

1. Estadísticos Descriptivos.

Análisis de caso: Las rentas de los socios y empleados de INFORMAX CONSULTING S.L.

Informax es una pequeña y pujante empresa de consultoría de sistemas nacida hace un par de años para dar respuesta a necesidades específicas de informatización en tiempo real que no eran cubiertas por las soluciones estándar. En nuestro pequeño archivo de personal tenemos registrado a cada socio/a y trabajador/a con un NUMERO de identificación, su sexo, su edad actual, y sus ingresos el primer (RENTA1) y el segundo año (RENTA2) de funcionamiento de la empresa expresados en miles. Nuestro objetivo es efectuar una descripción estadística de estos datos.

1. Obtención de estadísticos descriptivos básicos.

BMDP Instruction File : C:\E3\1D001.INP
 BMDP Program Output File: C:\E3\1D001.OUT

A los archivos de instrucciones se les suele poner la extensión .INP, a los archivos de resultados se les suele poner la extensión :OUT

INSTRUCCIONES. PROGRAMA 1D.

Introducción al uso del BMDP Programa 1D

/INPUT

VAR=5.
 FORMAT=FREE.
 CASE=10.
 TITLE='DESCRIPCION BASICA Y LISTADO DE DATOS'.

En el párrafo INPUT se describen los datos, se dice que el número de variables es 5 y el formato libre. Opcionalmente se añade que hay 10 casos y se le pone un título al análisis. Las instrucciones se ordenan en párrafos que empiezan por / Cada instrucción acaba en punto.

/VARIABLES

NAMES=NUMERO, SEXO, EDAD, RENTA1, RENTA2.
 USE= EDAD TO RENTA2.
 MINIMUM= (SEXO)0, (EDAD)18, (RENTA1)1500, (RENTA2)1500.
 MAXIMUM= (RENTA1)10000, (RENTA2)12000.

En el párrafo variables se pone nombre a la variables (el orden ha de coincidir con el orden en los datos) mediante NAMES, se restringe el análisis a aquellas variables citadas en USE (desde EDAD hasta (TO) RENTA2) y se dan mínimos y máximos para los valores de algunas variables. Por ejemplo, el mínimo de RENTA1 es 1500 y el máximo de RENTA2 es 12000. Los nombres de variables son arbitrarios, de 8 caracteres como máximo, y salvo el signo _ no pueden incluir caracteres especiales ni espacios en blanco.

/PRINT

DATA.
 LINESIZE=80.

En este párrafo /PRINT, que es opcional, se solicita que el output contenga una lista de los datos (DATA) y se formatea la anchura del output a 80 columnas (80 columnas significa 80 caracteres por línea). En general no es necesario modificar la anchura con LINESIZE.

/END

1 1 18 1500 1550
 2 0 27 1900 1850
 3 1 43 6500 6530
 4 0 58 9770 9790
 5 1 60 1900 1900
 6 0 34 7800 7600
 7 0 35 3500 3510
 8 1 38 6500 3500
 9 0 29 2000 3340
 10 1 47 4500 4510
 /END

Después del /END los datos en formato "libre", es decir, cada dato separada por un blanco. Cada caso comienza línea nueva, y cada caso ocupa el mismo número de líneas.

OUTPUT SELECCIONADO

BMDP1D - SIMPLE DATA DESCRIPTION
 NUMBER OF CASES READ. 10

Nº y nombre de las variables y estadísticos descriptivos (por orden): Frecuencia, media, desviación típica, error típico de la media, coeficiente de variación, valor más pequeño en puntuaciones directas y en puntuaciones típicas, valor más grande en directas y en típicas.

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
3	EDAD	10	38.900	13.370	4.2281	.34371	18.000	-1.56	60.000	1.58
4	RENTA1	10	4587.0	2912.8	921.09	.63500	1500.0	-1.06	9770.0	1.78
5	RENTA2	10	4408.0	2737.4	865.65	.62101	1550.0	-1.04	9790.0	1.97

CASE NO.	2 SEXO	3 EDAD	4 RENTA1	5 RENTA2
1	1.00	18.00	1500.00	1550.00
2	0.00	27.00	1900.00	1850.00
3	1.00	43.00	6500.00	6530.00
4	0.00	58.00	9770.00	9790.00
5	1.00	60.00	1900.00	1900.00
6	0.00	34.00	7800.00	7600.00
7	0.00	35.00	3500.00	3510.00
8	1.00	38.00	6500.00	3500.00
9	0.00	29.00	2000.00	3340.00
10	1.00	47.00	4500.00	4510.00

Este es el listado de los datos producido por la instrucción DATA del párrafo /PRINT.

2. Cómo incluir variables alfanuméricas.

Los paquetes estadísticos están mejor diseñados para tratar datos numéricos, que es su función primordial, no obstante permiten también utilizar variables alfanuméricas (datos en forma de letras y palabras) con ciertas restricciones. En BMDP se pueden usar variables alfanuméricas extensas con pocas restricciones en un tipo de formato profesional que se llama FORMATO FIJO. En el formato libre (FORMAT FREE) solo puede haber hasta dos variables alfanuméricas en el mismo archivo de datos y los datos de cada una de ellas no pueden presentar más de 4 caracteres. Hay que definir esas variables como "etiquetas" (LABEL) para que el programa no trate de leerlas como datos numéricos. No obstante, hay un "truco" para introducir hasta 8 caracteres seguidos en cada caso: consiste en declarar 2 variables alfanuméricas seguidas, como se hace en el ejemplo siguiente. El programa las trata adecuadamente como una sola en los listados de datos. Si se necesita incluir variables alfanuméricas más largas o más variables alfanuméricas entonces debe consultarse el uso de formatos fijos de datos (FORMAT FIXED) en el manual del programa.

INSTRUCCIONES. PROGRAMA 1D.

```

/INPUT
  VAR=7.
  FORMAT=FREE.
  CASE=10.
  TITLE='DESCRIPCION BASICA'.
/VARIABLES
  NAMES=NUMERO, NOMBRE1, NOMBRE2, SEXO, EDAD, RENTA1, RENTA2.
  LABELS= NOMBRE1, NOMBRE2.
  USE= SEXO TO RENTA2.
  MINIMUM= (SEXO)0, (EDAD)18, (REN1)1500, (REN2)1500.
  MAXIMUM= (REN1)10000, (REN2)12000.
/PRINT
  MAXIMUM.
  MINIMUM.
  MISSING.
/END
1 GOMEZ 1 18 1500 1550
2 KREUS 0 27 1900 1850
3 FERNAN 1 43 6500 6530
4 GRAS 0 58 9770 9790
5 LUJAN 1 60 1900 1900
6 HUERTAS 0 34 7800 7600
7 RIUS 0 35 3500 3510
8 CASAN 1 38 6500 3500
9 BASAURI 0 29 2000 3340
10 MARTIN 1 47 4500 4510
/END
    
```

Aquí se declara que dos variables NOMBRE1 y NOMBRE2 han de leerse como etiquetas (LABELS). En realidad actúan reservando 8 caracteres para letras en esa posición de los datos.

Se solicita que se haga un listado SOLAMENTE de aquellos casos que presenten algún dato faltante (MISSING) o que exceda el máximo o el mínimo marcado para las variables. Esto es muy útil al comienzo de un análisis de datos para detectar errores en los datos y casos en los que faltan datos. Los máximos y los mínimos se han especificado, opcionalmente, en el párrafo /VARIABLES para aquellas variables que

Los datos contienen ahora el apellido de los 10 casos, con un máximo de 8 caracteres, en la posición en que están NOMBRE1 y NOMBRE2 en la lista de variables: después de NUMERO y antes de SEXO.

OUTPUT SELECCIONADO

CASE NO.	LABEL	4 SEXO	5 EDAD	6 RENTA1	7 RENTA2
1	GOMEZ	1.00	18.00	1500.00	1550.00
2	KREUS	0.00	27.00	1900.00	1850.00
3	FERNAN	1.00	43.00	6500.00	6530.00
4	GRAS	0.00	58.00	9770.00	9790.00
5	LUJAN	1.00	60.00	1900.00	1900.00
6	HUERTAS	0.00	34.00	7800.00	7600.00
7	RIUS	0.00	35.00	3500.00	3510.00
8	CASAN	1.00	38.00	6500.00	3500.00
9	BASAURI	0.00	29.00	2000.00	3340.00
10	MARTIN	1.00	47.00	4500.00	4510.00

El programa lista siempre "por defecto" (es decir, como opción preasignada) los 10 primeros casos, que en este ejemplo son todos. Esto es útil para verificar que hemos dado bien las instrucciones y el programa está leyendo los datos correctamente. En este caso, observamos que ha leído bien las "etiquetas" que contienen el apellido de las 10 personas de la empresa. Este listado NO es el listado selectivo de valores faltantes y fuera de rango: En este caso, como no los hay no producirá tal listado

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
4 SEXO	10	.50000	.52705	.16667	1.0541	0.0000	-0.95	1.0000	0.95
5 EDAD	10	38.900	13.370	4.2281	.34371	18.000	-1.56	60.000	1.58
6 RENTA1	10	4587.0	2912.8	921.09	.63500	1500.0	-1.06	9770.0	1.78
7 RENTA2	10	4408.0	2737.4	865.65	.62101	1550.0	-1.04	9790.0	1.97

3. Obtención de un listado de casos con datos faltantes y fuera de rango.

Con el fin de ilustrar un listado de casos con valores faltantes y fuera de rango volvemos a solicitarlo después de manipular los datos para introducir intencionadamente un valor faltante (que se escribe en los datos como un *) y algunos valores fuera de los límites de máximos y mínimos que le hemos dado a las variables.

INSTRUCCIONES. PROGRAMA 1D.

```

/INPUT VAR=7. FORMAT=FREE. CASE=10.
      TITLE='DESCRIPCION BASICA CON VALORES FALTANTES Y FUERA DE RANGO'.
/VARIABLES
      NAMES=NUMERO, NOMBRE1, NOMBRE2, SEXO, EDAD, RENTA1, RENTA2.
      LABELS= NOMBRE1, NOMBRE2.
      USE= SEXO TO RENTA2.
      MINIMUM= (SEXO)0, (EDAD)18, (RENTA1)1500, (RENTA2)1500.
      MAXIMUM= (RENTA1)10000, (RENTA2)12000.

/PRINT
      MAXIMUM.
      MINIMUM.
      MISSING.

/END
1 GOMEZ 1 18 1450 1550
2 KREUS 0 * 1900 1850
3 FERNAN 1 43 6500 6530
4 GRAS 0 58 9770 12510
5 LUJAN 1 60 1900 1900
6 HUERTAS 0 34 7800 7600
7 RIUS 0 35 3500 3510
8 CASAN 1 38 6500 3500
9 BASAURI 0 29 2000 3340
10 MARTIN 1 47 4500 4510
/END
    
```

Algunos datos han sido cambiados para introducir faltantes (*) y valores por encima del máximo o por debajo del mínimo, para ilustrar el efecto del listado solicitado en el párrafo /PRINT

OUTPUT SELECCIONADO

C A S E	4	5	6	7
NO. LABEL	SEXO	EDAD	RENTA1	RENTA2
1 GOMEZ	1.00	18.00	LT MIN	1550.00
2 KREUS	0.00	MISSING	1900.00	1850.00
3 FERNAN	1.00	43.00	6500.00	6530.00
4 GRAS	0.00	58.00	9770.00	GT MAX
5 LUJAN	1.00	60.00	1900.00	1900.00
6 HUERTAS	0.00	34.00	7800.00	7600.00
7 RIUS	0.00	35.00	3500.00	3510.00
8 CASAN	1.00	38.00	6500.00	3500.00
9 BASAURI	0.00	29.00	2000.00	3340.00
10 MARTIN	1.00	47.00	4500.00	4510.00

En el caso 1 en la variable RENTA1 aparece un dato menor que el mínimo para esa variable. En el caso 2 el dato de la variable EDAD es faltante. En el caso 4 la variable RENTA2 presenta un valor más alto que el máximo que hemos dado a esa variable en las instrucciones del párrafo /VARIABLE. Para esas tres variables los análisis descriptivos se refieren solo a nueve casos (columna TOTAL FREQ) y los estadísticos por tanto ahora se calculan excluidos esos

NUMBER OF CASES READ. 10

DESCRIPTIVE STATISTICS OF DATA

VARIABLE	TOTAL	STANDARD	ST.ERR	COEFF	SMALLEST	LARGEST
NO. NAME	FREQ.	MEAN	DEV.	OF MEAN	VALUE	Z-SCR
4 SEXO	10	50000	.52705	.16667	1.0541	0.0000 -0.95
5 EDAD	9	40.222	13.470	4.4900	.33489	18.000 -1.65 60.000 1.47
6 RENTA1	9	4930.0	2867.2	955.75	.58159	1900.0 -1.06 9770.0 1.69
7 RENTA2	9	3810.0	2099.3	699.77	.55100	1550.0 -1.08 7600.0 1.81

4. Cómo producir listados de datos ordenados según ciertas variables.

En muchas ocasiones es necesario obtener listados de los datos ordenados en función de uno o más criterios. En el ejemplo siguiente aprenderemos a obtener listados de datos ordenados en función de una o más variables.

INSTRUCCIONES. PROGRAMA 1D.

```

/INPUT
  TITLE='LISTADO DE DATOS ORDENADOS'.
  VAR=5.
  FORMAT=FREE.
  CASE=10.
  SORT=SEXO, EDAD.

/VARIABLES
  NAMES=NUMERO, SEXO, EDAD, RENTA1, RENTA2.

/PRINT
  DATA.

/END
1 1 18 1500 1550
2 0 27 1900 1850
3 1 43 6500 6530
4 0 58 9770 9790
5 1 60 1900 1900
6 0 34 7800 7600
7 0 35 3500 3510
8 1 38 6500 3500
9 0 29 2000 3340
10 1 47 4500 4510
/END
    
```

Se solicita presentar una lista de los datos ordenados primero por la variable SEXO, y dentro de cada categoría de la variable SEXO ordenados por su EDAD. Los datos pueden ordenarse por una, dos o más variables.

Aquí se solicita el listado de los DATOS.

OUTPUT SELECCIONADO

(El resto del output, incluidos estadísticos es como en los casos anteriores)

```

SORTING ON
  SEXO      ASCENDING
  EDAD      ASCENDING

ALL AVAILABLE STORAGE USED FOR INTERNAL SORT AREA
INTERNAL SORT CAPACITY 26576 CASES
*****
    
```

En este apartado el programa informa del orden en que ordenara cada variable (ascendente significa de menos a más) y de cuantos casos de este tipo podría ordenar como máximo.

CASE NO.	1 NUMERO	2 SEXO	3 EDAD	4 RENTA1	5 RENTA2
2	2.00	0.00	27.00	1900.00	1850.00
9	9.00	0.00	29.00	2000.00	3340.00
6	6.00	0.00	34.00	7800.00	7600.00
7	7.00	0.00	35.00	3500.00	3510.00
4	4.00	0.00	58.00	9770.00	9790.00
1	1.00	1.00	18.00	1500.00	1550.00
8	8.00	1.00	38.00	6500.00	3500.00
3	3.00	1.00	43.00	6500.00	6530.00
10	10.00	1.00	47.00	4500.00	4510.00
5	5.00	1.00	60.00	1900.00	1900.00

Los casos están ahora ordenados. Primero todos los que tienen SEXO codificado como 0 (que son los varones de la empresa), y dentro de ellos de menos a más edad. Después las mujeres (SEXO = 1) también ordenadas de menos a más edad. Obsérvese que, naturalmente, el número de caso aparece ahora en desorden.

5. Cómo crear nuevas variables a partir de variables existentes.

Cualquier cambio sistemático efectuado sobre los datos se denomina en general una transformación. Un tipo de transformación muy útil es la creación de nuevas variables a partir de variables existentes. En nuestro análisis estamos interesados en crear una nueva variable, que llamaremos INCREMENT, que represente la diferencia entre RENTA2 y RENTA1 es decir que conteste a la cuestión de cuanto más (o menos) ha ganado cada miembro de la empresa el segundo año. También crearemos una segunda transformada que sea un índice que relativice las ganancias en los dos años a la edad de la persona. Dado que las ganancias suelen estar relacionadas con la edad, este índice nos da la unidad de ganancia que ha obtenido una persona por año de edad. Estas transformadas nos ayudarán a entender diversos aspectos de la situación de remuneración en nuestra empresa.

INSTRUCCIONES. PROGRAMA 1D.

```

/INPUT
  TITLE='INTRODUCCION DE VARIABLES TRANSFORMADAS'.
  VAR=5.
  FORMAT=FREE.
  CASE=10.
/VARIABLES
  NAMES=NUMERO, SEXO, EDAD, RENTA1, RENTA2.
  USE=EDAD, RENTA1, RENTA2, INCREMENT, INDICE1.
/TRANS
  INCREMENT=RENDA2 - RENTA1.
  INDICE1 = (RENDA1 + RENTA2)/EDAD.
/PRINT
  DATA.
/END
1 1 18 1500 1550
2 0 27 1900 1850
3 1 43 6500 6530
4 0 58 9770 9790
5 1 60 1900 1900
6 0 34 7800 7600
7 0 35 3500 3510
8 1 38 6500 3500
9 0 29 2000 3340
10 1 47 4500 4510
/END
    
```

El párrafo /TRANS permite crear cuantas variables transformadas se desee utilizando operadores aritméticos. Los más usuales:
 + suma (a + b expresa sumar a más b)
 - resta
 * multiplicación
 / división
 ** potencia (x**b significa x elevado a b).
 También hay disponibles muchas funciones, por ejemplo:
 sqrt(x) significa raíz cuadrada de x
 log(c) significa logaritmo decimal de c
 También es posible utilizar operadores lógicos y funciones resumen que permiten hacer cualquier transformación necesaria con una gran flexibilidad.
 (Para una lista completa de funciones ver BMDP user's Digest (pags 11 a 14) y

OUTPUT SELECCIONADO

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO.	TOTAL NAME	FREQ.	MEAN	STANDARD DEV.	ST.ERR OF MEAN	COEFF OF VAR	SMALLEST VALUE	Z-SCR	LARGEST VALUE	Z-SCR
3	EDAD	10	38.900	13.370	4.2281	.34371	18.000	-1.56	60.000	1.58
4	RENDA1	10	4587.0	2912.8	921.09	.63500	1500.0	-1.06	9770.0	1.78
5	RENDA2	10	4408.0	2737.4	865.65	.62101	1550.0	-1.04	9790.0	1.97
6	INCREMENT	10	-179.00	1081.2	341.92	-6.0404	-3000.0	-2.61	1340.0	1.40
7	INDICE1	10	230.42	111.47	35.250	.48378	63.333	-1.50	452.94	2.00

CASE NO.	3 EDAD	4 RENTA1	5 RENTA2	6 INCREMENT	7 INDICE1
1	18.00	1500.00	1550.00	50.00	169.44
2	27.00	1900.00	1850.00	-50.00	138.89
3	43.00	6500.00	6530.00	30.00	303.02
4	58.00	9770.00	9790.00	20.00	337.24
5	60.00	1900.00	1900.00	0.00	63.33
6	34.00	7800.00	7600.00	-200.00	452.94
7	35.00	3500.00	3510.00	10.00	200.29
8	38.00	6500.00	3500.00	-3000.00	263.16
9	29.00	2000.00	3340.00	1340.00	184.14
10	47.00	4500.00	4510.00	10.00	191.70

Los datos muestran caso a caso las variaciones en resta y el índice de ingresos por años. Por ejemplo, el caso número 8 ha visto reducidos sus ingresos en 3 millones (recuérdese que estas variables renta están registradas en miles), mientras que la persona 9 ha ganado el segundo año 1.340.000 más.
 La persona que más gana para la edad que tiene es la número 6, una persona de 34 años con INDICE1 igual a 452'94 (es decir, ha ingresado casi medio millón por año en esos dos años). La persona con el INDICE1 más bajo es la número 5, una persona de 60 años con solo 63.330 por año de vida.
 La media de INCREMENT revela que en promedio cada persona de la empresa ha ganado 179.000 menos el segundo año.
 En nuestra empresa el INDICE1 promedio es 230.420, lo que significa que, en función de su edad, una persona de 18 años, por ejemplo, podría esperar ingresar en nuestra empresa aproximadamente unas $18 * 230.420 / 2$ por año, es decir 2.073.780 anuales.

6. Cómo obtener estadísticos por grupos.

Es muy frecuente que necesitemos describir separadamente subgrupos (submuestras o subpoblaciones) dentro del conjunto total de datos. Por ejemplo, podemos estar interesados en obtener la media y la desviación típica de los varones, por un lado, y de las mujeres, por otro. En este ejemplo veremos como utilizar el programa 1D con este propósito. En el ejemplo se introduce una nueva variable llamada MBA que refleja si cada miembro de la consultora ostenta (en cuyo caso MBA vale 1) ó no (en cuyo caso MBA vale 0) un título de MBA. Esto permitirá analizar los ingresos y los índices separadamente también para los dos grupos que forma esta variable: Masters frente a No-Masters. Adicionalmente, el ejemplo muestra como obtener un listado de los datos ordenados por si se tiene o no MBA, en orden A ascendente (primero los que No y luego los que SI, del 0 al 1), y dentro de cada uno de estos grupos, ordenados por edad en orden D descendente (del más mayor al más joven).

El ejemplo se aprovecha para ilustrar también diversas cuestiones sobre el modo de escribir las instrucciones.

- 1) Mayúsculas y minúsculas: Se muestra que las instrucciones y nombres de variables pueden darse indistintamente en mayúsculas o minúsculas.
- 2) Número de instrucciones por línea. En una línea puede haber una sentencia de instrucciones o más de una, siempre terminadas en un punto. Es más compacto poner todas las instrucciones seguidas, pero se gana en claridad poniendo una en cada línea.
- 3) Uso de =, is y are. Son tres modos de decir igual, puede usarse cualquiera de ellos, incluso si es gramaticalmente incorrecto en inglés, como se ejemplifica en las instrucciones de este ejemplo.
- 4) Cómo introducir comentarios en las instrucciones. Cualquier información en las instrucciones situada a la derecha del signo # es ignorada por el programa, considerándola un comentario a las instrucciones. Así la primera línea de las instrucciones:
 # Introduccion al uso del BMDP Programa P1D
 es un comentario toda ella y será ignorada. Y en la segunda línea del párrafo /INPUT el texto a la derecha del signo # es un comentario, mostrando como se pueden poner comentarios al final de una línea de instrucciones.
 Los comentarios son muy útiles para recordar unos días después de un análisis cual era el sentido de una instrucción o de una transformación. En la fase de aprendizaje puede servir para dejarse "dentro del archivo de instrucciones" explicaciones útiles y personales sobre el significado de las instrucciones.

INSTRUCCIONES. PROGRAMA 1D.

```
# Introduccion al uso del BMDP Programa P1D
/INPUT
TITLE is 'ESTADISTICOS POR GRUPOS Y DATOS ORDENADOS'.
VAR are 6.  FORMAT is FREE. # pueden mezclarse mayusculas y minusculas
sort are mba, edad.
ORDER = A, D.

/VARIABLES
NAMES is NUMERO, SEXO, EDAD, RENTA1, RENTA2, MBA.
USE=SEXO TO INDICE1.
/TRANS
INCREMEN = RENTA2 - RENTA1.
INDICE1 = (REN1 + RENTA2)/EDAD.

/GROUP
VARIABLE = SEXO, MBA.
CODES (SEXO, MBA) = 0, 1.
NAMES (SEXO) = VARON, MUJER.
NAMES (MBA) = no, si.

/PRINT DATA.

/END
1 1 18 1500 1550 0
2 0 27 1900 1850 0
3 1 43 6500 6530 1
4 0 58 9770 9790 1
5 1 60 1900 1900 0
6 0 34 7800 7600 1
7 0 35 3500 3510 0
8 1 38 6500 3500 0
9 0 29 2000 3340 0
10 1 47 4500 4510 1
/END
```

Se especifica las variables que ordenan los datos jerárquicamente (la primera es el primer criterio de ordenación dentro del cual opera la segunda). Para cada una si el orden es Ascendente o

El párrafo /GROUP es el que determina que se analice por grupos. La sentencia VARIABLE=SEXO, MBA. especifica que se analizarán los casos separadamente para los grupos que forma cada una de estas variables (ver output). La instrucción CODES establece que en las variables SEXO y MBA se han usado dos códigos, 0 y 1. (Si fueran distintos para cada variable usaríamos una instrucción CODE separada para cada una de ellas.) La instrucción NAMES le pone nombre (hasta 8 caracteres) a cada uno de los CODE, por orden. Así, NAMES(SEXO)=VARON, MUJER. establece que al 0, primer código de esta variable en CODES, le corresponde la etiqueta VARON, y al 1 la etiqueta MUJER, dando así nombre a los grupos en el output. En el caso de la variable MBA la sentencia NAMES asigna el titulo NO al 0 y la etiqueta SI al 1. Las instrucciones CODES y NAMES no son aquí imprescindibles para hacer el análisis por grupos, pero ayudan mucho a la inteligibilidad del output al poner nombre a los diversos grupos para los que se van a obtener estadísticos separados para cada una de las variables mencionadas en la sentencia USE, excluidas aquellas que determinan los grupos (es decir excepto SEXO y MBA en este caso).

OUTPUT SELECCIONADO

```

NUMBER OF CASES READ. . . . . 10
GROUPING VARIABLE. . . SEXO
                                CATEGORY  FREQUENCY
                                -----  -
                                VARON     5
                                MUJER     5
GROUPING VARIABLE. . . MBA
                                CATEGORY  FREQUENCY
                                -----  -
                                si         4
                                no         6
    
```

Aquí se mencionan las variables agrupadoras y se expresa la frecuencia (el número de casos) que hay de cada uno de los grupos. En los datos hay 5 varones y 5 mujeres. De los 10 casos, 4 tienen un MBA y 6 no lo tienen.

DESCRIPTIVE STATISTICS OF DATA

A continuación se muestran los estadísticos descriptivos de los 10 casos tomados conjuntamente. (Este análisis no hay que solicitarlo explícitamente, forma parte siempre del output.)

VARIABLE NO.	NAME	TOTAL FREQ.	MEAN	STANDARD DEV.	ST. ERR OF MEAN	COEFF OF VAR	SMALLEST VALUE	Z-SCR	LARGEST VALUE	Z-SCR
3	EDAD	10	38.900	13.370	4.2281	.34371	18.000	-1.56	60.000	1.58
4	RENTA1	10	4587.0	2912.8	921.09	.63500	1500.0	-1.06	9770.0	1.78
5	RENTA2	10	4408.0	2737.4	865.65	.62101	1550.0	-1.04	9790.0	1.97
7	INCREMEN	10	-179.00	1081.2	341.92	-6.0404	-3000.0	-2.61	1340.0	1.40
8	INDICE1	10	230.42	111.47	35.250	.48378	63.333	-1.50	452.94	2.00

PRINT SUMMARY STATISTICS OVER ALL CASES AND
BROKEN DOWN BY INDIVIDUAL CATEGORY ON SEXO
AND MBA

SORTING ON

```

MBA          ASCENDING
EDAD        DESCENDING
    
```

Aquí se refleja que se ha solicitado además un listado de los datos ordenados por las variables MBA (ascendente) y EDAD (en orden descendente).

ALL AVAILABLE STORAGE USED FOR INTERNAL SORT AREA
INTERNAL SORT CAPACITY 19921 CASES

CASE NO.	2 SEXO	3 EDAD	4 RENTA1	5 RENTA2	6 MBA	7 INCREMEN	8 INDICE1
5	MUJER	60.00	1900.00	1900.00	no	0.00	63.33
8	MUJER	38.00	6500.00	3500.00	no	-3000.00	263.16
7	VARON	35.00	3500.00	3510.00	no	10.00	200.29
9	VARON	29.00	2000.00	3340.00	no	1340.00	184.14
2	VARON	27.00	1900.00	1850.00	no	-50.00	138.89
1	MUJER	18.00	1500.00	1550.00	no	50.00	169.44
4	VARON	58.00	9770.00	9790.00	si	20.00	337.24
10	MUJER	47.00	4500.00	4510.00	si	10.00	191.70
3	MUJER	43.00	6500.00	6530.00	si	30.00	303.02
6	VARON	34.00	7800.00	7600.00	si	-200.00	452.94

Este es el listado de los datos ordenados. Primero están los 6 casos NO en MBA de más a menos edad. Después están los 4 casos SI en MBA de más a menos edad. En el listado el programa utiliza las etiquetas de SEXO y MBA en lugar de los números 0 y 1.

A continuación viene el análisis estadístico por grupos. Se calculan los estadísticos descriptivos media, desviación típica, valor menor y valor mayor, de cada variable mencionada en el USE (y si no hay sentencia USE de todas) siempre excepto las que sirven para formar los grupos. Primero se vuelven a mostrar los resultados para los 10 casos juntos. Después se menciona la primera variable agrupadora (en este caso SEXO) y se dan los resultados para su grupos (en este caso para VARONES y MUJERES separadamente). Después se menciona la segunda variable agrupadora (MBA) y se muestran los estadísticos para cada uno de sus grupos (los que SI y los No tienen un MBA). Aquí podemos ver que las mujeres de la empresa tienen una media de edad de 41 años y los varones de 36, aproximadamente. Los que tienen MBA son más mayores (unos 45 años de media) que los que no tienen (unos 34 años de media).

VARIABLE NO.	GROUPING NAME	TOTAL FREQ.	MEAN	STANDARD DEVIATION	SMALLEST VALUE	Z-SC	LARGEST VALUE	Z-SC
3	EDAD	10	38.900	13.370	18.000	-1.56	60.000	1.58
	SEXO							
	VARON	5	36.600	12.422	27.000	-0.77	58.000	1.72
	MUJER	5	41.200	15.320	18.000	-1.51	60.000	1.23
	MBA							
	no	6	34.500	14.293	18.000	-1.15	60.000	1.78
si	4	45.500	9.9499	34.000	-1.16	58.000	1.26	
4	RENTA1	10	4587.0	2912.8	1500.0	-1.06	9770.0	1.78
	SEXO							
	VARON	5	4994.0	3586.5	1900.0	-0.86	9770.0	1.33
	MUJER	5	4180.0	2410.8	1500.0	-1.11	6500.0	0.96
	MBA							
	no	6	2883.3	1902.0	1500.0	-0.73	6500.0	1.90
si	4	7142.5	2216.0	4500.0	-1.19	9770.0	1.19	
5	RENTA2	10	4408.0	2737.4	1550.0	-1.04	9790.0	1.97
	SEXO							
	VARON	5	5218.0	3330.3	1850.0	-1.01	9790.0	1.37
	MUJER	5	3598.0	2032.0	1550.0	-1.01	6530.0	1.44
	MBA							
	no	6	2608.3	931.70	1550.0	-1.14	3510.0	0.97
si	4	7107.5	2199.9	4510.0	-1.18	9790.0	1.22	
7	INCREMEN	10	-179.00	1081.2	-3000.0	-2.61	1340.0	1.40
	SEXO							
	VARON	5	224.00	630.02	-200.00	-0.67	1340.0	1.77
	MUJER	5	-582.00	1351.8	-3000.0	-1.79	50.000	0.47
	MBA							
	no	6	-275.00	1438.5	-3000.0	-1.89	1340.0	1.12
si	4	-35.000	110.30	-200.00	-1.50	30.000	0.59	
8	INDICE1	10	230.42	111.47	63.333	-1.50	452.94	2.00
	SEXO							
	VARON	5	262.70	129.56	138.89	-0.96	452.94	1.47
	MUJER	5	198.13	92.555	63.333	-1.46	303.02	1.13
	MBA							
	no	6	169.87	66.545	63.333	-1.60	263.16	1.40
si	4	321.23	107.57	191.70	-1.20	452.94	1.22	

Notas de interpretación de los resultados de la página anterior:

Los resultados del análisis de datos por grupos son muy interesantes y reveladores, hasta el punto que podrían desprenderse conclusiones muy importantes para rediseñar la política salarial de la consultora y para el diseño de carrera de un miembro de la misma que desee mejorar sus ingresos.

El análisis revela que las mujeres de la empresa ganan menos que los hombres tanto el primer año como el segundo. El primer año ganaron en promedio (4994 - 4180 =) unas 814.000 menos que los varones (dado que los datos están en miles). El segundo año han ganado (5218 - 3598) del orden de 1.620.000 menos que los varones, por lo que, como muestra el índice INCREMEN, la situación todavía ha tendido a empeorar para ellas. Así mientras que el segundo año los varones ganaban en promedio 224.000 más que el primero, las mujeres perdían en promedio 582.000.

Además las rentas de los varones presentan mayor variabilidad (ver desviaciones típicas), lo que significa que para ellos hay más oportunidades de ganar más, mientras que el grupo de mujeres recibe de modo más homogéneo resultados económicos más bajos.

Estas diferencias entre varones y mujeres no son fruto de la edad, pues el índice de ganancias por edad es 138'89 en promedio para los varones, pero solo de 63'33 para las mujeres: es decir, además las mujeres en nuestra empresa tienen peor expectativa de ganancia en promedio a lo largo de su vida laboral.

La mujer que más ha perdido ha llegado a perder 3.000.000 del primer al segundo año, mientras que la que más ha ganado solo ha incrementado sus ingresos en 50.000. El varón que más ha visto decrementados sus ingresos solo ha perdido 200.000 mientras que el que más ha incrementado sus ingresos lo ha hecho en 1.340.000. (ver mínimos y máximos).

Una mirada atenta a los datos revelará que de las 5 mujeres de la empresa, una mantiene los mismos ingresos los dos años, 3 obtienen ligeras aumentos entre 10.000 y 50.000, pero una obtiene resultados espectacularmente peores llegando a perder 3.000.000. Este valor extremo supone que las mujeres presenten en conjunto unos resultados manifiestamente peores de los que hubieran obtenido sin considerar este caso particular.

Los resultados del análisis por grupos para la variable MBA son igualmente interesantes.

Los que tienen un MBA ganan en la empresa el primer año unas 2'5 veces más que los que no lo tienen (7142'5/2883'3=2'4772), y el segundo año esa relación ha crecido (7107'5/2608'3=2'725) hasta unas 2'75 veces más. En conjunto cada persona de la empresa ha ganado en el segundo año un promedio de 179.000 menos que el primero. Pero el descenso de ingresos no ha afectado igual a los que tienen MBA. Así, los titulados MBA solo han visto reducidos sus ingresos en unas 35.000 en promedio, mientras que los no titulados MBA han sufrido un descenso promedio de 275.000, siendo la dispersión para los titulados mucho menor (desviación típica de 110.300).

Además, los titulados MBA presentan un índice de ingresos por año considerablemente mayor (321.230) que los no titulados (169.870), con una dispersión mucho mayor para los titulados (107.570) que revela mayores oportunidades de ganancia para algunos de ellos.

Podría pensarse que la situación global de las mujeres se debe a que no tienen MBA, pero esto se puede descartar completamente comprobando en el listado de datos ordenados anterior que hay igual número de mujeres que de hombres en cada grupo formado con la variable MBA.

La información de todo este comentario de análisis está contenida en las tablas de resultados del output del análisis por grupos anterior, poniéndose de manifiesto el gran interés aplicado que tiene este tipo de análisis, y como es revelador de muchos aspectos muy importantes que pasarían inadvertidos en los

7. Cómo utilizar datos tabulados para los análisis.

En ocasiones disponemos como punto de partida de datos ya tabulados, es decir, de datos en los que nos consta la frecuencia (el número de casos) de cada configuración de puntuaciones.

En los datos del ejemplo siguiente se ha simulado esta situación incluyendo una nueva variable F que en realidad expresa la frecuencia del tipo de datos de ese registro.

INSTRUCCIONES. PROGRAMA 1D.

```

/INPUT
  TITLE = 'ESTADISTICOS PARA DATOS CON VARIABLE DE FRECUENCIAS'.
  VAR = 7.
  FORMAT= FREE.
  CASE=10.

/VARIABLES
  NAMES= TIPO, SEXO, EDAD, RENTA1, RENTA2, MBA, F.
  USE= RENTA1, RENTA2.
  FREQ= F.

/PRINT CASE=0.
/END
1 1 18 1500 1550 0 10
2 0 27 1900 1850 0 2
3 1 43 6500 6530 1 7
4 0 58 9770 9790 1 13
5 1 60 1900 1900 0 4
6 0 34 7800 7600 1 17
7 0 35 3500 3510 0 13
8 1 38 6500 3500 0 11
9 0 29 2000 3340 0 16
10 1 47 4500 4510 1 10
/END
    
```

Se crea una nueva variable, que se ha llamado aquí F, para contener las frecuencias.

Mediante la sentencia FREQ=F. se declara que la variable F en realidad contiene frecuencias para todas las demás.

CASE permite controlar cuantos casos muestra el listado del output. CASE=2, por ejemplo, significaría que presente un listado con solo los 2 primeros casos. CASE=0 supone que no muestra dato alguno, siempre que se evite también la instrucción DATA, lo que se utiliza para abreviar el output.

Esta columna es la variable F, y contiene las frecuencias, como hemos declarado mediante FREQ=F. Por ejemplo, la línea 1, dice que hay 10 casos que tengan sexo 1, edad 18, RENTA1 de 1500, RENTA2 de 1550 y que no tienen un MBA. Por ejemplo, la fila 5 dice que hay 4 mujeres de 60 años sin MBA que ganaron 1900 el primer año y 1900 el segundo año. Este es por tanto un modo de introducir muchos casos iguales sin necesidad de escribir una línea para cada uno de ellos.

OUTPUT SELECCIONADO

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO.	NAME	TOTAL FREQ.	MEAN	STANDARD DEV.	ST. ERR OF MEAN	COEFF OF VAR	SMALLEST VALÚE	Z-SCR	LARGEST VALÚE	Z-SCR
4	RENTA1	103	5102.0	2843.0	280.12	.55722	1500.0	-1.27	9770.0	1.64
5	RENTA2	103	4967.5	2632.5	259.39	.52995	1550.0	-1.30	9790.0	1.83

El análisis se reduce a las variables RENTA1 y RENTA2 porque así lo hemos establecido en la instrucción USE del párrafo /VAR. Obsérvese que en la columna TOTAL FREQ. se constata que los estadísticos de RENTA1 y de RENTA2 han sido calculados para 103 casos. 103 casos es la suma de las frecuencias que hemos reflejado en la variable de frecuencias F.

8. Mantener los datos separados en un archivo distinto.

Los datos para un análisis estadístico pueden tener muchas fuentes. Pueden provenir de un procesador de textos (guardados en ascii o modo solo texto con saltos de línea), de una hoja de cálculo, de una base de datos, o de otros paquetes estadísticos. Existen procedimientos para enviar datos de unos programas a otros y convertir archivos de unos en aptos para otros.

En cuanto al BMDP hay tres modos de presentarle los datos:

- 1) El más sencillo, como hemos venido haciendo en todos los ejemplos anteriores, consiste en colocar los datos a continuación de las instrucciones. Esto es lo más práctico cuando el número de datos es pequeño.
- 2) Es posible sin embargo mantener los datos en un archivo aparte, dedicado solo a los datos, y "llamarlo" desde el párrafo INPUT de las instrucciones. Esto es lo más práctico cuando tenemos masas de datos más grandes, pues copiar los datos en cada archivo de instrucciones sería costoso en términos de espacio, dado que lo usual en un proceso de análisis es que unos mismos datos sean objeto de varios análisis utilizando diversas variantes de archivos de instrucciones para diferentes programas.
- 3) Si una masa de datos de cierto tamaño va a ser analizada y reanalizada varias veces, como es usual, todavía hay un tercer modo de guardar los datos y llamarlos por los archivos de instrucciones. Consiste en guardarlos en un archivo tipo BMDP, que es un modo particular en que BMDP puede guardar unos datos para un uso más eficiente en sucesivos reanálisis. Este procedimiento profesional más sofisticado tiene ventajas de eficiencia con análisis de datos donde las variables o los casos son muy numerosas y los análisis a hacer muchos.

En el ejemplo siguiente ilustraremos el segundo método, que es suficiente para un trabajo en análisis de datos ocasional, aunque se trate de grandes masas de datos.

Para ello hemos guardado los datos utilizando el mismo editor del BMDP en un archivo llamado 1D008.DAT que está en el directorio E3 del disco C:. La extensión .DAT es la usual para los archivos de datos que pueden crearse también con cualquier otro programa, tal como un procesador de textos, capaz de guardar en ASCII, lo que en algunos procesadores de texto se llama "solo texto", o mejor "solo texto con saltos de línea". En ese archivo solo están los datos, separados por blancos, de la misma forma en que hasta ahora los escribíamos después de /END. No hay ni títulos, ni nombres de variables, ni líneas en blanco al principio, ni ninguna otra cosa. Es conveniente recordar dar un retorno de carro después de la última línea de datos.

En el párrafo /INPUT se llama a este archivo de datos con la sentencia FILE='C:\E3\1D008.DAT'. que describe el "path" del archivo de datos. Las demás instrucciones no varían, y, obviamente, no volvemos a colocar los datos después de /END.

INSTRUCCIONES. PROGRAMA 1D.

```
/INPUT TITLE = 'Como utilizar datos en otros archivos tipo ASCII'.
VAR = 7. FORMAT= FREE. CASE=10 file = 'c:\e3\1D008.dat'.
/VARIABLES NAMES= TIPO, SEXO, EDAD, RENTA1, RENTA2,
MBA, F. USE= RENTA1, RENTA2. FREQ= F.
/END
```

Esta es la instrucción que llama al archivo de datos para que BMDP sepa sobre que datos ejecuta las instrucciones. Por otra parte este archivo de instrucciones muestra como se pueden escribir las instrucciones de modo más compacto poniéndolas unas a continuación de otras en pocas líneas. Intencionadamente la enumeración de variables se ha partido en 2 líneas, para apreciar como puede escribirse fácilmente una enumeración muy larga, simplemente cambiando de línea después de una coma y siguiendo en la línea siguiente (y si es necesario en 3 o

Archivo de datos: c:\e3\1D008.dat

```
1 1 18 1500 1550 0 10
2 0 27 1900 1850 0 2
3 1 43 6500 6530 1 7
4 0 58 9770 9790 1 13
5 1 60 1900 1900 0 4
6 0 34 7800 7600 1 17
7 0 35 3500 3510 0 13
8 1 38 6500 3500 0 11
9 0 29 2000 3340 0 16
10 1 47 4500 4510 1 10
```

OUTPUT SELECCIONADO

(El output no difiere en nada del que hubiéramos obtenido con los datos después del párrafo /END)

NUMBER OF CASES READ. 10

DESCRIPTIVE STATISTICS OF DATA

VARIABLE	TOTAL	STANDARD	ST.ERR	COEFF	SMALLEST	LARGEST
NO. NAME	FREQ.	MEAN	DEV.	OF MEAN	VALUE	Z-SCR
4 RENTA1	103	5102.0	2843.0	280.12	.55722	1500.0 -1.27 9770.0 1.64
5 RENTA2	103	4967.5	2632.5	259.39	.52995	1550.0 -1.30 9790.0 1.83

9. Un análisis por grupos en el programa 1D según el estado civil.

El siguiente ejemplo, sobre otro conjunto de datos, muestra un análisis descriptivo por grupos para la variable estado civil (llamada estado) en la que a sus tres códigos se asignan respectivamente los nombres identificadores de los grupos. El análisis por grupos permite conocer el comportamiento de estos tres colectivos en las otras variables, concretamente en edad y sueldo.

INSTRUCCIONES. PROGRAMA 1D.

```

/input var are 3. case are 10. format is free.
/var names are edad, sueldo, estado. use are estado, sueldo.
/group variable=estado. codes(estado)=0,1,2. names(estado)=solteros, casados, divorc.
/end
19 70 0
55 450 2
35 200 1
40 175 2
50 350 0
21 110 1
25 120 2
29 140 1
33 170 0
39 250 1
/end
    
```

Los nombres siempre se asignan a los códigos por orden.

OUTPUT SELECCIONADO

CASE NO.	3 estado	2 sueldo
1	solteros	70.00
2	divorc	450.00
3	casados	200.00
4	divorc	175.00
5	solteros	350.00
6	casados	110.00
7	divorc	120.00
8	casados	140.00
9	solteros	170.00
10	casados	250.00

NUMBER OF CASES READ. 10

GROUPING VARIABLE. . . estado

CATEGORY	FREQUENCY
solteros	3
casados	4
divorc	3

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO.	TOTAL NAME	FREQ.	MEAN	STANDARD DEV.	ST.ERR OF MEAN	COEFF OF VAR	SMALLEST VALUE	Z-SCR	LARGEST VALUE	Z-SCR
2	sueldo	10	203.50	117.33	37.104	.57657	70.000	-1.14	450.00	2.10

PRINT SUMMARY STATISTICS OVER ALL CASES AND
BROKEN DOWN BY INDIVIDUAL CATEGORY ON estado

En el ejemplo, los divorciados son el grupo que presenta mayores ingresos medios, aunque también presentan la mayor variabilidad. La presencia de un divorciado con altos ingresos puede justificar estos resultados con tan pocos datos

VARIABLE NO.	GROUPING VAR/LEVL	TOTAL FREQ.	MEAN	STANDARD DEVIATION	SMALLEST VALUE	Z-SC	LARGEST VALUE	Z-SC
2	sueldo	10	203.50	117.33	70.000	-1.14	450.00	2.10
	estado							
	solteros	3	196.67	141.89	70.000	-0.89	350.00	1.08
	casados	4	175.00	62.450	110.00	-1.04	250.00	1.20
	divorc	3	248.33	176.80	120.00	-0.73	450.00	1.14

Modos de ejecutar un programa BMDP.

El BMDP permite diferentes modos de llamar a los programas, ejecutarlos y ver los resultados.

Modo 1. Entrar en el menú general de BMDP .

Paso 1. Llamar al BMDP mediante la instrucción BMDP (si es la versión DYNAMIC la instrucción es BMDPDYN).

En DOS esto puede hacerse desde dentro del directorio donde esté el BMDP, o, desde el directorio raíz (esto último puede hacerse si se ha establecido el path para BMDP en el AUTOEXEC.BAT, lo que suele ser lo usual si se ha permitido a BMDP modificar el AUTOEXEC.BAT durante el proceso de instalación).

Paso 2. Una vez dentro del menú general de BMDP podemos optar por:

Opción 1. Seleccionar RUN, seleccionar un programa, escribir las instrucciones (ó abrir un archivo de instrucciones tipo .INP con F2 opción OPEN) y hacer correr el programa mediante la tecla F6 (RUN). Después podemos guardar el archivo de instrucciones con F2 opción SAVE AS. Para salir F7 (QUIT).

Opción 2. Seleccionar EDIT que abre una hoja para instrucciones (bien una nueva, o bien la última en que habíamos trabajado). Después de escribir las instrucciones podemos llamar a un programa mediante F6 (RUN) que, en este caso, nos mostrará primero la lista de programas para que escojamos el que deseamos hacer funcionar.

En general, tanto desde la opción 1 como 2, si deseamos un archivo de instrucciones nuevo y vacío podemos obtenerlo mediante F2 opción NEW. Si deseamos otro archivo de instrucciones anterior basta abrirlo con F2 opción OPEN.

Hay que tener en cuenta que todos los programas BMDP comparten una misma sintaxis de instrucciones y que en los párrafos comunes básicamente las instrucciones son las mismas para cada uno de ellos. En algunos casos exactamente las mismas instrucciones funcionarían en varios programas. Para hacer funcionar un nuevo programa en la mayoría de casos basta con adaptar un archivo de instrucciones anterior añadiendo algunas instrucciones y suprimiendo otras (o simplemente convirtiéndolas en comentario temporalmente colocando el signo # al comienzo de las mismas para que sean ignoradas por el nuevo programa sin perderlas del archivo). Por tanto al cambiar de programa no hay que reescribir completamente el archivo de instrucciones ni mucho menos los datos. Basta con abrir el archivo de instrucciones anterior con F2 opción OPEN, hacer los cambios oportunos en pantalla, y guardar el nuevo archivo modificado con otro nombre (si se desea conservar la versión anterior) con F2 opción SAVE AS. esto puede hacerse tanto desde la opción 1 como desde la 2.

Modo 2. Llamar directamente a un programa concreto desde DOS.

Si escribimos en DOS por ejemplo:

BMDP 1D (seguido siempre de retorno de carro, como en cualquier instrucción)
el programa entra directamente en el 1D, en modo interactivo.

Llamar a un programa concreto y especificar el archivo de output:

Si escribimos:

BMDP 2D OUT='c:\E3\prueba.out'

entrará en el 2D en modo interactivo y enviará el output al archivo prueba.out del directorio llamado E3, que, obviamente, tiene que existir primero.

Modo 3. Llamar directamente a un archivo de instrucciones para que se ejecute.

Supongamos que tenemos en el directorio E3 un archivo llamado k3.inp que contiene instrucciones para el 1D.

Si llamamos:

BMDP 1D IN='c:\E3\k3.inp'

entrará en 1D y ejecutará en modo Batch k3.inp (en modo batch significa que ejecutará directamente k3.inp sin que podamos intervenir sobre el proceso, sin que nos de opción antes a ver las instrucciones y a modificarlas).

Llamar a un archivo de instrucciones y especificar el archivo de output.

Dada desde DOS la instrucción:

BMDP 2D IN=k3.inp OUT=k5.out

entra en 2D ejecuta k3.inp en batch y guarda los resultados en k5.out.

Al no especificar un path (una ubicación en disco y directorios) para los archivos estos han de estar en el directorio preestablecido por defecto (configurable desde SINGLE CONFIGURE).

En el manual BMDP User's Guide pueden encontrarse otras especificaciones en el modo de llamar a los programas y de configurar el BMDP.

¿Donde va el output del BMDP?

El output puede ir a un archivo con nombre especificado, si así está configurado nuestro BMDP, o bien a un archivo temporal llamado BMDPOUT&.OUT. Al comenzar a ejecutar un programa, BMDP menciona donde va a enviar el output que nos presenta en pantalla. Si el output va a BMDPOUT&.OUT al acabar el programa hay que salir (F7), y después de ver el output con VIEW (en el menú general) si lo deseamos, salir de BMDP (con EXIT en el menú general) y cambiarle el nombre al archivo temporal BMDPOUT&.OUT. Si deseamos conservar el output que acabamos de realizar hay que renombrar BMDPOUT&.OUT porque, de lo contrario, un nuevo RUN hará que se pierda al volver a guardar con ese mismo nombre el output del nuevo trabajo.

Ver el output con VIEW y editarlo con EDIT.

Una vez ejecutado un programa desde el menú general BMDP podemos ver el output desde VIEW y desde EDIT. VIEW solo permite leer el output, pero no modificarlo. EDIT permite leerlo y modificarlo.

Para ver el output con VIEW inmediatamente después de un run (una ejecución de un programa) basta con seleccionar VIEW que está automáticamente dispuesto a abrir el último output producido.

Para ver el output con EDIT hay que entrar en EDIT y abrir el output mediante F2 opción OPEN. Si deseamos guardar modificaciones en el output antes de salir de EDIT hay que hacer F2 opción SAVE.

¿Cómo configurar el destino del output?

En SINGLE CONFIGURE (menú principal del BMDP) podemos variar muchos aspectos importantes del modo en que se ejecutan los programas. En la línea llamada OUTPUT FILE podemos escoger tres opciones para el archivo de output:

Prompt : Significa que antes de cada RUN nos preguntará el nombre, y path si lo deseamos, del archivo para output (llamamos un RUN a cada ejecución de un programa que se produce precisamente después de la instrucción RUN, generalmente dada con la tecla F6). Generalmente un archivo para output le damos la extensión .OUT.

input.OUT : significa que llamará al output con el nombre del archivo de input pero cambiando la extensión .INP de este por .OUT propia de un output.

Temp File : Los resultados irán a un archivo temporal llamado BMDPOUT&.OUT.

En los menús de SINGLE CONFIGURE se pueden variar otros muchos aspectos, incluidos los directorios en que se desea que opere BMDP por defecto, y las extensiones por defecto que deseamos para cada tipo de archivo. Para que las nuevas elecciones funcionen hay que contestar YES a SAVE THESE SETTINGS antes de salir de los menús de SINGLE CONFIGURE, salir del BMDP y volver a entrar.

Abrir y guardar output con un procesador de textos.

Los archivos de output del BMDP son de tipo texto y pueden abrirse con cualquier procesador de textos (p.e. con el WORD) y, después guardarlos con otro nombre.

Al abrir con un procesador de textos un archivo de output se recomienda seleccionar todo el archivo, y cambiar el tipo de letra a Courier new 8 puntos, quitando también todo sangrado de párrafo: así de golpe se recompondrá el formato de todo el output. (Un usuario más avanzado de un procesador de textos puede crear un estilo con estas características para hacer esto de modo más fácil muchas veces.)

Evidentemente poder tomar el output dentro de un procesador de textos es una ventaja inestimable para añadir comentarios, eliminar aspectos no deseados y, en resumen, poder producir informes.

BMDP en Windows.

BMDP es una aplicación DOS, pero funciona perfectamente llamada desde WINDOWS 3.1 o desde WINDOWS 95 (en algunos ordenadores portátiles con configuraciones particulares de memoria puede funcionar incluso mejor).

La versión DYNAMIC de BMDP lleva incluso un icono para ser ubicado en WINDOWS.

En WINDOWS 95 abrir la carpeta donde se desee ver el icono del programa BMDP. En el menú de Archivo de la carpeta seleccionar NUEVO y dentro de NUEVO seleccionar ACCESO DIRECTO, mediante EXAMINAR seleccionar el programa BMDP.EXE de la carpeta donde esté el BMDP y seguir instrucciones.

Ver el manual de otras versiones de WINDOWS para seguir los pasos para ubicar una aplicación DOS disponible desde WINDOWS.

Para usuarios avanzados en WINDOWS es posible cambiar toda una serie de parámetros sobre el modo en que se ejecuta el programa (p.e. el tamaño de la ventana de ejecución). La ayuda de WINDOWS puede ser muy útil.

Ayuda dentro de BMDP.

En muchos contextos BMDP puede ofrecer ayuda mediante la tecla F1.

Ejemplos de archivos de instrucciones BMDP.

El programa BMDP aporta al menos un ejemplo de instrucciones para cada uno de sus programas. Estos ejemplos normalmente se habrán instalado en la carpeta donde se instalará BMDP y resultan de gran ayuda como punto de partida para construir uno mismo sus propios archivos.

Por ejemplo es posible abrir el archivo 1D.INP que contiene un ejemplo para el 1D, hacer los cambios que sean oportunos para nuestros propios datos y guardarlo con otro nombre para no sacrificar el ejemplo original que siempre puede volver a ser de ayuda.

Generar los archivos de instrucciones mediante menús (la opción MENTOR).

MENTOR permite manejar mediante menús algunos de los programas BMDP más utilizados.

A MENTOR se accede desde el menú principal del BMDP. Es una utilidad que permite generar fácilmente archivos de instrucciones. Sin ningún conocimiento previo del lenguaje de instrucciones del BMDP. Es decir, no hay que recordar instrucciones, basta con hacer elecciones en los menús y MENTOR ira escribiendo el archivo de instrucciones por nosotros.

MENTOR es bastante autoexplicativo, así que puede probarse a intentar un análisis sencillo entrando en él desde el menú general de BMDP. Además, como en otros puntos, puede pedir ayuda desde dentro del programa mediante F1 (HELP). Para salir, como de costumbre, F7 (QUIT).

Para un tutorial detallado de MENTOR ver el manual BMDP User's Guide.

Este método de manejar BMDP resulta más fácil para principiantes si no se conoce el lenguaje de instrucciones del BMDP, pero una vez que se tiene un conocimiento básico, la generación de cada instrucción mediante menús puede resultar demasiado tediosa. Una vez que se sabe manejar BMDP en sus aspectos básicos no es usual recurrir a este método de trabajo.

Versiones BMDP.

BMDP es uno de los grandes paquetes estadísticos entre los que pueden mencionarse al SPSS, al SYSTAT, y al SAS. Existen unos pocos grandes paquetes estadísticos, de los que los anteriores pueden considerarse los más utilizados e importantes. Además existen virtualmente miles de programas estadísticos diseñados para objetivos concretos o diseñados precisamente para la enseñanza.

El BMDP, junto con el SPSS, es sin duda uno de los grandes paquetes de elección profesional más frecuente para el tratamiento profesional estadístico de los datos. En líneas generales estos grandes paquetes son equivalentes, aunque cada uno de ellos puede incorporar algunos tipos de análisis o variantes de análisis que otros no, especialmente en niveles estadísticos científicos o profesionales avanzados y sofisticados.

Hay un paquete BMDP ó SPSS instalado en virtualmente todos los tipos de grandes MAINFRAMES (grandes multiprocesadores de análisis de datos) como VAX, UNIVAC, IBMs etc. Desde los años 80 estos paquetes están también disponibles en versiones para Pcs.

En lo esencial el paquete BMDP no ha cambiado desde finales de los 70, siendo desde el punto de vista estadístico un programa muy probado y avalado por su uso en la investigación internacional, prácticamente en cualquier campo científico. Por ejemplo, básicamente las mismas instrucciones que hacían funcionar al BMDP en el multiprocesador SPERRY/UNIVAC de un gran Centro de Calculo en la década de los 70 hacen funcionar las versiones para PC actuales. En general la estadística no cambia según los avances de la informática y un paquete profesional, con algoritmos de cálculo bien probados que satisfacen las necesidades de los cálculos estadísticos, por tanto, no tiene porque cambiar de versión frecuentemente como sucede con los juegos o con los procesadores de texto.

Esta estabilidad en el modo de manejo es importante en términos prácticos porque facilita la utilidad de la herramienta por largos periodos de tiempo de la vida profesional y a través de diversos tipos de versiones y de sistemas informáticos. De ese modo aprender a manejar una versión del paquete es un esfuerzo que seguirá siendo básicamente útil a través de diversos sistemas informáticos y diversas versiones del paquete estadístico. Trabajar desde el principio con un paquete profesional es muy pronto más rentable que aprender primero un programa didáctico para la enseñanza para después tener que aprender una herramienta profesional.

En concreto, del BMDP hay tres tipos de versiones para Pcs:

-versión BMDP PC

-versión BMDP PC DYNAMIC (básicamente permite archivos de mayor tamaño que la anterior).

-BMDP NEW SYSTEM (que opera en formato WINDOWS y en conexión con alguna de las otras).

Cada versión presenta el paquete estadístico completo.

Actualmente la relativa dificultad de aprendizaje del BMDP (necesidad de usar instrucciones) hacen que se tienda a preferir en la enseñanza por razones didácticas, que no por razones estadísticas o profesionales, versiones WINDOWS. Sin embargo, una vez superada esta dificultad el BMDP ofrece un abanico de posibilidades que los estadísticos aplicados, los científicos y los analistas de datos profesionales aprecian considerablemente. Una vez que se sabe manejar el BMDP clásico basado en instrucciones, curiosamente toda una serie de análisis son más sencillos y más rápidos de ordenar desde este formato clásico que desde versiones WINDOWS. Para una comprensión de la versatilidad y la potencia de cálculo estadístico del paquete es necesario revisar los volúmenes de la documentación original del programa.

En la práctica, además, se pueden combinar perfectamente diferentes paquetes estadísticos dado que existen procedimientos para intercambiar archivos de datos entre paquetes con relativa facilidad, lo que permite al profesional optar por la mejor solución y la mejor presentación para cada análisis concreto. Esta intercambiabilidad de los archivos de datos también tiene una gran importancia práctica, porque permitirá en el desempeño profesional reanalizar con BMDP datos preparados para otros paquetes y viceversa.