

# Estadística Descriptiva con ordenador

## 1. Tablas y gráficas estadísticas con ordenador

El cálculo de tablas de frecuencias y porcentajes así como su representación gráfica puede realizarse de manera sencilla con la ayuda de herramientas informáticas. Estudios estadísticos simples pueden realizarse mediante hojas de cálculo (tipo Microsoft Excel o OpenOffice Calc). Análisis más complejos requieren el uso de herramientas más sofisticadas, como el software especializado en estadística SPSS o R.

En esta sección aprenderemos a obtener tablas y gráficas mediante hojas de cálculo. Utilizaremos el programa OpenOffice Calc, que es la versión de software libre de hoja de cálculo. El programa puede obtenerse de forma gratuita de <http://es.openoffice.org/> y se instala fácilmente en cualquier sistema operativo. La versión utilizada en los siguientes ejemplos es la 3.0.

### Ejemplo 1

Consideramos los siguientes datos obtenidos de la web del Instituto Nacional de Estadística. Calcularemos la tabla de frecuencias y porcentajes y haremos varias representaciones gráficas.

| Estadísticas judiciales 2005                                |     |
|---|-----|
| Estadística de lo Penal. Condenados. Resultados autonómicos |     |
| Condenados según edad y sexo                                |     |
| Unidades: nº de condenados                                  |     |
| Ambos sexos   |     |
| Baleares (Illes)  |     |
| De 18 a 20 años   | 155 |
| De 21 a 25 años   | 543 |
| De 26 a 30 años   | 653 |
| De 31 a 35 años   | 619 |
| De 36 a 40 años   | 515 |
| De 41 a 50 años   | 636 |
| De 51 a 60 años   | 248 |
| De 60 y más   | 100 |

FUENTE: Instituto Nacional de Estadística

Los pasos a seguir para calcular las tablas de frecuencias y porcentajes son los siguientes:

1. Abrir la aplicación OpenOffice Calc.

Se abrirá una ventana como la que se muestra en la figura 1.

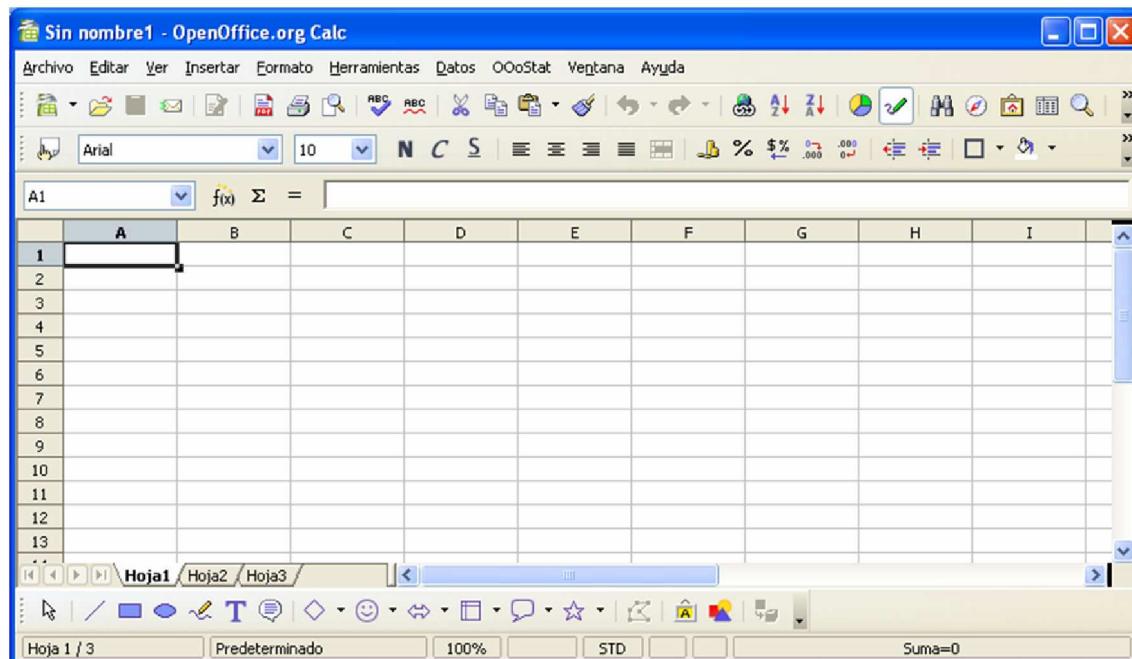


Figura 1: Ventana de inicio de OpenOffice Calc

2. Para introducir los datos del problema nos situamos sobre la casilla A1 (columna A, fila 1) moviéndonos con el cursor del ratón y escribimos en ella el título de la tabla: *Condenados Illes Balears por edad (año 2005)*. La fila 2 la dejamos en blanco para facilitar la lectura de la tabla.

A continuación escribimos en la casilla A3 *Edad (años)*, y en las posiciones inferiores de la misma columna: 18 – 20, 21 – 25, …, 51 – 60, más de 60. Para desplazarnos de una casilla a la siguiente podemos utilizar el ratón, las flechas del teclado o la tecla *Tab*.

Repetiremos la operación en la columna B. En la casilla B3 escribiremos *Nº condenados* y en las casillas inferiores: 155, 543, …, 100.

Si en algún momento deseamos rectificar alguno de los datos introducidos deberemos hacer doble clic sobre la casilla correspondiente y reintroducir el valor.

Después de este paso la hoja de cálculo tendrá el aspecto que se muestra en la figura 2.

|    | A  | B   | C | D |
|----|--|-----|---|---|
| 1  | Condenados Illes Balears por edad (año 2005) |     |   |   |
| 2  |  |     |   |   |
| 3  | Edad (años) Nº condenados                    |     |   |   |
| 4  | 18-20  | 155 |   |   |
| 5  | 21-25  | 543 |   |   |
| 6  | 26-30  | 653 |   |   |
| 7  | 31-35  | 619 |   |   |
| 8  | 36-40  | 515 |   |   |
| 9  | 41-50  | 636 |   |   |
| 10 | 51-60  | 248 |   |   |
| 11 | Más de 60                                    | 100 |   |   |
| 12 |  |     |   |   |

Figura 2: Hoja de cálculo trás la introducción de los datos del ejemplo 1

3. Los valores de la columna B (*nº* condenados) son las frecuencias absolutas de la variable *Edad*. Deseamos calcular las frecuencias relativas y los porcentajes. Además, como la variable *Edad* es cuantitativa podemos calcular también las frecuencias y porcentajes acumulados.

Empezamos por dar nombre a las columnas que mostrarán los valores calculados. Nos situamos sobre la casilla *C3* y escribimos *Frecuencia relativa*. Utilizando la tecla *Tab* o el ratón nos desplazaremos a la siguiente casilla de la misma fila (casilla *D4*) y escribiremos *Porcentaje*. Repitiendo el proceso escribiremos en las casillas *E5* a *H5* los valores: *Frecuencia absoluta acumulada*, *Frecuencia relativa acumulada* y *Porcentaje acumulado*.

Si el tamaño del texto escrito es mayor que la anchura de la columna el texto se sobreescibirá sobre las columnas vecinas. Para evitarlo podemos aumentar la anchura de las columnas situándonos sobre las líneas que separan las letras de las columnas



y desplazándolas con el cursor.

También podemos ajustar el texto automáticamente al tamaño de la columna situándonos sobre la columna a modificar y siguiendo los siguientes pasos: acceder a la opción *Formato* del menú principal, hacer clic sobre la opción *Celdas...*, se abrirá una nueva ventana en la que seleccionaremos la pestaña *Alineación* y haremos clic sobre la opción *Ajustar texto automáticamente* dentro del campo *Propiedades*.

Tras estos ajustes la fila 3 de la hoja de cálculo contiene los siguientes valores:

| 3 | Edad (años) | Nº condenados | Frecuencia relativa | Porcentaje | Frecuencia absoluta acumulada | Frecuencia relativa acumulada | Porcentaje acumulado |
|---|-------------|---------------|---------------------|------------|-------------------------------|-------------------------------|----------------------|
|   |             |               |                     |            |                               |                               |                      |

4. Calcularemos primero la suma de los valores de frecuencias absolutas, es decir, el número total de condenados. Escribiremos este valor al final de la

columna *B* (casilla *B12*). Para ello nos situaremos sobre esta casilla, escribiremos  $=SUMA(B4:B11)$  y pulsaremos la tecla *Enter*. El valor 3469 se mostrará en la casilla. La función **SUMA** es una función de Calc que permite sumar los valores de las casillas que se le indican (en nuestro caso desde la casilla *B4* hasta la *B11*).

5. Para calcular las frecuencias relativas debemos dividir las frecuencias absolutas entre la suma de las frecuencias. Para ello nos situaremos sobre la casilla *C4*, escribiremos  $=B4/$B$12$  y pulsaremos *Enter*. En la casilla aparece el valor 0,04, resultado de dividir el valor de las casillas *B4* y *B12*.

Podemos repetir la operación con el resto de las casillas de la columna pero Calc ofrece una manera más sencilla de hacer estas operaciones. Basta situarnos con el cursor sobre la esquina inferior derecha de la casilla *C4*, hacer clic con el botón izquierdo del ratón y, manteniendo el botón pulsado, arrastrar el cursor hasta la casilla *C11*. Al soltar el botón aparecen en las casillas los valores calculados (ver columna *C* en la figura 3), ya que Calc reescribe automáticamente la fórmula de la primera casilla para adaptarla a las casillas seleccionadas.

6. Los porcentajes se obtienen multiplicando las frecuencias relativas por 100. Para ello nos situamos sobre la casilla *D4*, escribimos  $=C4*100$  y pulsamos *Enter*. A continuación nos situamos con el cursor en la esquina inferior derecha de la casilla y, manteniendo el botón izquierdo del ratón pulsado, arrastramos el cursor hasta la casilla *D11*. Al soltar el botón los resultados se escriben en las casillas correspondientes (ver columna *D* en la figura 3).
7. Las frecuencias absolutas acumuladas se calculan sumando a la frecuencia absoluta del valor considerado las frecuencias absolutas de los valores anteriores. La frecuencia absoluta acumulada del primer valor (18–20) es igual a su frecuencia absoluta, por lo que en la casilla *E4* escribimos  $=B4$  y pulsamos *Enter*. En la casilla siguiente, *E5*, escribimos  $=E4+B5$  y pulsamos *Enter*. Las restantes casillas se calcularán automáticamente si situamos el cursor en la esquina inferior derecha de la casilla *E5* y, manteniendo el botón izquierdo del ratón pulsado, arrastramos el cursor hasta la casilla *E11*. Al soltar el botón se muestran los valores calculados (ver columna *E* en la figura 3).
8. Las frecuencias relativas acumuladas se calculan dividiendo las frecuencias absolutas acumuladas entre la frecuencia absoluta total. Para ello escribimos en la casilla *F4*  $=E4/$B$12$  y pulsamos *Enter*. Repitiendo el procedimiento explicado en los casos anteriores extendemos el cálculo hasta la casilla *F11* (el resultado se muestra en la columna *F* en la figura 3).
9. Finalmente, los porcentajes acumulados se obtienen multiplicando por 100 las frecuencias relativas acumuladas. Para ello escribimos  $=F4*100$  y pulsamos

*Enter* en la casilla *G4*. Repitiendo el procedimiento explicado en los casos anteriores extendemos el cálculo hasta la casilla *G11*.

El tabla final obtenida se muestra en la figura 3.

|    | A  | B             | C                   | D          | E                             | F                             | G                    |  |
|----|--|---------------|---------------------|------------|-------------------------------|-------------------------------|----------------------|--|
| 1  | Condenados Illes Balears por edad (año 2005) |               |                     |            |                               |                               |                      |  |
| 2  |  |               |                     |            |                               |                               |                      |  |
| 3  | Edad (años)                                  | Nº condenados | Frecuencia relativa | Porcentaje | Frecuencia absoluta acumulada | Frecuencia relativa acumulada | Porcentaje acumulado |  |
| 4  | 18-20  | 155           | 0,04                | 4,47       | 155                           | 0,04                          | 4,47                 |  |
| 5  | 21-25  | 543           | 0,16                | 15,65      | 698                           | 0,2                           | 20,12                |  |
| 6  | 26-30  | 653           | 0,19                | 18,82      | 1351                          | 0,39                          | 38,94                |  |
| 7  | 31-35  | 619           | 0,18                | 17,84      | 1970                          | 0,57                          | 56,79                |  |
| 8  | 36-40  | 515           | 0,15                | 14,85      | 2485                          | 0,72                          | 71,63                |  |
| 9  | 41-50  | 636           | 0,18                | 18,33      | 3121                          | 0,9                           | 89,97                |  |
| 10 | 51-60  | 248           | 0,07                | 7,15       | 3369                          | 0,97                          | 97,12                |  |
| 11 | Más de 60                                    | 100           | 0,03                | 2,88       | 3469                          | 1                             | 100                  |  |
| 12 |  | 3469          |                     |            |                               |                               |                      |  |
| 13 |  |               |                     |            |                               |                               |                      |  |

Figura 3: Frecuencias y porcentajes obtenidos a partir de los datos del ejemplo 1

10. Podemos imprimir la tabla calculada o guardarla como un fichero .pdf para su posterior impresión utilizando los iconos y respectivamente, del menú de Calc.

En todo caso, la visualización de la tabla mejora si separamos las filas y las columnas mediante líneas. Para ello, antes de imprimir o guardar el fichero seleccionaremos todas las casillas de la tabla situándonos sobre la casilla *A3* y, manteniendo pulsado el botón izquierdo del ratón, arrastrando el cursor hasta la casilla *G12*. A continuación accederemos a la opción *Formato* del menú principal, haremos clic sobre la opción *Celdas...*, se abrirá una nueva ventana en la que seleccionaremos la pestaña *Borde* y haremos clic sobre el ícono

Si imprimimos la tabla o visualizamos el fichero .pdf en la pantalla obtendremos el resultado de la figura 4:

La tabla anterior puede incluirse fácilmente en informes escritos con Microsoft Word o OpenOffice Writer. Para ello seleccionaremos con el cursor todas las casillas que componen la tabla y utilizaremos la combinación de teclas *Ctrl+C*. La tabla queda copiada en el portapapeles de Windows y puede ser pegada en otros documentos mediante la combinación de teclas *Ctrl+V*.

A continuación explicamos como representar gráficamente los valores calculados

1. **Creación de un diagrama de barras.** Obtendremos en primer lugar un diagrama de barras que represente el número de condenados en función de su edad. Los pasos a seguir son los siguientes:

Condenados Illes Balears por edad (año 2005)

| Edad (años) | Nº condenados | Frecuencia relativa | Porcentaje | Frecuencia absoluta acumulada | Frecuencia relativa acumulada | Porcentaje acumulado |
|-------------|---------------|---------------------|------------|-------------------------------|-------------------------------|----------------------|
| 18-20       | 155           | 0,04                | 4,47       | 155                           | 0,04                          | 4,47                 |
| 21-25       | 543           | 0,16                | 15,65      | 698                           | 0,2                           | 20,12                |
| 26-30       | 653           | 0,19                | 18,82      | 1351                          | 0,39                          | 38,94                |
| 31-35       | 619           | 0,18                | 17,84      | 1970                          | 0,57                          | 56,79                |
| 36-40       | 515           | 0,15                | 14,85      | 2485                          | 0,72                          | 71,63                |
| 41-50       | 636           | 0,18                | 18,33      | 3121                          | 0,9                           | 89,97                |
| 51-60       | 248           | 0,07                | 7,15       | 3369                          | 0,97                          | 97,12                |
| Más de 60   | 100           | 0,03                | 2,88       | 3469                          | 1                             | 100                  |
|             | 3469          |                     |            |                               |                               |                      |

Figura 4: Tabla final del ejemplo 1

- a) Seleccionamos la opción *Insertar* del menú principal y hacemos clic sobre  **Gráfico...**.
- b) Se abrirá una ventana titulada *Asistente de gráficos* (ver Figura 5). En primer lugar debemos seleccionar el tipo de diagrama, en nuestro caso un diagrama de barras con texto, por lo que hacemos clic sobre el icono  **Columna**. En la parte derecha de la ventana se elige una variante del diagrama de barras, nosotros nos quedamos con la opción por defecto (*Normal*) y pulsamos *Siguiente*.
- c) Se muestra una nueva ventana donde seleccionaremos las casillas de datos a representar. Para ello escribiremos **A3:A11;B3:B11** en *Rango de datos*. Los datos de la primera columna seleccionada se representarán sobre el eje horizontal y los de la segunda sobre el eje vertical. Haciendo clic sobre el botón *Siguiente* pasamos a la siguiente ventana (*Series de datos*). Aplicamos las opciones por defecto por lo que volvemos a pulsar *Siguiente*.
- d) En la última ventana debemos añadir el texto de la gráfica. Escribimos el *Título del diagrama*: “Condenados Illes Balears (año 2005)”, los títulos de los ejes X e Y (“Edad” y “Nº condenados”, respectivamente) y desactivamos la opción *Mostrar leyenda*. Finalmente pulsamos el botón *Finalizar*.

El gráfico creado se muestra sobre la hoja de cálculo. Podemos variar su posición y tamaño mediante el ratón. El resultado final se muestra en la figura 6.

Al igual que para la tabla de frecuencias este diagrama puede imprimirse o bien guardarse como un fichero .pdf. Además, haciendo clic sobre el mismo y utilizando la combinación de teclas *Ctrl+C* es posible copiarlo en el portapapeles de Windows. De esta forma puede ser pegado fácil-

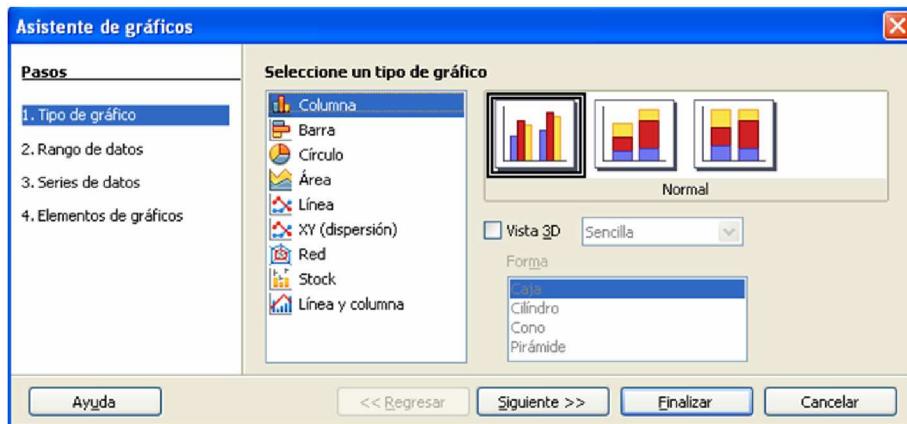


Figura 5: Inserción de un diagrama de barras. Ventana inicial del *Asistente de gráficos*

mente (combinación de teclas *Ctrl+V*) en un documento de Microsoft Word o OpenOffice Writer para la elaboración de un informe.

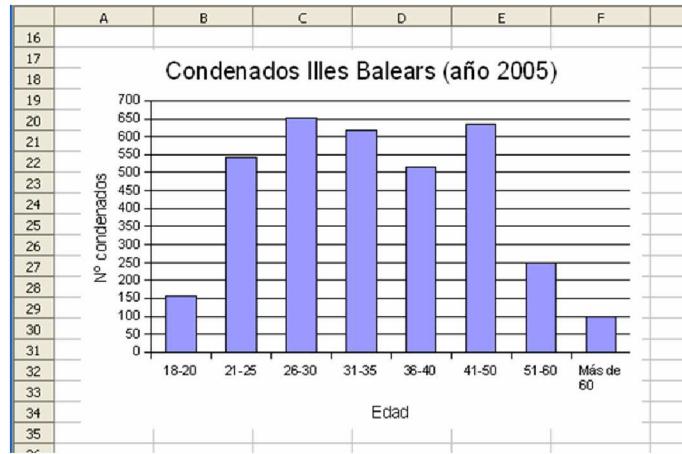


Figura 6: Diagrama de barras de frecuencias absolutas del ejemplo 1

2. **Creación de un diagrama de tarta.** Obtendremos a continuación un diagrama de tarta que represente los porcentajes de condenados para cada intervalo de edad. Los pasos a seguir son prácticamente idénticos a los del diagrama de barras, con las siguientes modificaciones
  - a) En la ventana inicial del *Asistente de gráficos* seleccionamos el tipo de diagrama haciendo clic sobre el icono (Círculo).
  - b) En la siguiente ventana el rango de datos es A3:A11;D3:D11.

- c) Una vez creado el diagrama es posible cambiar el tamaño del texto de la leyenda haciendo doble clic sobre el diagrama, seleccionando la opción *Formato* del menú principal y a continuación la opción *Leyenda*. Aparece una nueva ventana en la que hay que escoger la pestaña *Caracteres* y el *Tamaño* deseado. El resultado final se muestra en la figura 7.

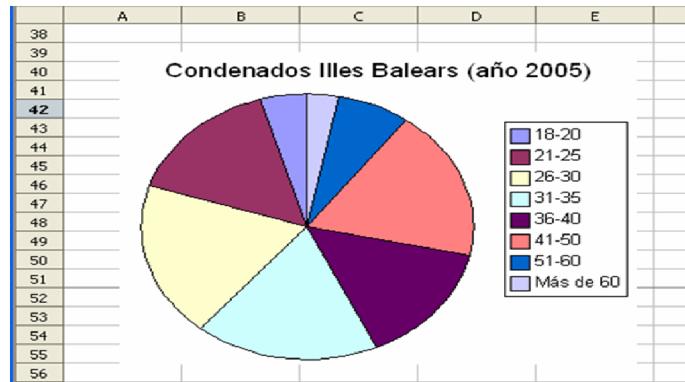


Figura 7: Diagrama de tarta de porcentajes del ejemplo 1

- d) Este diagrama puede ser imprimido o insertado en otro documento al igual que el diagrama de barras.

## Ejemplo 2

En este ejemplo aprenderemos a crear un diagrama de barras dobles a partir de los siguientes datos sobre población reclusa menor de edad en Baleares:

| Edad (años) | Varón | Mujer |
|-------------|-------|-------|
| 14          | 62    | 6     |
| 15          | 78    | 10    |
| 16          | 134   | 21    |
| 17          | 332   | 29    |

1. Introducimos los datos en una hoja de cálculo tal como se ha explicado para el ejemplo 1. Supongamos por ejemplo que los datos de *Edad* ocupan las casillas *A4* a *A7*, los de *Varón* las casillas *B4* a *B7* y los de *Mujer* de *C4* a *C7* (las casillas *A3*, *B3* y *C3* contienen los títulos de las columnas).
2. Seguimos los pasos explicados para la creación de diagramas de barras en el ejemplo 1 pero seleccionando ahora las tres columnas de datos (rango de valores *A3:A7;B3:B7;C3:C7*). Las opciones a elegir son las mismas que en el caso del ejemplo 1 con la diferencia de que en la última ventana seleccionamos la opción *Mostrar leyenda* y que los títulos del diagrama y los ejes X e Y son, respectivamente: “Población reclusa menor de edad en Baleares”, “Edad” y “Nº reclusos”.

Al pulsar sobre el botón *Finalizar* obtenemos el resultado que se muestra en la figura 8.

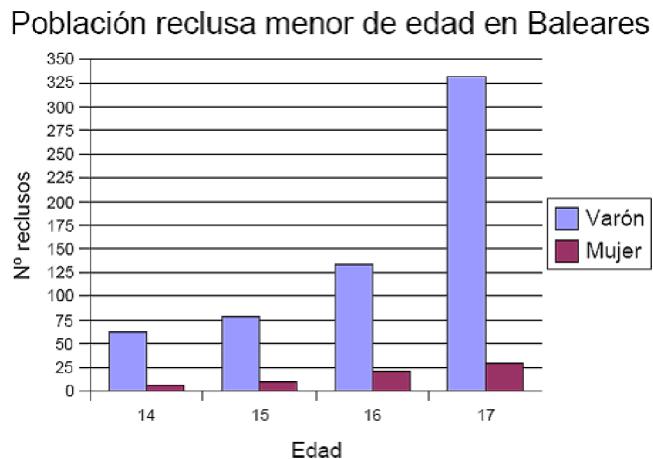


Figura 8: Diagrama de barras dobles del ejemplo 2

### Ejemplo 3

En este ejemplo mostramos cómo calcular un histograma de frecuencias absolutas a partir de los datos siguientes sobre el peso de un grupo de personas:

| Peso (Kg) | Frecuencia absoluta |
|-----------|---------------------|
| 45-49     | 20                  |
| 50-54     | 35                  |
| 55-59     | 40                  |
| 60-64     | 55                  |
| 65-69     | 45                  |
| 70-74     | 50                  |
| 75-79     | 35                  |
| 80-84     | 30                  |
| 85-89     | 25                  |
| 90-94     | 15                  |
| 95-99     | 5                   |

1. En primer lugar creamos un documento OpenOffice Calc con los datos de la tabla anterior, tal como se ha explicado en el ejemplo 1. Supongamos que los datos sobre *Peso* ocupan las casillas *A4* a *A14* y los de frecuencia absoluta las casillas *B4* a *B14* (ver figura 9).
2. OpenOffice Calc no proporciona ninguna herramienta para la creación automática de histogramas en un caso general. Sólo en el caso de que todos los intervalos de valores sean de la misma amplitud (como en este ejemplo) es posible crear un histograma de manera sencilla.

En el caso del ejemplo todos los intervalos son de longitud 5 y podemos representar el histograma como un diagrama de barras modificado. En primer lugar debemos calcular la altura de las barras.

Sabemos que el área de las barras del histograma es igual al valor representado (en este caso la frecuencia absoluta). Por ejemplo, la primera barra debe tener área 20, como su anchura es 5 su altura deberá ser  $\frac{20}{5} = 4$ . Razonando de la misma manera podemos calcular el resto de alturas. Podemos hacerlo de forma automática con Calc: escribimos en la casilla *C4* =*B4*/5, pulsamos *Enter* y a continuación extendemos el cálculo hasta la casilla *C14* utilizando el método explicado en el ejemplo 1. Al final de esta operación en la columna *C* aparecen los valores de altura calculados (ver figura 9).

3. Ahora el histograma se puede calcular como un diagrama de barras. Seguimos el procedimiento descrito para el ejemplo 3 seleccionando las casillas *A3* a *A14* y *C3* a *C14* en el rango de datos. Los títulos del diagrama y del eje X son, respectivamente, “Histograma pesos” y “Pesos (Kg)” y la opción *Mostrar leyenda* no se selecciona.

|    | A         | B                   | C                        | D |
|----|-----------|---------------------|--------------------------|---|
| 1  |           |                     |                          |   |
| 2  |           |                     |                          |   |
| 3  | Peso (Kg) | Frecuencia absoluta | Altura barras histograma |   |
| 4  | 45-49     | 20                  | 4                        |   |
| 5  | 50-54     | 35                  | 7                        |   |
| 6  | 55-59     | 40                  | 8                        |   |
| 7  | 60-64     | 55                  | 11                       |   |
| 8  | 65-69     | 45                  | 9                        |   |
| 9  | 70-74     | 50                  | 10                       |   |
| 10 | 75-79     | 35                  | 7                        |   |
| 11 | 80-84     | 30                  | 6                        |   |
| 12 | 85-89     | 25                  | 5                        |   |
| 13 | 90-94     | 15                  | 3                        |   |
| 14 | 95-99     | 5                   | 1                        |   |
| 15 |           |                     |                          |   |

Figura 9: Tabla de datos del ejemplo 3

Al crear el diagrama obtenemos un diagrama de barras con las barras separadas. Para unir las barras y darle la forma típica de un histograma debemos hacer doble clic sobre una de las barras del diagrama hasta que aparece la ventana que se muestra en la figura 10. Escogemos la pestaña *Opciones* y ponemos a 0 % el valor de *Espacio* en la opción *Configuración*. Al pulsar sobre el botón *Aceptar* de esta ventana obtenemos un histograma como el que se muestra en la figura 11.

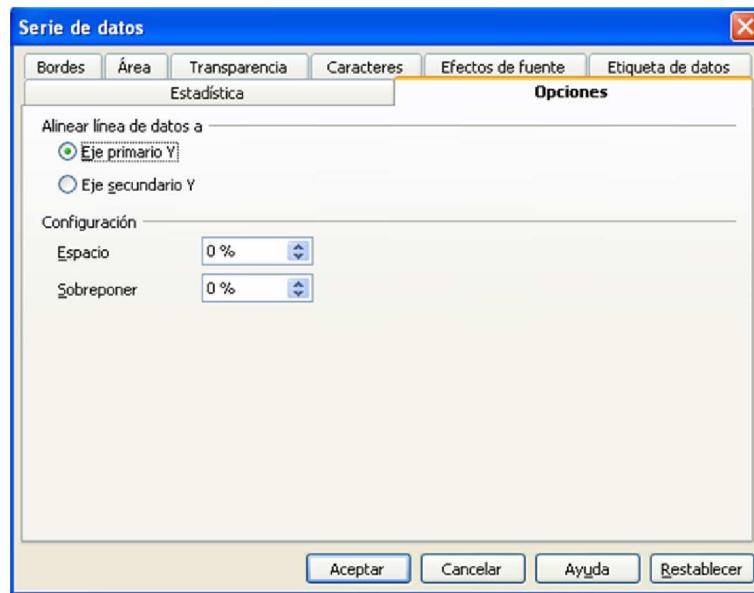


Figura 10: Ventana de diálogo para ajustar la anchura de las barras del diagrama de barras

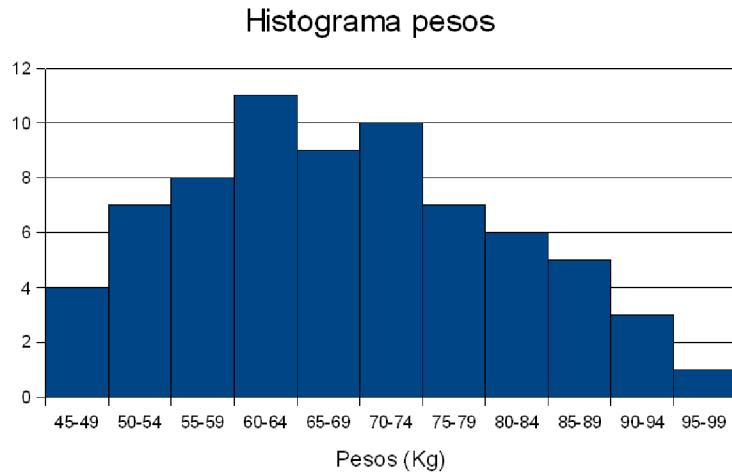


Figura 11: Histograma del ejemplo 3

#### Ejemplo 4

En este ejemplo mostramos cómo crear un diagrama lineal que representa la evolución del PIB español (en miles de millones de dólares) desde 1992 hasta 2007. Los datos proceden del Fondo Monetario Internacional.

| Año  | PIB  |
|------|------|
| 1992 | 612  |
| 1993 | 513  |
| 1994 | 515  |
| 1995 | 597  |
| 1996 | 622  |
| 1997 | 573  |
| 1998 | 601  |
| 1999 | 618  |
| 2000 | 582  |
| 2001 | 609  |
| 2002 | 688  |
| 2003 | 885  |
| 2004 | 1045 |
| 2005 | 1131 |
| 2006 | 1231 |
| 2007 | 1414 |

1. En primer lugar creamos un documento OpenOffice Calc con los datos de la tabla anterior, tal como se ha explicado en el ejemplo 1. Supongamos que los datos sobre *Años* ocupan las casillas *A4* a *A19* y los de PIB las casillas *B4* a *B19* (en *A3* y *B3* se encuentran los nombres de las columnas).
2. Creamos un diagrama siguiendo el procedimiento explicado en anteriores ejemplos. Las casillas de datos a seleccionar son de *A3* a *A19* y *B3* a *B19*. Seleccionamos el tipo de diagrama representado por el icono y la variante representada por el icono .

No seleccionamos la opción de *Mostrar leyenda* y los títulos del diagrama y los ejes X e Y son, respectivamente, “Evolución PIB de España”, “Año” y

“PIB (miles millones dólares)”. La gráfica obtenida se muestra en la figura 12.

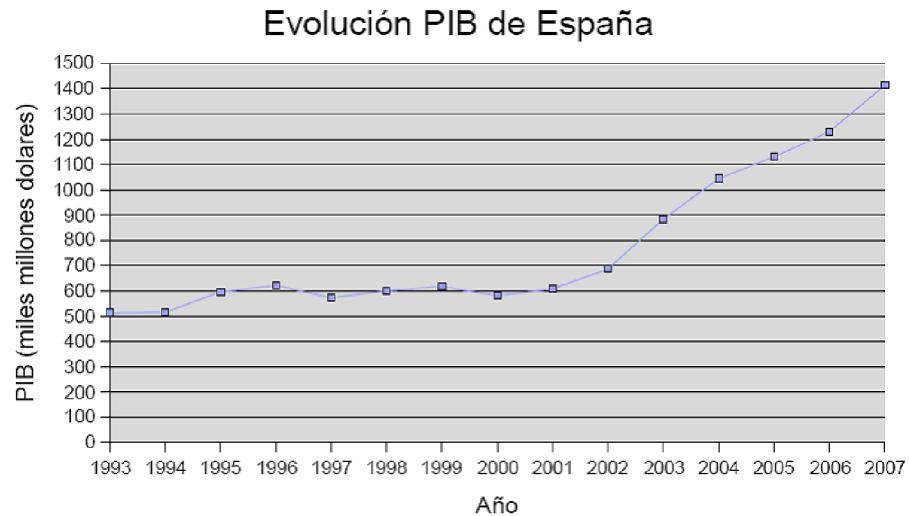
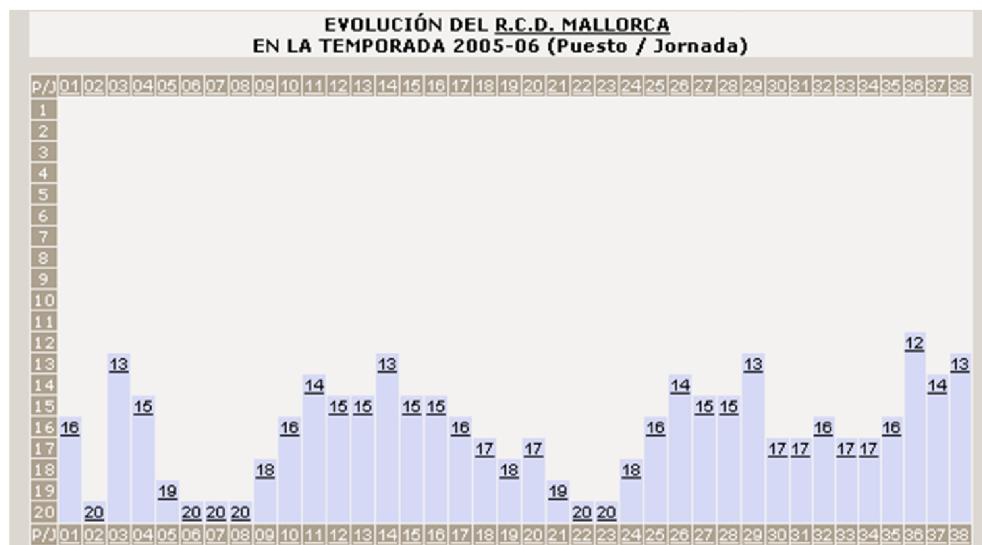


Figura 12: Diagrama lineal del ejemplo 4

### Ejemplo 5

En todos los ejemplos anteriores hemos partido de datos de frecuencias absolutas a partir de las cuales hemos calculado frecuencias acumuladas, porcentajes, etc. Es habitual sin embargo disponer de datos *en bruto* que deben organizarse primero en tablas de frecuencias absolutas antes de realizar cualquier otro cálculo. En este último ejemplo explicamos cómo organizar este tipo de datos.

Partimos de los datos que se muestran en el siguiente gráfico (fuente LFP):



Para la variable “clasificación del RCD Mallorca durante la temporada 2005-2006” deseamos calcular las frecuencias absolutas y acumuladas. Los pasos a seguir son los siguientes:

1. Creamos un documento OpenOffice Calc y escribimos en la primera columna los datos *brutos* del gráfico. Los datos ocupan las casillas A1 a A38. El resultado se muestra en la figura 13.

|    | A  | B             | C                   |
|----|----|---------------|---------------------|
| 1  | 12 | Clasificación | Frecuencia absoluta |
| 2  | 13 | 12            |                     |
| 3  | 13 | 13            |                     |
| 4  | 13 | 14            |                     |
| 5  | 13 | 15            |                     |
| 6  | 14 | 16            |                     |
| 7  | 14 | 17            |                     |
| 8  | 14 | 18            |                     |
| 9  | 15 | 19            |                     |
| 10 | 15 | 20            |                     |
| 11 | 15 |               |                     |
| 12 | 15 |               |                     |
| 13 | 15 |               |                     |
| 14 | 15 |               |                     |
| 15 | 15 |               |                     |
| 16 | 16 |               |                     |
| 17 | 16 |               |                     |
| 18 | 16 |               |                     |
| 19 | 16 |               |                     |
| 20 | 16 |               |                     |
| 21 | 16 |               |                     |
| 22 | 17 |               |                     |
| 23 | 17 |               |                     |
| 24 | 17 |               |                     |
| 25 | 17 |               |                     |
| 26 | 17 |               |                     |
| 27 | 17 |               |                     |
| 28 | 18 |               |                     |
| 29 | 18 |               |                     |
| 30 | 18 |               |                     |
| 31 | 19 |               |                     |
| 32 | 19 |               |                     |
| 33 | 20 |               |                     |
| 34 | 20 |               |                     |
| 35 | 20 |               |                     |
| 36 | 20 |               |                     |
| 37 | 20 |               |                     |
| 38 | 20 |               |                     |

Figura 13: Izquierda: documento OpenOffice Calc con los datos *en bruto* del ejemplo 5 (paso 1). Derecha: documento preparado para el cálculo de las frecuencias absolutas (paso 2)

2. A continuación creamos dos nuevas columnas en el documento (por ejemplo, las columnas *B* y *C*). En la parte superior de la primera columna escribimos el

nombre de la variable (“Clasificación”) y a continuación escribimos, en orden creciente, los distintos valores que toma la variable. En la parte superior de la segunda columna escribimos “Frecuencia absoluta”, que calcularemos a continuación.

Para facilitar la tarea de escribir en orden creciente los distintos valores de la variable podemos **ordenar** los valores *brutos* del siguiente modo:

- a) Seleccionamos las casillas  $A1$  a  $A38$  manteniendo el botón izquierdo del ratón pulsado.
- b) Pulsamos el icono . Los valores se ordenan de menor a mayor.  
(Con la opción los valores se ordenarían de mayor a menor).

Ahora es sencillo ver qué valores toma la variable y escribirlos de forma ordenada en la columna  $C$ .

Al final de este paso el documento Calc tiene la forma que se muestra en la figura 13-derecha.

3. Para calcular las frecuencias absolutas nos situamos en la casilla  $C2$ , correspondiente a la frecuencia absoluta del valor 12. Escribimos  
`=CONTAR.SI($A$1:$A$38; "="&B2)`<sup>1</sup> y pulsamos *Enter*. Un 1 aparece escrito en la casilla, lo que significa que el valor 12 aparece un única vez en la lista de datos *brutos* (su frecuencia absoluta es 1). Extendemos el cálculo a las casillas  $C3$  a  $C10$  del siguiente modo: situamos el cursor en la esquina inferior derecha de la casilla  $C5$  y, manteniendo el botón izquierdo del ratón pulsado, arrastramos el cursor hasta la casilla  $C10$ . Al soltar el botón se mostrarán los valores calculados. Al final de este paso la hoja de cálculo tiene el aspecto que se muestra en la figura 14-izquierda.
4. Las frecuencias acumuladas se calculan siguiendo el procedimiento explicado en los ejemplos anteriores. La tabla final se muestra en la figura 14.

---

<sup>1</sup>La instrucción `=CONTAR.SI(A1:A38; "="&B2)` examina las columnas  $A1$  a  $A38$  y cuenta cuántas de ellas tienen un valor igual al de la casilla  $B2$

| B             | C                   |
|---------------|---------------------|
| Clasificación | Frecuencia absoluta |
| 12            | 1                   |
| 13            | 4                   |
| 14            | 3                   |
| 15            | 7                   |
| 16            | 6                   |
| 17            | 6                   |
| 18            | 3                   |
| 19            | 2                   |
| 20            | 6                   |

| Clasificación | Frecuencia absoluta | Frecuencia acumulada |
|---------------|---------------------|----------------------|
| 12            | 1                   | 1                    |
| 13            | 4                   | 5                    |
| 14            | 3                   | 8                    |
| 15            | 7                   | 15                   |
| 16            | 6                   | 21                   |
| 17            | 6                   | 27                   |
| 18            | 3                   | 30                   |
| 19            | 2                   | 32                   |
| 20            | 6                   | 38                   |

Figura 14: Izquierda: frecuencias absolutas del ejemplo 5 (paso 3). Derecha: tabla final.

## 1.1. Utilización de la base de datos del INE

El Instituto Nacional de Estadística (INE) ofrece a través de su página web gran cantidad de información sobre distintos temas: Educación, Cultura, Salud, Economía, Justicia, etc. Los datos de los ejemplos de la sección anterior se han obtenido de esta web. En esta sección explicamos cómo utilizar la base de datos del INE.

Por ejemplo, supongamos que deseamos hacer un estudio sobre Educación. Los pasos a seguir para obtener los datos del INE son los siguientes:

1. abrir en el navegador la dirección <http://www.ine.es>
2. hacer clic sobre la opción  que aparece a la izquierda de la página principal
3. las diferentes opciones de la base de datos se muestran en la nueva página (ver Figura 15)
4. para acceder a los datos sobre Educación hacemos clic sobre *Educación* (bajo el epígrafe *Sociedad*) en el menú principal
5. en la nueva página se ofrecen varios estudios estadísticos relacionados con la educación. Supongamos que deseamos conocer los datos sobre *Gasto público en educación*, haremos clic sobre este concepto.
6. en la nueva página que se abre se explica en qué consisten los datos recopilados y se permite al usuario acceder a la información de un año en concreto. En el menu desplegable que aparece al hacer clic sobre *Seleccione un año* escogemos por ejemplo la opción *2004-2005*
7. se nos ofrecen varios tipos de datos (ver Figura 16). Elegimos por ejemplo la opción *Becas e importe de las mismas por universidad en la que está matriculado el becario, número, entidad que las concede y tipo de beca*, bajo el epígrafe *Enseñanzas universitarias*
8. la nueva página que se abre nos permite escoger los datos a mostrar y la manera de mostrarlos mediante una serie de menus (ver Figura 17). Podemos seleccionar por ejemplo las universidades Autónoma de Barcelona, Complutense de Madrid, Illes Balears, Pública de Navarra, Sevilla y Deusto.

Para ello haremos clic sobre las opciones del menú *Universidad en la que está matriculado el becario*, manteniendo pulsada la tecla *Ctrl*, lo que nos permitirá la selección simultánea de varias opciones.

A continuación seleccionaremos las opciones Becas e Importe en el menú *Número* (manteniendo pulsada la tecla *Ctrl*), la opción Todas las Administraciones en el menú *Entidad que las concede* y Total en *Tipo de beca*.

|   |  |
|---|--|
| <b>Entorno físico y medio ambiente</b>                          | <b>Economía</b>  |
| <a href="#">Entorno físico</a>                                  | <a href="#">Empresas</a>   |
| <a href="#">Estadísticas sobre el medio ambiente</a>            | <a href="#">Cuentas económicas</a>                                       |
| <a href="#">Cuentas ambientales</a>                             | <a href="#">Estadísticas financieras y monetarias</a>                    |
| <a href="#">Indicadores ambientales</a>                         | <a href="#">Comercio exterior</a>  |
| <a href="#">Otros estudios ambientales</a>                      | <a href="#">Información tributaria</a>                                   |
| <b>Demografía y población</b>                                   | <b>Ciencia y tecnología</b>  |
| <a href="#">Cifras de población</a>                             | <a href="#">Investigación y desarrollo tecnológico</a>                   |
| <a href="#">- Padrón municipal</a>                              | <a href="#">Nuevas tecnologías de la información y la comunicación</a>   |
| <a href="#">- Estimaciones y proyecciones</a>                   |  |
| <a href="#">- Censos de Población</a>                           |  |
| <a href="#">- Datos históricos</a>                              |  |
| <a href="#">Movimiento natural de la población</a>              |  |
| <a href="#">Migraciones</a>                                     |  |
| <a href="#">Análisis y estudios demográficos</a>                |  |
| <b>Sociedad</b>   | <b>Industria y construcción</b>  |
| <a href="#">Educación</a>                                       | <a href="#">Industria</a>  |
| <a href="#">Cultura y ocio</a>                                  | <a href="#">Energía</a>  |
| <a href="#">Salud</a>   | <a href="#">Construcción y vivienda</a>                                  |
| <a href="#">Justicia</a>  |  |
| <a href="#">Nivel, calidad y condiciones de vida.(IPC, ...)</a> |  |
| <a href="#">Mercado laboral</a>                                 | <b>Servicios</b>   |
| <a href="#">Análisis sociales</a>                               | <a href="#">Encuestas globales del sector servicios</a>                  |
| <a href="#">Elecciones</a>                                      | <a href="#">Comercio</a>   |
|   | <a href="#">Transporte y actividades conexas, comunicaciones</a>         |
|   | <a href="#">Hostelería y turismo</a>                                     |
|   | <a href="#">Otros servicios empresariales, personales y comunitarios</a> |
|   | <b>Clasificaciones</b>   |
|   | <a href="#">Clasificaciones nacionales</a>                               |
|   | <a href="#">Clasificaciones internacionales</a>                          |
|   | <a href="#">Proceso de revisión de clasificaciones</a>                   |
|   | <b>Internacional</b>   |
|   | <a href="#">Internacional</a>  |
|   | <b>Historia</b>  |
|   | <a href="#">Fondo documental</a>   |

Figura 15: Opciones de la base de datos del INE

|   |
|---|
| <b>Becas y ayudas. Curso 2004-2005</b>  |
| <b>Todas las enseñanzas</b>   |
| <a href="#">■ 1.1 Becas, ayudas, becarios, beneficiarios e importe de las mismas por CCAA de destino, entidad que las concede, número y tipo de enseñanza.</a>          |
| <a href="#">■ 1.2 Becas, ayudas e importe de las mismas concedidas por administración educativa financiadora, número y tipo de beca o ayuda (1).</a>                    |
| <b>Enseñanzas obligatorias, educación infantil y educación especial</b>   |
| <a href="#">■ 2.1 Ayudas e importe de las mismas concedidas por administración educativa financiadora, número y tipo de ayuda.</a>                                      |
| <a href="#">■ 2.2 Ayudas e importe de las mismas por CCAA de destino, número, entidad que las concede y tipo de ayuda.</a>  |
| <a href="#">■ 2.3 Ayudas, beneficiarios e importe de las mismas concedidas por CCAA de destino, número, entidad que las concede y nivel educativo del beneficiario.</a> |
| <b>Enseñanzas postobligatorias no universitarias</b>  |
| <a href="#">■ 3.1 Becas e importe de las mismas concedidas por administración educativa financiadora, número y tipo de beca.</a>  |
| <a href="#">■ 3.2 Becas e importe de las mismas por CCAA de destino, número, entidad que las concede y tipo de beca.</a>  |
| <a href="#">■ 3.3 Becas, becarios e importe de las mismas por CCAA de destino, número, entidad que las concede y nivel educativo del becario.</a>                       |
| <b>Enseñanzas universitarias</b>  |
| <a href="#">■ 4.1 Becas e importe de las mismas por administración educativa financiadora, número y tipo de beca.</a>   |
| <a href="#">■ 4.2 Becas e importe de las mismas por universidad en la que está matriculado el becario, número, entidad que las concede y tipo de beca.</a>              |
| <a href="#">■ 4.3 Becas, becarios e importe de las mismas (1) por universidad en la que está matriculado el becario, entidad que las concede y número.</a>              |

Figura 16: Datos sobre gasto público en educación en la base de datos del INE

Finalmente elegiremos la forma de visualizar los datos. Podemos elegir qué variables se mostrarán por filas y cuáles por columnas. Por defecto el menú nos ofrece visualizar por filas la variable *Universidad en la que está matriculado el becario* y por columnas el número de becas, la entidad que las concede y el tipo. Aceptamos las opciones por defecto y generamos el resultado pulsando el botón **Consultar selección**. Obtendremos el resultado que se muestra en la figura 18

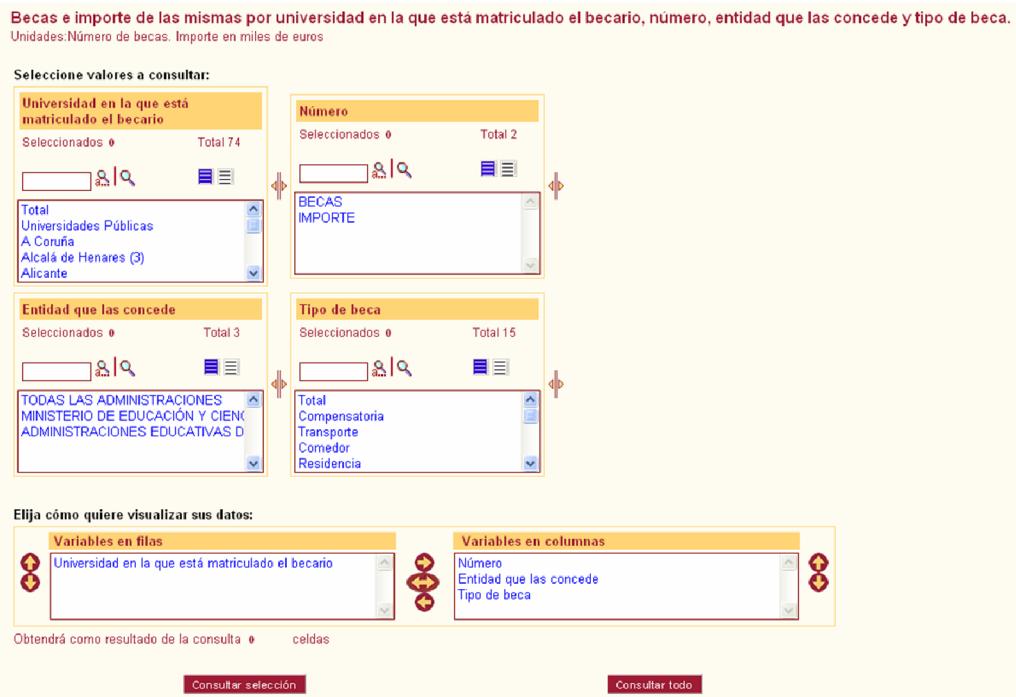


Figura 17: Menus para la selección de datos sobre número de becas e importe de las mismas en centros universitarios

- la tabla obtenida se puede imprimir pulsando el icono de la página de resultados. También se puede guardar en distintos formatos para ser utilizada por distintos programas estadísticos pulsando sobre el botón **Descargar como:** .

Nosotros elegiremos descargar como fichero Excel y guardaremos el fichero obtenido (extensión .xls) en alguna carpeta de nuestro ordenador. Este tipo de ficheros pueden leerse desde la aplicación Excel de Microsoft y también desde la aplicación Calc de OpenOffice, que es la que utilizamos en nuestros ejemplos.

Al abrir el fichero obtenido con Calc los nombres de las universidades seleccionadas aparecen en la columna A de la hoja de cálculo mientras que el número de becas obtenidas por cada universidad así como su importe apa-

| Becas y ayudas. Curso 2004-2005  |                            |                            |
|--|----------------------------|----------------------------|
| Enseñanzas universitarias  |                            |                            |
| Becas e importe de las mismas por universidad en la que está matriculado el becario, número, entidad que las concede y tipo de beca. |                            |                            |
| Unidades:Número de becas. Importe en miles de euros  |                            |                            |
|  | BECAS                      | IMPORTE                    |
|  | TODAS LAS ADMINISTRACIONES | TODAS LAS ADMINISTRACIONES |
| Total  | Total                      | Total                      |
| Autónoma de Barcelona  | 13.528                     | 10.514,4                   |
| Complutense de Madrid (3)  | 24.416                     | 26.248,1                   |
| Illes Balears (2)  | 8.237                      | 6.146,6                    |
| Pública de Navarra   | 3.267                      | 3.200,5                    |
| Sevilla  | 33.692                     | 32.506,0                   |
| Deusto   | 2.622                      | 1.609,8                    |

Notas:

1) No se contabilizan como becas las de exención de precios académicos a familias numerosas de 3 hijos que afectan a 102.670 alumnos y ascienden a 33.912,4 miles de euros.

2) Se ha supuesto que los becarios y por tanto las becas son para los alumnos de la Universidad de Baleares.

3) En la Comunidad de Madrid, en el caso de las becas de exención de precios concedidas a alumnos con discapacidad superior al 50%, figura su importe, pero se desconoce el número de ayudas.

Fuente: Ministerio de Educación y Ciencia

Figura 18: Tabla de datos sobre número de becas e importe de las mismas en centros universitarios

recen en las columnas B y C. Siguiendo los pasos explicados en la sección anterior podremos calcular las tablas y gráficos asociados a estos datos.

De manera similar se puede acceder a datos sobre la evolución anual de diferentes parámetros (población, ocupación turística, producción industrial, etc) haciendo clic sobre la opción  **Banco de series temporales** que aparece a la izquierda de la página principal del INE.

## 1.2. Ejercicios propuestos

### Ejercicio 1

A partir de los datos de la siguiente tabla calcular las frecuencias relativas y porcentajes.

- ¿Es posible calcular frecuencias y porcentajes acumulados? Calcularlos en caso de respuesta afirmativa.
- Dibujar un diagrama de tarta que muestre los porcentajes.

**Estadísticas judiciales 2005**  
**Estadística de lo Penal. Menores. Resultados autonómicos y provinciales**

**Menores según infracción cometida**

Unidades:nº de menores

|                           | <b>BALEARS (ILLES)</b> |
|---------------------------|------------------------|
| Homicidio                 | 0                      |
| Lesiones                  | 31                     |
| Contra la libertad sexual | 12                     |
| Hurto                     | 69                     |
| Robo                      | 306                    |
| Contra la salud pública   | 10                     |

Fuente:Instituto Nacional de Estadística

## Ejercicio 2

A partir de los datos de la siguiente tabla calcular las frecuencias relativas y porcentajes:

- a) ¿Es posible calcular frecuencias y porcentajes acumulados? Calcularlos en caso de respuesta afirmativa.
- b) Dibujar un diagrama de barras que muestre las frecuencias absolutas.
- c) Agrupar los datos en periodos de 2 años y dibujar el histograma de porcentajes.

**Pruebas de acceso a la universidad. Año 2006**  
**Serie anual del alumnado matriculado y aprobado**

**Evolución del alumnado por tipo de indicador, convocatoria y años.**  
 Unidades:especificadas en la variable 'tipo de indicador'

|       | <b>TOTALES DE ALUMNADO APROBADO</b> |
|-------|-------------------------------------|
| Total |                                     |
| 1997  | 244.006                             |
| 1998  | 237.310                             |
| 1999  | 227.282                             |
| 2000  | 206.830                             |
| 2001  | 198.098                             |
| 2002  | 190.544                             |
| 2003  | 190.387                             |
| 2004  | 182.325                             |
| 2005  | 181.849                             |
| 2006  | 182.698                             |

Fuente:Instituto Nacional de Estadística

## Ejercicio 3

Acceder a la base de datos del INE y dibujar un diagrama de barras doble con los datos de la encuesta de migraciones del año 2003, en valor absoluto, para los

grupos de edad “de 0 a 9” hasta “de 65 y más”. Representar en una barra los datos correspondientes a varones y en la otra los correspondientes a mujeres.

**Ejercicio 4**

Acceder a la base de datos del INE y dibujar un diagrama de tarta de porcentajes con los datos de divorcios en España en el año 2005 clasificados según la duración del matrimonio para cualquier edad del esposo.

**Ejercicio 5**

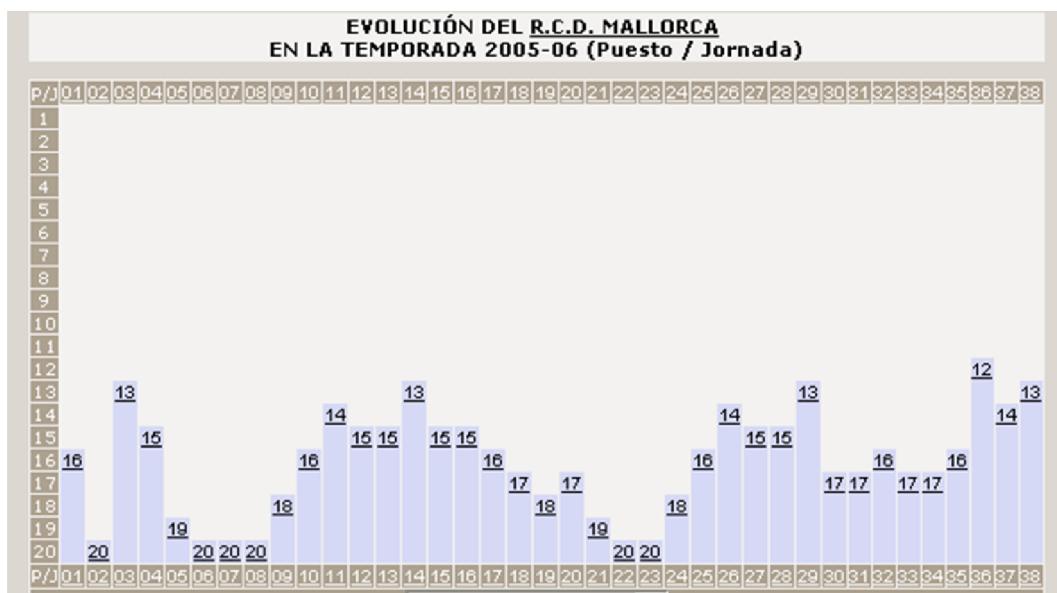
Acceder al banco de series temporales del INE (Tempus) y representar mediante un diagrama de línea los datos de población de ambos sexos en Baleares desde 1996.

## 2. Cálculo de moda, mediana, media y percentiles con ordenador

Las operaciones matemáticas que deben realizarse para calcular los estadísticos explicados en este tema son muy sencillas y pueden realizarse con una simple calculadora. No obstante, si la cantidad de datos es elevada, programas como el OpenOffice Calc permiten calcular fácilmente los estadísticos más habituales de una distribución. Explicaremos cómo hacerlo en los siguientes ejemplos.

### Ejemplo 1

Calcular la media, la moda, la mediana y los cuartiles 1º y 3º de la clasificación del RCD Mallorca durante la temporada 2005-2006 a partir de los datos del siguiente gráfico (fuente LFP):



OpenOffice Calc permite calcular de forma muy sencilla la moda, mediana y media de un conjunto de datos *brutos*, es decir, no organizados en un tabla de frecuencias. Para resolver este ejemplo seguiremos los pasos siguientes:

1. Al igual que en el ejemplo 5 del tema anterior, en primer lugar escribiremos los datos *brutos* en la primera columna de la hoja de cálculo (casillas A1 a A38). El resultado se muestra en la figura 13-izquierda del tema anterior.
  2. A continuación escribimos las palabras “Moda”, “Mediana”, “Media”, “1er cuartil” y “3er cuartil” en las casillas C15, C16, C17, C18 y C19 de la hoja de cálculo (o en otras casillas cualesquiera)

3. **Moda.** Nos situamos en la casilla  $D15$  y escribimos  $=Moda(A1:A38)$ . Al pulsar *Enter* obtenemos el valor de la moda.
4. **Mediana.** Nos situamos en la casilla  $D16$  y escribimos  $=Mediana(A1:A38)$ . Al pulsar *Enter* obtenemos el valor de la mediana.
5. **Media.** Nos situamos en la casilla  $D17$  y escribimos  $=Promedio(A1:A38)$ . Al pulsar *Enter* obtenemos el valor de la media.
6. **1<sup>er</sup> cuartil.** Nos situamos en la casilla  $D18$  y escribimos  $=Cuartil(A1:A38;1)$ . Al pulsar *Enter* obtenemos el valor del 1<sup>er</sup> cuartil.
7. **3<sup>er</sup> cuartil.** Nos situamos en la casilla  $D19$  y escribimos  $=Cuartil(A1:A38;3)$ . Al pulsar *Enter* obtenemos el valor del 3<sup>er</sup> cuartil.

El resultado obtenido es:

|                         |       |
|-------------------------|-------|
| Moda                    | 15    |
| Mediana                 | 16    |
| Media                   | 16,34 |
| 1 <sup>er</sup> cuartil | 15    |
| 3 <sup>er</sup> cuartil | 18    |

Los tres valores centrales (moda, media y mediana) son muy similares, aunque se cumple que media > mediana > moda, por lo que podemos afirmar que se trata de una distribución casi simétrica con una ligera asimetría hacia la derecha.

**Nota:** al utilizar las fórmulas *Mediana* y *Cuartil* de Calc para el cálculo de la mediana y el primer y tercer cuartiles los resultados obtenidos pueden ser ligeramente diferentes a los que obtendríamos con el método que se explica en los ejemplos siguientes. La razón es que Calc utiliza unas fórmulas diferentes a las nuestras para el cálculo de cuartiles de datos *brutos*.

## Ejemplo 2

Repetir los cálculos del ejemplo anterior pero a partir de los datos organizados en la tabla de frecuencias siguiente (correspondiente al ejemplo 5 del tema 2):

| Clasificación | Frecuencia absoluta |
|---------------|---------------------|
| 12            | 1                   |
| 13            | 4                   |
| 14            | 3                   |
| 15            | 7                   |
| 16            | 6                   |
| 17            | 6                   |
| 18            | 3                   |
| 19            | 2                   |
| 20            | 6                   |

En este caso el cálculo de la media es muy sencillo pero el cálculo de moda y mediana es algo más complicado. A continuación explicamos cómo hacerlo:

1. En primer lugar escribimos los valores de la variable y sus frecuencias en una tabla de OpenOffice Calc. A continuación calculamos las frecuencias y los porcentajes acumulados, tal como se ha explicado en el tema anterior. La tabla que obtenemos es como la de la figura 19.

| A<br>Clasificación | B<br>Frecuencia absoluta | C<br>Frecuencia acumulada | D<br>Porcentaje acumulado |
|--------------------|--------------------------|---------------------------|---------------------------|
| 2                  | 12                       | 1                         | 2,63                      |
| 3                  | 13                       | 4                         | 13,16                     |
| 4                  | 14                       | 3                         | 21,05                     |
| 5                  | 15                       | 7                         | 39,47                     |
| 6                  | 16                       | 6                         | 55,26                     |
| 7                  | 17                       | 6                         | 71,05                     |
| 8                  | 18                       | 3                         | 78,95                     |
| 9                  | 19                       | 2                         | 84,21                     |
| 10                 | 20                       | 6                         | 100                       |
| 11                 |                          |                           |                           |
| 12                 |                          |                           |                           |

Figura 19: OpenOffice Calc con los del ejemplo 2 y las frecuencias y porcentajes acumulados calculadas

2. **Moda.** En casos como el del ejemplo, en los que la tabla de datos es pequeña basta observar los valores de frecuencia absoluta para descubrir para qué valor de la variable tenemos un máximo. En este ejemplo la moda es 15, ya que tiene el valor de frecuencia absoluta mayor (7).

Cuando la tabla de datos es muy grande, la moda se puede encontrar siguiendo el siguiente procedimiento:

- a) Primero copiamos los datos originales en unas nuevas columnas para evitar perderlos:
  - 1) seleccionamos, manteniendo el botón del ratón pulsado, las casillas con los valores de la variable y sus frecuencias absolutas (en nuestro ejemplo las casillas *A2 a A10 y B2 a B10*),
  - 2) pulsamos la combinación de teclas *Ctrl-C* para copiar estos datos,
  - 3) situamos el cursor en alguna otra casilla del documento (por ejemplo, *A14*),
  - 4) pulsamos el botón derecho del ratón y seleccionamos la opción *Pegado especial...*,
  - 5) activamos la opción *Números* y desactivamos todas las demás,
  - 6) finalmente pulsamos *Aceptar*. Los datos (sólo los valores numéricos, no las fórmulas) quedan copiados en las casillas *A14 a A22 y B14 a B22*.
- b) Seleccionamos las casillas *A14 a A22 y B14 a B22* manteniendo el botón izquierdo del ratón pulsado.

- c) En el menú principal escogemos la opción *Datos*, y a continuación *Ordenar...*
- d) Se abre una ventana en la que se definen los criterios de ordenación. En nuestro caso escogemos las opciones: *Ordenar según: Columna B* y *Ascendente*. Pulsamos *Aceptar*.
- e) En la columna A (casillas A14 a A22) aparecen los valores de la variable ordenados en orden decreciente de frecuencia absoluta (casillas B14 a B22), ver figura 20. El primer valor de la columna es la moda de la distribución. En nuestro caso: moda=15. Si la distribución fuera bimodal (el máximo ocurre en dos valores de la variable), deberíamos tomar como moda los dos primeros valores de la columna A.

|    | A  | B |
|----|----|---|
| 13 |    |   |
| 14 | 15 | 7 |
| 15 | 20 | 6 |
| 16 | 16 | 6 |
| 17 | 17 | 6 |
| 18 | 13 | 4 |
| 19 | 14 | 3 |
| 20 | 18 | 3 |
| 21 | 19 | 2 |
| 22 | 12 | 1 |
| 23 |    |   |

Figura 20: Ilustración del cálculo de la moda en el ejemplo 2

- 3. **Mediana.** La mediana es el primer valor de la columna *Clasificación* cuyo porcentaje acumulado es superior o igual al 50 %. Observando la tabla de frecuencias de la figura 19 vemos que la mediana es 16, ya que su porcentaje acumulado es 55, 26 %.
- 4. **1<sup>er</sup> cuartil.** El cálculo es similar al de la mediana. El 1<sup>er</sup> cuartil es el primer valor de la columna *Clasificación* cuyo porcentaje acumulado es superior o igual al 25 %. Observando la tabla de frecuencias de la figura 19 vemos que el 1<sup>er</sup> cuartil es 15, ya que su porcentaje acumulado es 39, 47 %.
- 5. **3<sup>er</sup> cuartil.** Igual que en el caso anterior pero buscando un porcentaje acumulado mayor o igual que 75 %. En este caso el 3<sup>er</sup> cuartil es 18, ya que su porcentaje acumulado es 78, 95 %.
- 6. **Media.** La media en este caso se calcula de manera muy sencilla del siguiente modo:
  - a) Situamos en cursor en una casilla vacía cualquiera, por ejemplo la casilla F2.
  - b) Escribimos la fórmula

=SUMA . PRODUCTO(A2:A10 ; B2:B10) / SUMA(B2:B10)<sup>2</sup> y pulsamos *Enter*. El valor de la media se escribe en la casilla *F2*. En este caso, media=16,34.

Por último comentar que los valores de moda, mediana y media obtenidos son los mismos que en el ejemplo anterior.

### Ejemplo 3

Calcular la mediana y la moda de la calificación de los alumnos de *Fonaments Matemàtics II* (Ingeniería Telemática) del curso 2003-2004, a partir de los datos de la siguiente tabla (fuente UIB). ¿Es posible calcular la media?

| Assignatura:      | 2485 - Fonaments Matemàtics II |               |      |
|-------------------|--------------------------------|---------------|------|
| Any acadèmic:     | 2003-04                        | Convocatòria: | Juny |
| Qualificació      | Núm. alumnes                   |               |      |
| Suspens           | 17                             |               |      |
| Aprovat           | 16                             |               |      |
| Notable           | 12                             |               |      |
| Excel·lent        | 1                              |               |      |
| Matricula d'honor | 3                              |               |      |

En este ejemplo, la variable “Qualificació” es una variable ordinal por lo que no es posible calcular su media, pero sí su moda y mediana.

Al tratarse de una tabla con muy pocos valores la moda se encuentra fácilmente observando la tabla: moda=*Suspens*, ya que es el valor con la máxima frecuencia absoluta.

Para calcular la mediana debemos calcular primero los porcentajes acumulados. Lo podemos hacer en una hoja de cálculo de OpenOffice Calc tal como se ha explicado en el tema anterior y obtendríamos el resultado de la figura 21.

A continuación seguimos el procedimiento explicado en el ejemplo 2 para el cálculo de la mediana: buscamos el primer valor de la columna *Qualificació* cuyo porcentaje acumulado sea superior o igual a 50 %. En este caso la solución es *Aprovat* (porcentaje acumulado=67,35).

Este ejemplo muestra como el valor de la moda no explica suficientemente bien la distribución de valores. En este caso la moda era *Suspens*, sin embargo más de la mitad de los estudiantes han aprobado (de hecho  $16 + 12 + 1 + 3 = 32$  estudiantes

---

<sup>2</sup>Esta fórmula calcula la media empleando la fórmula vista en clase: multiplica los valores de las casillas *A2* y *B2*, *A3* y *B3*, etc, suma los productos y finalmente divide el total por la suma de los valores de las casillas *B2* a *B10*

|   | A            | B              | C               | D                    |  |
|---|--------------|----------------|-----------------|----------------------|--|
| 1 | Qualificació | Frec. Absoluta | Frec. acumulada | Porcentaje acumulado |  |
| 2 | Suspens      | 17             | 17              | 34,69                |  |
| 3 | Aprovat      | 16             | 33              | 67,35                |  |
| 4 | Notable      | 12             | 45              | 91,84                |  |
| 5 | Excellent    | 1              | 46              | 93,88                |  |
| 6 | Matricula H. | 3              | 49              | 100                  |  |
| 7 |              |                |                 |                      |  |
|   |              |                |                 |                      |  |

Figura 21: Izquierda: frecuencias absolutas y acumuladas para el ejemplo 3.

aprueban, exactamente el doble de alumnos suspendidos). La mediana describe de manera mejor los valores de la variable al dar un valor de *Aprovat*.

#### Ejemplo 4

Calcular la moda y la mediana de la edad de los condenados en Illes Balears en 2005 a partir de los datos de la siguiente tabla. Calcular los cuartiles primero y tercero. ¿Es posible calcular la media? Si la respuesta es negativa, ¿cómo estimarías de manera aproximada el valor de la media?

| Estadísticas judiciales 2005                                |     |
|---|-----|
| Estadística de lo Penal. Condenados. Resultados autonómicos |     |
| Condenados según edad y sexo                                |     |
| Unidades: nº de condenados                                  |     |
| Ambos sexos   |     |
| Balears (Illes)   |     |
| De 18 a 20 años   | 155 |
| De 21 a 25 años   | 543 |
| De 26 a 30 años   | 653 |
| De 31 a 35 años   | 619 |
| De 36 a 40 años   | 515 |
| De 41 a 50 años   | 636 |
| De 51 a 60 años   | 248 |
| De 60 y más   | 100 |

Fuente: Instituto Nacional de Estadística

Se trata de datos agrupados en forma de intervalos, de manera que calcularemos los estadísticos siguiendo el procedimiento explicado en clase:

1. **Moda.** Observando la tabla vemos que el valor máximo se da en el intervalo 26 – 30. La moda se calcula como el valor medio del intervalo, es decir:  $\text{moda} = \frac{26+30}{2} = 28$ .
2. **Mediana.** Calculamos las frecuencias acumuladas para cada intervalo y seguimos el procedimiento descrito en el ejemplo 2 para hallar la mediana. En este caso, la mediana se encuentra en el intervalo 31 – 35 (ver figura 22).

|    | A     | B              | C               | D                    |  |
|----|-------|----------------|-----------------|----------------------|--|
| 1  | Edad  | Frec. Absoluta | Frec. Acumulada | Porcentaje acumulado |  |
| 2  | 18-20 | 155            | 155             | 4,47                 |  |
| 3  | 21-25 | 543            | 698             | 20,12                |  |
| 4  | 26-30 | 653            | 1351            | 38,94                |  |
| 5  | 31-35 | 619            | 1970            | 56,79                |  |
| 6  | 36-40 | 515            | 2485            | 71,63                |  |
| 7  | 41-50 | 636            | 3121            | 89,97                |  |
| 8  | 51-60 | 248            | 3369            | 97,12                |  |
| 9  | 60-70 | 100            | 3469            | 100                  |  |
| 10 |       |                |                 |                      |  |

Figura 22: Tabla de frecuencias absolutas y acumuladas y porcentajes acumulados para el ejemplo 3.

Para calcular de modo más preciso la mediana utilizamos la fórmula dada en clase:

$$\text{mediana} = L_i + \frac{50\% \cdot N - N_{i-1}}{n_i} \cdot (L_{i+1} - L_i)$$

donde  $L_i$  y  $L_{i+1}$  denotan los límites inferior y superior del intervalo,  $n_i$  es la frecuencia del intervalo,  $N_{i-1}$  es la frecuencia acumulada en el intervalo anterior y  $N$  es la suma de todas las frecuencias

En nuestro caso:  $L_i = 31$ ,  $L_{i+1} = 35$ ,  $n_i = 619$ ,  $N_{i-1} = 1351$  y  $N = 3469$ , por tanto

$$\begin{aligned} \text{mediana} &= 31 + \frac{50\% \cdot 3469 - 1351}{619} \cdot (35 - 31) = \\ &= 31 + \frac{1734,5 - 1351}{619} \cdot 4 = 33,48 \end{aligned}$$

El cálculo de los cuartiles es muy similar al de la mediana. Primero hallamos los intervalos en los que se encuentra cada uno de ellos, siguiendo un procedimiento similar al explicado en el ejemplo 2. A continuación aplicamos la fórmula de los cuartiles:

- **Primer cuartil.**  $L_i = 26$ ,  $L_{i+1} = 30$ ,  $n_i = 653$ ,  $N_{i-1} = 698$  y  $N = 3469$

$$26 + \frac{25\% \cdot 3469 - 698}{653} \cdot (30 - 26) = 26 + \frac{867,25 - 698}{653} \cdot 4 = 27,04$$

- **Tercer cuartil.**  $L_i = 41$ ,  $L_{i+1} = 50$ ,  $n_i = 636$ ,  $N_{i-1} = 2485$  y  $N = 3469$

$$41 + \frac{75\% \cdot 3469 - 2485}{636} \cdot (50 - 41) = 41 + \frac{2601,75 - 2485}{636} \cdot 9 = 42,65$$

3. **Media.** Para calcular la media de unos valores agrupados en intervalos el primer paso consiste en calcular el valor medio de cada intervalo. En este

ejemplo sin embargo tenemos el problema de que para el último intervalo no podemos calcular el valor medio, ya que está definido como *60 y más* y no conocemos el límite superior:

En estos casos podemos calcular la media de manera aproximada haciendo alguna suposición razonable sobre el valor máximo del intervalo desconocido. A continuación explicamos el procedimiento a seguir si suponemos que el valor máximo del intervalo es 70:

- Partimos de un documento OpenOffice Calc en el que hemos creado una tabla de frecuencias absolutas y acumuladas como la que se muestra en la figura 22.
- Insertamos una nueva columna a la derecha de la columna *A*. Para ello situamos el cursor sobre la parte superior de la columna *B*, hacemos clic en el botón derecho del ratón y elegimos la opción *insertar columnas*. Una nueva columna *B* aparece y las columnas *B*, *C*, *D*, etc se desplazan hacia a la derecha (ver figura 23-arriba).
- En la primera casilla de la nueva columna escribimos *Edad media* y en las casillas inferiores escribimos las fórmulas que calculan los valores medios de los intervalos:  $=(18+20)/2$ , *Enter* (casilla *B2*);  $=(21+25)/2$ , *Enter* (casilla *B3*); etc. Finalmente, para la casilla *B9* *suponemos* que el valor máximo del intervalo es 70 y escribimos  $=(60+70)/2$ . La tabla resultante tiene la forma que se muestra en la figura 23-abajo.

|    | A     | B | C              | D               | E                    |
|----|-------|---|----------------|-----------------|----------------------|
| 1  | Edad  |   | Frec. Absoluta | Frec. Acumulada | Porcentaje acumulado |
| 2  | 18-20 |   | 155            | 155             | 4,47                 |
| 3  | 21-25 |   | 543            | 698             | 20,12                |
| 4  | 26-30 |   | 653            | 1351            | 38,94                |
| 5  | 31-35 |   | 619            | 1970            | 56,79                |
| 6  | 36-40 |   | 515            | 2485            | 71,63                |
| 7  | 41-50 |   | 636            | 3121            | 89,97                |
| 8  | 51-60 |   | 248            | 3369            | 97,12                |
| 9  | 60-70 |   | 100            | 3469            | 100                  |
| 10 |       |   |                |                 |                      |

|    | A     | B          | C              | D               | E                    |
|----|-------|------------|----------------|-----------------|----------------------|
| 1  | Edad  | Edad media | Frec. Absoluta | Frec. Acumulada | Porcentaje acumulado |
| 2  | 18-20 | 19         | 155            | 155             | 4,47                 |
| 3  | 21-25 | 23         | 543            | 698             | 20,12                |
| 4  | 26-30 | 28         | 653            | 1351            | 38,94                |
| 5  | 31-35 | 33         | 619            | 1970            | 56,79                |
| 6  | 36-40 | 38         | 515            | 2485            | 71,63                |
| 7  | 41-50 | 45,5       | 636            | 3121            | 89,97                |
| 8  | 51-60 | 55,5       | 248            | 3369            | 97,12                |
| 9  | 60-70 | 65         | 100            | 3469            | 100                  |
| 10 | ...   |            |                |                 |                      |

Figura 23: Tablas de frecuencias y porcentajes para el ejemplo 3. Arriba, inserción de una nueva columna. Abajo, datos insertados con los valores centrales de los intervalos

- El cálculo de la media se hace ahora de manera similar al ejemplo 2:
  - Situamos en cursor en una casilla vacía cualquiera, por ejemplo la casilla *A12*.

- 2) Escribimos la fórmula  
 $=\text{SUMA}.\text{PRODUCTO}(\text{B2:B9};\text{C2:C9})/\text{SUMA}(\text{C2:C9})$  y  
 pulsamos *Enter*. El valor de la media se escribe en la casilla *A12*.  
 En este caso, media=35,43.

Como comentario final decir que otras suposiciones razonables sobre el valor máximo del intervalo 60 *y más* hubieran producido resultados similares. Por ejemplo, para la suposición 60 – 65 hubiéramos obtenido media=35,36; para 60 – 75, media=35,51; para 60 – 80, media=35,58, etc. Lo importante es no suponer valores absurdos (por ejemplo 60 – 150, pues es muy poco probable que haya personas de 150 años que cometan delitos).

### Ejemplo 5

Calcular los estadísticos de tendencia central asociados a la variable “tipo de infracción” a partir de los datos de la siguiente tabla:

| Estadísticas judiciales 2005  |                 |
|---|-----------------|
| Estadística de lo Penal. Menores. Resultados autonómicos y provinciales |                 |
| Menores según infracción cometida                                       |                 |
| Unidades:nº de menores  |                 |
|   | BALEARS (ILLES) |
| Homicidio   | 0               |
| Lesiones  | 31              |
| Contra la libertad sexual   | 12              |
| Hurto   | 69              |
| Robo  | 306             |
| Contra la salud pública   | 10              |

Fuente:Instituto Nacional de Estadística

Se trata de una variable cualitativa por lo que el único estadístico que podemos calcular es la moda. Observando la tabla vemos que el valor máximo de frecuencia es 306, que corresponde al valor *Robo*. Por lo que concluimos que: moda=*Robo*.

### Ejemplo 6

Calcular los estadísticos de tendencia central asociados a la variable “autobuses matriculados por mes durante el año 2006” a partir de los datos de la siguiente tabla (fuente DGT):

| MATRICULACIONES POR MES Y TIPO DE VEHÍCULO |         |                         |                                      |           |          |              |                        |                 |
|--|---------|-------------------------|--------------------------------------|-----------|----------|--------------|------------------------|-----------------|
| Meses                                      | Total   | Camiones MMA > 3.500 kg | Camiones MMA ≤ 3.500 kg y furgonetas | Autobuses | Turismos | Motocicletas | Tractores Industriales | Otros vehículos |
| Enero                                      | 158.712 | 1.579                   | 25.601                               | 170       | 115.490  | 14.402       | 1.013                  | 457             |
| Febrero                                    | 178.150 | 1.852                   | 29.646                               | 274       | 128.831  | 15.888       | 1.217                  | 442             |
| Marzo                                      | 243.927 | 2.274                   | 39.265                               | 359       | 176.075  | 23.389       | 1.791                  | 774             |
| Abril                                      | 187.706 | 2.003                   | 29.014                               | 427       | 131.631  | 21.985       | 1.936                  | 710             |
| Mayo                                       | 224.821 | 2.106                   | 35.772                               | 376       | 155.805  | 28.302       | 1.783                  | 677             |
| Junio                                      | 246.787 | 2.301                   | 36.496                               | 364       | 171.028  | 33.877       | 1.906                  | 815             |
| Julio                                      | 240.910 | 2.228                   | 34.286                               | 278       | 169.034  | 32.703       | 1.722                  | 659             |
| Agosto                                     | 158.200 | 1.746                   | 25.308                               | 158       | 105.190  | 24.002       | 1.377                  | 419             |
| Septiembre                                 | 158.949 | 1.478                   | 24.497                               | 634       | 107.510  | 21.908       | 2.578                  | 344             |
| Octubre                                    | 186.334 | 1.858                   | 30.055                               | 282       | 128.178  | 23.135       | 2.397                  | 429             |
| Noviembre                                  | 193.677 | 1.949                   | 33.655                               | 235       | 135.134  | 20.231       | 2.000                  | 473             |
| Diciembre                                  | 186.483 | 1.486                   | 31.106                               | 290       | 136.721  | 15.096       | 1.368                  | 416             |

Los datos de la tabla son valores *en bruto*, por lo que aplicamos el procedimiento explicado en el ejemplo 1:

1. Escribimos los datos *brutos* en la primera columna de la hoja de cálculo (casillas A1 a A12). El resultado se muestra en la figura 24.

|    | A   |  |
|----|-----|--|
| 1  | 170 |  |
| 2  | 274 |  |
| 3  | 359 |  |
| 4  | 427 |  |
| 5  | 376 |  |
| 6  | 364 |  |
| 7  | 278 |  |
| 8  | 158 |  |
| 9  | 634 |  |
| 10 | 282 |  |
| 11 | 235 |  |
| 12 | 290 |  |
| 13 |     |  |

Figura 24: Datos brutos del ejemplo 6

2. A continuación escribimos las palabras “Moda”, “Media”, “Mediana”, “1er cuartil” y “3er cuartil” en las casillas C15, C16, C17, C18 y C19 de la hoja de cálculo (o en otras casillas cualesquiera)
3. **Moda.** Nos situamos en la casilla D15 y escribimos =Moda(A1:A12). Al pulsar *Enter* obtenemos el valor de la moda. En este caso obtenemos un mensaje de error pues todos los valores ocurren una única vez. Esto significa que el cálculo de la moda no tiene sentido en este problema.
4. **Media.** Nos situamos en la casilla D16 y escribimos =Promedio(A1:A12). Al pulsar *Enter* obtenemos el valor de la media.

5. **Mediana.** Nos situamos en la casilla *D17* y escribimos =Mediana(A1:A12). Al pulsar *Enter* obtenemos el valor de la mediana.
6. **1<sup>er</sup> cuartil.** Nos situamos en la casilla *D18* y escribimos =Cuartil(A1:A12;1). Al pulsar *Enter* obtenemos el valor del primer cuartil.
7. **3<sup>er</sup> cuartil.** Nos situamos en la casilla *D19* y escribimos =Cuartil(A1:A12;3). Al pulsar *Enter* obtenemos el valor del tercer cuartil.

El resultado obtenido es:

|                         |        |
|-------------------------|--------|
| Media                   | 320,58 |
| Mediana                 | 286    |
| 1 <sup>er</sup> cuartil | 264,25 |
| 3 <sup>er</sup> cuartil | 367    |

**Nota:** si para el cálculo de la mediana y cuartiles hubiéramos calculado primero la tabla de frecuencias absolutas de cada valor y a continuación hubéramos hecho el cálculo con el procedimiento explicado en los ejemplos anteriores, los resultados hubieran sido ligeramente diferentes: mediana=282, 1<sup>er</sup> cuartil=235 y 3<sup>er</sup> cuartil=364. La razón es que Calc utiliza unas fórmulas diferentes a las nuestras para el cálculo de cuartiles para datos *brutos*.

## 2.1. Ejercicios propuestos

### Ejercicio 1

A partir de los datos de la siguiente tabla calcular los estadísticos de tendencia central para la variable “Edad de víctimas de accidentes en 2006”. Calcular también el primer y el tercer cuartiles.

| Edad (años) | Nº víctimas |
|-------------|-------------|
| 0 a 4       | 343         |
| 5 a 14      | 1172        |
| 15 a 17     | 333         |
| 18 a 24     | 918         |
| 25 a 64     | 5026        |
| 65 y más    | 2947        |

Edad de víctimas de accidentes en 2006 (fuente DGT)

### Ejercicio 2

Calcular los estadísticos de tendencia central y los cuartiles 1º y 3º asociados a la variable “turismos matriculados por mes durante el año 2006” a partir de los datos de la tabla del ejemplo 6.

### Ejercicio 3

Calcular los estadísticos de tendencia central asociados a la variable “tipo de alojamiento” a partir de los datos de la siguiente tabla (fuente INE):

| Encuesta de ocupación en alojamientos de turismo rural 2006           |       |
|---|-------|
| Datos nacionales por modalidades. Oferta                              |       |
| Grado de ocupación por plazas en fin de semana por modalidad y meses. |       |
| Unidades: %   |       |
|   | Total |
| Hotel Rural   | 40,50 |
| Apartamento Rural   | 30,19 |
| Casa Rural  | 32,49 |
| Albergue Rural  | 24,25 |

### Ejercicio 4

Calcular los estadísticos de tendencia central y los cuartiles 1º y 3º de la variable “terminación del cupón de la ONCE en el período 30 de octubre a 27 de noviembre de 2007” a partir de los datos de la siguiente tabla (fuente ONCE). ¿Qué tipo de simetría presenta la distribución?

| Terminación | Frecuencia |
|-------------|------------|
| 0           | 0          |
| 1           | 4          |
| 2           | 0          |
| 3           | 2          |
| 4           | 2          |
| 5           | 3          |
| 6           | 3          |
| 7           | 6          |
| 8           | 3          |
| 9           | 1          |

### Ejercicio 5

Calcular los estadísticos de tendencia central del grado de satisfacción de los clientes de un determinado establecimiento a partir de los datos siguientes:

| Grado satisfacción  | <i>Nº</i> clientes |
|---------------------|--------------------|
| Muy satisfecho      | 50                 |
| Bastante satisfecho | 80                 |
| Satisfecho          | 100                |
| Poco satisfecho     | 40                 |
| Nada satisfecho     | 10                 |