponderaciones de cada uno de esos valores. En esas circunstancias basta definir P como la frecuencia de X para que la media aritmética de X sea en realidad su media ponderada por P.

INSTRUCCIONES. PROGRAMA 2D.

```
#Media Ponderada.
/input var=2. case=5. format=free.
/var names=x, p. freq=p. #p actúa como una ponderación de los casos.
/end
2 1
3 1
4 1
5 2
6 3
/end
```

OUTPUT SELECCIONADO

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES  
Release: 7.0 (BMDP/DYNAMIC)

```
#Media Ponderada.
/input var=2. case=5. format=free.
/var names=x, p. freq=p. #p actúa como una ponderacion de los casos.
/end
```

CASE NO.	2 p	1 x
1	1.00	2.00
2	1.00	3.00
3	1.00	4.00
4	2.00	5.00
5	3.00	6.00

NUMBER OF CASES READ. . . . . 5

\*\*\*\*\*  
\* x \*  
\*\*\*\*\*

VARIABLE NUMBER	1	MAX	VALUE	ZSCORE	CASE #
NUMBER OF DISTINCT VALUES	5	MIN	2.0000000	-1.743	1
NUMBER OF VALUES COUNTED	8				
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS	-0.54	VALUE/S.E.	
		KURTOSIS	-1.41	-0.815	

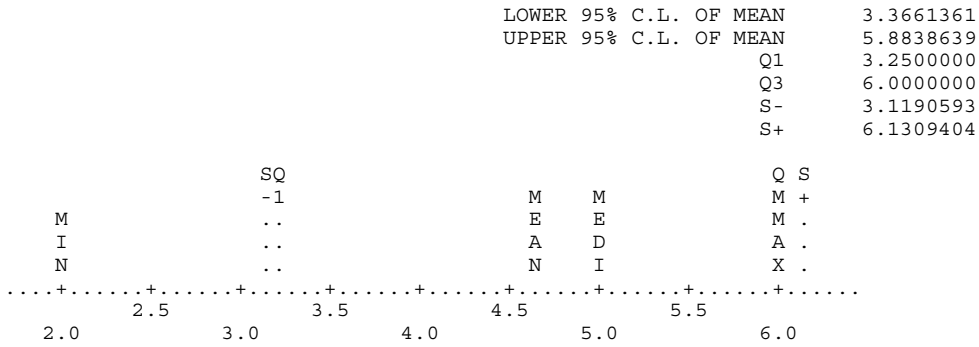
EACH 'H' REPRESENTS 1 RESPONSE.

```

          H
        H H
      H H H
    .+. .+. .+. .+.
     2   4   6
```

La media (mean=4.625) es la media de X ponderada por P dado que hemos tratado P como si fueran las frecuencias de X.  
Con este procedimiento para calcular la media ponderada hay que tener cuidado pues los otros estadísticos calculados también han sufrido la ponderación de X, lo que puede hacer que no sea los que nos interesen.  
Para obtener los demás estadísticos sin ponderación basta suprimir "freq=p" en el input y hacer un nuevo run solo para X.

	ESTIMATE	ST. ERROR	ESTIMATE
MEAN	4.6250000	0.5324304	ST.DEV. 1.5059407
MEDIAN	5.0000000	0.8660258	VARIANCE 2.2678573
MODE	6.0000000		RANGE 4.0000000
			(Q3-Q1)/2 1.3750000



PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
2.	1	12.5	12.5	5.	2	25.0	62.5
3.	1	12.5	25.0	6.	3	37.5	100.0
4.	1	12.5	37.5				

La descripción frecuencial, gráficos y otros estadísticos utilizan las "frecuencias" aportadas por P, lo que hace que no sean las indicadas si simplemente queremos las frecuencias de X.

10. Un segundo procedimiento para obtener la Media ponderada mediante transformaciones.

Vamos a introducirnos en el uso de las transformaciones con operadores aritméticos en BMDP mediante un segundo procedimiento para obtener la media de X ponderada por P que no hace uso de la instrucción FREQ y que requiere calculadora en los últimos dos pasos. Supongamos que, como en el ejemplo anterior, X contiene la variable cuya media ponderada deseamos calcular, y p las ponderaciones que deseamos para cada caso. Puede utilizarse una variable transformada para calcular la media ponderada de X. Para ello se pueden seguir los siguientes pasos:

1) Se crea una nueva variable transformada, que aquí llamaremos T, del tipo X multiplicado por P, añadiendo el párrafo /TRANS a las instrucciones:

```
/TRANS T=X*P.
```

(La transformación dice: Crea una variable T que será igual a X por P).

En el ejemplo anterior los valores de T serían:

X	P	T
2	1	2
3	1	3
4	1	4
5	2	10
6	3	18

Los valores de T podrían listarse en BMDP con la instrucción en 1D:

```
/PRINT data.
```

2) Mediante un programa descriptivo (p.e., el 1D) se obtiene la media aritmética de estas variables. En el ejemplo estás serían:

X	P	T
2	1	2
3	1	3
4	1	4
5	2	10
6	3	18
Suma	20	8 37
Media	4	1'6 7'4

El programa da las medias de X, P y T, pero no la suma de sus valores para todos los casos.

Ninguna de las tres medias es la media de X ponderada por P, que es la que buscamos. Sin embargo, las medias de P y de T nos ayudaran a obtener las sumas de las puntuaciones de P y de T, y de ahí la media de T, que es la media ponderada de X. Veámoslo.

3) Obtener la sumas de P y T a partir de sus medias aritméticas:

Si la media de P es 1'6, habiendo 5 casos, entonces la suma de los valores de P es  $5 \cdot 1'6 = 8$ .

Si la media aritmética de T es 7'4, con 5 casos, entonces la suma de T es 7'4 por 5 igual a 37.

4) Obtener la media de X ponderada por P como la suma de T dividida por la suma de P:

La media ponderada de X es igual a la suma de T dividida por la suma de P, por tanto la media ponderada es

$$37/8 = 4'625.$$

#### 11. Cómo obtener medias ponderadas a partir de datos tabulados.

Supongamos en los datos anteriores, además de X y sus ponderaciones P tuviéramos además una auténtica columna de frecuencias llamada F.

En ese caso necesitaríamos crear una nueva transformada que podemos llamar W tal que sea el producto de F por P.

/TRANS W=F\*P.

Ahora basta con establecer que W son las frecuencias mediante:

FREQ=W.

en el párrafo /VAR.

Téngase en cuenta que en ese caso la media de X será su media ponderada por P teniendo en cuenta las frecuencias F, pero que todos los demás estadísticos también estarán afectados por esta ponderación P, lo que no siempre será lo que deseemos por lo que probablemente habrá que repetir el run utilizando como frecuencias F (datos de X sin ponderar por P).

## Utilización de Transformaciones con Operadores Lógicos en BMDP.

12. Cómo introducir ponderaciones cuando el archivo de datos es muy grande.

Para cinco, diez o 20 casos introducir una columna de datos P con ponderaciones es relativamente cómodo y rápido. Si tuviéramos que ponderar los, por ejemplo, seis posibles valores que puede adoptar X en un archivo de diez mil casos es poco operativo pretender hacerlo escribiendo una nueva columna de datos. Afortunadamente para crear P podemos utilizar transformaciones, dado que los valores de P dependen de los valores de X (ó de cualquier otra variable previamente introducida).

Por ejemplo si el patrón de ponderaciones P fuera, como en el ejemplo anterior:

```
P vale: si X es ...
1 si X es igual o menor a 4.
2 si X es igual a 5.
3 si X es igual a 6.
```

podíamos crear una variable de ponderaciones P para todos los casos del siguiente modo:

```
/TRANS
IF (X LE 4) THEN (P = 1).
IF (X EQ 5) THEN (P = 2).
IF (X EQ 6) THEN (P = 3).
```

La primera línea dice, "Si (X es menor o igual que 4) entonces (la variable P vale 1).

La segunda dice: "Si (X es igual a 5) entonces (P vale 2).

La tercera: "Si (X es igual a 6) entonces (P = 3).

Estas instrucciones llevan implícito la creación de una nueva variable P, que no requiere más instrucciones. (Si P ya existiera las instrucciones anteriores la recodificarían, es decir, le cambiarían los valores si fuera necesario).

Una vez que dispusiéramos mediante las instrucciones anteriores de la nueva variable P podríamos aplicar cualquiera de las estrategias anteriores para obtener la media de X ponderada por P.

### 13. Introducción a las transformaciones con operadores lógicos.

Las instrucciones anteriores son transformaciones que utilizan operadores lógicos. En el primer paréntesis de un IF pueden usarse los siguientes operadores lógicos:

```
LE significa "menor o igual que"
LT significa "menor que"
GE significa "mayor o igual que"
GT significa "menor que"
NE significa "no igual que"
EQ significa "igual que"
AND significa "y"
OR significa "ó"
NOT significa "no"
```

Las transformaciones se ejecutan automáticamente para cada caso al ejecutar el archivo de instrucciones. Es decir, en el ejemplo:

```
/TRANS
IF (X LE 4) THEN (P = 1).
IF (X EQ 5) THEN (P = 2).
IF (X EQ 6) THEN (P = 3).
```

para el caso 1 la primera instrucción verifica si su valor X es menor o igual que 4 en cuyo caso le asigna un 1 a P, la segunda verifica si su valor X es 5, y en su caso le asigna un 2 a P, la tercera verifica si su valor en X es 6, en cuyo caso asigna un 3 a P. Y así para el caso 2, el 3, etc.

Las transformaciones lógicas permiten resolver muchas dificultades de análisis de datos de otro modo no solubles. Pueden ser simples como las anteriores o mucho más complejas. Por ejemplo puede escribirse una transformación como la siguiente:

```
IF (X GT S OR X LT 5.0073) THEN (X = SQRT(S)*3.134/S. P = 14.)
```

Dice: "Si (el valor de la variable X es mayor que el de la variable S, ó bien el valor de la variable X es menor que el valor 5'0073) entonces (el nuevo valor de X es la raíz cuadrada de lo que valga la variable S, multiplicado por 3'134 y dividido por el valor de S. Por otra parte, el valor de P será 14. )

En la instrucción hay dos condiciones unidas por un "ó lógico": es decir, basta que se de una de ambas para que se den las 2 consecuencias que siguen al THEN. Como puede ver las instrucciones pueden combinar variables (S, X), constantes, (5^0073, 14, 3^134), funciones (SQRT( ) es decir, raíz cuadrada), operadores lógicos (GT, OR, LT, solo en la parte de las condiciones que sigue al IF), y operaciones (producto \*, división / etc.)

Dominar las transformaciones es fundamental en análisis de datos profesional. En el volumen 1 del manual del BMDP se dedica un capítulo a una explicación detallada de todas las opciones, operadores, funciones de diversos tipos, etc.

Operadores disponibles:

Operadores:	Significado:	Ejemplo:
+	Adición	X = S + P. (S y P son dos variables previas)
-	sustracción	X = S - P.
*	producto	Índice = Calidad * S.
/	división	Indice2 = Producc/tiempo.
**	Exponenciación	X = J**3.
a MOD b	Resto de a/b	meses = edad MOD 12.
y = a	asignación	Y = 12. (Se asigna 12 a Y en todos los casos).

**Estadísticos de tendencia central robustos.**

14. Media recortada al 15%, estimador de Hampel y estimador Biweight.

**INSTRUCCIONES. PROGRAMA 2D.**

```





1 9 1 10
2 10 1 7
3 11 0 1
4 13 1 9
5 12 1 15
5 7 0 3
6 9 1 18
6 9 0 2
6 12 1 17
7 12 0 1
8 13 1 19
/end
    
```

Solicita la media recortada al 15%, y los estimadores de tendencia central Hampel y Biweight.

**OUTPUT SELECCIONADO**

```

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES





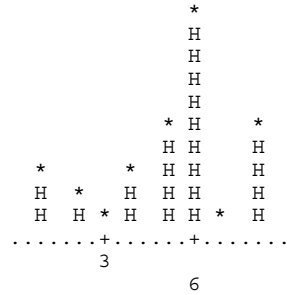

CASE NO. 4 f 1 capital 2 interes 3 tipo
-----
1 10.00 1.00 9.00 1.00
2 7.00 2.00 10.00 1.00
3 1.00 3.00 11.00 0.00
4 9.00 4.00 13.00 1.00
5 15.00 5.00 12.00 1.00
6 3.00 5.00 7.00 0.00
7 18.00 6.00 9.00 1.00
8 2.00 6.00 9.00 0.00
9 17.00 6.00 12.00 1.00
10 1.00 7.00 12.00 0.00

NUMBER OF CASES READ. . . . . 11
    
```

```

*****
* capital *
*****

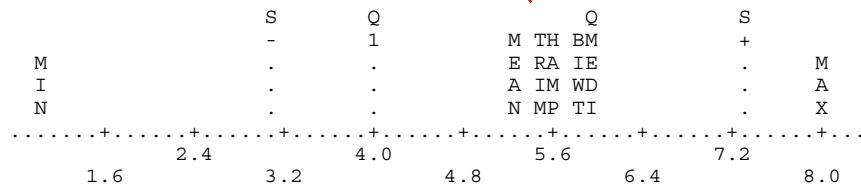
          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 1      MAX      8.0000000  1.321   11
          MIN      1.0000000 -2.024    1
NUMBER OF DISTINCT VALUES . 8
NUMBER OF VALUES COUNTED. . 102
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  -0.64   -2.653
          KURTOSIS  -0.36   -0.745
'H' REPRESENTS 4 RESPONSES, '*' REPRESENTS LESS THAN 4 RESPONSES.
    
```



Estadísticos de tendencia Central Robustos:  
 Media recortada al 15% ,  
 Estimador de Hampel, y  
 Estimador Biweight.  
 Estos estimadores de la tendencia central pueden ser más adecuados que la media con distribuciones con largas colas o valores extremos en las colas.

	ESTIMATE	ST. ERROR	ST. DEV.	ESTIMATE
MEAN	5.2352943	0.2072162		2.0927811
MEDIAN	6.0000000	0.2886753	VARIANCE	4.3797326
MODE	6.0000000		RANGE	7.0000000
			(Q3-Q1) / 2	1.0000000
TRIM (.15)	5.4761906		LOWER 95% C.L. OF MEAN	4.8242331
HAMPEL	5.5522470		UPPER 95% C.L. OF MEAN	5.6463556
BWEIGHT	5.8547192		Q1	4.0000000
			Q3	6.0000000
			S-	3.1425133
			S+	7.3280754

El gráfico permite apreciar la posición de todos estimadores de tendencia central, robustos y no robustos. Obsérvese que todos ellos se ubican entre la Media aritmética y la Mediana.



VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	10	9.8	9.8	5.	18	17.6	44.1
2.	7	6.9	16.7	6.	37	36.3	80.4
3.	1	1.0	17.6	7.	1	1.0	81.4
4.	9	8.8	26.5	8.	19	18.6	100.0

\*\*\*\*\*  
 \* interes \*  
 \*\*\*\*\*

			VALUE	ZSCORE	CASE #
		MAX	13.0000000	1.070	4
VARIABLE NUMBER . . . . .	2	MIN	7.0000000	-2.306	6
NUMBER OF DISTINCT VALUES . . . . .	6				
NUMBER OF VALUES COUNTED . . . . .	102		VALUE	VALUE/S.E.	
NUMBER OF VALUES NOT COUNTED . . . . .	0	SKEWNESS	-0.48	-1.987	
		KURTOSIS	-1.21	-2.486	

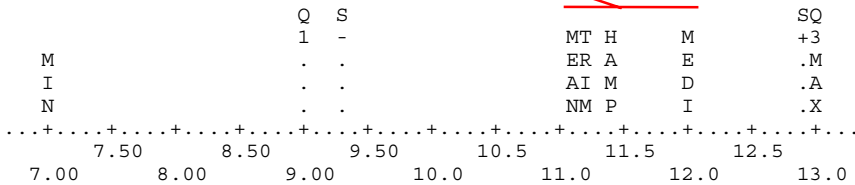
'H' REPRESENTS 4 RESPONSES, '\*' REPRESENTS LESS THAN 4 RESPONSES.

```

      *
      H
    *  H
    H  H H
    H  H H
    H  H H
    H  H H
    H  H H
    H * H H
    *  H H * H H
    .....+.....+.....
      9.
      12
    
```

	ESTIMATE	ST. ERROR		ESTIMATE
MEAN	11.0980396	0.1759738	ST. DEV.	1.7772478
MEDIAN	12.0000000	0.0000000	VARIANCE	3.1586096
MODE	12.0000000		RANGE	6.0000000
			(Q3-Q1)/2	2.0000000
TRIM(.15)	11.2240896		LOWER 95% C.L. OF MEAN	10.7489548
HAMPEL	11.3897057		UPPER 95% C.L. OF MEAN	11.4471245
BWEIGHT	11.1462650		Q1	9.0000000
			Q3	13.0000000
			S-	9.3207922
			S+	12.8752871

Estimadores robustos ubicados entre la Media y la Mediana.

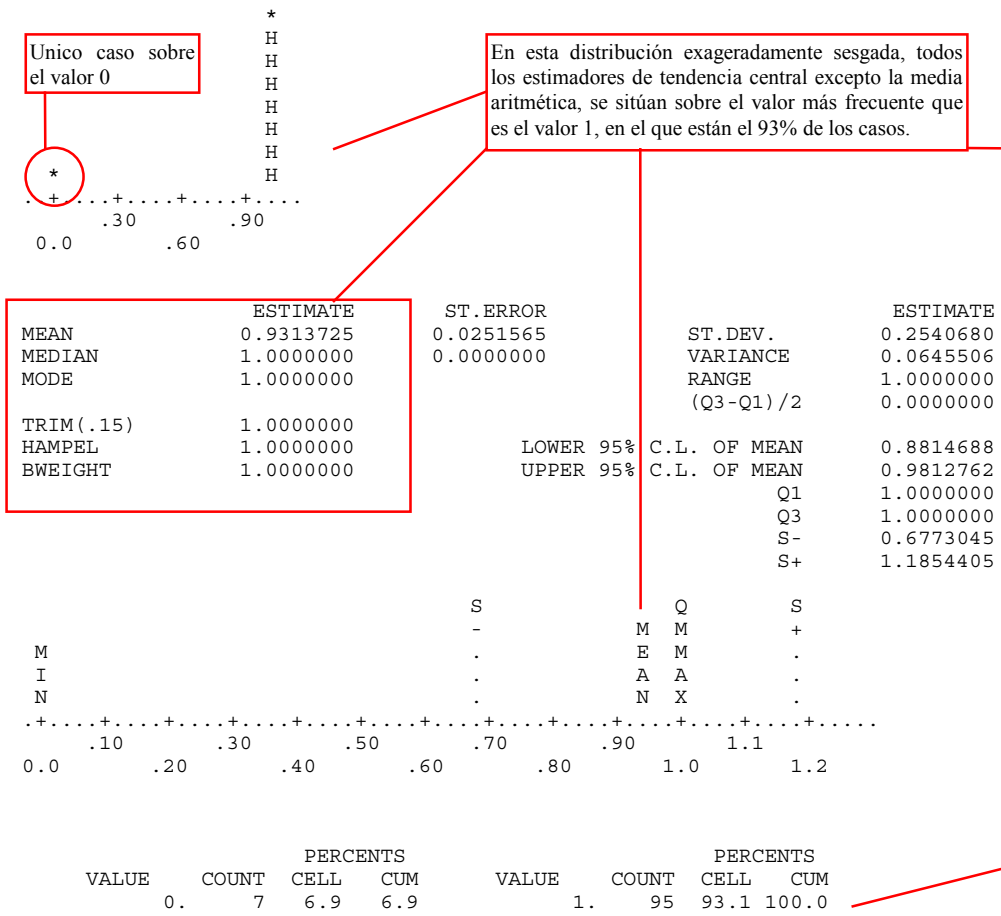


PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
7.	3	2.9	2.9	11.	1	1.0	40.2
9.	30	29.4	32.4	12.	33	32.4	72.5
10.	7	6.9	39.2	13.	28	27.5	100.0

\*\*\*\*\*  
 \* tipo \*  
 \*\*\*\*\*

			VALUE	ZSCORE	CASE #
		MAX	1.0000000	0.270	1
VARIABLE NUMBER . . . . .	3	MIN	0.0000000	-3.666	3
NUMBER OF DISTINCT VALUES . . . . .	2				
NUMBER OF VALUES COUNTED . . . . .	102		VALUE	VALUE/S.E.	
NUMBER OF VALUES NOT COUNTED . . . . .	0	SKEWNESS	-3.36	-13.864	
		KURTOSIS	9.40	19.375	

'H' REPRESENTS 12 RESPONSES, '\*' REPRESENTS LESS THAN 12 RESPONSES.



BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES  
PROGRAM TERMINATED



**Análisis de caso: Créditos devueltos y fallidos en una entidad bancaria.**

En una sucursal bancaria se pretende analizar algunas características de los créditos que han resultado devueltos y fallidos en un determinado periodo. Se dispone de la siguiente información:

Capital: Principal por el que se concedió el crédito (cantidad que se presto, en millones).

Interés: Interes nominal del crédito, aquí se asume constante para la vida del crédito.

Tipo: Clasifica a os créditos en fallidos (=0, es decir no devueltos en su plazo), y devueltos (=1).

f: Es la frecuencia de cada combinación de circunstancias anteriores que se ha dado en la vida comercial de la sucursal en ese periodo.

15. Media geométrica, media armónica y desviación media.

Vamos a efectuar el análisis descriptivo por grupos (devueltos versus fallidos) introduciendo diversos estadísticos de tendencia central.

En particular estamos interesados en obtener la media geométrica, la media armónica y la desviación media de la variable interes.

**INSTRUCCIONES. PROGRAMA 2D.**

```
/input
case=11.
Var=4.
format=free.
```

```
/var
name=capital, interes, tipo, f.
freq=f.
```

Define los nombres de las variables y expresa que la variable F contiene en realidad las frecuencias.

```
/group
var=tipo.
codes(tipo)=0, 1.
names(tipo)=fallidos, devuelto.
```

Define el análisis por grupos, según la variable TIPO

```
/trans
il=log(interres).
```

Crea una variable IL que es el logaritmo decimal de la variable INTERES.

```
ia=1/interres.
```

Crea una variable IA que contendrá el inverso de cada valor de la variable INTERES.

```
dmi=abs(interres-11.2631578).
```

Crea una variable DMI que es la diferencia en valores absolutos de cada valor de INTERES a su media

```
/print estimate.
```

Solicita la media recortada al 15%, el estimador de Hampel y el Biweight.

```
/end
1 9 1 10
2 10 1 7
3 11 0 1
4 13 1 9
5 12 1 15
5 7 0 3
6 9 1 18
6 9 0 2
6 12 1 17
7 12 0 1
8 13 1 19
/end
```

**OUTPUT SELECCIONADO**

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES

CASE NO.	4 f	1 capital	2 interes	3 tipo	5 il	6 ia	7 dmi
1	10.00	1.00	9.00	devuelto	0.95	0.11	2.10
2	7.00	2.00	10.00	devuelto	1.00	0.10	1.10
3	1.00	3.00	11.00	fallidos	1.04	0.09	0.10
4	9.00	4.00	13.00	devuelto	1.11	0.08	1.90
5	15.00	5.00	12.00	devuelto	1.08	0.08	0.90
6	3.00	5.00	7.00	fallidos	0.85	0.14	4.10
7	18.00	6.00	9.00	devuelto	0.95	0.11	2.10
8	2.00	6.00	9.00	fallidos	0.95	0.11	2.10
9	17.00	6.00	12.00	devuelto	1.08	0.08	0.90
10	1.00	7.00	12.00	fallidos	1.08	0.08	0.90

NUMBER OF CASES READ. . . . . 11

GROUPING VARIABLE. . . tipo

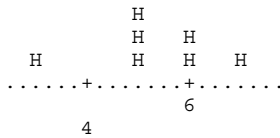
CATEGORY	FREQUENCY
fallidos	7
devuelto	95

Las frecuencias totales de fallidos y devueltos son -como puede esperarse- muy distintas.

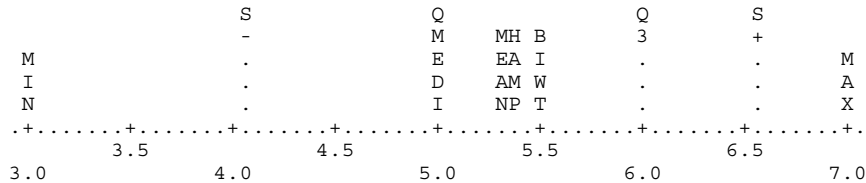
\*\*\*\*\*  
 \* ANALYSIS OF capital FOR GROUP fallidos \*  
 \*\*\*\*\*

VARIABLE NUMBER		MAX	MIN	VALUE	ZSCORE	CASE #
1		7.0000000	3.0000000		1.368	10
4					-1.823	3
7				VALUE	VALUE/S.E.	
0	SKEWNESS			-0.45	-0.489	
	KURTOSIS			-0.89	-0.481	

EACH 'H' REPRESENTS 1 RESPONSE.



	ESTIMATE	ST. ERROR		ESTIMATE
MEAN	5.2857141	0.4738036	ST. DEV.	1.2535664
MEDIAN	5.0000000	0.2886753	VARIANCE	1.5714287
MODE	5.0000000		RANGE	4.0000000
			(Q3-Q1)/2	0.5000000
TRIM(.15)	5.3979592			
HAMPEL	5.3833332		LOWER 95% C.L. OF MEAN	4.1266065
BWEIGHT	5.5241761		UPPER 95% C.L. OF MEAN	6.4448218
			Q1	5.0000000
			Q3	6.0000000
			S-	4.0321479
			S+	6.5392804

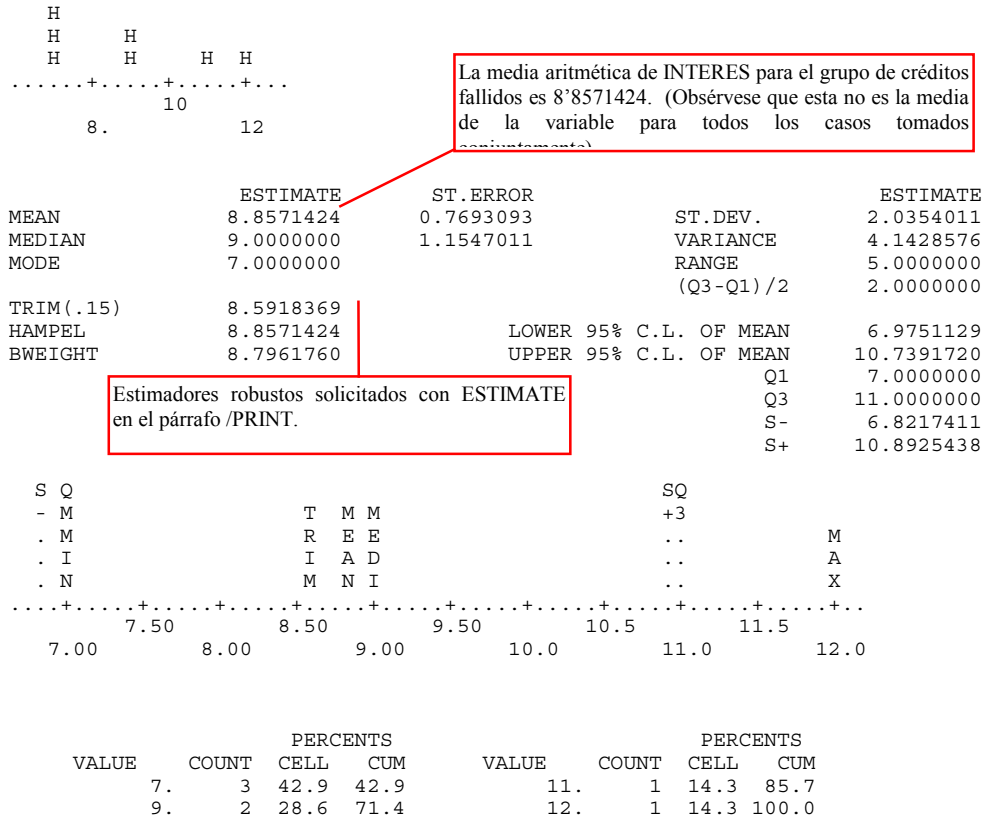


PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
3.	1	14.3	14.3	6.	2	28.6	85.7
5.	3	42.9	57.1	7.	1	14.3	100.0

\*\*\*\*\*  
 \* ANALYSIS OF interes FOR GROUP fallidos \*  
 \*\*\*\*\*

VARIABLE NUMBER		MAX	MIN	VALUE	ZSCORE	CASE #
2		12.0000000	7.0000000		1.544	10
4					-0.912	6
7				VALUE	VALUE/S.E.	
0	SKEWNESS			0.37	0.397	
	KURTOSIS			-1.72	-0.926	

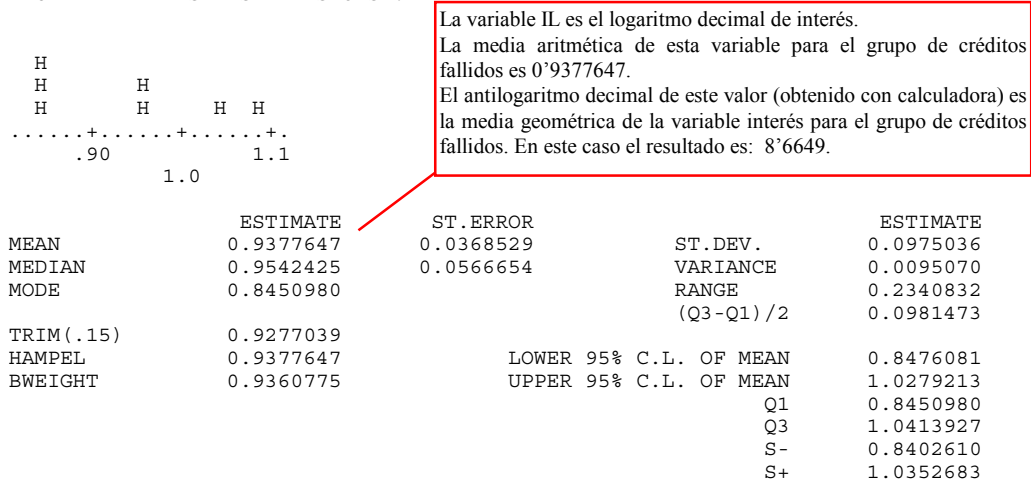
EACH 'H' REPRESENTS 1 RESPONSE.



\*\*\*\*\*  
 \* ANALYSIS OF il FOR GROUP fallidos \*  
 \*\*\*\*\*

	VALUE	ZSCORE	CASE #
VARIABLE NUMBER . . . . .	5	MAX 1.0791812	1.450 10
NUMBER OF DISTINCT VALUES .	4	MIN 0.8450980	-0.950 6
NUMBER OF VALUES COUNTED . .	7	VALUE	VALUE/S.E.
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS 0.24	0.260
		KURTOSIS -1.84	-0.991

EACH 'H' REPRESENTS 1 RESPONSE.



SQ	T	M	M	S	Q
-M				+ 3	
.M	R	E	E	.	M
.I	I	A	D	.	A
.N	M	N	I	.	X

PERCENTS	PERCENTS
CELL	CUM
.870	.930
.990	1.05

VALUE	COUNT	PERCENTS	VALUE	COUNT	PERCENTS
CELL	CUM	CELL	CUM	CELL	CUM
0.8451	3	42.9	1.0414	1	14.3
0.9542	2	28.6	1.0792	1	14.3
		71.4			100.0

\*\*\*\*\*  
 \* ANALYSIS OF ia FOR GROUP fallidos \*  
 \*\*\*\*\*

VARIABLE NUMBER	6	MAX	0.1428571	ZSCORE	0.982	CASE #	6
NUMBER OF DISTINCT VALUES	4	MIN	0.0833333		-1.357		10
NUMBER OF VALUES COUNTED	7			VALUE			
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS	-0.13	VALUE/S.E.	-0.136		
		KURTOSIS	-1.93		-1.045		

EACH 'H' REPRESENTS 1 RESPONSE.

H	H	H	H
.09			.15
	.12		

La variable IA contiene el recíproco de interés.  
 La media de IA (es decir, la media de los recíprocos de INTERES ) en el grupo de fallidos es 0'1178623.  
 El recíproco de este valor es la media armónica de INTERES para este grupo de fallidos. En este caso la media armónica de INTERES es 8'4845.

	ESTIMATE	ST. ERROR	ESTIMATE
MEAN	0.1178623	0.0096166	ST. DEV. 0.0254431
MEDIAN	0.1111111	0.0149961	VARIANCE 0.0006474
MODE	0.1428571		RANGE 0.0595238
			(Q3-Q1)/2 0.0259740
TRIM (.15)	0.1198280		
HAMPEL	0.1178623	LOWER 95% C.L. OF MEAN	0.0943364
BWEIGHT	0.1180701	UPPER 95% C.L. OF MEAN	0.1413882
		Q1	0.0909091
		Q3	0.1428571
		S-	0.0924192
		S+	0.1433054

QS	M	M	T	S
1-	E	E	R	M
.M	D	A	I	M
.I	I	N	M	A
.N				X

PERCENTS	PERCENTS
CELL	CUM
.085	.105
.095	.115
.105	.125
.115	.135
.125	.145

VALUE	COUNT	PERCENTS	VALUE	COUNT	PERCENTS
CELL	CUM	CELL	CUM	CELL	CUM
0.08333	1	14.3	0.11111	2	28.6
0.09091	1	14.3	0.14286	3	42.9
		28.6			100.0

```
*****
* ANALYSIS OF dmi      FOR GROUP fallidos *
*****
                VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 7      MAX      4.0980396  0.970    6
                MIN      0.0980396 -1.456    3
NUMBER OF DISTINCT VALUES . 4
NUMBER OF VALUES COUNTED . . 7
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  -0.18    -0.198
                KURTOSIS  -1.85    -1.000
```

EACH 'H' REPRESENTS 1 RESPONSE.

```

          H
        H H
      H H H
    +.....+.....+.....
      2
    0          4
```

La variable transformada DMI contiene los valores de distancia a la media de INTERES tomados en valores absolutos. La media aritmética de esos valores es precisamente la desviación media. Por tanto la desviación media para el grupo fallidos de la variable INTERES es 2'4985998.

```

                ESTIMATE      ST.ERROR      ESTIMATE
MEAN            2.4985998      0.6231156      ST.DEV.      1.6486089
MEDIAN          2.0980396      0.9226291      VARIANCE     2.7179115
MODE            4.0980396
                RANGE         4.0000000
                (Q3-Q1)/2     1.5980396
TRIM(.15)      2.6620650
HAMPEL         2.4985998      LOWER 95% C.L. OF MEAN 0.9742168
BWEIGHT        2.5189426      UPPER 95% C.L. OF MEAN 4.0229826
                Q1           0.9019604
                Q3           4.0980396
                S-           0.8499908
                S+           4.1472087
```

```

          SQ          M      M T          QS
        -1          E      E R          M+
M        ..          D      A I          M.
I        ..          I      N M          A.
N        ..          X.
+.....+.....+.....+.....+.....+.....+.....+.....+.....+.....+.....
    .50          1.5          2.5          3.5          4.5
0.0          1.0          2.0          3.0          4.0
```

```

                PERCENTS          PERCENTS
VALUE  COUNT  CELL  CUM      VALUE  COUNT  CELL  CUM
0.09804  1  14.3  14.3      2.09804  2  28.6  57.1
0.90196  1  14.3  28.6      4.09804  3  42.9  100.0
```

```
*****
* ANALYSIS OF capital  FOR GROUP devuelto *
*****
Comienzan los análisis para el grupo de créditos devueltos.
```

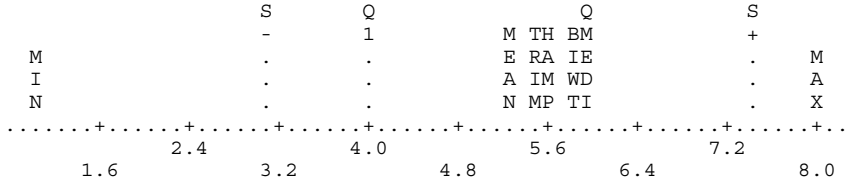
```

                VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 1      MAX      8.0000000  1.290    11
                MIN      1.0000000 -1.972    1
NUMBER OF DISTINCT VALUES . 6
NUMBER OF VALUES COUNTED . . 95
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  -0.63    -2.509
                KURTOSIS  -0.46    -0.914
'H' REPRESENTS      4 RESPONSES, '*' REPRESENTS LESS THAN 4 RESPONSES.
```

```

          *
          H
          H
          H
          H      *
        * H      H
      *   * H H  H
    H *   H H H  H
    H H   H H H  H
    +.....+.....+.....
      3
          6
```

	ESTIMATE	ST. ERROR		ESTIMATE
MEAN	5.2315788	0.2201764	ST. DEV.	2.1460137
MEDIAN	6.0000000	0.2886753	VARIANCE	4.6053748
MODE	6.0000000		RANGE	7.0000000
			(Q3-Q1)/2	1.0000000
TRIM (.15)	5.4812031			
HAMPEL	5.5712500	LOWER 95% C.L. OF MEAN		4.7944136
BWEIGHT	5.8842535	UPPER 95% C.L. OF MEAN		5.6687441
			Q1	4.0000000
			Q3	6.0000000
			S-	3.0855651
			S+	7.3775926

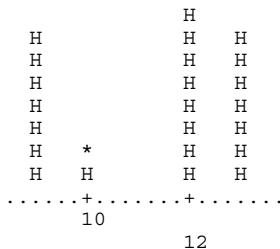


PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	10	10.5	10.5	5.	15	15.8	43.2
2.	7	7.4	17.9	6.	35	36.8	80.0
4.	9	9.5	27.4	8.	19	20.0	100.0

\*\*\*\*\*  
 \* ANALYSIS OF interes FOR GROUP devuelto \*  
 \*\*\*\*\*

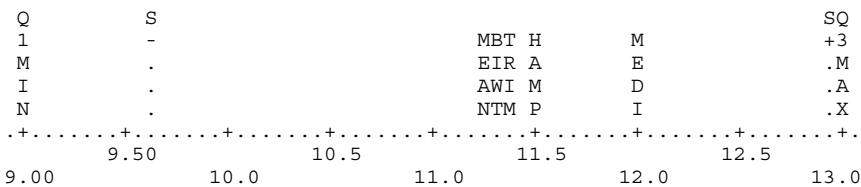
VARIABLE NUMBER		MAX	VALUE	ZSCORE	CASE #
2		13.0000000	13.0000000	1.052	4
4		9.0000000	9.0000000	-1.370	1
95					
0					
		SKEWNESS	VALUE	VALUE/S.E.	
		-0.42	-0.42	-1.666	
		KURTOSIS	-1.56	-3.108	

'H' REPRESENTS 4 RESPONSES, '\*' REPRESENTS LESS THAN 4 RESPONSES.



Estadísticos de tendencia central de la variable INTERES para el grupo devueltos.

	ESTIMATE	ST. ERROR		ESTIMATE
MEAN	11.2631578	0.1694537	ST. DEV.	1.6516305
MEDIAN	12.0000000	0.0000000	VARIANCE	2.7278833
MODE	12.0000000		RANGE	4.0000000
			(Q3-Q1)/2	2.0000000
TRIM (.15)	11.3759403			
HAMPEL	11.4985075	LOWER 95% C.L. OF MEAN		10.9267035
BWEIGHT	11.3046989	UPPER 95% C.L. OF MEAN		11.5996122
			Q1	9.0000000
			Q3	13.0000000
			S-	9.6115274
			S+	12.9147882



PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
9.	28	29.5	29.5	12.	32	33.7	70.5
10.	7	7.4	36.8	13.	28	29.5	100.0

\*\*\*\*\*  
 \* ANALYSIS OF il FOR GROUP devuelto \*  
 \*\*\*\*\*

VARIABLE NUMBER	MAX	VALUE	ZSCORE	CASE #
5	1.1139433	1.1139433	1.011	4
4	0.9542425	0.9542425	-1.393	1

NUMBER OF DISTINCT VALUES	VALUE	VALUE/S.E.
4	-0.48	-1.902
95	-1.55	-3.077

NUMBER OF VALUES COUNTED . . . . . 95  
 NUMBER OF VALUES NOT COUNTED . . . . . 0  
 SKEWNESS . . . . . -0.48  
 KURTOSIS . . . . . -1.55

'H' REPRESENTS 4 RESPONSES, '\*' REPRESENTS LESS THAN 4 RESPONSES.

H	H	H
H	H	H
H	H	H
H	H	H
H	H	H
H	H	H
H	*	H
H	H	H
H	H	H

ESTIMATE	ST. ERROR	ESTIMATE
MEAN 1.0467684	0.0068170	0.0664437
MEDIAN 1.0791812	0.0000000	0.0044148
MODE 1.0791812		0.1597008
		(Q3-Q1)/2 0.0798504
TRIM(.15) 1.0522008		LOWER 95% C.L. OF MEAN 1.0332332
HAMPEL 1.0609311		UPPER 95% C.L. OF MEAN 1.0603037
BWEIGHT 1.0488955		Q1 0.9542425
		Q3 1.1139433
		S- 0.9803247
		S+ 1.1132121

La media geométrica de la variable INTERES para el grupo de créditos devueltos es 11'137

Q	S	MBT	H	M	SQ
1	-	EIR	A	E	+3
M	.	AWI	M	D	.M
I	.	NTM	P	I	.A
N	.				.X

ESTIMATE	ST. ERROR	ESTIMATE
MEAN 1.0467684	0.0068170	0.0664437
MEDIAN 1.0791812	0.0000000	0.0044148
MODE 1.0791812		0.1597008
		(Q3-Q1)/2 0.0798504
TRIM(.15) 1.0522008		LOWER 95% C.L. OF MEAN 1.0332332
HAMPEL 1.0609311		UPPER 95% C.L. OF MEAN 1.0603037
BWEIGHT 1.0488955		Q1 0.9542425
		Q3 1.1139433
		S- 0.9803247
		S+ 1.1132121

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
0.9542	28	29.5	29.5	1.0792	32	33.7	70.5
1.0000	7	7.4	36.8	1.1139	28	29.5	100.0

```

*****
* ANALYSIS OF ia          FOR GROUP devuelto *
*****
          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 6      MAX      0.1111111  1.413    1
          MIN      0.0769231 -0.972    4
NUMBER OF DISTINCT VALUES . 4
NUMBER OF VALUES COUNTED. . 95          VALUE      VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  0.53      2.111
          KURTOSIS -1.53      -3.035
'H' REPRESENTS      4 RESPONSES, '*' REPRESENTS LESS THAN 4 RESPONSES.

      H
H   H           H
H   H           H
H   H           H
H   H           H
H   H           H
H   H           *   H
H   H           H   H
...+...+...+...+...+...
      .09      .11
      .08      .10

          ESTIMATE      ST. ERROR      ESTIMATE
MEAN      0.0908592      0.0014709      ST. DEV.      0.0143361
MEDIAN    0.0833333      0.0000000      VARIANCE      0.0002055
MODE      0.0833333      RANGE          0.0341880
          (Q3-Q1) / 2      0.0170940
TRIM (.15) 0.0895058
HAMPEL    0.0862165      LOWER 95% C.L. OF MEAN 0.0879388
BWEIGHT   0.0902898      UPPER 95% C.L. OF MEAN 0.0937796
          Q1          0.0769231
          Q3          0.1111111
          S-          0.0765231
          S+          0.1051953

      SQ          S          Q
-1      M   H   TBM      +   3
.M      E   A   RIE      .   M
.I      D   M   IWA      .   A
.N      I   P   MTN      .   X
...+...+...+...+...+...+...+...+...+...
      .076      .084      .088      .092      .096      .100      .104      .108      .112

          PERCENTS          PERCENTS
          VALUE  COUNT  CELL  CUM      VALUE  COUNT  CELL  CUM
0.07692  28  29.5  29.5      0.10000  7  7.4  70.5
0.08333  32  33.7  63.2      0.11111  28  29.5  100.0
    
```

Por tanto, la media armónica de INTERES para el grupo de créditos devueltos es 11'006

```

*****
* ANALYSIS OF dmi        FOR GROUP devuelto *
*****
          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 7      MAX      2.0980396  1.002    1
          MIN      0.9019604 -1.241    5
NUMBER OF DISTINCT VALUES . 4
NUMBER OF VALUES COUNTED. . 95          VALUE      VALUE/S.E.
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS -0.32      -1.277
          KURTOSIS -1.81      -3.608
    
```



'H' REPRESENTS 4 RESPONSES, '\*' REPRESENTS LESS THAN 4 RESPONSES.

```
H
H          H H
H          H H
H          H H
H          H H
H          H H
H *        H H
H H        H H
...+.....+.....+.....
      1.5
1.0          2.0
```

	ESTIMATE	ST.ERROR		ESTIMATE
MEAN	1.5636739	0.0547098	ST.DEV.	0.5332452
MEDIAN	1.9019604	0.0000000	VARIANCE	0.2843505
MODE	0.9019604		RANGE	1.1960793
			(Q3-Q1)/2	0.5980396
TRIM(.15)	1.5909625			
HAMPEL	1.8174474	LOWER 95% C.L. OF MEAN		1.4550462
BWEIGHT	1.9999988	UPPER 95% C.L. OF MEAN		1.6723015
			Q1	0.9019604
			Q3	2.0980396
			S-	1.0304286
			S+	2.0969191

```
Q      S
M      -           M T           H M B +
M      .           E R           A E I M
I      .           A I           M D W A
N      .           N M           P I T X
...+.....+.....+.....+.....+.....+.....+.....+.....+.....
.90          1.1          1.3          1.5          1.7          1.9          2.1
      1.0          1.2          1.4          1.6          1.8          2.0
```

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
0.9020	32	33.7	33.7	1.9020	28	29.5	70.5
1.0980	7	7.4	41.1	2.0980	28	29.5	100.0

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES  
PROGRAM TERMINATED

16. Medias cuadráticas y medias de Foster.

En el ejemplo anterior hemos visto como obtener medias geométricas y medias armónicas utilizando variables transformadas. De modo análogo podemos obtener una media cuadrática, o, en general, cualquier media de Foster.

Supongamos que en el ejemplo anterior estuviésemos interesados en obtener la media cuadrática de la variable intereses, para ello en el párrafo /TRANS crearíamos una transformada, que vamos a llamar i2, equivalente a la variable interes elevada al cuadrado, introduciendo la siguiente línea:

i2 = interes\*\*2.

El programa (el 2D o incluso el 1D) nos proporcionaría la media aritmética de i2. La raíz cuadrada (obtenida con calculadora) de esa media aritmética será la media cuadrática de la variable interes.

El procedimiento puede generalizarse para cualquier media de Foster. Basta cambiar el exponente a que se eleva y hace después la raíz correspondiente. Por ejemplo, para una media cúbica:

i3 = interes\*\*3.

Se obtiene la media aritmética de i3 y se calcula la raíz cúbica de esa media, que será precisamente la media cúbica de la variable interes.

17. Desviación media a la mediana.

También hemos visto en el ejemplo anterior como una transformada dmi permitía calcular las desviaciones a la media aritmética (previamente calculada) de la variable interes. Bastaba con calcular la media aritmética de la transformada para obtener la desviación media.

El mismo procedimiento puede utilizarse para calcular la desviación media a la mediana. Supongamos que deseamos obtener la mediana de una variable llamada X. Primero, mediante 2D obtenemos la mediana de la variable. Segundo, creamos una transformada que recoja las diferencias de los valores a la mediana en valores absolutos, por ejemplo:

```
dmnx = abs(X - valor numérico de la mediana de X).
```

Por supuesto en la expresión anterior del párrafo /TRANS "valor numérico de la mediana de X" es un dato concreto que hemos obtenido previamente mediante una descripción de X con el programa 2D.

La media aritmética de dmnx será la desviación media a la mediana.

La desviación media a la mediana es un estadístico de dispersión de muy raro uso, pero permite ilustrar como las transformaciones pueden ser utilizadas para obtener estadísticos poco habituales.

18. Test de Normalidad W de Shapiro y Wilks.

En muchas ocasiones es útil poder conocer si la distribución de puntuaciones de una variable difiere significativamente de una distribución normal. Para averiguar esto puede utilizarse un test de normalidad. El programa 2D permite contrastar si una variable difiere significativamente de una distribución normal mediante el estadístico W de Shapiro y Wilks.

Para que se aprecie el modo en que el test de normalidad actúa he elaborado 6 variables cuyas distribuciones presentan una distribución con una forma característica. también se solicitan los estadísticos robustos para verificar su comportamiento sobre estos tipos de distribuciones.

INSTRUCCIONES. PROGRAMA 1D.
-----------------------------

```
/input var=6. format=free.
/print
```

WSTAT.
--------

El estadístico de contraste de la normalidad W se solicita en el párrafo print con esta instrucción.
--

ESTIMATE.
-----------

Solicitamos los estadísticos robustos.
--

```
/end
1 1 1 1 2 1
1 2 2 2 3 2
2 3 2 3 3 2
2 4 2 3 3 1
2 5 2 4 4 3
2 6 2 4 4 3
2 7 3 4 4 2
2 8 4 5 4 4
2 8 5 5 5 4
2 8 6 5 5 5
2 8 7 5 5 5
2 8 8 5 5 5
2 8 8 5 5 1
3 9 8 6 6 8
4 9 8 6 6 7
5 9 8 6 6 6
6 9 8 7 6 9
7 9 2 7 7 7
8 9 2 8 7 6
9 9 8 9 8 3
/end
```

OUTPUT SELECCIONADO

BMDP2D - DETAILED DATA DESCRIPTION INCLUDING FREQUENCIES

CASE NO.	1	2	3	4	5	6
	A1	A2	A3	A4	A5	A6
1	1.00	1.00	1.00	1.00	2.00	1.00
2	1.00	2.00	2.00	2.00	3.00	2.00
3	2.00	3.00	2.00	3.00	3.00	2.00
4	2.00	4.00	2.00	3.00	3.00	1.00
5	2.00	5.00	2.00	4.00	4.00	3.00
6	2.00	6.00	2.00	4.00	4.00	3.00
7	2.00	7.00	3.00	4.00	4.00	2.00
8	2.00	8.00	4.00	5.00	4.00	4.00
9	2.00	8.00	5.00	5.00	5.00	4.00
10	2.00	8.00	6.00	5.00	5.00	5.00

De los 20 casos, como de costumbre solo muestra como opción preestablecida, los datos de

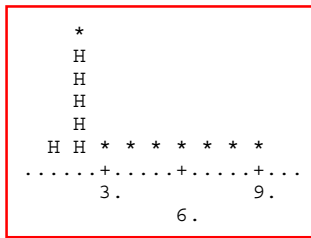
NUMBER OF CASES READ. . . . . 20

\*\*\*\*\*  
\* A1 \*  
\*\*\*\*\*

VARIABLE NUMBER		MAX	VALUE	ZSCORE	CASE #
	1	9.0000000	9.0000000	2.389	20
	9	1.0000000	1.0000000	-0.964	1
NUMBER OF DISTINCT VALUES	9				
NUMBER OF VALUES COUNTED	20				
NUMBER OF VALUES NOT COUNTED	0				

	VALUE	VALUE/S.E.
SKEWNESS	1.16	2.123
KURTOSIS	-0.10	-0.092

'H' REPRESENTS 2 RESPONSES, '\*' REPRESENTS LESS THAN 2 RESPONSES.



Como puede apreciarse en el histograma, ésta es una variable con una fuerte asimetría, muy sesgada a la derecha. Su forma es patentemente distinta del de una distribución normal. La media aritmética es el estadístico más sensible a los valores alejados por la cola derecha. Es interesante ver el orden en que se ubican los estadísticos para una distribución como ésta. El estadístico W es significativo (es decir, el valor en su nivel de significación SIGNIFICANCE LEVEL es menor o igual a 0.05), lo que significa que la distribución difiere

	ESTIMATE	ST. ERROR	ESTIMATE
MEAN	3.3000000	0.5336074	2.3863647
MEDIAN	2.0000000	0.2886753	5.6947365
MODE	2.0000000		8.0000000
TRIM(.15)	2.7142856		1.3750000
HAMPEL	2.0000000		
BWEIGHT	2.0000000		

TEST OF NORMALITY  
W STATISTIC 0.7519  
SIGNIFICANCE LEVEL 0.0001

	ESTIMATE
LOWER 95% C.L. OF MEAN	2.1831481
UPPER 95% C.L. OF MEAN	4.4168515
Q1	2.0000000
Q3	4.7500000
S-	0.9136353
S+	5.6863647

S	Q	T	M	Q	S
-	M			3	+
.M	E	R	E	.	.
.I	D	I	A	.	.
.N	I	M	N	.	.

VALUE	COUNT	PERCENTS CELL	PERCENTS CUM
1.	2	10.0	10.0
2.	11	55.0	65.0
3.	1	5.0	70.0
4.	1	5.0	75.0
5.	1	5.0	80.0

```

*****
* A2 *
*****

          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 2      MAX      9.0000000    0.799    14
          MIN      1.0000000   -2.320     1
NUMBER OF DISTINCT VALUES . . . . . 9
NUMBER OF VALUES COUNTED . . . . . 20
NUMBER OF VALUES NOT COUNTED . . . . . 0
          SKEWNESS  -1.07      -1.950
          KURTOSIS  -0.32      -0.291

EACH 'H' REPRESENTS 1 RESPONSE.
    
```

```

          H
        H H
        H H
        H H
        H H
        H H
        H H
    H H H H H H H H
    .....+.....+.....+.....
        3.          9.
          6.
    
```

Esta variable es también marcadamente asimétrica, y por tanto esta lejos de describir una curva normal en su histograma. Obsérvese como se ubican los estadísticos sobre la distribución. El test de normalidad es significativo, (Nivel de significación menor o igual a 0'05) y por tanto se dice que la distribución difiere significativamente de una distribución normal.

MEAN	ESTIMATE	ST. ERROR	ST. DEV.	ESTIMATE
MEDIAN	8.0000000	0.5773506	VARIANCE	6.5763159
MODE	9.0000000		RANGE	8.0000000
			(Q3-Q1) / 2	1.8750000
TRIM (.15)	7.5714288		LOWER 95% C.L. OF MEAN	5.7498097
HAMPEL	7.8894739		UPPER 95% C.L. OF MEAN	8.1501904
BWEIGHT	8.3148451		Q1	5.2500000
			Q3	9.0000000
			S-	4.3855667
			S+	9.5144329

TEST OF NORMALITY  
W STATISTIC 0.7824  
SIGNIFICANCE LEVEL 0.0003

```

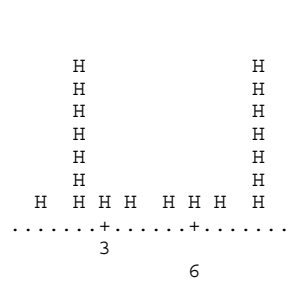
          S      Q
          -      1
M          .      .      M      T      H M B      Q      S
I          .      .      E      R      A E I      M      .
N          .      .      A      I      M D W      A      .
          .      .      N      M      P I T      X      .
    .....+.....+.....+.....+.....+.....+.....+.....
        1          2          3          4          5          6          7          8          9
    
```

PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	1	5.0	5.0	6.	1	5.0	30.0
2.	1	5.0	10.0	7.	1	5.0	35.0
3.	1	5.0	15.0	8.	6	30.0	65.0
4.	1	5.0	20.0	9.	7	35.0	100.0
5.	1	5.0	25.0				

\*\*\*\*\*  
\* A3 \*  
\*\*\*\*\*

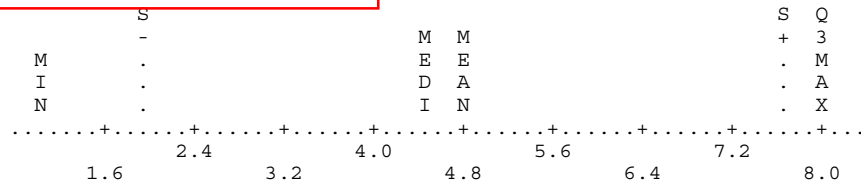
			VALUE	ZSCORE	CASE #
		MAX	8.0000000	1.134	12
		MIN	1.0000000	-1.347	1
VARIABLE NUMBER . . . . .	3				
NUMBER OF DISTINCT VALUES .	8				
NUMBER OF VALUES COUNTED . .	20		VALUE	VALUE/S.E.	
NUMBER OF VALUES NOT COUNTED	0	SKEWNESS	0.06	0.109	
		KURTOSIS	-1.89	-1.723	

EACH 'H' REPRESENTS 1 RESPONSE.



La variable A3 es casi simétrica y esto se refleja en los estadísticos de tendencia central que quedan todos muy próximos entre si (robustos y no robustos). Sin embargo, la distribución presenta una forma bimodal de U muy alejada de una distribución normal. El test W refleja perfectamente este hecho. W es siempre positivo, con un máximo de 1. Si la distribución es normal W es mayor de 0'9 y ese valor tiende a crecer con un N muestral mayor. Un valor de W por debajo de 0'9 muestra alejamiento de la normalidad. En este caso puede rechazarse la hipótesis nula de normalidad al nivel de significación del 0'001 (el nivel de significación exacto es 0'0004).

MEAN	ESTIMATE	ST. ERROR	ST. DEV.	ESTIMATE
MEDIAN	4.8000002	0.6307890	VARIANCE	2.8209741
MODE	4.5000000	1.7320516	RANGE	7.9578948
	NOT UNIQUE		(Q3-Q1) / 2	7.0000000
TRIM (.15)	4.7857141			3.0000000
HAMPEL	4.8000002	LOWER 95% C.L. OF MEAN		3.4797451
BWEIGHT	4.7888970	UPPER 95% C.L. OF MEAN		6.1202550
TEST OF NORMALITY		Q1		2.0000000
W STATISTIC	0.7917	Q3		8.0000000
SIGNIFICANCE LEVEL	0.0004	S-		1.9790261
		S+		7.6209745



PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	1	5.0	5.0	5.	1	5.0	55.0
2.	7	35.0	40.0	6.	1	5.0	60.0
3.	1	5.0	45.0	7.	1	5.0	65.0
4.	1	5.0	50.0	8.	7	35.0	100.0

```

*****
* A4 *
*****

          VALUE   ZSCORE  CASE #
          MAX     9.000000  2.055   20
VARIABLE NUMBER . . . . . 4     MIN     1.000000  -2.055   1
NUMBER OF DISTINCT VALUES .
NUMBER OF VALUES COUNTED . . 20
NUMBER OF VALUES NOT COUNTED 0     SKEWNESS  0.00    0.000
          KURTOSIS  -0.41   -0.374

EACH 'H' REPRESENTS 1 RESPONSE.
    
```

```

      H
      H
      H
    H H H
  H H H H H
H H H H H H H H
.....+.....+.....+.....
      3.          9.
          6.
    
```

Aquí tenemos una distribución simétrica, cuya forma semeja la de una distribución normal (aunque no es exactamente una distribución normal, y difícilmente puede serlo exactamente debido a los pocos casos y a la naturaleza discreta de las puntuaciones de la variable, lo que impone saltos bruscos en el histograma). Los estadísticos reflejan bien la forma de la distribución y el test W no es significativo, lo que permite afirmar que esta distribución no difiere significativamente de una distribución normal (lo que no prueba que sea exactamente una distribución normal, pero permitiría tratarla razonablemente como si fuera una distribución

	ESTIMATE	ST. ERROR	ST. DEV.	ESTIMATE
MEAN	5.0000000	0.4352857		1.9466571
MEDIAN	5.0000000	0.5773506	VARIANCE	3.7894738
MODE	5.0000000		RANGE	8.0000000
			(Q3-Q1)/2	1.0000000
TRIM (.15)	5.0000000		LOWER 95% C.L. OF MEAN	4.0889378
HAMPEL	5.0000000		UPPER 95% C.L. OF MEAN	5.9110627
BWEIGHT	4.9999995		Q1	4.0000000
TEST OF NORMALITY			Q3	6.0000000
W STATISTIC	0.9750		S-	3.0533428
SIGNIFICANCE LEVEL	0.8409		S+	6.9466572

```

      S      Q      M      Q      S
      -      1      .      3      +
M      .      .      E      .      .
I      .      .      A      .      .
N      .      .      N      .      .
.....+.....+.....+.....+.....+.....+.....+.....+.....+.....
      1          3          5          7          9
          2          4          6          8
    
```

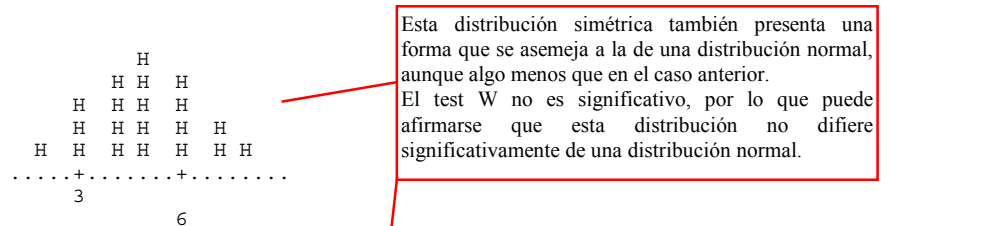
PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	1	5.0	5.0	6.	3	15.0	80.0
2.	1	5.0	10.0	7.	2	10.0	90.0
3.	2	10.0	20.0	8.	1	5.0	95.0
4.	3	15.0	35.0	9.	1	5.0	100.0
5.	6	30.0	65.0				

```

*****
* A5 *
*****

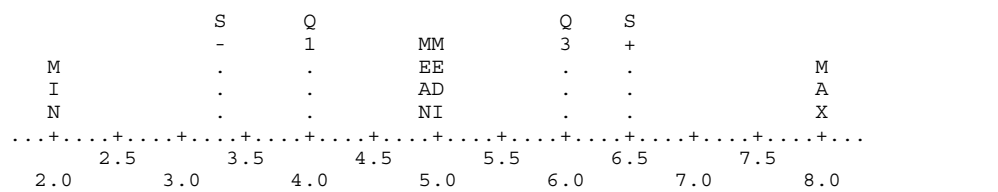
          VALUE      ZSCORE  CASE #
VARIABLE NUMBER . . . . . 5      MAX      8.0000000  1.997  20
          MIN      2.0000000 -1.868   1
NUMBER OF DISTINCT VALUES . 7
NUMBER OF VALUES COUNTED . 20
NUMBER OF VALUES NOT COUNTED 0      SKEWNESS  0.08    0.140
          KURTOSIS -0.85    -0.778

EACH 'H' REPRESENTS 1 RESPONSE.
    
```



Esta distribución simétrica también presenta una forma que se asemeja a la de una distribución normal, aunque algo menos que en el caso anterior. El test W no es significativo, por lo que puede afirmarse que esta distribución no difiere significativamente de una distribución normal.

	ESTIMATE	ST. ERROR	ST. DEV.	ESTIMATE
MEAN	4.9000001	0.3471690		1.5525869
MEDIAN	5.0000000	0.5773506	VARIANCE	2.4105260
MODE	5.0000000		RANGE	6.0000000
			(Q3-Q1) / 2	1.0000000
TRIM (.15)	4.8571429		LOWER 95% C.L. OF MEAN	4.1733680
HAMPEL	4.8692307		UPPER 95% C.L. OF MEAN	5.6266322
BWEIGHT	4.8854132		Q1	4.0000000
TEST OF NORMALITY			Q3	6.0000000
W STATISTIC	0.9666		S-	3.3474131
SIGNIFICANCE LEVEL	0.6778		S+	6.4525871



PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
2.	1	5.0	5.0	6.	4	20.0	85.0
3.	3	15.0	20.0	7.	2	10.0	95.0
4.	4	20.0	40.0	8.	1	5.0	100.0
5.	5	25.0	65.0				

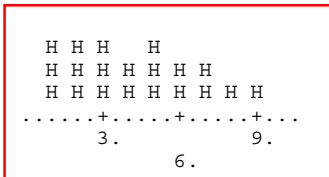
```

*****
* A6 *
*****

          VALUE   ZSCORE  CASE #
VARIABLE NUMBER . . . . . 6     MAX     9.0000000  1.984   17
          MIN     1.0000000 -1.323   1
NUMBER OF DISTINCT VALUES . 9
NUMBER OF VALUES COUNTED . 20
NUMBER OF VALUES NOT COUNTED 0

          VALUE   VALUE/S.E.
SKEWNESS    0.31    0.562
KURTOSIS    -1.14   -1.039
    
```

EACH 'H' REPRESENTS 1 RESPONSE.



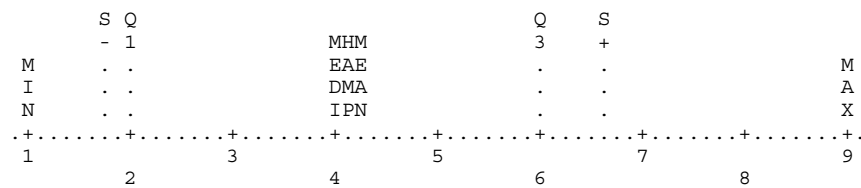
En los dos casos anteriores teníamos distribuciones que intencionadamente se asemejaban considerablemente a una normal. Aquí tenemos una distribución platocúrtica y ligeramente asimétrica cuyo histograma no semeja aparentemente el de una curva normal. Sin embargo, la diferencia entre esta distribución y una normal no es tan fuerte que podamos rechazar la hipótesis nula de normalidad. El test W no es significativo, de modo que la distribución no difiere significativamente de una normal. El ejemplo ha sido escogido cuidadosamente para que se pueda apreciar que la hipótesis de normalidad no pueda rechazarse en muchos casos en que

```

          ESTIMATE   ST. ERROR   ST. DEV.   ESTIMATE
MEAN      4.1999998  0.5409543  2.4192212
MEDIAN    4.0000000  0.5773506  5.8526311
MODE      NOT UNIQUE
          RANGE      (Q3-Q1)/2    8.0000000
          2.0000000
TRIM(.15) 4.0714288
HAMPEL    4.0999999
BWEIGHT   4.1306696
          LOWER 95% C.L. OF MEAN  3.0677707
          UPPER 95% C.L. OF MEAN  5.3322287
          Q1          2.0000000
          Q3          6.0000000
          S-         1.7807786
          S+         6.6192207
    
```

```

TEST OF NORMALITY
W STATISTIC    0.9444
SIGNIFICANCE LEVEL 0.3016
    
```



PERCENTS				PERCENTS			
VALUE	COUNT	CELL	CUM	VALUE	COUNT	CELL	CUM
1.	3	15.0	15.0	6.	2	10.0	80.0
2.	3	15.0	30.0	7.	2	10.0	90.0
3.	3	15.0	45.0	8.	1	5.0	95.0
4.	2	10.0	55.0	9.	1	5.0	100.0
5.	3	15.0	70.0				

END OF INSTRUCTIONS  
PROGRAM TERMINATED



### 19. Transformaciones no lineales.

Algunas técnicas estadísticas requieren que las variables analizadas (especialmente la variable o variables dependientes) cumplan determinados supuestos, entre ellos, la normalidad. Esto es una dificultad con distribuciones particularmente sesgadas, como algunas de las vistas anteriormente.

En ocasiones, para tratar de resolver esta dificultad se recurre a transformaciones no-lineales que alteran la forma de la distribución.

A continuación se presentan algunas de las más habituales y el modo de escribirlas en el párrafo /TRANS siendo X la variable original a transformar y Z la variable resultado de la transformación de X:

#### **Transformaciones basadas en la raíz cuadrada.**

Se utilizan para datos frecuenciales que sigan aproximadamente una distribución de Poisson.

Raíz cuadrada de la variable:

$$Z = \text{SQRT}(X).$$

Con valores pequeños de X puede ensayarse:

$$Z = \text{SQRT}(X + 0.5).$$

Las transformaciones basadas en la raíz cuadrada y las basadas en logaritmos son las más utilizadas para tratar de transformar en simétricas las distribuciones asimétricas sesgadas a la derecha (es decir con una cola izquierda gruesa y una cola derecha delgada o inusualmente larga).

#### **Transformaciones basadas en el Arcoseno.**

Se utiliza especialmente para datos negativamente sesgados y datos en forma de proporciones. Se puede ensayar el arcoseno de la variable X,

$$Z = \text{ASIN}(X).$$

el arcoseno de la raíz de la variable X,

$$Z = \text{ASIN}(\text{SQRT}(X)).$$

ó, para valores de X muy pequeños el arcoseno de la raíz de la (variable X más N/2):

$$Z = \text{ASIN}(\text{SQRT}(X + N/2)). \text{ (Sustituyendo N por el número de casos del análisis).}$$

#### **Transformaciones logarítmicas.**

Se utilizan para datos asimétricos positivamente sesgados.

Logaritmo decimal de la variable:

$$Z = \text{LOG}(X).$$

Logaritmo de la variable más 1:

$$Z = \text{LOG}(X + 1).$$

#### **Transformaciones basadas en recíprocos.**

Para datos basados en el tiempo:

$$Z = 1/X.$$

Para valores pequeños de X puede ensayarse:

$$Z = 1/(X + 1).$$

La aplicación de una transformación no garantiza que la distribución adopte una forma aceptable para un test estadístico determinado. Frecuentemente es necesario ensayar diversas transformaciones antes de hallar aquella que parece más adecuada para unos datos.

## 20. Redondear y truncar los datos de las variables.

En ocasiones si una variable continua tiene muchos valores diferentes y el número de casos es alto la descripción frecuencial del programa 2D se vuelve extraordinariamente larga y puede ser poco práctica. Para resolver esto en el programa 2D puede redondearse y truncarse los valores de la variable.

Supongamos que tenemos casos con los siguientes datos en una variable X:  
345, 389, 297. 303.

Si especificamos:

```
/COUNT ROUND=(X)5.
```

Los casos anteriores se redondean al valor más próximo divisible por 5. Lo que supone que son tratados respectivamente como:

345, 390, 295, 305.

Si hubiéramos escrito:

```
/COUNT ROUND=(X)100.
```

Los casos anteriores se redondean al valor más próximo divisible por 100. Lo que supone que son tratados respectivamente como:

300, 400, 300, 300.

Se puede redondear a cualquier valor que se desee.

Supongamos que tenemos los mismos casos con los mismos datos en una variable X:  
345, 389, 297. 303.

Si especificamos:

```
/COUNT TRUNCATE=(X)100.
```

Los casos anteriores se truncan al valor más próximo más bajo divisible por 100. Lo que supone que son tratados respectivamente como:

300, 300, 200, 300.

Puede truncarse a 10, 100 y cualquier valor que se desee. En el ejemplo puede verse la diferencia entre redondear (lo asimila al más próximo divisible por esa unidad de redondeo) y truncar (lo asimila al más próximo por abajo divisible por esa unidad de truncamiento).

Debe advertirse que redondear o truncar actúa tratando los datos en esos intervalos y esto afecta a todos los estadísticos calculados por 2D. Sin embargo, los datos no por eso quedan almacenados en esos intervalos, y por tanto otros programas que no pueden utilizar el párrafo /COUNT no podrán tratar los datos agrupados en intervalos según este procedimiento. Si se desea esto último puede utilizarse un variable transformada mediante IF ... THEN que puede usarse a través de todos los programas. No obstante debe recordarse que, salvo fines de simplificación de la descripción, en general no es una estrategia recomendable tratar los datos agrupados en intervalos.

21. Gráficos de "tallos y hojas" (Steam and leaf).

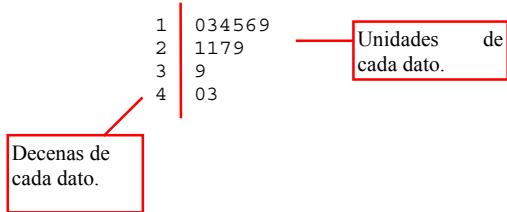
La visión sintética que produce un truncamiento en la tabulación puede conseguirse de forma visual, sin necesidad de utilizar intervalos, mediante un diagrama de tallo y hojas.

Un diagrama de tallo y hojas es un histograma (o, más exactamente un diagrama de barras horizontales) en que los últimos dígitos de los datos se utilizan para formar las barras del diagrama, reservándose el primero o primeros para expresar el eje.

Sean los datos:

10 13 14 15 16 19 21 21 27 29 32 26 26 39 40 43

Representados en un diagrama de tallo y hojas tendrían la siguiente apariencia:



El programa 2D puede obtener este tipo de gráficos añadiendo la instrucción:

/PRINT STEM.

El formato que 2D utiliza para el gráfico de tallo y hojas puede ser ligeramente distinto según el rango de la variable y según la frecuencia de casos de cada tipo. En particular cuando hay muchos casos con los mismos datos el formato anterior (que es el clásico de un steam and leaf) no es viable porque necesitaría líneas muy largas. En ese caso 2D sitúa en el eje de abcisas las unidades y el en interior del diagrama muestra las frecuencias de cada combinación de decenas y unidades (de modo semejante a una tabla de contingencia). En ese último caso el diagrama pierde parte de poder intuitivo para mostrar los datos, aunque sigue ayudando a comprender la estructura de la distribución.

**Tablas de Correlación y de Contingencia.**

**INSTRUCCIONES:**

**#Tablas de Correlación y de Contingencia. Programa 4F.**

```

 /input
    case=15. Var=2. format=free.
 /var
    names=descuent, ventas.
 /table
    col=descuent. # Especifica la variable en las columnas de la tabla.
    row=ventas. # Especifica la variable en las filas de la tabla.
 /print
    perc=row, col, tot. #Solicita tablas de porcentajes, por filas(row),
    #por columnas(col) y por el total N (tot).
/end
5 1
3 1
3 2
5 5
5 5
5 6
6 7
6 8
6 7
4 3
4 3
5 1
4 4
3 2
5 3
/end
    
```

Este signo indica que comienza un comentario. Todo lo que hay entre # y el final de la línea es ignorado por el BMDP.

col=descuent.  
row=ventas.

# Especifica la variable en las columnas de la tabla.  
# Especifica la variable en las filas de la tabla.

perc=row, col, tot.

#Solicita tablas de porcentajes, por filas(row),  
#por columnas(col) y por el total N (tot).

Opcional si se desea estas tablas, se puede incluir uno o varios tipos.

**RESULTADOS:**

BMDP Program Output File: D:\EBMDP\E121\_1.OUT  
BMDP4F - TWO-WAY AND MULTIWAY FREQUENCY TABLES

NUMBER OF CASES READ. . . . . 15

\*\*\* OBSERVED FREQUENCY (Tabla de Frecuencias)

ventas	descuent				TOTAL
	3	4	5	6	
1	1	0	2	0	3
2	2	0	0	0	2
3	0	2	1	0	3
4	0	1	0	0	1
5	0	0	2	0	2
6	0	0	1	0	1
7	0	0	0	2	2
8	0	0	0	1	1
TOTAL	3	3	6	3	15

Distribución condicionada. Distribución de la variable Ventas condicionada a que descuent = 4

Distribución marginal de la variable ventas

Distribución marginal de la variable descuent

\*\*\* PERCENTS OF ROW TOTALS (Porcentajes sobre el total de cada fila)

ventas	descuent				
	3	4	5	6	TOTAL
1	33.3	0.0	66.7	0.0	100.0
2	100.0	0.0	0.0	0.0	100.0
3	0.0	66.7	33.3	0.0	100.0
4	0.0	100.0	0.0	0.0	100.0
5	0.0	0.0	100.0	0.0	100.0
6	0.0	0.0	100.0	0.0	100.0
7	0.0	0.0	0.0	100.0	100.0
8	0.0	0.0	0.0	100.0	100.0
TOTAL	20.0	20.0	40.0	20.0	100.0

Cada celdilla contiene un porcentaje sobre el total de su fila respectiva

\*\*\* PERCENTS OF COLUMN TOTALS (Porcentajes sobre el total de cada Columna)

ventas	descuent				
	3	4	5	6	TOTAL
1	33.3	0.0	33.3	0.0	20.0
2	66.7	0.0	0.0	0.0	13.3
3	0.0	66.7	16.7	0.0	20.0
4	0.0	33.3	0.0	0.0	6.7
5	0.0	0.0	33.3	0.0	13.3
6	0.0	0.0	16.7	0.0	6.7
7	0.0	0.0	0.0	66.7	13.3
8	0.0	0.0	0.0	33.3	6.7
TOTAL	100.0	100.0	100.0	100.0	100.0

Cada celdilla expresa el porcentaje sobre el total de la columna

\*\*\* PERCENTS OF THE TABLE TOTAL (Porcentajes sobre el N total de la tabla)

ventas	descuent				
	3	4	5	6	TOTAL
1	6.7	0.0	13.3	0.0	20.0
2	13.3	0.0	0.0	0.0	13.3
3	0.0	13.3	6.7	0.0	20.0
4	0.0	6.7	0.0	0.0	6.7
5	0.0	0.0	13.3	0.0	13.3
6	0.0	0.0	6.7	0.0	6.7
7	0.0	0.0	0.0	13.3	13.3
8	0.0	0.0	0.0	6.7	6.7
TOTAL	20.0	20.0	40.0	20.0	100.0

Cada celdilla expresa el porcentaje sobre el N total de la tabla.