

1 Prácticas de Tablas de Contingencia. Programa 4F.

1.1 Introducción.

Como sabemos, las tablas de frecuencias se emplean para resumir resultados acerca de datos de tipo categórico. En cuanto a las categorías, pueden venir dadas de múltiples formas:

- Desordenadas (varón-hembra, respuestas a una pregunta con elecciones múltiples para las respuestas,...).
- Ordenadas por rangos de preferencias.
- Por agrupaciones de datos continuos en intervalos (intervalos de edad, peso,...).

El programa **4F** de BMDP permite analizar tablas de frecuencias de dos vías (o sea, indexadas por dos variables categóricas), multivía (indexadas por más de dos variables categóricas) o secciones cruzadas de una tabla multivía.

Por ejemplo, consideremos las variables *Empleo* y *Fumador*. Supongamos que la primera viene dada por tres categorías (privado, público, ninguno) y la segunda por dos (sí y no). Entonces la tabulación de *Empleo* por *Fumador* es un ejemplo de una tabla de dos vías. Si esta tabulación se hace por separado para cada sexo, cada una de las tablas de dos vías resultante es una sección cruzada de una tabla de tres vías (*Empleo* por *Fumador* y por *Sexo*). De forma similar, una tabla de cuatro vías puede separarse en secciones cruzadas de tres vías o de dos vías.

4F consta de tres partes bien diferenciadas:

- La primera permite examinar las distintas tablas de porcentajes (por fila, columna y totales), Asimismo, para tablas de doble entrada, permite estudiar la dependencia o no de las variables mediante el contraste de χ^2 .
- La segunda está centrada prácticamente en el análisis de tablas de dos vías, mostrando 26 medidas de asociación distintas, así como diversos contrastes basados en dichas medidas.
- La tercera parte permite ajustar modelos log-lineales a tablas multivía.

Nosotros nos vamos a centrar exclusivamente en el primero de estos apartados puesto que sólo estamos preocupados en *preparar* tablas de doble entrada para un posterior Análisis de Correspondencias Simple. Dentro de las características principales merecen destacarse las siguientes cuestiones:

- En cuanto a los datos, éstos pueden ser introducidos, principalmente, de dos formas
 - Introduciendo los datos originales.
 - Introduciendo la tabla ya formada.
- La salida del programa incluye:
 - Tablas de frecuencias observadas.
 - Un breve estudio descriptivo de las variables consideradas.
 - Lista de datos no incluidos en el estudio, mostrando las razones para su no inclusión.
 - Tablas de porcentajes por fila, por columna y totales.
- Asimismo, **4F** permite:
 - Formar subtablas marginales de una tabla multivía.
 - Agrupar variables combinando algunas categorías en una sola.

1.2 Órdenes específicas de 4F.

En cada párrafo se comentan las principales opciones y comentarios que sobre las instrucciones de cada uno incluye **4F**.

1. Párrafo INPUT.

- **TABLE=List. #**: Se emplea cuando se introduce una tabla multivía. Los valores de la lista corresponden al número de niveles de cada una de las variables, o sea, los índices que definen la tabla. Su producto debe coincidir con el número de celdillas que hay en la tabla. El primer valor es el número de niveles del índice que varía más rápido y que suele coincidir con el número de columnas de la tabla.
- **VARIABLES=#**: El número de variables ha de ser igual al de índices por lo menos.
- **CONTENT= (DATA, COUNT#)**: La primera de las opciones permite leer filas de datos desde un fichero en ASCII, mientras que la segunda lee la tabla de frecuencias ya formada de un fichero en formato BMDP, siendo # el número de secuencia en el que la tabla fue grabada en el fichero BMDP. El valor por defecto es **DATA**.

2. Párrafo VARIABLE

- **FREQUENCY=Var.:** Var es el número de orden de una variable o el nombre de una variable que contiene la frecuencia de los casos, si fuera preciso su empleo. Todo caso con frecuencia menor o igual que cero es excluido de los cálculos posteriores. Se emplea cuando la tabla se introduce indicando los índices de la celdilla correspondiente y la frecuencia observada en dicha celdilla.

3. Párrafo CATEGORY (o GROUP)

- **CODES(var)=List.:** Indica los códigos de las variables categóricas.
- **CUTPOINTS(var)=List.:** Indica los puntos de corte para definir intervalos de agrupación a partir de variables numéricas. Cada intervalo se encuentra cerrado por la derecha.
- **NAMES(var)=List.:** Asigna nombres a las categorías de cada variable.

4. Párrafo TABLE

Este párrafo es obligatorio y puede ser repetido varias veces. Define las tablas a formar a partir de los datos. Para formar tablas de doble entrada se pueden emplear las instrucciones **ROW** y **COLUMN**, mientras que para tablas multivía se utilizan las órdenes **INDEX** o **CATVAR**.

- **COLUMN=List:** Indica los nombres o números de orden de las variables que definen las categorías columna. **4F** forma una tabla separada para cada par de variables en la lista de **ROW** y **COLUMN**. Si una variable **COLUMN** tiene más de 10 categorías, deben especificarse sus códigos o puntos de corte en el párrafo **CATEGORY**.
- **ROW=List:** Indica los nombres o números de orden de las variables que definen las categorías fila. **4F** forma una tabla separada para cada par de variables en la lista de **ROW** y **COLUMN**. Si una variable **ROW** tiene más de 10 categorías, deben especificarse sus códigos o puntos de corte en el párrafo **CATEGORY**.
- **INDICES=List:** Indica los nombres o números de orden de las variables que definen una tabla multivía, teniendo en cuenta que el primer valor corresponde al índice que varía de forma más rápida y forma las columnas de la tabla.
- **PAIR. o CROSS.:**
 - **CROSS.:** Forma todas las posibles combinaciones de las variables fila y columna para definir tablas de doble entrada. Cada variable columna se cruza con todas las variables fila.
 - **PAIR.:** Forma tablas de doble entrada emparejando la primera variable columna con la primera variable fila y así sucesivamente. Es el valor por defecto.
- **CONDITION=List:** Indica los nombres o números de orden de las variables para estratificar tablas. Se emplea para formar tablas separadas para cada categoría de la variable **CONDITION**.
- **COUNT=List:** Se emplea cuando se introducen las frecuencias de cada celdilla directamente. Indica los nombres o números de orden de las variables que contienen tales frecuencias, formándose tablas separadas para cada variable de la lista.
- **STACK=List:** Indica los nombres o números de orden de las variables que son usadas para formar una nueva variable cuyas categorías consisten en todas las posibles combinaciones de los niveles de las variables empleadas para ello.

5. Párrafo PRINT

El párrafo PRINT es opcional. Si el párrafo se omite, **4F** imprime automáticamente las tablas de frecuencias observadas y parte de la lista de casos excluidos. Debe repetirse tras cada párrafo **TABLE**.

- **OBSERVED.:** Imprime las tablas de frecuencias observadas mientras no se especifique **NO OBSERVED**.
- **EXCLUDED.:** Mientras no se especifique **NO EXCLUDED.**, imprime una tabla de doble entrada con las frecuencias de los datos excluidos así como las razones para dicha exclusión.
- **PERCENT=(NONE, ROW, COL, TOTAL.):** Pueden elegirse una o más de estas opciones.
 - **NONE.:** No imprime ninguna tabla.
 - **ROW.:** Imprime los porcentajes por fila.
 - **COL.:** Imprime los porcentajes por columna.
 - **TOTAL.:** Imprime los porcentajes totales.

El valor por defecto es **NONE**.

- **MARGINAL=#.:** Se emplea para imprimir todas las subtablas marginales para todas las dimensiones hasta la especificada. La tabla marginal más grande es la tabla observada y se imprime por defecto.
- **LIST=#.:** Lista el número de casos que no deben incluirse en una o más tablas. El valor por defecto es 5.

- **BAR='Carácter'**.: Indica el carácter a imprimir en los diagramas verticales para las tablas de doble entrada y multivía. El valor por defecto es una barra vertical.

6. Párrafo SAVE

El párrafo SAVE permite la creación de un fichero en formato BMDP que contenga una o varias de las tablas estudiadas.

- **CONTENT=(DATA,COUNT)**: Especifica la matriz a grabar en el fichero. Cuando se graban frecuencias (**COUNT**), se genera un número de orden para cada tabla almacenada. El valor por defecto es **DATA**.

1.3 Ejemplos de aplicación.

1.3.1 Ejemplo 1.

Los datos correspondientes a tablas de doble entrada pueden venir dados de forma bruta, esto es, las respuestas concretas de los individuos a los cuestionarios planteados, o bien elaborados, esto es, las frecuencias observadas para cada pareja de categorías de las variables tratadas. Trabajaremos en esta práctica con tablas del segundo tipo.

El objetivo de esta práctica es aprender la lectura de datos con el capítulo **4F** de **BMDP**. Para ello vamos, en primer lugar, a realizar cuatro ejemplos haciendo uso de una tabla de doble entrada ya elaborada. En ellos veremos:

1. Entrada de la tabla de frecuencias elaborada.
2. Entrada de la tabla dando los índices de las celdillas y las frecuencias correspondientes a cada celdilla.
3. Grabación en un fichero **BMDP** de la tabla de frecuencias.
4. Lectura de la tabla de frecuencias desde un fichero **BMDP** previamente grabado.

El ejemplo que vamos a considerar es el siguiente:

*Se clasifican 6800 varones atendiendo a dos características, el color del pelo (variable **PELO**) y el color de los ojos (variable **OJOS**), con la idea de estudiar si existe o no dependencia entre el color del cabello y el color de los ojos. Los datos resultantes se resumen en la tabla siguiente:*

		PELO			
OJOS		Rubio	Moreno	Rojo	Cano
	Azul	1768	807	189	47
	Verde	946	1387	746	53
	Pardo	115	438	288	16

1. Entrada de la tabla elaborada.

En este ejemplo tenemos una tabla de doble entrada formada por cuatro columnas y tres filas. Los datos, bien introducidos al final de las instrucciones del programa, bien introducidos en un fichero, estarán dispuestos de la siguiente forma en formato libre:

```

1768  807  189  47
 946 1387  746  53
 115  438  288  16

```

Si se introducen en el fichero A:\4F\4F1-1.ASC, el fichero de órdenes quedaría de la forma siguiente:

```

/PROBLEM      TITLE IS 'PRACTICA 1 DE 4F. EJEMPLO 1.'.
/INPUT        VARIABLES = 2.
              FILE IS 'A:\4F\4F1-1.ASC'.
              FORMAT IS FREE.
              TABLE = 4,3.
/VARIABLE NAMES ARE PELO, OJOS.
/CATEGORY NAMES(PELO) = RUBIO, MORENO, ROJO, CANO.
              NAMES(OJOS) = AZUL, VERDE, PARDO.
/TABLE        ROW = OJOS.
              COL = PELO.
/PRINT        PERCENT = ROW, COL, TOTAL.
/END

```

Si se incluyesen al final de las instrucciones, debería eliminarse la línea:

FILE IS 'A:\4F\4F1-1.ASC'.

(Observemos que los datos no son los valores de la encuesta sino las frecuencias observadas, por lo que en el párrafo CATEGORY no hay que especificar los códigos de las respuestas).

- *Crea un fichero llamado 4F1-1.ASC y almacena en él los datos anteriores.*
- *Crea un fichero llamado 4F1-1.INP y almacena en él las órdenes anteriores.*
- *Ejecuta el programa en modo interactivo.*

2. Entrada de la tabla dando los índices de las celdillas y las frecuencias correspondientes a cada celdilla.

Los datos, en este caso, estarán dispuestos de la siguiente forma, en el fichero A:\4F\4F1-2.ASC:

1	1	1768
2	1	807
3	1	189
4	1	47
1	2	946
2	2	1387
3	2	746
4	2	53
1	3	115
2	3	438
3	3	288
4	3	16

El fichero de órdenes quedaría de la forma siguiente:

```
/PROBLEM
      TITLE IS 'PRACTICA 1 DE 4F. EJEMPLO 2.'.
/INPUT  VARIABLES = 3.
      FILE IS 'A:\4F\4F1-2.ASC'.
      FORMAT IS FREE.
/VARIABLE NAMES ARE PELO, OJOS, FREQ.
/CATEGORY NAMES(PELO) = RUBIO, MORENO, ROJO, CANO.
      NAMES(OJOS) = AZUL, VERDE, PARDO.
/TABLE  ROW = OJOS.
      COL = PELO.
      COUNT = FREQ.
/PRINT  PERCENT = ROW, COL, TOTAL.
/END
```

- *Crea un fichero llamado 4F1-2.ASC y almacena en él los datos anteriores.*
- *Crea un fichero llamado 4F1-2.INP y almacena en él las órdenes anteriores.*
- *Ejecuta el programa en modo interactivo.*

3. Grabación en un fichero BMDP de la tabla de frecuencias.

Nos proponemos ahora salvar en un fichero **BMDP** la tabla de frecuencias anterior. Ello puede ser de utilidad al emplear estos datos en estudios posteriores, bien con 4F, bien con otro capítulo distinto de **BMDP**.

El fichero de órdenes quedaría de la forma siguiente:

```
/PROBLEM
      TITLE IS 'PRACTICA 1 DE 4F. EJEMPLO 3.'.
/INPUT  VARIABLES = 2
      FILE IS 'A:\4F\4F1-1.ASC'.
      FORMAT IS FREE.
      TABLE = 4,3.
/VARIABLE NAMES ARE PELO, OJOS.
/CATEGORY NAMES(PELO) = RUBIO, MORENO, ROJO, CANO.
      NAMES(OJOS) = AZUL, VERDE, PARDO.
/TABLE  ROW = OJOS.
      COL = PELO.
```

```

/SAVE      CONTENT = COUNT.
          FILE IS 'A:\4F\4F1.BMD'.
          CODE IS '4F1'.
          NEW.

/END

```

- Crea un fichero llamado **4F1-3.INP** y almacena en él las órdenes anteriores.
- Ejecuta el programa en modo interactivo.

4. Lectura de la tabla de frecuencias desde un fichero BMDP previamente grabado.

Volvemos a realizar el análisis, leyendo ahora los datos desde el fichero **BMDP** creado anteriormente.

El fichero de órdenes quedaría de la forma siguiente:

```

/PROBLEM
TITLE IS 'PRACTICA 1 DE 4F. EJEMPLO 4.'.
/INPUT  FILE IS 'A:\4F\4F1.BMD'.
        CODE IS '4F1'.
        CONTENT IS COUNT1.
/TABLE  ROW = OJOS.
        COL = PELO.
        COUNT = FREQ.
/PRINT  PERCENT = ROW, COL, TOTAL.
/END

```

- Crea un fichero llamado **4F1-4.INP** y almacena en él las órdenes anteriores.
- Ejecuta el programa en modo interactivo.

1. A la vista de dichas tablas responde a las siguientes cuestiones:

- De las personas con pelo rubio, ¿qué porcentaje tienen los ojos azules?
- De las personas con ojos azules, ¿qué porcentaje tienen pelo rubio?
- De las personas con pelo cano, ¿qué porcentaje tienen los ojos pardos?
- De las personas con los ojos pardos, ¿qué porcentaje tienen el pelo cano?
- ¿Qué porcentaje de personas, de entre el total de la población, tienen el pelo cano y los ojos pardos simultáneamente?

2. ¿Es admisible la hipótesis de independencia de los dos caracteres estudiados, al nivel de significación $\alpha = 0.05$? ¿Por qué? Relaciona la respuesta dada con lo observado en las tablas de frecuencias anteriores.

Salida del Programa proporcionada por BMDP

```

*****
* TABLE PARAGRAPH 1 *
*****

```

***** OBSERVED FREQUENCY TABLE 1

```

VARIABLE 3 FREQ USED AS COUNT VARIABLE.
*****

```

OJOS	PELO				
	RUBIO	MORENO	ROJO	CANO	TOTAL
AZUL	1768	807	189	47	2811
VERDE	946	1387	746	53	3132
PARDO	115	438	288	16	857
TOTAL	2829	2632	1223	116	6800

***** PERCENTS OF ROW TOTALS -- TABLE 1

VARIABLE 3 FREQ USED AS COUNT VARIABLE.

OJOS	PELO				TOTAL
	RUBIO	MORENO	ROJO	CANO	
AZUL	62.9	28.7	6.7	1.7	100.0
VERDE	30.2	44.3	23.8	1.7	100.0
PARDO	13.4	51.1	33.6	1.9	100.0
TOTAL	41.6	38.7	18.0	1.7	100.0

***** PERCENTS OF COLUMN TOTALS -- TABLE 1

VARIABLE 3 FREQ USED AS COUNT VARIABLE.

OJOS	PELO				TOTAL
	RUBIO	MORENO	ROJO	CANO	
AZUL	62.5	30.7	15.5	40.5	41.3
VERDE	33.4	52.7	61.0	45.7	46.1
PARDO	4.1	16.6	23.5	13.8	12.6
TOTAL	100.0	100.0	100.0	100.0	100.0

***** PERCENTS OF THE TABLE TOTAL -- TABLE 1

VARIABLE 3 FREQ USED AS COUNT VARIABLE.

OJOS	PELO				TOTAL
	RUBIO	MORENO	ROJO	CANO	
AZUL	26.0	11.9	2.8	0.7	41.3
VERDE	13.9	20.4	11.0	0.8	46.1
PARDO	1.7	6.4	4.2	0.2	12.6
TOTAL	41.6	38.7	18.0	1.7	100.0

***** ANALYSIS OF OBSERVED FREQUENCY TABLE 1

MINIMUM ESTIMATED EXPECTED VALUE IS 14.62

STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	1073.508	6	0.0000

- Observando la tabla de porcentajes por columna se observa que el 62.5% de las personas con el pelo rubio tienen los ojos azules.
- Observando la tabla de porcentajes por fila se observa que el 62.9% de las personas con los ojos azules tienen el pelo rubio.
- Observando la tabla de porcentajes por columna se observa que el 13.8% de las personas con el pelo cano tienen los ojos pardos.
- Observando la tabla de porcentajes por fila se observa que el 1.9% de las personas con los ojos pardos tienen el pelo cano.

- Observando la tabla de porcentajes por fila se observa que el 0.2% del total de la población, tienen el pelo cano y los ojos pardos.
- A la vista de los resultados obtenidos mediante el test de χ^2 , no hay razones empíricas suficientes para admitir la independencia de los dos caracteres estudiados. Esta situación puede corroborarse, por ejemplo, observando la tabla de porcentajes por columna; si en dicha tabla nos fijamos en la distribución del color de los ojos condicionada al color del pelo rubio, observamos como va disminuyendo la proporción de individuos a medida que se va oscureciendo el color de los ojos. Por otra parte, fijándonos ahora en la tabla de porcentajes por fila observamos un fenómeno parecido en la distribución del color del pelo condicionada al color azul de los ojos. Estas dos cuestiones hacen ver la dependencia que tiene que haber entre ambas características.

1.3.2 Ejemplo 2

En este ejemplo vamos a partir de los datos originales, estableciendo distintas categorías en las variables mediante la agrupación por intervalos.

Concretamente los datos se refieren a 46 personas sobre las cuales se han medido las siguientes variables:

- **Peso** (medido en kg.)
- **Altura** (medida en cm.)
- **Sexo** (0: Mujer; 1: Varón)

Peso	Altura	Sexo	Peso	Altura	Sexo
52	160	0	49	160	0
50	159	0	65	170	0
50	160	0	96	198	1
57	163	0	82	178	1
64	171	1	66	170	1
67	174	1	74	185	1
83	185	1	50	165	0
59	171	1	55	165	0
75	170	1	57	168	0
58	162	0	79	162	1
56	163	0	80	183	1
52	165	0	72	176	1
56	167	0	55	164	0
77	165	0	74	172	1
74	179	1	80	176	1
73	178	1	82	172	1
60	170	0	61	163	0
112	187	1	74	176	1
60	167	0	58	164	0
57	161	0	66	175	1
43	164	0	63	165	0
50	160	0	51	160	0
47	163	0	70	180	1

Las categorías que se van a establecer para las variables **Altura** y **Peso** son las siguientes:

- **Peso:** ≤ 60 , $60 - 80$, > 80 .
- **Altura:** ≤ 165 , $165 - 180$, > 180 .

Los pasos que vamos a seguir en esta práctica son los siguientes:

- Grabar los datos anteriores en un fichero llamado **A:\4F\4F2.ASC**.
- Elaborar un fichero de órdenes que realice las siguientes tareas:
 - Leer los datos del fichero **A:\4F\4F2.ASC**.
 - Categorizar las variables atendiendo a las especificaciones anteriores.
 - Obtener las tablas observadas y de frecuencias por filas, por columnas y totales para las tres variables de forma conjunta, así como para todas las subtablas de orden dos.
 - Almacenar las tablas de frecuencias observadas en un fichero con formato BMDP.

- Elaborar otro fichero de órdenes que permita obtener las tablas conjuntas del **Peso** y la **Altura** condicionadas a los valores de la variable **Sexo**.

Los ficheros de órdenes respectivos los llamaremos **4F2-1.INP** y **4F2-2.INP** y sus contenidos serán los siguientes:

- **Fichero 4F2-1.INP**

```

/INPUT
    TITLE IS 'PRACTICA 2 DE 4F. EJEMPLO 1.'.
    VARIABLES = 3.
    FORMAT IS FREE.
    FILE IS 'A:\4F\4F2.ASC'.

/VARIABLE
    NAMES=PESO,ALTURA,SEXO.

/CATEGORY
    CUTPOINT(PESO)=60,80.
    NAMES(PESO) = '<= 60','60 - 80', '> 80'.
    CUTPOINT(ALTURA)=165,180.
    NAMES(ALTURA) ='<= 165','165 - 180','> 180'.
    CODE(SEXO)=0,1.
    NAMES(SEXO) = MUJER,VARON.

/TABLE
    ROW =PESO,ALTURA.
    COL =SEXO,ALTURA.
    CROSS.

/PRINT
    PERCENT=ROW,COL,TOTAL.

/TABLE
    INDICES=PESO,ALTURA,SEXO.

/PRINT
    PERCENT=ROW,COL,TOTAL.

/SAVE
    CONTENT=COUNT.
    FILE IS 'A:\4F\4F2.BMD'.
    CODE='4F2'.
    NEW.

/END

```

- **Fichero 4F2-2.INP**

```

/INPUT
    TITLE IS 'PRACTICA 2 DE 4F. EJEMPLO 2.'.
    FILE IS 'A:\4F\4F2.BMD'.
    CODE ='4F2'.
    CONTENT= COUNT5.

/TABLE
    ROW=PESO.
    COL=ALTURA.
    CONDITION=SEXO.
    COUNT=FREQ.

/PRINT
    PERCENT=ROW,COL,TOTAL.

/END

```

Notemos que en el primer ejemplo la tabla se ha obtenido en quinto lugar, por lo que es necesario especificar en el párrafo **INPUT** del segundo ejemplo la orden **COUNT=5**.

Salida del Programa proporcionada por BMDP

```

*****      IN THE  1-TH TABLE PARAGRAPH, THE    4-TH TABLE IS
              OMITTED SINCE THE SAME VARIABLE APPEARS AS TWO INDICES.

```

```

BMDP FILE IS BEING WRITTEN
CODE. . . IS      4F2

```


FILE NAME IS A:\4F\4F2.BMD
CONTENT . IS COUNT1
VARIABLES ARE
1 SEXO 2 PESO 3 FREQ

BMDP FILE HAS BEEN COMPLETED.
FILE NAME IS A:\4F\4F2.BMD

BMDP FILE IS BEING WRITTEN
CODE. . . IS 4F2
FILE NAME IS A:\4F\4F2.BMD
CONTENT . IS COUNT2
VARIABLES ARE
1 SEXO 2 ALTURA 3 FREQ

BMDP FILE HAS BEEN COMPLETED.
FILE NAME IS A:\4F\4F2.BMD

BMDP FILE IS BEING WRITTEN
CODE. . . IS 4F2
FILE NAME IS A:\4F\4F2.BMD
CONTENT . IS COUNT3
VARIABLES ARE
1 ALTURA 2 PESO 3 FREQ

BMDP FILE HAS BEEN COMPLETED.
FILE NAME IS A:\4F\4F2.BMD

BMDP FILE IS BEING WRITTEN
CODE. . . IS 4F2
FILE NAME IS A:\4F\4F2.BMD
CONTENT . IS COUNT5
VARIABLES ARE
1 PESO 2 ALTURA 3 SEXO 4 FREQ

BMDP FILE HAS BEEN COMPLETED.
FILE NAME IS A:\4F\4F2.BMD

* TABLE PARAGRAPH 1 *

***** OBSERVED FREQUENCY TABLE 1

PESO	SEXO		TOTAL
	MUJER	VARON	
< 60	21	1	22
60 - 80	4	15	19
> 80	0	5	5
TOTAL	25	21	46

***** ANALYSIS OF OBSERVED FREQUENCY TABLE 1

MINIMUM ESTIMATED EXPECTED VALUE IS 2.28

STATISTIC VALUE D.F. PROB.

PEARSON CHISQUARE 29.425 2 0.0000

***** OBSERVED FREQUENCY TABLE 2

ALTURA	SEXO		
-----	-----		
	MUJER	VARON	TOTAL
-----	-----	-----	-----
< 165	20	1	21
165 - 18	5	15	20
> 180	0	5	5
-----	-----	-----	-----
TOTAL	25	21	46

***** ANALYSIS OF OBSERVED FREQUENCY TABLE 2

MINIMUM ESTIMATED EXPECTED VALUE IS 2.28

STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	27.047	2	0.0000

***** OBSERVED FREQUENCY TABLE 3

PESO	ALTURA			
-----	-----			
	< 165	165 - 18	> 180	TOTAL
-----	-----	-----	-----	-----
< 60	17	5	0	22
60 - 80	4	13	2	19
> 80	0	2	3	5
-----	-----	-----	-----	-----
TOTAL	21	20	5	46

***** ANALYSIS OF OBSERVED FREQUENCY TABLE 3

MINIMUM ESTIMATED EXPECTED VALUE IS 0.54

STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	28.028	4	0.0000

 * TABLE PARAGRAPH 2 *

***** OBSERVED FREQUENCY TABLE 5

SEXO	ALTURA	PESO			
-----	-----	-----			
		< 60	60 - 80	> 80	TOTAL
-----	-----	-----	-----	-----	-----
MUJER	< 165	17	3	0	20
	165 - 18	4	1	0	5
	> 180	0	0	0	0
	-----	-----	-----	-----	-----
	TOTAL	21	4	0	25
VARON	< 165	0	1	0	1
	165 - 18	1	12	2	15
	> 180	0	2	3	5
	-----	-----	-----	-----	-----
	TOTAL	1	15	5	21

***** PERCENTS OF ROW TOTALS -- TABLE 5

SEXO	ALTURA	PESO			
-----	-----		-----		
		< 60	60 - 80	> 80	TOTAL
-----	-----	-----	-----	-----	-----
MUJER	< 165	85.0	15.0	0.0	100.0
	165 - 18	80.0	20.0	0.0	100.0
	> 180	0.0	0.0	0.0	0.0
	-----	-----	-----	-----	-----
	TOTAL	84.0	16.0	0.0	100.0
VARON	< 165	0.0	100.0	0.0	100.0
	165 - 18	6.7	80.0	13.3	100.0
	> 180	0.0	40.0	60.0	100.0
	-----	-----	-----	-----	-----
	TOTAL	4.8	71.4	23.8	100.0

***** PERCENTS OF COLUMN TOTALS -- TABLE 5

SEXO	ALTURA	PESO			
-----	-----		-----		
		< 60	60 - 80	> 80	TOTAL
-----	-----	-----	-----	-----	-----
MUJER	< 165	81.0	75.0	0.0	80.0
	165 - 18	19.0	25.0	0.0	20.0
	> 180	0.0	0.0	0.0	0.0
	-----	-----	-----	-----	-----
	TOTAL	100.0	100.0	0.0	100.0
VARON	< 165	0.0	6.7	0.0	4.8
	165 - 18	100.0	80.0	40.0	71.4
	> 180	0.0	13.3	60.0	23.8
	-----	-----	-----	-----	-----
	TOTAL	100.0	100.0	100.0	100.0

***** PERCENTS OF THE TABLE TOTAL -- TABLE 5

SEXO	ALTURA	PESO			
-----	-----		-----		
		< 60	60 - 80	> 80	TOTAL
-----	-----	-----	-----	-----	-----
MUJER	< 165	37.0	6.5	0.0	43.5
	165 - 18	8.7	2.2	0.0	10.9
	> 180	0.0	0.0	0.0	0.0
	-----	-----	-----	-----	-----
	TOTAL	45.7	8.7	0.0	54.3
VARON	< 165	0.0	2.2	0.0	2.2
	165 - 18	2.2	26.1	4.3	32.6
	> 180	0.0	4.3	6.5	10.9
	-----	-----	-----	-----	-----
	TOTAL	2.2	32.6	10.9	45.7

1.4 1º Ejercicio propuesto

Se desea estudiar la posible influencia, en el cultivo de un determinado producto agrícola de invernadero, del tipo de tierra donde se cultiva, la clase de abono empleado y el sistema de riego utilizado. Para ello se eligen 300 parcelas de terreno y, al final de la experiencia, se encuesta al agricultor responsable de cada parcela sobre las siguientes cuestiones:

1. **Tipo de tierra empleada.** Dicha variable será notada (**TIERRA**) y toma los valores (**A**), (**B**) y (**C**).
2. **Tipo de abono utilizado.** Esta variable se llamará (**ABONO**) y toma dos valores: (**A1**) y (**A2**).
3. **Sistema de riego.** A esta variable se le llamará (**RIEGO**) y toma los valores 1 para denotar el riego por aspersión y 2 el riego por goteo.
4. **Nivel de producción.** Esta variable, llamada **PROD**, mide el número de kilos, expresado en miles, obtenidos y está categorizada de la siguiente forma:

- 1: Menor o igual que 1000 kg.

- 2: Mayor que 1000 y menor o igual que 3000.
- 3: Mayor o igual que 3000.

Los resultados que se obtuvieron fueron:

ABONO TIERRA RIEGO			PROD.		
A1			≤ 1	1-3	> 3
	A	Asper.	6	5	2
		Goteo	4	10	0
	B	Asper.	1	3	3
		Goteo	9	44	10
	C	Asper.	1	1	1
		Goteo	4	10	8
A2					
	A	Asper.	2	6	2
		Goteo	1	17	10
	B	Asper.	4	12	8
		Goteo	6	28	25
	C	Asper.	1	7	10
		Goteo	3	16	20

Pasos a realizar en el ejercicio

- Almacena la tabla anterior en un fichero llamado **4F3.ASC**.
- Elabora un fichero de órdenes llamado **4F3-1.INP** que permita grabar la tabla anterior en formato BMDP.
- Elabora un fichero de órdenes llamado **4F3-2.INP** con el cual obtener:
 - Tablas de frecuencias por filas, columnas y totales a partir de la tabla original.
 - Todas las subtablas marginales de orden inferior o igual a tres.
 - Copia las tablas de frecuencias por filas, columnas y totales.
 - A la vista de dichas tablas responde a las siguientes cuestiones:
 - * De entre las parcelas tratadas con abono **A2** y tierra **A**, ¿que porcentaje tiene una producción superior a 3000 kg.? ¿Podrías haber deducido este dato observando otra tabla de las que se obtienen en la salida? ¿Cuál sería esa tabla?
 - * De entre las parcelas tratadas con abono **A1**, tierra **B** y riego por aspersión, ¿que porcentaje presenta una producción inferior o igual a 1000 kg.?
 - * De entre las parcelas tratadas con abono **A2**, tierra **C** y que producen más de 3000 kg, ¿que porcentaje posee riego por goteo?
 - * ¿Que porcentaje, del total de la parcelas estudiadas, producen más de 3000 kg, están tratadas con el abono **A1**, poseen tierra del tipo **C** y son regadas por aspersión?
 - * ¿Cuántas parcelas están tratadas con la tierra **C**?
 - * ¿Cuántas parcelas producen entre 1000 y 3000 kg. y son regadas por aspersión?
- A continuación nos proponemos estudiar la dependencia de la producción sobre cada una de las variables restantes. Para ello, confecciona un fichero de órdenes llamado **4F3-3.INP** y estudia la dependencia o no de las variables en cada uno de los emparejamientos producidos y comenta los resultados obtenidos.
- Como ejercicio final nos planteamos la posibilidad de combinar las variables **RIEGO** y **TIERRA** en una sola. Para ello, crea un fichero llamado **4F3-4.INP** que permita tal fusión. A la vista de los resultados obtenidos responde a las siguientes cuestiones
 - ¿Cuántas parcelas tienen una producción superior a 3000 kg. y están tratadas con el abono **A2**? De entre ellas, ¿que porcentaje tienen tierra de la clase **B** y riego por goteo?
 - ¿Cuántas parcelas están tratadas con el abono **A1**, poseen tierra de la clase **C** y son regadas por goteo? De entre ellas, ¿que porcentaje posee una producción entre 1000 y 3000 kg.?

1.5 2º Ejercicio propuesto

De un estudio de 465 mujeres realizado por Fidell and Prather se estudian las siguientes variables:

- Autoestima, (**Autoestima**): Valorada en una escala de 1 a 22.
Se establecen tres modalidades: Baja (**A_baja**), Media (**A_media**) y Alta (**A_alta**), para valores menores o iguales a 11, entre 11 y 16, y mayores de 16 respectivamente.
- Satisfacción en su matrimonio, (**Sa_matrimonio**): Valorada de 1 a 48.
Se establecen tres modalidades: Baja (**S_baja**), Media (**S_media**) y Alta (**S_alta**), para valores menores o iguales a 33, entre 33 y 41, y mayores de 41 respectivamente.
- Actitud hacia el papel de la mujer (**Papel_mujer**): Valorada de 1 a 38.
Se establecen tres modalidades: **Conservadora**, **Moderada** y **Liberal**, para valores menores o iguales a 13 entre 13 y 26, y mayores de 26.
- Nivel de estudios (**Estudios**): Valorada de 1 a 25.
Las categorías son: **Primarios**, **Bachillerato**, **Diplomada** y **Licenciada**, para valores menores o iguales a 11, entre 11 y 12, entre 12 y 15, y mayores a 15 respectivamente.
- Tipo de trabajo (**Trabajo**): De 0 a 3.
Las categorías son: Fuera de casa (**Fuera**) y En casa (**Casa**), cero para trabajo fuera del hogar y mayor que cero para trabajos en el hogar.
- Estado civil (**E_civil**): **Soltera** con el código 1 y **Casada** codificada con 2.

Los datos están organizados en un registro por mujer en el fichero 4F4.ASC, con 9 variables por registro, siendo las seis utilizadas las números 1, 2, 3, 5, 6 y 7. Dicho fichero se le proporcionará al alumno.

Se pide:

- Elabora el fichero de órdenes 4F4-1.INP que lea la tabla de orden 6, formada por las variables: Sa_matrimonio, Papel_mujer, Trabajo, Estudios, Autoestima y E_civil, del fichero de datos 4F4.ASC y la grabe en formato BMDP en el fichero 4F4.BMD
- Elabora el fichero de órdenes 4F4-2.INP que lea dicha tabla del formato BMDP y observando únicamente la tabla de porcentajes por columnas te permita contestar a las siguientes cuestiones, utilizando tantos párrafos TABLE y órdenes INDICES como sean necesarias.
- Elabora el fichero 4F4-3.INP que te permita contestar a las mismas cuestiones pero observando únicamente la tabla de porcentajes por filas.
 1. ¿Cuántas mujeres y qué porcentaje de entre las que trabajan fuera del hogar, tienen estudios primarios?
 2. ¿Cuántas mujeres y qué porcentaje de entre las que trabajan fuera del hogar y tienen alta autoestima son licenciadas?
 3. ¿Cuántas mujeres y qué porcentaje de entre las que trabajan fuera del hogar, tienen una baja autoestima y están casadas?
 4. ¿Cuántas mujeres y qué porcentaje de entre las que trabajan fuera del hogar, tienen una baja autoestima, están casadas y tienen estudios de bachillerato?
 5. ¿Cuántas mujeres licenciadas y qué porcentaje trabajan en el hogar, tienen una alta autoestima y están solteras?
 6. De entre las mujeres licenciadas que trabajan fuera del hogar y tienen una baja autoestima, ¿cuántas y qué porcentaje están casadas y cuántas solteras?
 7. De entre las mujeres licenciadas que tienen una alta autoestima y trabajan fuera del hogar, ¿cuántas y qué porcentaje están casadas?
 8. ¿Cuántas y qué porcentaje de entre las licenciadas que trabajan fuera, son solteras con baja autoestima?
 9. ¿Cuántas y qué porcentaje de entre las licenciadas que trabajan en el hogar, están casadas y tienen una alta autoestima?
 10. De entre las que tienen una alta satisfacción en el matrimonio, ¿cuántas y qué porcentaje tienen una alta autoestima, son licenciadas, trabajan fuera del hogar y su actitud hacia el papel de la mujer es conservador?

Respuestas a las cuestiones

Cuest.	Frec. sobre	Porcent.	Indices	Tabla de %	Stack
a	19 de 246	7.7%	Trab, Estud	Columna	No
a	19 de 246	7.7%	Estud, Trab	Fila	No
b	26 de 91	28.6%	Trab, Estud, Autoes	Columna	No
b	26 de 91	28.6%	Estud, Autoes, Trab	Fila	No
c	34 de 246	13.8%	Trab, E_ci, Autoes	Columna	E_ci, Autoes
c	34 de 246	13.8%	E_ci, Autoes, Trab	Fila	E_ci, Autoes
d	17 de 246	6.9%	Trab, Estud, E_ci, Autoes	Columna	Estud, E_ci, Autoes
d	17 de 246	6.9%	Estud, E_ci, Autoes, Trab	Fila	Estud, E_ci, Autoes
e	0 de 465	0%	Estud, E_ci, Autoes, Trab	Totales	No
f	3, 3 de 6	50%, 50%	Estud, E_ci, Autoes, Trab	Columna	No
f	3, 3 de 6	50%, 50%	E_ci, Autoes, Trab, Estud	Fila	No
g	18 de 26	69.2%	Estud, E_ci, Trab, Autoes	Columna	No
g	18 de 26	69.2%	E_ci, Trab, Autoes, Estud	Fila	No
h	3 de 51	5.9%	Estud, Autoes, E_ci, Trab	Columna	Autoes, E_ci
h	3 de 51	5.9%	Autoes, E_ci, Trab, Estud	Fila	Autoes, E_ci
i	10 de 25	40%	Estud, Autoes, E_ci, Trab	Columna	Autoes, E_ci
i	10 de 25	40%	Autoes, E_ci, Trab, Estud	Fila	Autoes, E_ci
j	0 de 151	0%	Sa_matri, Papel_mu, Trab Estud, Autoes	Columna	Papel_mu, Trab, Estud Autoes
j	0 de 151	0%	Papel_mu, Trab, Estud Autoes, Sa_matri	Fila	Papel_mu, Trab, Estud Autoes