CAPÍTULO 8
ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

1. INTERROGANTES CENTRALES DEL CAPÍTULO

a) ¿Qué es la Estadística?

b) ¿Cómo se organizan los datos de una muestra extraída de una población?

c) ¿Cuáles son las representaciones gráficas más adecuadas para extraer conclusiones sobre el comportamiento real de la variable estadística?

d) ¿Cómo se sintetiza la información aportada por un conjunto de datos en una serie de valores que nos fijen el comportamiento global del experimento aleatorio?

2. CONTENIDOS FUNDAMENTALES DEL CAPÍTULO

2.1. Introducción a la Estadística

2.1.1. *Definición de Estadística*

Estadística: ciencia que se ocupa de recoger, clasificar, representar y resumir los datos de muestras, y de hacer inferencias (extraer conclusiones) acerca de las poblaciones de las que éstas proceden.

1. *Estadística descriptiva*: parte de la estadística que se ocupa de recoger, clasificar, representar y resumir los datos de las muestras.
2. *Estadística inferencial*: parte de la estadística que se ocupa de llegar a conclusiones (inferencias) acerca de las poblaciones a partir de los datos de las muestras extraídas de ellas.

2.1.2. *Conceptos generales*

– *Población*: conjunto de individuos con propiedades comunes sobre los que se realiza una investigación de tipo estadístico.

– *Muestra*: subconjunto de la población.

– *Tamaño muestral*: número de individuos que forman la muestra.

– *Muestreo*: proceso de obtención de muestras representativas de la población.

– *Variable*: propiedad o cualidad que puede manifestarse bajo dos o más formas distintas en un individuo de una población.

– *Modalidades, categorías o clases*: distintas formas en que se manifiesta una variable.

– Las variables se clasifican en:

1. *Cuantitativas*: se expresan numéricamente. Se clasifican en:
   (a) *Discretas*: toman valores numéricos aislados, por lo que, fijados dos consecutivos, no pueden tomar ningún valor intermedio.
   (b) *Continuas*: pueden tomar cualquier valor dentro de unos límites, por lo que entre dos valores cualesquiera, por próximos que sean, siempre pueden encontrarse valores intermedios.
2. *Cualitativas*: no se expresan numéricamente. Se clasifican en:
   (a) *Ordinales*: admiten una ordenación de menor a mayor aunque sus resultados no son numéricos.
   (b) *Nominales*: no admiten una ordenación de menor a mayor.

2.2. Tabulación de los datos

2.2.1. *Variables cualitativas*

Ejemplo de recogida (no ordenada) de unos datos cualitativos:

francés	inglés	francés	inglés	francés	alemán	ruso	español	francés	inglés
francés	inglés	español	francés	español	francés	alemán	inglés	español	inglés
inglés	español	inglés	francés	español	ruso	alemán	francés	inglés	español
alemán	inglés	español	francés	alemán	inglés	inglés	inglés	español	francés

- *Frecuencia absoluta* de la clase i-ésima: f_i =número de observaciones contenidas dentro de ella.
- *Frecuencia relativa* de la clase i-ésima: $h_i = \dfrac{f_i}{n}$, siendo n el número total de observaciones.
- *Porcentaje* de la clase i-ésima: $\%_i = 100\, h_i$.
- Se verifican las propiedades siguientes:

$$f_1 + f_2 + \cdots + f_k = n,$$
$$h_1 + h_2 + \cdots + h_k = 1,$$
$$\%_1 + \%_2 + \cdots + \%_k = 100,$$

  siendo k el número de clases.

- *Distribución de frecuencias*: tabla conteniendo las distintas clases y las frecuencias correspondientes a cada una de ellas.

La distribución de frecuencias de los datos cualitativos del ejemplo anterior es:

clases	f_i	h_i	$\%_i$
alemán	5	$0'125$	$12'5$
español	9	$0'225$	$22'5$
francés	11	$0'275$	$27'5$
inglés	13	$0'325$	$32'5$
ruso	2	$0'050$	$5'0$
suma	40	$1'000$	$100'0$

2.2.2. *Variables discretas*

Ejemplo de recogida (no ordenada) de unos datos cuantitativos discretos:

14	13	3	13	7	12	13	11	13	12	11	13
7	10	12	13	14	11	13	12	4	12	10	13
9	12	13	11	13	14	10	12	11	13	15	9
12	11	13	10	13	11	12	5	9	12	13	15

Los mismos criterios usados para el caso cualitativo sirven para el caso cuantitativo discreto a la hora de presentar tabularmente los datos. Además se pueden calcular:

- *Frecuencia absoluta acumulada* de la clase i-ésima: $F_i = f_1 + f_2 + \cdots + f_i$=número de individuos que caen dentro de dicha clase o cualquier clase anterior (una vez ordenadas las clases de menor a mayor).
- *Frecuencia relativa acumulada* de la clase i-ésima: $H_i = h_1 + h_2 + \cdots + h_i = \dfrac{F_i}{n}$.

La distribución de frecuencias de los datos cuantitativos discretos del ejemplo anterior es:

x_i	f_i	h_i	F_i	H_i
3	1	$0'0208$	1	$0'0208$
4	1	$0'0208$	2	$0'0417$
5	1	$0'0208$	3	$0'0625$
7	2	$0'0417$	5	$0'1042$
9	3	$0'0625$	8	$0'1667$
10	4	$0'0833$	12	$0'2500$
11	7	$0'1458$	19	$0'3958$
12	10	$0'2083$	29	$0'6042$
13	14	$0'2917$	43	$0'8958$
14	3	$0'0625$	46	$0'9583$
15	2	$0'0417$	48	$1'0000$

2.2.3. *Variables continuas*

Los datos procedentes de una variable continua se pueden tabular de la misma forma que los datos de una variable discreta, pero lo usual en el caso de variables continuas es dividir el intervalo de valores posibles en intervalos

contiguos llamados *intervalos de clase*. Una vez agrupados los datos en intervalos, éstos se tabulan de forma análoga al caso de variables discretas.

- Número adecuado de intervalos (Regla de Sturges):

$$k = 1 + 3'322 \log n,$$

siendo n el número total de observaciones.

- *Amplitud* del intervalo de clase (ℓ_i, ℓ_{i+1}): $d = \ell_{i+1} - \ell_i$.
- *Marca de clase* del intervalo (ℓ_i, ℓ_{i+1}): $x_i = \dfrac{\ell_i + \ell_{i+1}}{2}$.

Ejemplo de recogida (no ordenada) de unos datos cuantitativos continuos:

$3'9$	$4'1$	$4'2$	$3'2$	$1'6$
$2'5$	$1'1$	$8'1$	$5'1$	$2'7$
$1'9$	$7'3$	$2'4$	$4'9$	$1'6$
$5'0$	$2'5$	$6'5$	$1'9$	$5'2$
$6'3$	$1'2$	$3'3$	$1'8$	$4'4$

Pasos para la agrupación en intervalos de clase de igual amplitud:

1º) Se calcula el recorrido de las observaciones (ver el apartado 2.5.1.):

$$R = x_{max} - x_{min} = 8'1 - 1'1 = 7\,.$$

2º) El número de intervalos de clase que podemos tomar para agrupar los datos es:

$$k = 1 + 3'322\ log\, 25 = 5'64\,,$$

que aproximamos por el número natural siguiente: $k = 6$.

3º) Por tanto, la amplitud de cada intervalo es:

$$d = \frac{R}{k} = \frac{7}{6} = 1'1667\,.$$

Al no ser exacta, aproximamos la cantidad anterior a un número ligeramente superior; por ejemplo, $d = 1'17$.

Como la amplitud de los intervalos la hemos tomado un poco mayor de lo que se obtiene en un principio, entonces el nuevo recorrido es:

$$R' = \text{número de intervalos} \cdot \text{ amplitud} = 6 \cdot 1'17 = 7'02\,.$$

Como el recorrido original es 7, entonces sobra $0'02$, con lo cual repartimos este sobrante restando la mitad a la observación mínima y sumando la otra mitad a la observación máxima; es decir:

$$x_{min} - 0'01, \qquad x_{max} + 0'01\,,$$

con lo que obtenemos los seis intervalos de clase determinados por los valores siguientes:

$$\begin{aligned} x_{min} - 0'01 &= 1'10 - 0'01 = 1'09,\\ &\quad 1'09 + 1'17 = 2'26,\\ &\quad 2'26 + 1'17 = 3'43,\\ &\quad 3'43 + 1'17 = 4'60,\\ &\quad 4'60 + 1'17 = 5'77,\\ &\quad 5'77 + 1'17 = 6'94,\\ &\quad 6'94 + 1'17 = 8'11 = x_{max} + 0'01\,. \end{aligned}$$

Los intervalos son: $(1'09, 2'26]$, $(2'26, 3'43]$, $(3'43, 4'60]$, $(4'60, 5'77]$, $(5'77, 6'94]$, $(6'94, 8'11]$. Agrupamos, pues, los datos en los intervalos anteriores y damos su distribución de frecuencias en la tabla siguiente:

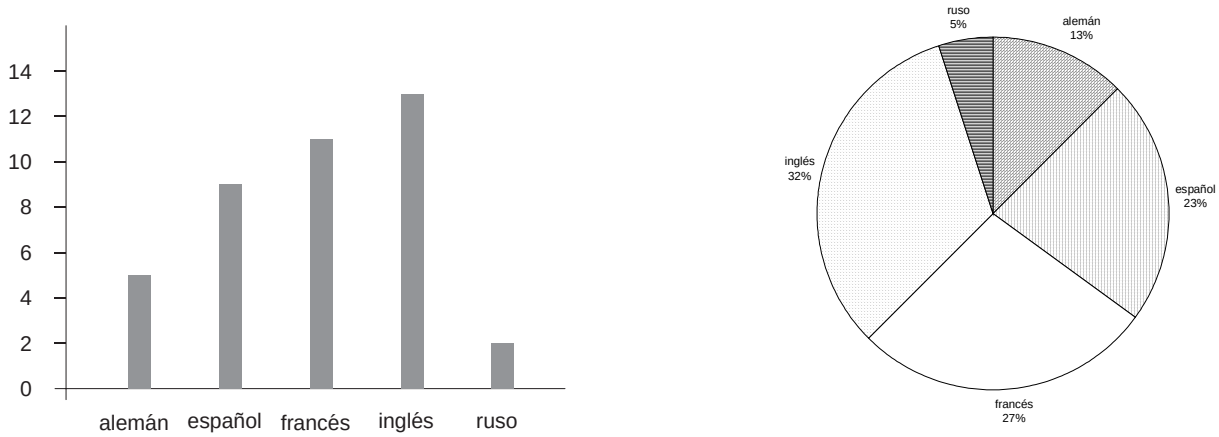
$(l_i, l_{i+1}]$	x_i	f_i	h_i	F_i	H_i
$(1'09, 2'26]$	$1'675$	7	$0'28$	7	$0'28$
$(2'26, 3'43]$	$2'845$	6	$0'24$	13	$0'52$
$(3'43, 4'60]$	$4'015$	4	$0'16$	17	$0'68$
$(4'60, 5'77]$	$5'185$	4	$0'16$	21	$0'84$
$(5'77, 6'94]$	$6'355$	2	$0'08$	23	$0'92$
$(6'94, 8'11]$	$7'525$	2	$0'08$	25	$1'00$

La agrupación en intervalos de clase también puede hacerse con una variable discreta si el número de datos distintos es muy grande.

2.3. Representaciones gráficas

2.3.1. *Variables cualitativas*

1. *Diagrama de barras*: se sitúan en el eje horizontal las clases y sobre cada una de ellas se levanta un segmento rectilíneo (o un rectángulo) de altura igual a la frecuencia (absoluta o relativa) de cada clase (Figura 8.1 (a)).
2. *Gráfico de sectores*: se divide el área de un círculo en sectores circulares de ángulos proporcionales a las frecuencias absolutas de las clases (Figura 8.1 (b)).



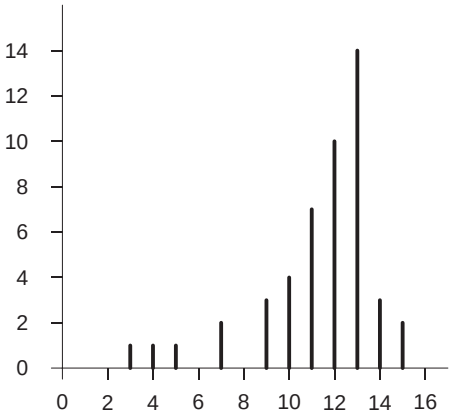
a) Diagrama de barras de frecuencias absolutas b) Gráfico de sectores

Figura 8.1: Representaciones gráficas de los datos cualitativos de la sección 2.2.1.

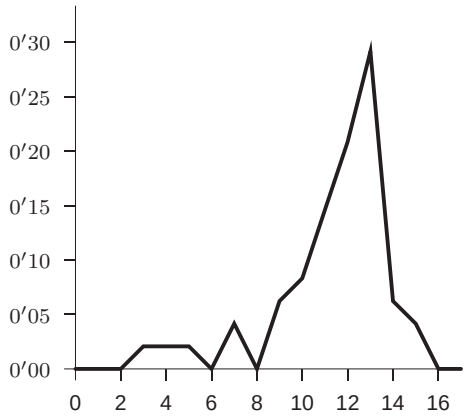
2.3.2. *Variables cuantitativas con datos no agrupados en intervalos*

1. *Diagrama de barras*: igual que en el caso de variables cualitativas (Figura 8.2 (a)).
2. *Polígono de frecuencias*: se sitúan los puntos que resultan de tomar en el eje horizontal los distintos valores de la variable y en el eje vertical sus correspondientes frecuencias (absolutas o relativas), uniendo después los puntos mediante segmentos rectilíneos (Figura 8.2 (b)).

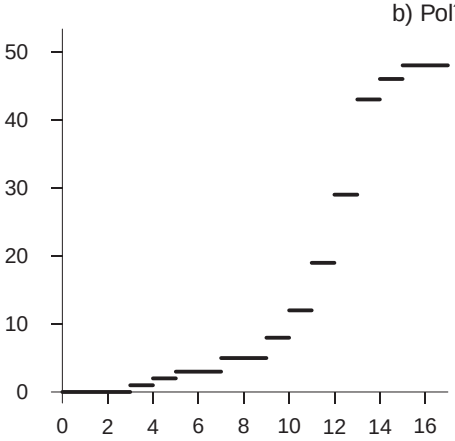
3. *Gráfico de frecuencias acumuladas*: es la representación gráfica de las frecuencias acumuladas (absolutas o relativas), para todo valor numérico. Indiquemos que la frecuencia acumulada (absoluta o relativa) de un valor numérico que no aparezca en la distribución de frecuencias es igual a la frecuencia acumulada (absoluta o relativa) de la observación inmediatamente anterior (ordenadas de menor a mayor). Por tanto, el gráfico de frecuencias acumuladas siempre tiene forma de "escalera" (Figura 8.2 (c)).



a) Diagrama de barras de frecuencias absolutas



b) Polígono de frecuencias relativas



c) Gráfico de frecuencias absolutas acumuladas

Figura 8.2: Representaciones gráficas de los datos cuantitativos discretos de la sección 2.2.2.

2.3.3. *Variables cuantitativas con datos agrupados en intervalos*

1. *Histograma de frecuencias*: se sitúan en el eje horizontal los intervalos de clase y sobre cada uno se levanta un rectángulo de área proporcional a la frecuencia absoluta.

(a) Si todos los intervalos tienen la misma amplitud, entonces basta con hacer los rectángulos con una altura igual a la frecuencia absoluta (Figura 8.4 (a)).

(b) Si los intervalos tienen distinta amplitud, la construcción del histograma presenta una importante variación. Una vez marcados sobre el eje horizontal los extremos de los intervalos, hay que calcular la altura de los rectángulos de forma que su área sea igual o proporcional a la frecuencia absoluta del intervalo. Veámoslo con un ejemplo.

Sea la siguiente distribución de frecuencias:

$intervalo$	f_i
$[0,3]$	11
$(3,5'5]$	10
$(5'5,6'5]$	2
$(6'5,8]$	1
$(8,10]$	1

Recordemos que la fórmula del *área* de un rectángulo es *base* $\times$ *altura* y consideremos que los rectángulos del histograma van a tener un área igual a la frecuencia absoluta. Por ejemplo, para averiguar la altura del primer rectángulo, tenemos en cuenta que la base es igual a 3 y el área del rectángulo ha de ser igual a 11, por lo que la altura debe ser igual a $11/3 = 3'6667$. Procediendo de forma análoga con el resto de los intervalos, se tiene que el histograma de la distribución de frecuencias dada por la tabla anterior es la Figura 8.3.

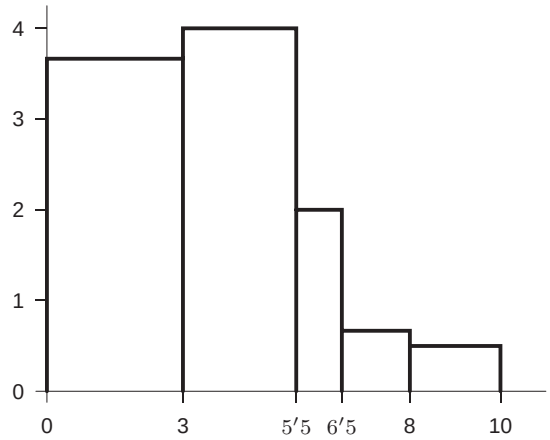


Figura 8.3: Histograma (intervalos de distinta amplitud)

2. *Polígono de frecuencias*: se sitúan los puntos que resultan de tomar en el eje horizontal las marcas de clase de los intervalos y en el eje vertical sus correspondientes frecuencias (absolutas o relativas), uniendo después los puntos mediante segmentos rectilíneos (Figura 8.4 (b)).

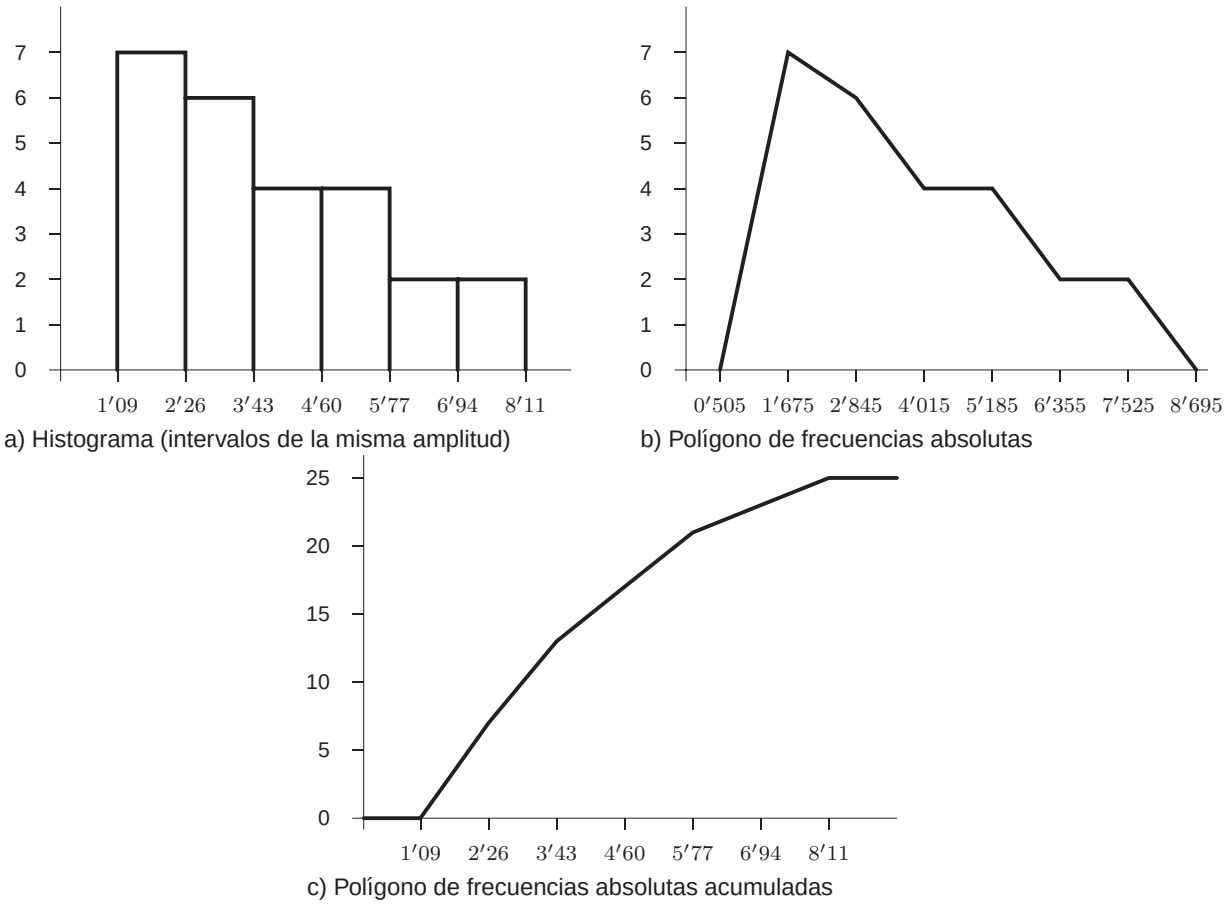
3. *Polígono de frecuencias acumuladas*: se sitúan los puntos que resultan de tomar en el eje horizontal los extremos superiores de los intervalos de clase y en el eje vertical sus correspondientes frecuencias acumuladas (absolutas o relativas), uniendo después los puntos mediante segmentos rectilíneos (Figura 8.4 (c)).

2.4. Medidas de posición

Son valores que nos sirven para indicar la posición alrededor de la cual se distribuyen las observaciones. Sólo se calculan cuando la variable es cuantitativa.

2.4.1. *Moda*

La denotaremos por M_o. No tiene que ser única.



a) Histograma (intervalos de la misma amplitud)

b) Polígono de frecuencias absolutas

c) Polígono de frecuencias absolutas acumuladas

Figura 8.4: Representaciones gráficas de los datos cuantitativos continuos de la sección 2.2.3.

1. Datos no agrupados en intervalos.
M_o es el dato (o datos) con mayor frecuencia absoluta.

2. Datos agrupados en intervalos.
Intervalo modal: el que tiene mayor frecuencia absoluta. No tiene que ser único.

(a) Intervalo modal no único.
Las modas son las marcas de clase de los intervalos modales.

(b) Intervalo modal único.

- Intervalos de la misma amplitud.

$$M_o = \ell_i + \frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} \, (\ell_{i+1} - \ell_i) \, ,$$

donde $(\ell_i, \ell_{i+1}]$ es el intervalo modal, f_i es su frecuencia absoluta, f_{i-1} es la frecuencia absoluta del intervalo anterior al modal, y f_{i+1} es la frecuencia absoluta del intervalo posterior al modal.

- Intervalos de distinta amplitud.

$$M_o = \ell_i + \frac{k_i - k_{i-1}}{2k_i - k_{i-1} - k_{i+1}} \, (\ell_{i+1} - \ell_i) \, ,$$

donde $(\ell_i, \ell_{i+1}]$ es el intervalo modal, k_i es la altura del rectángulo del histograma que tiene de base al intervalo modal, k_{i-1} es la altura del rectángulo del histograma que tiene de base al intervalo anterior al modal, y k_{i+1} es la altura del rectángulo del histograma que tiene de base al intervalo posterior al modal.

2.4.2. *Mediana*

La denotaremos por M_e. Es el valor que tiene la propiedad de dejar a su izquierda el 50% de las observaciones y a su derecha el 50% restante, siempre que hayamos ordenado los datos de menor a mayor. Por tanto, la frecuencia absoluta acumulada de la mediana es igual a $n/2$, siendo n el número total de datos.

1. Datos no agrupados en intervalos.

(a) Si en la distribución de frecuencias no aparece ninguna frecuencia absoluta acumulada igual a $n/2$ entonces se toma como mediana el valor cuya frecuencia absoluta acumulada sea la más próxima a $n/2$ por exceso. Un caso en que esto ocurre es cuando el número total de datos es impar; en cuyo caso también se puede hallar la mediana como el dato central, una vez que los datos están ordenados de menor a mayor.

Por ejemplo, sea la distribución de frecuencias siguiente:

x_i	f_i	F_i
2	2	2
4	3	5
6	5	10
8	3	13

En este caso se verifica que $n/2 = 13/2 = 6'5$, por lo que no hay ningún dato cuya frecuencia acumulada sea igual a $n/2$. Entonces, se toma como mediana el siguiente dato; es decir, el dato cuya frecuencia acumulada es 10. Por tanto, $M_e = 6$. Esto se puede observar más claramente si recurrimos al gráfico de frecuencias absolutas acumuladas, que es la Figura 8.5.

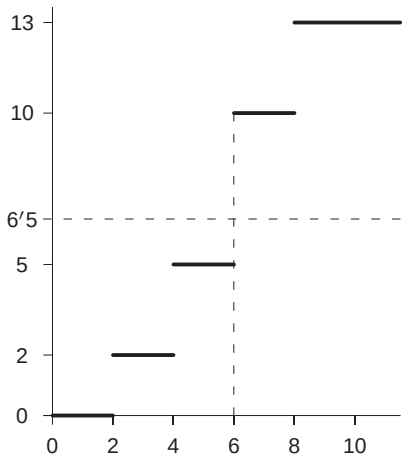


Figura 8.5: Gráfico de frecuencias absolutas acumuladas

(b) Si en la distribución de frecuencias aparece la frecuencia absoluta acumulada igual a $n/2$ entonces ocurre que hay todo un intervalo $[a, b)$ de valores cuya frecuencia absoluta acumulada es igual a $n/2$. En este caso se toma como mediana el valor $M_e = \frac{a+b}{2}$.

Para que esto ocurra es necesario que n sea par; en cuyo caso también se puede calcular la mediana ordenando los datos de menor a mayor, y luego hallando el punto medio de los dos datos centrales.

Por ejemplo, sea la distribución de frecuencias siguiente:

x_i	f_i	F_i
2	1	1
4	5	6
6	4	10
8	2	12

Todos los valores del intervalo $[4, 6)$ tienen la frecuencia absoluta acumulada igual a $n/2 = 6$, de forma que la mediana es $M_e = \frac{4+6}{2} = 5$. Esto se puede observar más claramente si recurrimos al gráfico de frecuencias absolutas acumuladas, que es la Figura 8.6.

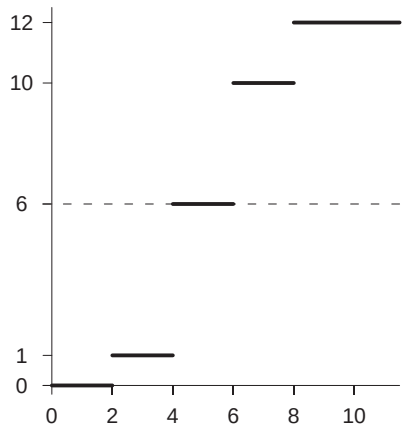


Figura 8.6: Gráfico de frecuencias absolutas acumuladas

2. Datos agrupados en intervalos.

Intervalo mediano: intervalo que contiene a la mediana. Es el primer intervalo cuya frecuencia absoluta acumulada (F_i) es igual o mayor que $n/2$.

$$M_e = \ell_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \left(\ell_{i+1} - \ell_i\right),$$

donde $(\ell_i, \ell_{i+1}]$ es el intervalo mediano, f_i es su frecuencia absoluta y F_{i-1} es la frecuencia absoluta acumulada del intervalo anterior al mediano.

2.4.3. *Percentil o cuantil*

El *percentil* (o *cuantil*) al r% es aquel valor que deja a su izquierda el r% de las observaciones y a su derecha el (100-r)% restante, siempre que hayamos ordenado los datos de menor a mayor. Se suele denotar por P_r (o por C_r).

El cálculo de los percentiles se hace de modo similar al cálculo de la mediana, teniendo en cuenta que el percentil al r% verifica que su frecuencia absoluta acumulada es igual a:

$$\frac{nr}{100}.$$

Cuando los datos están agrupados en intervalos de clase, la fórmula del percentil al r% es:

$$P_r = \ell_i + \frac{\frac{nr}{100} - F_{i-1}}{f_i} \left(\ell_{i+1} - \ell_i\right),$$

donde $(\ell_i, \ell_{i+1}]$ es el intervalo que contiene a P_r, f_i es su frecuencia absoluta y F_{i-1} es la frecuencia absoluta acumulada del intervalo anterior.

Algunos percentiles especiales son:

- *Cuartiles*: Primer cuartil $= Q_1 = P_{25}$, Segundo cuartil $= Q_2 = P_{50} = Me$ y Tercer cuartil $= Q_3 = P_{75}$.
- *Deciles*: Primer decil $= D_1 = P_{10}$, Segundo decil $= D_2 = P_{20}$, ..., Noveno decil $= D_9 = P_{90}$.

2.4.4. *Media aritmética*

Si $x_1, x_2, \ldots, x_n$ son los n valores de la muestra, su *media aritmética* es:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}\,.$$

Si los valores de los datos son $x_1, x_2, \ldots, x_k$, y ellos aparecen con frecuencias absolutas respectivas f_1, f_2, ..., f_k (con $f_1 + f_2 + \cdots + f_k = n$) entonces la expresión de la *media aritmética* es:

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{n} = \frac{\sum_{i=1}^{k} x_i f_i}{n}\,.$$

Si los datos están agrupados en intervalos de clase, la fórmula de la media aritmética es la misma, salvo que x_i representa la marca de clase del intervalo $i-$ésimo.

Dado que la media aritmética es la más común, en adelante la llamaremos sólo *media*.

Si disponemos de los datos de toda la población, entonces representamos la media aritmética por la letra griega μ (que se lee *mu*).

Propiedades de la media

1. Si $y_i = a + bx_i$, siendo a y b constantes, entonces la media de la nueva variable es $\overline{y} = a + b\overline{x}$.
2. Si $y_i = x_i - \overline{x}$, entonces $\overline{y} = 0$.

Cálculo de la media por *codificación* (para datos agrupados en intervalos de la misma amplitud)

Si a es la marca de clase del intervalo central (o una de las centrales, cuando el número de ellos es par) y b es igual a la amplitud de los intervalos, entonces realizamos la transformación siguiente:

$$z_i = \frac{x_i - a}{b},$$

para todas las marcas de clase x_i, y luego hallamos la media $\overline{z}$. Como $x_i = a + bz_i$ entonces por la propiedad 3 se cumple que la media deseada es $\overline{x} = a + b\overline{z}$.

2.4.5. *Otras medias*

De los otros tipos de medias, la de mayor interés práctico es la *media ponderada*. Ésta consiste en asignar a cada valor x_i de los datos un peso p_i que depende de su importancia relativa bajo algún criterio. La definición de la

media ponderada es:

$$\overline{x}_p = \frac{x_1p_1 + x_2p_2 + \cdots + x_kp_k}{p_1 + p_2 + \cdots + p_k} = \frac{\displaystyle\sum_{i=1}^{k} x_ip_i}{\displaystyle\sum_{i=1}^{k} p_i}\,.$$

Si los datos de la muestra son $x_1, x_2, \ldots, x_k$ y ellos aparecen con frecuencias absolutas respectivas f_1, f_2, ..., f_k (con $f_1 + f_2 + \cdots + f_k = n$), entonces se definen:

• *Media geométrica*:

$$\overline{x}_g = \sqrt[n]{x_1^{f_1}x_2^{f_2}\cdots x_k^{f_k}}\,.$$

• *Media cuadrática*:

$$\overline{x}_c = \sqrt{\frac{x_1^2f_1 + x_2^2f_2 + \cdots + x_k^2f_k}{n}} = \sqrt{\frac{\displaystyle\sum_{i=1}^{k} x_i^2f_i}{n}}\,.$$

• *Media armónica*:

$$\overline{x}_a = \frac{n}{\dfrac{f_1}{x_1} + \dfrac{f_2}{x_2} + \cdots + \dfrac{f_k}{x_k}} = \frac{n}{\displaystyle\sum_{i=1}^{k} \frac{f_i}{x_i}}\,.$$

En las tres definiciones anteriores, si los datos están agrupados en intervalos entonces x_i representa la marca de clase del intervalo i-ésimo.

Fijada una muestra cualquiera, siempre se verifica:

$$\overline{x}_a \leqslant \overline{x}_g \leqslant \overline{x} \leqslant \overline{x}_c\,.$$

2.5. Medidas de dispersión

Son valores que miden el grado de separación de las observaciones entre sí o con respecto a ciertas medidas de posición. Sólo se calculan cuando la variable es cuantitativa.

2.5.1. *Recorrido*

Es una medida de dispersión global que se define como la diferencia entre la observación mayor, x_{max}, y la observación menor, x_{min}, y se denota por R; es decir:

$$R = x_{max} - x_{min}\,.$$

Si el recorrido es pequeño entonces los datos están poco dispersos.

2.5.2. *Recorrido intercuartílico*

Se denota por RI y se define como la diferencia entre el tercer cuartil y el primer cuartil; es decir:

$$RI = Q_3 - Q_1 \, .$$

Si el recorrido intercuartílico es pequeño entonces los datos están cerca de la mediana; en caso contrario, los datos están alejados de ella.

2.5.3. *Desviación mediana*

Si los datos de la muestra son $x_1, x_2, \ldots, x_k$ y ellos aparecen con frecuencias absolutas respectivas f_1, f_2, ..., f_k (con $f_1 + f_2 + \cdots + f_k = n$), entonces:

$$DM_e = \frac{\sum_{i=1}^{k} |x_i - M_e| f_i}{n} \, .$$

Si los datos están agrupados en intervalos, x_i representa la marca de clase del intervalo $i-$ésimo.

Cuando DM_e es pequeña, entonces los datos están cerca de la mediana. En caso contrario, los datos están alejados de la mediana.

2.5.4. *Mediana de las desviaciones absolutas respecto de la mediana*

La *mediana de las desviaciones absolutas respecto de la mediana* (MDA) es la mediana de los valores absolutos de las diferencias entre los valores de la muestra y la mediana de todos los datos. Para determinar MDA se calcula la mediana de todos los datos (M_e); se hallan las desviaciones de los datos respecto de la mediana:

$$x_1 - M_e, x_2 - M_e, \ldots, x_n - M_e \, ,$$

se toman los valores absolutos para eliminar los signos:

$$|x_1 - M_e|, |x_2 - M_e|, \ldots, |x_n - M_e| \, ,$$

y luego se calcula la mediana de estos valores.

Cuando MDA es pequeña, entonces los datos están cerca de la mediana. En caso contrario, los datos están alejados de la mediana.

Si los datos están agrupados en intervalos, en vez de calcular MDA se calcula DM_e, pues su interpretación descriptiva es la misma.

2.5.5. *Desviación media*

Si los datos de la muestra son $x_1, x_2, \ldots, x_k$ y ellos aparecen con frecuencias absolutas respectivas f_1, f_2, ..., f_k (con $f_1 + f_2 + \cdots + f_k = n$), entonces:

$$D\overline{x} = \frac{\sum_{i=1}^{k} |x_i - \overline{x}| f_i}{n} .$$

Si los datos están agrupados en intervalos, x_i representa la marca de clase del intervalo $i-$ésimo.

Cuando $D\overline{x}$ es pequeña, entonces los datos están cerca de la media. En caso contrario, los datos están alejados de la media.

2.5.6. *Varianza y desviación típica*

Si los datos de la muestra son $x_1, x_2, \ldots, x_k$ y ellos aparecen con frecuencias absolutas respectivas f_1, f_2, ..., f_k (con $f_1 + f_2 + \cdots + f_k = n$), entonces se definen:

• *Varianza* (algunos autores la llaman *varianza sesgada*):

$$s_x^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{n} = \frac{\sum_{i=1}^{k} x_i^2 f_i}{n} - \overline{x}^2 .$$

• *Desviación típica*: raíz cuadrada de la varianza:

$$s_x = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{n}} = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 f_i}{n} - \overline{x}^2} .$$

• *Cuasivarianza* (algunos autores la llaman *varianza insesgada*, *varianza corregida* o sólo *varianza*):

$$S_x^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{n-1} = \frac{\left(\sum_{i=1}^{k} x_i^2 f_i\right) - n\overline{x}^2}{n-1} .$$

• *Cuasidesviación típica*: raíz cuadrada de la cuasivarianza:

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{n-1}} = \sqrt{\frac{\left(\sum_{i=1}^{k} x_i^2 f_i\right) - n\overline{x}^2}{n-1}} .$$

En consecuencia, la varianza y la cuasivarianza están relacionadas de la siguiente forma:

$$(n-1) S_x^2 = n s_x^2 ,$$

por lo cual se puede calcular una de ellas a partir de la otra.

Si los datos están agrupados en intervalos de clase, las fórmulas anteriores son las mismas, salvo que x_i representa la marca de clase del intervalo $i-$ésimo.

Cuando la desviación típica (o la cuasidesviación típica) es pequeña, entonces los datos están cerca de la media. En caso contrario, los datos están alejados de la media.

Si disponemos de los datos de toda la población, la varianza se denota por σ^2 y la desviación típica por σ (letra griega que se lee *sigma*).

Propiedad de la varianza

Si $y_i = a + bx_i$, siendo a y b constantes, entonces la varianza de la nueva variable es $s_y^2 = b^2 s_x^2$ y por tanto la desviación típica es $s_y = |b| s_x$.

Cálculo de la varianza por *codificación* (para datos agrupados en intervalos de la misma amplitud)

Si a es la marca de clase del intervalo central (o una de las centrales, cuando el número de ellos es par) y b es igual a la amplitud de los intervalos, entonces realizamos la transformación siguiente:

$$z_i = \frac{x_i - a}{b}$$

y después calculamos la varianza de la nueva variable, s_z^2. Como $x_i = a + bz_i$ entonces la varianza deseada es $s_x^2 = b^2 s_z^2$ y por tanto la desviación típica es $s_x = |b| s_z$.

2.5.7. *Coeficiente de variación*

$$CV = \frac{s_x}{|\overline{x}|} \qquad \text{ó} \qquad CV = \frac{s_x}{|\overline{x}|}\, 100\% \,.$$

Algunos autores sustituyen s_x por S_x en la fórmula anterior. Este coeficiente nos sirve para comparar la dispersión relativa de dos muestras distintas. La muestra que tenga un coeficiente de variación más grande es la más heterogénea (sus datos están más dispersos).

2.6. Momentos

Si los datos de la muestra son $x_1, x_2, \ldots, x_k$ y ellos aparecen con frecuencias absolutas respectivas f_1, f_2, …, f_k (con $f_1 + f_2 + \cdots + f_k = n$), entonces se definen:

• *Momento de orden k respecto del origen*:

$$a_k = \frac{\sum_{i=1}^{k} x_i^k f_i}{n} \qquad \text{para } k = 1, 2, 3, \ldots \,.$$

• *Momento de orden k respecto de la media*:

$$m_k = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^k f_i}{n} \qquad \text{para } k = 1, 2, 3, \ldots \,.$$

Algunos casos particulares son:

$$a_1 = \overline{x} = \text{media.}$$
$$m_1 = 0.$$
$$m_2 = s_x^2 = \text{varianza.}$$

Desarrollando el binomio $(x_i - \overline{x})^k$ se puede comprobar que existe una relación entre los momentos respecto al origen y los momentos respecto de la media; por ejemplo:

$$\begin{aligned} m_2 &= a_2 - a_1^2 \\ m_3 &= a_3 - 3a_2a_1 + 2a_1^3 \\ m_4 &= a_4 - 4a_3a_1 + 6a_2a_1^2 - 3a_1^4 \end{aligned}$$

2.7. Medidas de forma

A través de las representaciones gráficas (histogramas, diagramas de barras, etc.) podemos hacernos una idea sobre la forma de las distribuciones, pero también resulta importante cuantificar esta característica a través de las medidas de forma.

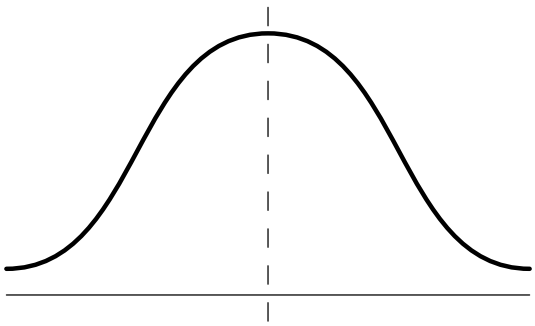


Figura 8.7: Curva normal

Las características bajo-ancho, alto-estrecho se miden respecto de una curva modelo llamada *curva Normal* (véase la Figura 8.7). Esta curva es la representación gráfica de la siguiente función:

$$y = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

donde x, μ y σ son números reales, siendo además $\sigma > 0$; e es la base de los logaritmos neperianos ($e = 2'718281829\ldots$) y π es la relación de la longitud de una circunferencia a su diámetro ($\pi = 3'141592654\ldots$).

Para cada par de valores de μ y σ tendremos una curva Normal distinta. Es decir, tenemos una familia de curvas. Pero todas ellas coinciden en algunas propiedades, como, por ejemplo:

a) Tiene un único máximo para $x = \mu$.

b) Es simétrica respecto al eje vertical que pasa por $x = \mu$.

c) Se acerca asintóticamente al eje horizontal. En otras palabras, se acerca más y más a ese eje, tanto por la derecha como por la izquierda, sin llegar a tocarlo en ningún punto.

2.7.1. *Asimetría*

Se dice que una distribución presenta una *asimetría positiva* o *por la derecha* cuando su polígono de frecuencias (absolutas o relativas) es similar a la Figura 8.8 (a). Análogamente, se dice que una distribución presenta una *asimetría negativa* o *por la izquierda* cuando su polígono de frecuencias (absolutas o relativas) tiene una forma parecida a la Figura 8.8 (b). Diremos que una distribución presenta *simetría* cuando su polígono de frecuencias (absolutas o relativas) es similar a la Figura 8.7.

(a) Distribución asimétrica por la derecha

(b) Distribución asimétrica por la izquierda

Figura 8.8: Asimetría

Para las distribuciones unimodales, como medida de asimetría se suele utilizar el *coeficiente de asimetría de Pearson*, que se define por la expresión:

$$CA = \frac{\overline{x} - M_o}{s_x},$$

que permite distinguir los casos:

a) $CA = 0$ (distribución simétrica),

b) $CA > 0$ (distribución asimétrica por la derecha),

c) $CA < 0$ (distribución asimétrica por la izquierda).

Cuando la distribución no es unimodal no se puede emplear el anterior coeficiente, por lo que se introduce el *coeficiente de asimetría de Fisher*, que viene dado por:

$$g_1 = \frac{m_3}{s_x^3},$$

que permite distinguir los casos:

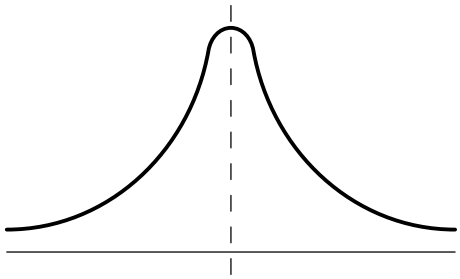
a) $g_1 = 0$ (distribución simétrica),

b) $g_1 > 0$ (distribución asimétrica por la derecha),

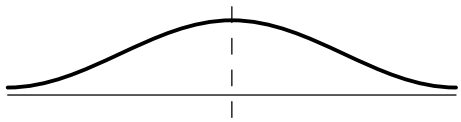
c) $g_1 < 0$ (distribución asimétrica por la izquierda).

2.7.2. *Apuntamiento*

Si el polígono de frecuencias (absolutas o relativas) es análogo a la curva Normal, entonces se dice que la distribución es *mesocúrtica* (véase la Figura 8.7); si es más elevado y estrecho que la curva Normal, entonces se llama distribución *leptocúrtica* (véase la Figura 8.9 (a)); y si es menos elevado y más ancho que la curva Normal, entonces se llama distribución *platicúrtica* (véase la Figura 8.9 (b)).



(a) Distribución muy apuntada o leptocúrtica



(b) Distribución poco apuntada o platicúrtica

Figura 8.9: Apuntamiento

Como medida del apuntamiento se utiliza el *coeficiente de curtosis*, dado por:

$$g_2 = \frac{m_4}{s_x^4} - 3\,,$$

permitiendo distinguir los casos:

a) $g_2 = 0$ (distribución mesocúrtica),

b) $g_2 > 0$ (distribución leptocúrtica),

c) $g_2 < 0$ (distribución platicúrtica).

3. ACTIVIDADES DE APLICACIÓN DE LOS CONOCIMIENTOS

A.8.1. La tabla siguiente muestra la puntuación obtenida por los alumnos de primero de una facultad en un examen "tipo test" realizado en una determinada asignatura:

$6'3$	$7'2$	$6'8$	$7'4$	$8'2$	$9'3$	$4'6$	$5'3$	$5'9$	$6'7$
$7'6$	$6'8$	$5'8$	$4'9$	$4'7$	$4'6$	$5'7$	$6'8$	$7'8$	$9'1$
$3'9$	$4'8$	$5'6$	$5'8$	$7'1$	$8'4$	$9'1$	$4'8$	$5'5$	$6'3$
$4'8$	$5'7$	$6'8$	$7'3$	$7'5$	$7'6$	$8'1$	$8'8$	$7'5$	$9'5$
$5'7$	$6'6$	$7'3$	$5'6$	$4'7$	$3'9$	$3'8$	$4'7$	$5'1$	$5'8$
$7'3$	$6'5$	$6'8$	$6'9$	$5'8$	$5'3$	$6'7$	$7'8$	$6'9$	$4'7$

a) Agrupar los datos en los siguientes intervalos: [0,5) (suspenso), [5,7) (aprobado), [7,9) (notable) y [9,10) (sobresaliente). A partir de esta agrupación hacer lo siguiente:

b) Dibujar el histograma de frecuencias absolutas.

c) Hallar la moda, la media y la mediana.

A.8.2. La tabla siguiente muestra los salarios de 34 trabajadores elegidos al azar en una determinada empresa.

salario mensual (en pts)	número de trabajadores
20000	1
30000	3
40000	3
50000	15
60000	6
70000	2
80000	4

Hacer el diagrama de barras de frecuencias absolutas y el gráfico de frecuencias acumuladas absolutas. Calcular la mediana y la desviación mediana. ¿Están los datos cerca de la mediana?

A.8.3. Los varones entre 20 y 60 años que contrajeron matrimonio durante el año 1961 en España presentan la siguiente distribución por edades.

edad	número de varones
[20,25]	41000
(25,30]	123000
(30,35]	44000
(35,40]	13000
(40,50]	7000
(50,60]	3000

Dibujar el polígono de frecuencias acumuladas absolutas. Calcular la moda, la mediana, los cuartiles, el segundo decil y el percentil al 37 por ciento. Calcular la desviación media y deducir si los datos están cerca de la media.

A.8.4. Se ha revisado un lote de 1000 piezas esmaltadas, obteniéndose el número de defectos de cada pieza, lo que se indica en la siguiente tabla:

$n^{\underline{o}}$ de defectos	frecuencia
0	600
1	310
2	75
3	13
4	2

Hacer el polígono de frecuencias absolutas. Determinar la media y la desviación típica. Deducir si los datos están cerca de la media.

A.8.5. Un fabricante de calzado quiere conocer la distribución de las tallas de los zapatos demandados por los hombres. Elige una muestra aleatoria de 50 hombres obteniendo los siguientes resultados.

talla	número de hombres
37	2
38	3
39	5
40	8
41	11
42	7
43	6
44	5
45	3

Se desea obtener:

a) El diagrama de barras y el gráfico de frecuencias acumuladas.

b) La media, moda y mediana.

c) La mediana de las desviaciones absolutas respecto de la mediana. ¿Están los datos cerca de la mediana?

d) La desviación típica. ¿Están los datos cerca de la media?

e) El coeficiente de asimetría de Pearson.

A.8.6. Un alumno "muy previsor"quiere escoger una asignatura de la que dan clases dos profesores diferentes A y B. Según datos del pasado año, la nota media de los alumnos que tuvieron el profesor A fue de $6'1$ y la desviación típica fue $0'95$. En el curso del profesor B, la media fue $5'2$ y la desviación típica $2'2$. Suponiendo que puede elegir al profesor ¿qué profesor debe elegir y por qué, según desee:

a) aprobar "sin demasiadas complicaciones'"?

b) conseguir la nota más alta?

A.8.7. A una competición de tiro concurren seis tiradores, A, B, C, D, E y F, en tres modalidades de tiro, que llamaremos 1, 2 y 3. El número de blancos obtenidos por cada uno, en cada modalidad de tiro, puede verse en la siguiente tabla.

tirador	1	2	3
A	44	29	12
B	44	27	13
C	46	28	13
D	43	27	17
E	47	25	14
F	46	26	15

Aparte de haber un primer premio en cada modalidad, el jurado debe conceder un premio especial al tirador que considere "globalmente mejor", lo cual es motivo de arduas discusiones (¿por qué?). Un miembro del jurado dice que "basándose en Estadística" este premio debe ser para el tirador D. ¿Por qué?

A.8.8. Una empresa debe cubrir un cierto número de puestos de trabajo de dos tipos, A y B. Se somete a los aspirantes a dos pruebas, ambas puntuadas de 0 a 50, diseñadas para valorar sus aptitudes en uno y en otro tipo de trabajo. En la prueba A la media de calificaciones ha sido 28 y la desviación típica $3'4$. En la prueba B la media ha sido 24 y la desviación típica $2'1$. ¿Qué tipo de puesto de trabajo asignaremos a un aspirante que hubiese obtenido 33 puntos en la prueba A y 28 en la B?

A.8.9. **a)** Sea $\{x_1, x_2, \ldots, x_{100}\}$ una muestra de media aritmética 21'5 y sean $x_{101} = 22$, $x_{102} = 19$, $x_{103} = 20'5$ tres observaciones más. Calcular la media aritmética de la nueva muestra $\{x_1, x_2, \ldots, x_{100}, x_{101}, x_{102}, x_{103}\}$.

b) Sea $\bar{x}$ la media aritmética de la muestra $\{x_1, x_2, \ldots, x_n\}$ y sea $\bar{y}$ la media aritmética de la muestra $\{y_1, y_2, \ldots, y_m\}$. ¿Cuál será la media aritmética de la unión de ambas muestras?

A.8.10. Se preguntó a varias personas, elegidas al azar, el número de periódicos distintos que leían trimestralmente, y se obtuvieron las siguientes respuestas:

Número de periódicos	0	1	2	3	4	5	6	7
Número de personas	7	13	18	15	11	6	4	2

Determinar la distribución de frecuencias relativas, acumuladas absolutas y acumuladas relativas. Hacer el diagrama de barras de frecuencias absolutas, el polígono de frecuencias relativas y el gráfico de frecuencias acumuladas absolutas. Hallar las medidas de posición, dispersión y forma e interpretar los resultados.

A.8.11. Se considera la variable cuantitativa peso, en kilogramos, de una muestra de 25 alumnos de una determinada facultad:

$53'5$	$55'5$	$53'0$	$76'5$	$52'5$
$51'5$	$56'0$	$73'5$	$51'0$	$49'5$
$57'0$	$53'0$	$72'0$	$52'5$	$51'5$
$68'5$	$68'0$	$65'0$	$53'0$	$68'5$
$53'5$	$55'5$	$53'0$	$51'5$	$53'0$

Agrupar los datos en intervalos de la misma amplitud. Con los datos agrupados en intervalos: determinar la distribución de frecuencias absolutas, relativas, acumuladas absolutas y acumuladas relativas. Hacer el histograma de frecuencias absolutas, el polígono de frecuencias relativas y el polígono de frecuencias acumuladas absolutas. Hallar las medidas de posición, dispersión y forma e interpretar los resultados.

A.8.12. A continuación se dan los datos correspondientes al tiempo de espera (en minutos) hasta que son atendidos 40 personas que visitan una determinada caja de ahorros:

11	13	15	20	21	24	17	9	10	12
20	18	16	14	13	11	10	13	15	19
24	20	21	22	18	16	11	14	8	6
15	18	19	20	17	15	16	13	12	14

Agrupar los datos en intervalos. Con los datos agrupados en intervalos: determinar la distribución de frecuencias absolutas, relativas, acumuladas absolutas y acumuladas relativas. Hacer el histograma de frecuencias absolutas y el polígono de frecuencias acumuladas relativas. Hallar las medidas de posición, dispersión y forma e interpretar los resultados.

A.8.13. Los datos siguientes corresponden al sueldo (en miles de pesetas) de una muestra de 25 empleados de una determinada empresa.

sueldo (en miles de pesetas)	número de empleados
[90,100]	12
(100,110]	7
(110,120]	4
(120,130]	2

Determinar la distribución de frecuencias relativas, acumuladas absolutas y acumuladas relativas. Hacer el histograma de frecuencias absolutas y el polígono de frecuencias relativas. Hallar las medidas de posición, dispersión y forma e interpretar los resultados.

4. ACTIVIDADES PRÁCTICAS DEL CAPÍTULO

4.1. Introducción

La práctica se va a realizar con el programa *STATISTIX for Windows*, versión 1.0 (en inglés). Para ejecutar el programa debemos acceder a la carpeta *STATISTIX* y ejecutar el programa *STATISTIX*. Una vez en el programa, aparecerá la pantalla de la Figura 8.10, con un menú en la parte superior que, de ahora en adelante, denominaremos *menú principal* (`File Edit Data Statistics Preferences Window Help`).

Todas las opciones del menú principal son ellas mismas menús, los cuales ofrecen variadas posibilidades de manipulación de datos y de procedimientos estadísticos.

El menú `File` incluye procedimientos para abrir y grabar ficheros de datos, así como la posibilidad de importar datos de otros programas o exportar datos en un formato predeterminado. El menú `Data` proporciona muchas funciones para manipular la hoja de datos, incluyendo las transformaciones de variables. El menú `Statistics` lista varios menús que nos permiten realizar análisis estadísticos básicos y avanzados. Por último, el menú `Windows` lista las ventanas abiertas dentro de *STATISTIX*, que incluye la hoja de datos y las ventanas de resultados.

Para salir del programa debe seleccionarse la opción `Exit` del menú `File`. Los usuarios de Windows 95 o Windows NT pueden abandonar el programa pulsando el botón de la esquina superior derecha de la ventana (botón ⊠).

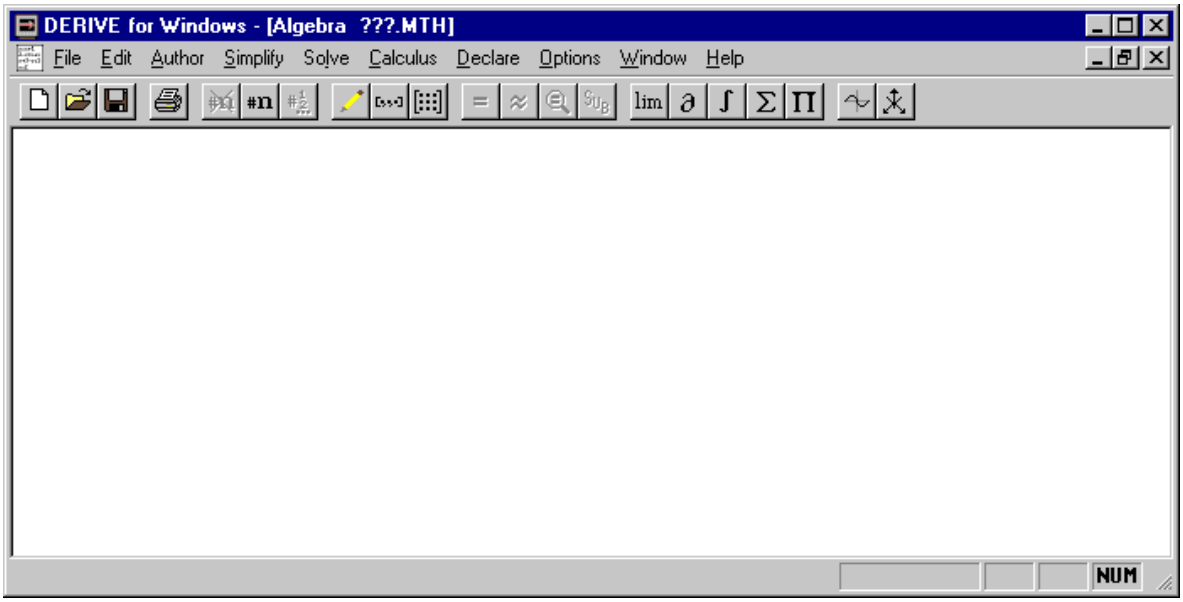


Figura 8.10: Pantalla inicial del programa donde aparece, además del menú principal del programa, una tabla en blanco donde introduciremos los datos.

Antes de realizar ninguna operación es necesario tener un conjunto de datos en uso, para lo cual podemos proceder de dos formas distintas: o introducirlos directamente a través del teclado o recuperar un fichero con datos previamente almacenado.

4.2. Entrada de datos a través del teclado

Los datos en *STATISTIX* pueden ser considerados tablas de datos, donde las columnas se denominan variables y las filas casos. En primer lugar debemos declarar las variables, para lo que seleccionamos las opciones `Data|Insert|Variables`. Antes de seguir adelante convendría hacer una aclaración: cuando digamos que se seleccionan varias opciones, como acabamos de hacer, se sobreentenderá que cada opción se encuentra en el menú correspondiente a la opción anterior, salvo la primera opción que se encuentra en el menú principal. Una vez elegidas las opciones anteriores aparece la ventana de introducción de variables (ver Figura 8.11).

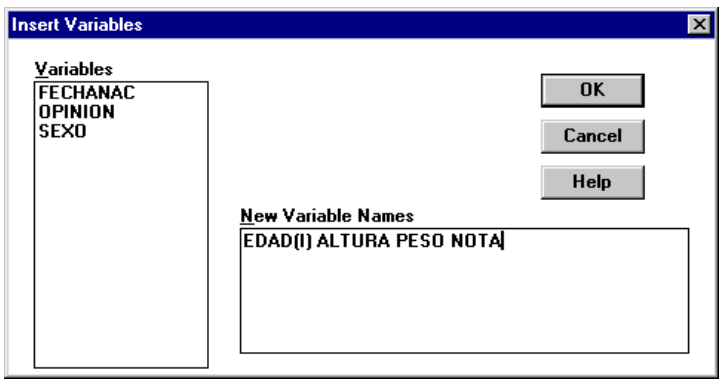


Figura 8.11: Pantalla de *STATISTIX* para la introducción de nuevas variables.

En el recuadro `New Variable Names` debemos escribir los nombres de las variables y su tipo, ya que *STATISTIX* distingue cuatro clases de datos: números reales, números enteros, fechas y cadenas de caracteres. Para indicar

los tres últimos tipos, el nombre de la variable se acompaña (entre paréntesis) de las letras `I`, `D` o `Sxx`, donde `xx` denota un número que indica la longitud máxima de las cadenas.

El nombre de una variable es una cadena de caracteres con una longitud variable entre uno y nueve caracteres, debe comenzar con una letra y sólo puede estar formado por letras, números y el carácter de subrayado. Solamente hay unas pocas palabras (como `CASE`, `PI`, `RANDOM`, etc.) que no se pueden utilizar como nombres de variables ya que están reservadas para otras tareas.

Supongamos que queremos introducir los datos de la siguiente tabla, que corresponde a una muestra de 25 alumnos procedentes de la población formada por todos los alumnos matriculados en la asignatura "Estadística" de una determinada licenciatura de la Universidad de Murcia.

Fecha de Nacimiento	Sexo	Opinión	Edad	Altura	Peso	Nota
15/02/79	M	2	18	1.62	53.5	3.9
17/09/78	M	3	19	1.64	55.5	4.1
23/05/77	M	1	20	1.63	53.0	4.2
02/01/72	V	1	25	1.78	76.5	3.2
31/10/75	M	0	22	1.62	52.5	1.6
12/09/77	M	1	20	1.65	51.5	2.5
09/11/77	M	1	20	1.64	56.0	1.1
21/12/78	V	4	19	1.74	73.5	8.1
16/06/78	M	3	19	1.63	51.0	5.1
12/07/78	M	2	19	1.61	49.5	2.7
05/08/77	M	1	20	1.64	57.0	1.9
03/04/78	M	3	19	1.62	53.0	7.3
19/02/76	V	2	21	1.75	72.0	2.4
14/05/78	M	3	19	1.66	52.5	4.9
29/06/77	M	2	20	1.63	51.5	1.6
17/09/78	V	4	19	1.70	68.5	5.0
23/03/73	V	2	24	1.72	68.0	2.5
22/10/78	V	3	19	1.74	65.0	6.5
01/09/78	M	1	19	1.66	53.0	1.9
09/07/78	V	3	19	1.71	68.5	5.2
12/10/78	M	3	19	1.65	53.5	6.3
13/11/74	M	2	23	1.66	55.5	1.2
25/04/77	M	2	20	1.64	53.0	3.3
28/07/78	M	3	19	1.60	51.5	1.8
15/11/78	M	2	19	1.61	53.0	4.4

Se ha realizado un experimento consistente en averiguar, de cada uno de los alumnos de la muestra, los siguientes datos:

`Fecha de Nacimiento`: Está en el formato `dd/mm/aa` (día/mes/año).

`Sexo`: M=Mujer, V=Varón.

`Opinión`: La opinión del alumno respecto de la asignatura está indicada según la siguiente equivalencia: 0=Mala, 1=Regular, 2=Normal, 3=Buena, 4=Excelente.

`Edad`: En años.

`Altura`: En metros.

`Peso`: En kilogramos.

`Nota`: La nota final de la asignatura es un número comprendido entre 0 y 10.

En el recuadro `New Variable Names` podemos declarar más de una variable, pero hemos de tener presente que los nombres de las variables deben ir separados entre sí por un espacio en blanco. Entonces una forma posible de declarar las variables anteriores sería la siguiente:

```
FechaNac(d) Sexo(s1) Opinion(i) Edad(i) Altura Peso Nota
```

Aparece entonces en la tabla de la pantalla los nombres de nuestras variables situados en la primera fila. Debajo de cada una de las variables, excepto de las cadenas de caracteres, aparece una letra `M` mayúscula, indicando que debemos completar dicha casilla. Para desplazarnos por la tabla de datos podemos utilizar la tecla de entrada ($\hookleftarrow$) o las flechas: $\uparrow$ y $\downarrow$ para desplazarnos por las columnas, y $\leftarrow$ y $\rightarrow$ para hacerlo en la misma fila. Una vez rellena, tendremos una tabla de datos como en la Figura 8.12.

4.3. Grabación y lectura de ficheros de datos

Una vez introducidos los datos, éstos pueden guardarse en un fichero para poder ser utilizados en cualquier otro momento. En realidad, los datos deberían guardarse muy a menudo, no sólo cuando hayamos terminado de introducirlos todos (¿Qué pasaría si tenemos que introducir 1000 datos y cuando ya hemos introducido 950 se produce un corte de energía eléctrica?)

Para grabar los datos se seleccionan las opciones `File|Save` o `File|Save As`. Por defecto, el programa nos ofrece almacenar el fichero en el directorio del programa (usualmente `c:\sxw`), pero podemos cambiar de directorio y guardarlo donde deseemos. Para distinguir los archivos creados con *STATISTIX* debemos recordar que tiene la extensión `sx`. Guardemos los datos anteriores en un archivo con nombre `Datos1.sx` para poder utilizarlos en las prácticas siguientes.

Para leer unos datos previamente almacenados debemos seleccionar las opciones `File|Open`. Aparece una ventana en la que hay que seleccionar la unidad, el directorio y el nombre del archivo que queremos leer. Una vez realizado lo anterior volvemos a tener disponibles los datos (ver Figura 8.12).

Statistix - [c:\datos1.sx]

File Edit Data Statistics Preferences Window Help

↓	FECHANAC	EDAD	ALTURA	PESO	NOTA	SEXO	OPINION
1	15/02/79	18	1.62	53.5	3.9	M	2
2	17/09/78	19	1.64	55.5	4.1	M	3
3	23/05/77	20	1.63	53.0	4.2	M	1
4	02/01/72	25	1.78	76.5	3.2	V	1
5	31/10/75	22	1.62	52.5	1.6	M	0
6	12/09/77	20	1.65	51.5	2.5	M	1
7	09/11/77	20	1.64	56.0	1.1	M	1
8	21/12/78	19	1.74	73.5	8.1	V	4
9	16/06/78	19	1.63	51.0	5.1	M	3
10	12/07/78	19	1.61	49.5	2.7	M	2
11	05/08/77	20	1.64	57.0	1.9	M	1
12	03/04/78	19	1.62	53.0	7.3	M	3
13	19/02/76	21	1.75	72.0	2.4	V	2

Modified | 7 variables. | 25 cases selected. 25 cases total.

Figura 8.12: Pantalla del programa donde aparece la tabla de datos `Datos1`.

4.4. Edición y modificación de los datos

Una vez que tenemos disponibles los datos del archivo `Datos1.sx`, es posible que detectemos un error, o bien que deseemos añadir nuevos datos y reorganizarlos. Todas estas operaciones se realizan seleccionando la opción `Data` del menú principal.

Para modificar un dato nos situaremos sobre él con el ratón o bien utilizando las teclas ↑, ↓, ← y →. Este quedará recuadrado y podremos modificarlo. Después de realizar cualquier modificación no hay que olvidar volver a grabar nuestra tabla de datos. Cualquier modificación de un dato no surge efecto hasta que no es pulsada la tecla de entrada (↵); en cualquier momento podemos abandonar la edición de una dato pulsando la tecla de escape (`ESC`).

Si queremos añadir una nueva fila de datos (lo que denominamos un nuevo `caso`) basta con que nos situemos en la primera fila disponible y comencemos a rellenar las diferentes casillas. Por el contrario, si deseamos borrar una fila completa de la tabla, o varias filas, podemos proceder de dos formas distintas. La manera rápida de hacerlo es seleccionar con el ratón las filas que deseamos eliminar (que deben ser contiguas) y pulsar la tecla `Ctrl+X`. La otra forma de hacerlo es seleccionar las opciones `Data|Delete|Cases` y nos aparece una ventana (ver Figura 8.13) solicitándonos los números de los casos primero y último que deseamos eliminar (se borrarán todos los casos entre el primero y el último, ambos incluidos). Una vez han sido borrados los casos, se vuelven a renumerar los casos, de forma que siempre sean consecutivos.

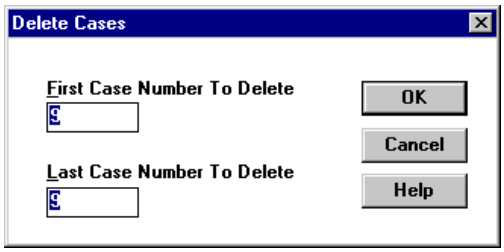


Figura 8.13: Pantalla del programa donde aparecen los casos que deseamos borrar. Debemos recordar que esto mismo se puede realizar con el uso de teclas calientes; concretamente debemos marcar las filas y pulsar `Ctrl+X`.

¿Qué ocurre si nos equivocamos al borrar los casos? Si no hemos guardado el archivo en el disco, podemos volver a leer el archivo original de datos. Sin embargo, si las modificaciones ya han sido almacenadas entonces sólo podemos insertar nuevos casos (líneas de la tabla) y volver a introducir los datos. Esto se consigue con las opciones `Data|Insert|Cases` con lo cual el programa nos solicita que introduzcamos el número de casos que queremos insertar y el caso por el que queremos comenzar (ver Figura 8.14).

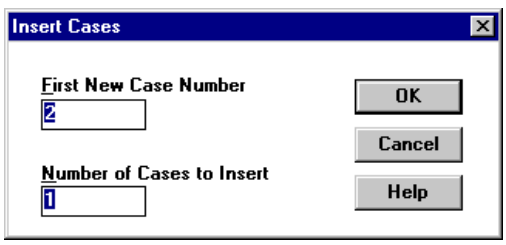


Figura 8.14: Pantalla del programa que nos permite insertar nuevos casos.

Los datos numéricos se pueden introducir de muchas maneras y con diferentes grados de exactitud. Pensemos, por ejemplo, en la altura. Como estamos considerando el metro como unidad de medida, es posible que aparezcan datos de la forma `2`, `1.8`, `1.75` o incluso `1.625`. Sin embargo, lo razonable es que todos los datos tengan un formato común, por ejemplo 2 decimales. Esto se puede conseguir con las opciones `Data|Column Formats` (ver Figura 8.15), que nos permite modificar el formato de las diferentes columnas, tanto numéricas como de

caracteres. En el caso de la variable `ALTURA` lo razonable es seleccionar el formato numérico `Decimal` con 3 dígitos significativos (`Significant Digits`).

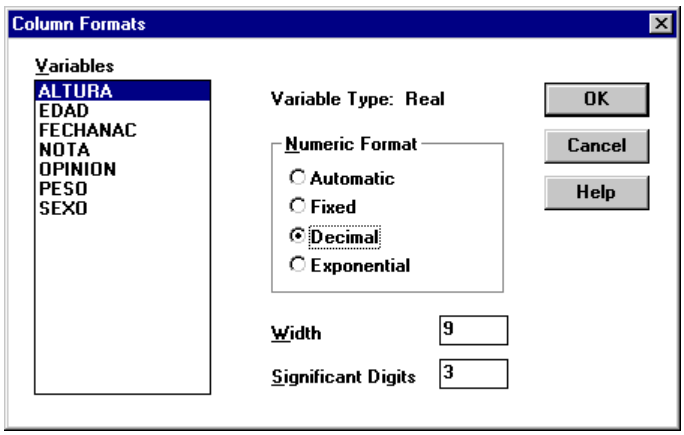


Figura 8.15: Pantalla del programa que nos permite modificar el formato de las diferentes columnas (tanto numéricas como de caracteres).

Otro problema que se nos puede plantear con los datos es la utilización de abreviaturas. Por ejemplo, en la columna `OPINION` hemos utilizado `0`, `1`, `2`, `3` y `4` para referirnos a 'Mala', 'Regular', 'Normal', 'Buena' y 'Excelente'; sin embargo, dentro de varios meses es posible que pensemos que dichos números se refieren a otros datos. Por tanto, sería conveniente poder asignarle a cada valor de la variable su significado; en nuestro caso, `0=Mala`, `1=Regular`, `2=Normal` y así sucesivamente. Para ello debemos seleccionar las opciones `Data|Labels|Value Labels` y rellenar los campos que nos ofrece (ver Figura 8.16). Por ejemplo, dentro de `Define Label` ponemos 4 en `Value` y 'Excelente' en `Label`; para que la asignación sea efectiva debemos pulsar el botón ▶, con lo que se almacena en la columna `Value Labels`.

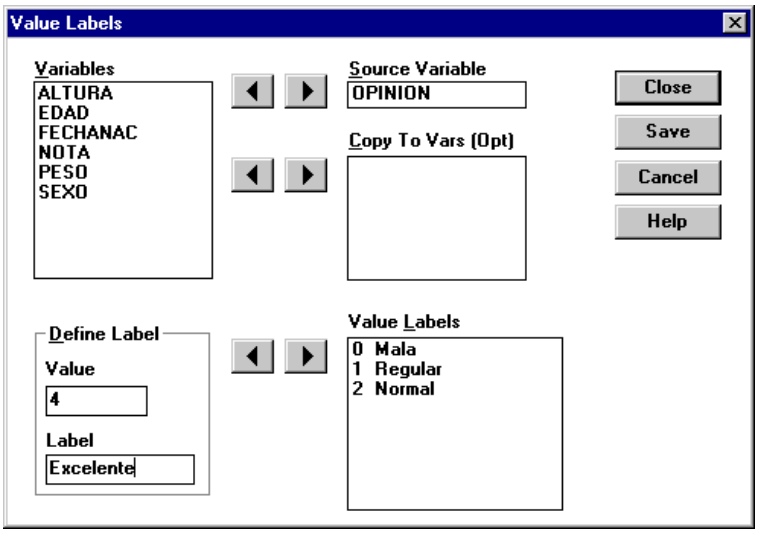


Figura 8.16: Pantalla del programa que nos permite asignar etiquetas a los diferentes valores numéricos de una variable. Esta posibilidad no está presente para variables de fechas y de caracteres.

4.5. Edición y modificación de las variables

Hay ocasiones en que los datos están correctos pero no nos termina de convencer el nombre que le hemos dado a las variables. También puede ocurrir que nos haya faltado una variable o, por el contrario, que hayamos introducido una variable de más. Para solucionar estos problemas se utilizan las opciones `Data|Rename Variables`, `Data|Insert|Variables` y `Data|Delete|Variables` (ver Figuras 8.11 y 8.17).

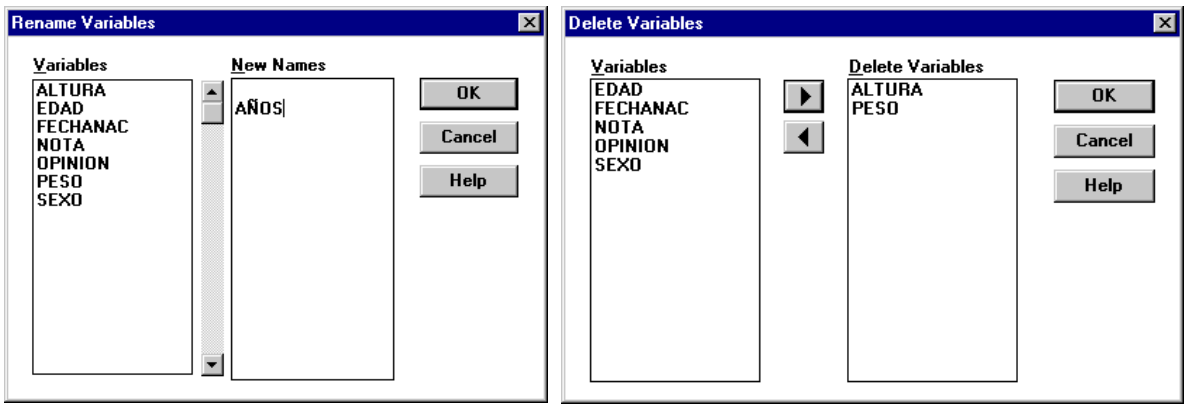


Figura 8.17: Pantallas del programa que permiten renombrar o borrar una variable.

Sin embargo, el nombre de las variables no refleja toda la información necesaria. Por ejemplo, el peso no indica (aunque se sobreentiende) que está computado en kilogramos y la altura está, como es fácil deducir, en metros. Puede haber ejemplos en los que no sea tan fácil deducirlo; en estos casos, lo conveniente es 'etiquetar' las variables. Para ello seleccionamos las opciones `Data|Labels|Variable Labels` y añadimos `en metros` y `en kilogramos` a las variables `ALTURA` y `PESO`, respectivamente (ver Figura 8.18).

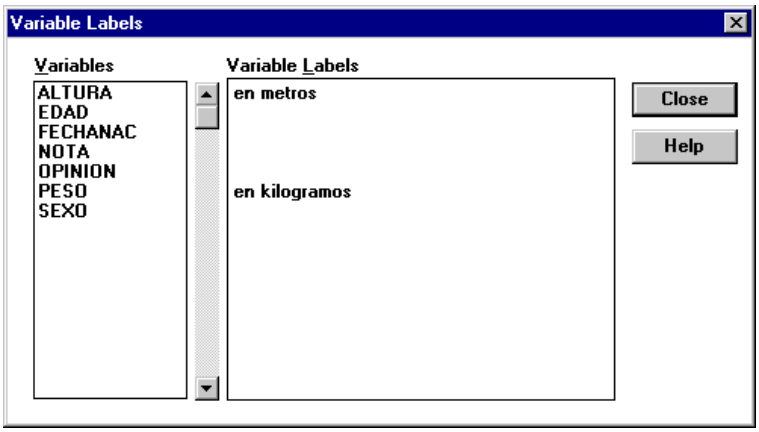


Figura 8.18: Pantalla del programa que permite etiquetar una variable.

Finalmente, otra modificación que podemos hacer con las variables es reordenarlas. Para ello seleccionamos las opciones `Data|Reorder Variables` y nos aparece una ventana similar a la ventana `Delete Variables` de la Figura 8.17. Ahora sólo es necesario ir seleccionando las variables en el nuevo orden.

Exponenciación	`A^B`
Multiplicación	`A * B`
División	`A / B`
Suma	`A + B`
Resta	`A - B`

Tabla 8.1: Operaciones aritméticas que se pueden realizar con *STATISTIX*.

4.6. Transformación aritmética de variables

En la Tabla 8.1 se encuentran recogidas las expresiones aritméticas que están permitidas. Debemos tener presente que las operaciones se evalúan de izquierda a derecha. Todas las expresiones entre paréntesis se evalúan antes que las que están fuera de los paréntesis y ante varios operadores en el mismo nivel, el orden de preferencia (de mayor a menor) es el que figura en la Tabla 8.1.

Para construir una nueva variable mediante transformaciones de otras ya existentes, debemos seleccionar las opciones `Data|Transformations` (basta pulsar `Ctrl+T`) y nos aparece la pantalla de la Figura 8.19. En esta ventana tenemos tres partes fundamentales: a la izquierda aparece la lista de variables existentes, a la derecha aparecen las funciones internas, y en el centro está el lugar destinado a la definición de la nueva variable.

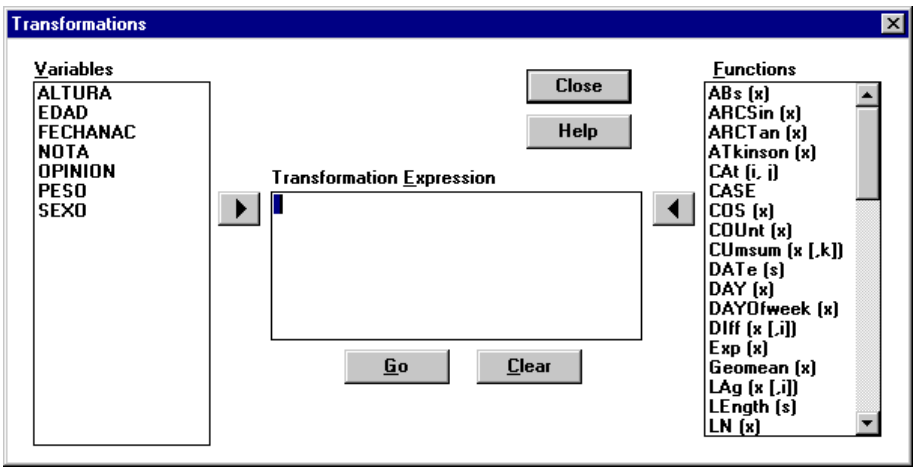


Figura 8.19: Pantalla del programa que permite definir nuevas variables (o modificar variables ya existentes) mediante operaciones.

La expresión de la transformación debe ser tecleada manualmente y sólo podemos hacer uso de las variables y de las funciones internas para insertarlas en los lugares adecuados (pulsando los correspondientes botones ▶ y ◀). Sólo está permitida una transformación, aunque ésta puede ocupar varias líneas.

Una vez hemos terminado de teclear la expresión, pulsamos en `Go`. Si *STATISTIX* encuentra un error en nuestra expresión nos lo indica convenientemente. Entonces debemos editar nuestra expresión y corregir el error, pulsando la tecla `Go` de nuevo.

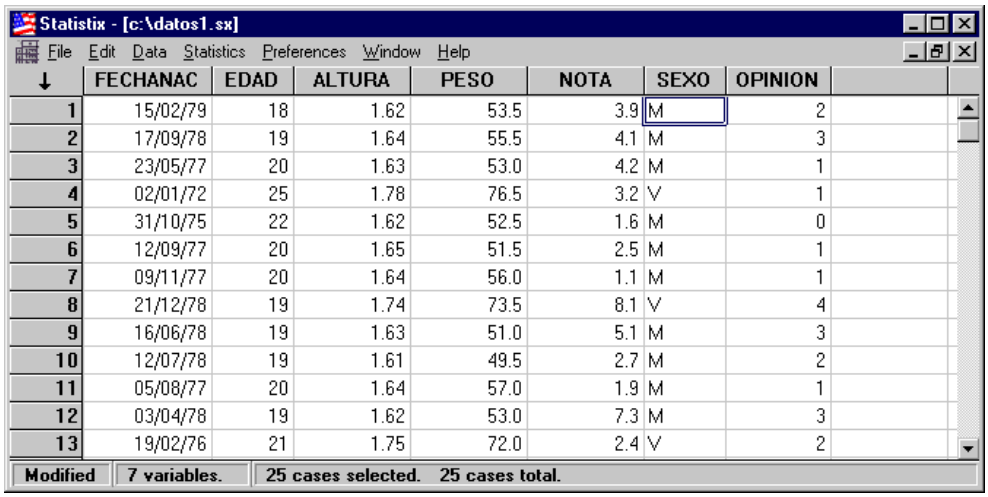
Supongamos que la variable `NUEVAVAR` es la variable que queremos crear para que almacene la transformación. Si la transformación consiste en la suma de otras tres variables (por ejemplo, `A`, `B` y `C`) entonces debemos escribir dentro del recuadro `Transformation Expression` la instrucción `NUEVAVAR = A+B+C`. Como *STATISTIX* admite cuatro tipos de variables, podemos indicar el tipo de `NUEVAVAR` de la forma usual.

Por ejemplo, con `datos1.sx` vamos a crear una nueva variable, que llamaremos `PesoTip` con los pesos tipificados. La transformación llamada tipificación consiste en restar a cada dato la media aritmética y después dividir por la desviación típica (o por la cuasidesviación típica). Para ello, en el recuadro `Transformation Expression`

escribimos: PesoTip=(Peso- MEAn(Peso))/ SD(Peso). Cuando la expresión utiliza una variable, se puede escribir su nombre o se puede trasladar dicho nombre desde el recuadro `Variables`. Análogamente, las expresiones como MEAn (media aritmética) o SD (cuasidesviación típica) se pueden teclear o se pueden seleccionar haciendo doble clic sobre la expresión del recuadro `Functions`.

4.7. Medidas descriptivas de los datos

En primer lugar leemos el archivo que contiene los datos (`datos1.sx`) utilizando las opciones `File|Open`. Entonces nos aparece la tabla de datos de la Figura 8.20.



	FECHANAC	EDAD	ALTURA	PESO	NOTA	SEXO	OPINION
1	15/02/79	18	1.62	53.5	3.9	M	2
2	17/09/78	19	1.64	55.5	4.1	M	3
3	23/05/77	20	1.63	53.0	4.2	M	1
4	02/01/72	25	1.78	76.5	3.2	V	1
5	31/10/75	22	1.62	52.5	1.6	M	0
6	12/09/77	20	1.65	51.5	2.5	M	1
7	09/11/77	20	1.64	56.0	1.1	M	1
8	21/12/78	19	1.74	73.5	8.1	V	4
9	16/06/78	19	1.63	51.0	5.1	M	3
10	12/07/78	19	1.61	49.5	2.7	M	2
11	05/08/77	20	1.64	57.0	1.9	M	1
12	03/04/78	19	1.62	53.0	7.3	M	3
13	19/02/76	21	1.75	72.0	2.4	V	2

Figura 8.20: Pantalla del programa donde aparece la tabla de datos `Datos1`.

Para poder obtener las medidas descriptivas de una muestra de datos cuantitativos debemos seleccionar las opciones `Statistics|Summary Statistics|Descriptive Statistics`. Entonces nos aparece la ventana mostrada en la Figura 8.21.

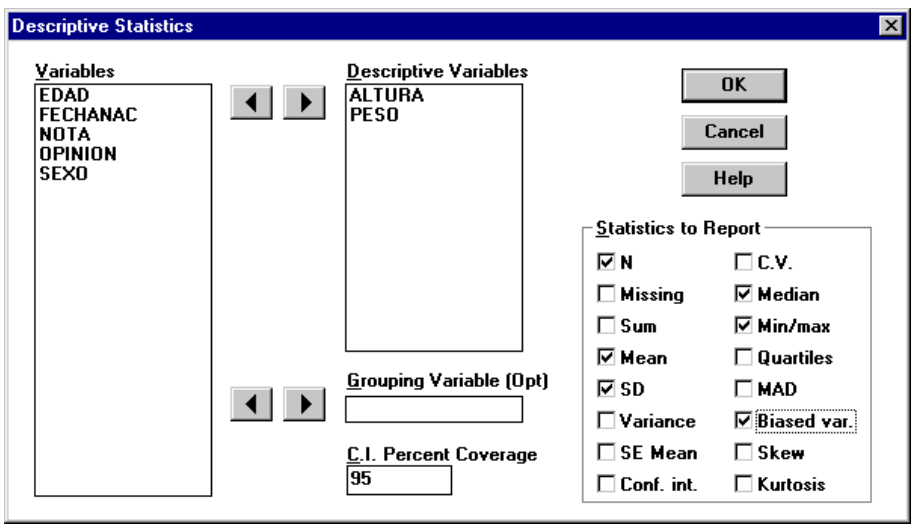


Figura 8.21: Pantalla del programa que permite elegir las variables a las que vamos a calcular sus principales medidas descriptivas.

En el recuadro `Descriptive Variables` seleccionamos las variables de las cuales queremos hallar las medidas descriptivas y en el recuadro `Statistics to Report` seleccionamos las medidas descriptivas que queremos calcular. El significado de las medidas descriptivas que aparecen viene dado en la Tabla 8.2.

`N`	Número total de observaciones (n)
`Missing`	Número de casos en los que se omite el valor de la variable
`Sum`	Suma de todos los datos ($\sum x_i$)
`Mean`	Media aritmética ($\overline{x}$)
`SD`	Cuasidesviación típica (S)
`Variance`	Cuasivarianza (S^2)
`SE Mean`	$S/\sqrt{n}$
`Conf. int.`	Intervalo de confianza para la media (no estudiado)
`C.V.`	Coeficiente de variación modificado ($S/\lvert\overline{x}\rvert$)
`Median`	Mediana (M_e)
`Min/max`	Mínimo/máximo valor de la variable
`Quartiles`	Cuartiles (Q_1 y Q_3)
`MAD`	Mediana de las desviaciones absolutas respecto de la mediana
`Biased var.`	Varianza (s^2)
`Skew`	Coeficiente de asimetría
`Kurtosis`	Coeficiente de curtosis o apuntamiento

Tabla 8.2: Significado de las medidas descriptivas que aparecen en el recuadro `Statistics to Report` de la ventana `Descriptive Statistics`.

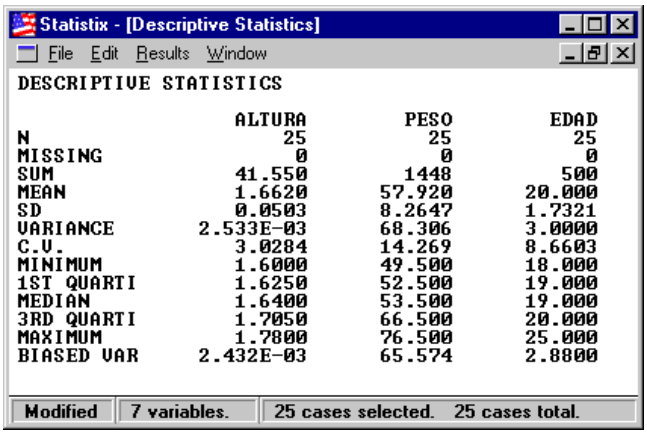


Figura 8.22: Pantalla del programa que muestra las medidas descriptivas seleccionadas de las variables elegidas.

Ejercicio. Calcula el número total de observaciones, la suma de todos los datos, la media aritmética, la cuasidesviación típica, la cuasivarianza , el coeficiente de variación modificado, la mediana, los valores máximo y mínimo, los cuartiles y la varianza de la variable `ALTURA`. En la Figura 8.23 aparece la solución.

Una opción interesante consiste en calcular las medidas descriptivas de un subconjunto de datos de la muestra original. Esto se consigue seleccionando la variable por la que queremos agrupar los datos, para lo cual debemos escribir dicha variable en el recuadro `Groupin Variable (Opt)`. En el ejercicio anterior podemos agrupar según la variable `OPINION` y el resultado obtenido (seleccionando las medidas `Mean`, `Median` y `Biased var.`) aparece en la Figura 8.24.

El programa nos permite calcular cualquier percentil, medida descriptiva que generaliza a la mediana. Para ello debemos seleccionar las opciones `Statistics|Summary Statistics|Percentiles`. Posteriormente debe-

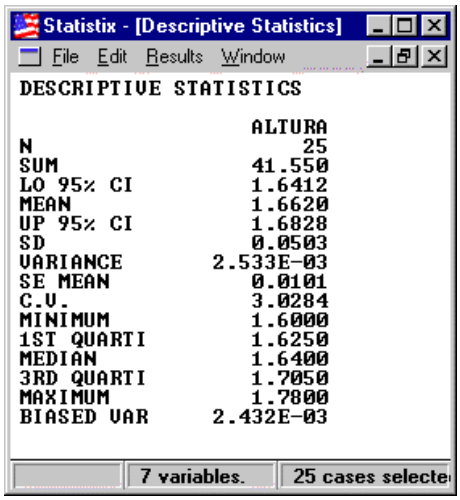


Figura 8.23: Pantalla del programa que muestra las medidas descriptivas seleccionadas de la variable ALTURA.

mos indicar, en el recuadro `Percentiles Variables`, las variables a las cuales vamos a calcular los percentiles y, en el recuadro `Percentiles`, aquellos que queremos determinar (ver Figura 8.25).

Ejercicio. Calcula el primer cuartil, el tercer cuartil y el segundo decil de la variable PESO.

4.8. Distribuciones de frecuencias

Con *STATISTIX* podemos calcular las distribuciones de frecuencias para una variable estadística discreta o continua. El programa determina la frecuencia absoluta, el porcentaje (y por tanto la frecuencia relativa), la frecuencia acumulada absoluta y el porcentaje acumulado (y por tanto la frecuencia relativa acumulada). Para conseguir la distribución de frecuencias debemos seleccionar las opciones `Statistics|Summary Statistics|Frequency Distribution` y nos aparece una ventana para seleccionar las variables estadísticas de las cuales queremos calcular sus distribuciones de frecuencias (ver Figura 8.26).

Si queremos agrupar los datos en intervalos de la misma amplitud, en el recuadro `Bin Size (Optional)` se especifica el valor mínimo de las observaciones (`Low`), el valor máximo (`High`) y la amplitud de los intervalos (`Step`). Hay que indicar que los intervalos son cerrados por la izquierda y abiertos por la derecha, salvo el último que es cerrado por ambos lados. Si el recuadro `Bin Size (Optional)` se deja en blanco, entonces los datos no se agruparán en intervalos. En la Figura 8.27 se muestra la distribución de frecuencias de la variable ALTURA (con datos sin agrupar en intervalos).

Ejercicio. Determina la distribución de frecuencias de la variable PESO cuando los datos se agrupan en intervalos de longitud 5 kilogramos comenzando en el valor 45 kilogramos.

4.9. Representaciones gráficas

Las posibilidades gráficas del programa *STATISTIX* son muy reducidas: sólo permite dibujar histogramas de frecuencias absolutas y polígono de porcentajes acumulados, ya que trata las variables cuantitativas como si fuesen

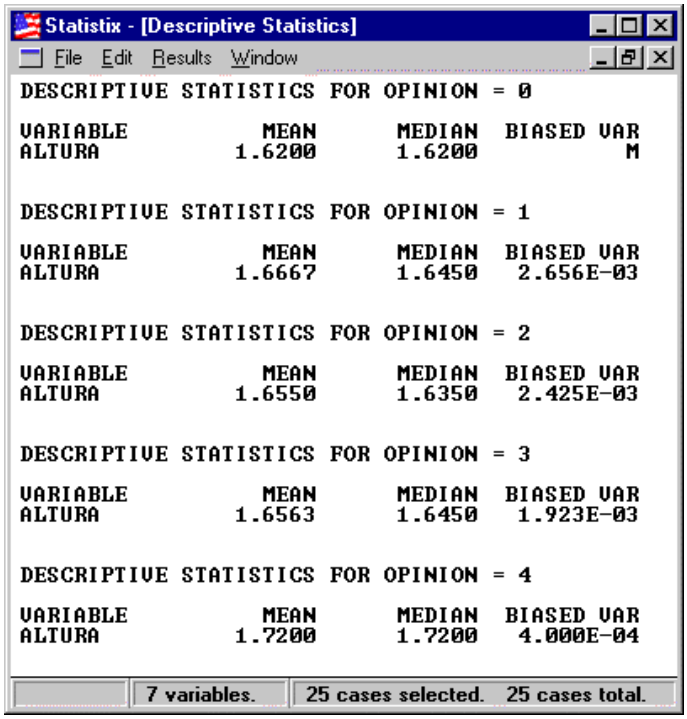


Figura 8.24: Pantalla del programa que muestra las medidas descriptivas seleccionadas de la variable `ALTURA` pero agrupadas según la variable `OPINION`.

continuas (aunque sean discretas). Para conseguir las representaciones gráficas debemos seleccionar las opciones `Statistics|Summary Statistics|Histogram` y nos aparece una ventana para seleccionar la variable estadística de la cual queremos representar gráficamente su histograma (ver Figura 8.28).

En el recuadro `X-Axis (Optional)` el usuario puede especificar el valor mínimo (`Low`), el valor máximo (`High`) y la amplitud de los intervalos (`Step`) que se van a utilizar en el histograma. En el recuadro `Graph Type` podemos seleccionar `Histogram` o `Cumulative Distribution`, según queramos dibujar el histograma de frecuencias absolutas o el polígono de porcentajes acumulados, respectivamente. Finalmente, si queremos que se superponga la gráfica de la curva normal, debemos seleccionar la opción `Display Normal Curve`. En la Figura 8.29 se muestra el histograma de frecuencias absolutas para la variable `EDAD`.

Ejercicio. Dibuja el histograma de frecuencias absolutas y el polígono de porcentajes acumulados de la variable `ALTURA`.

5. BIBLIOGRAFÍA DEL CAPÍTULO

CANDEL, J.; MARIN, A. y RUIZ, J.M. *Estadística aplicada I: Estadística descriptiva*. Barcelona: DM–PPU, 1991. Lecciones 1 y 2.

6. PREGUNTAS DE EVALUACIÓN

E.8.1. Para obtener información acerca del porcentaje de albúmina en el suero proteico de personas normales, se analizaron muestras de 40 personas, entre 2 y 40 años de edad, con los siguientes resultados:

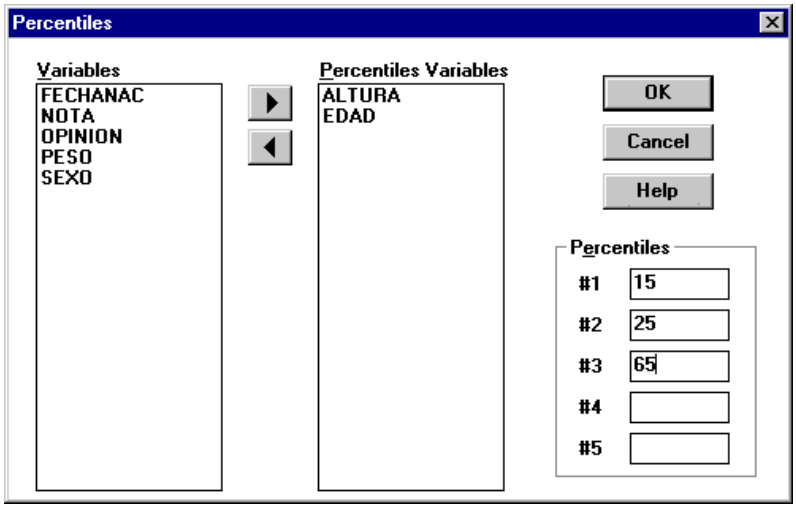


Figura 8.25: Pantalla del programa que permite seleccionar las variables estadísticas para las que vamos a calcular determinados percentiles.

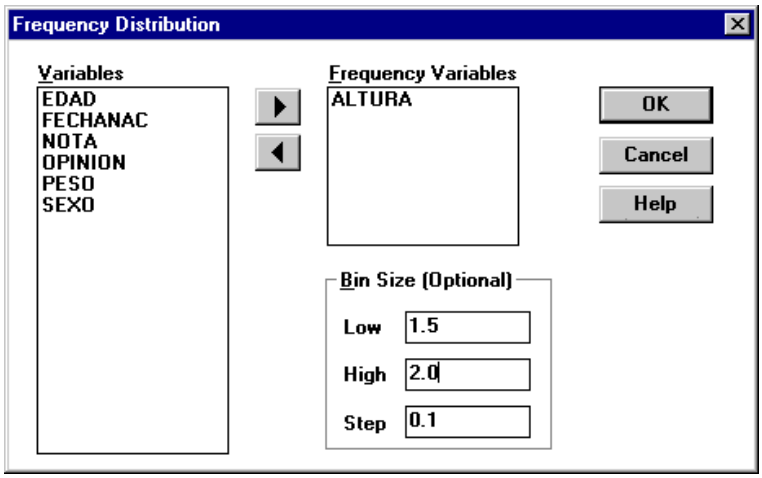


Figura 8.26: Pantalla del programa que permite seleccionar las variables para las que vamos a calcular las distribuciones de frecuencias.

70′2	62′4	62′8	68′4	65′2	66′3	70′7	72′8	68′7	71′9
64′4	60′4	67′0	62′9	65′9	67′5	66′6	67′8	73′5	63′1
72′3	73′4	70′1	70′0	68′3	65′0	72′1	66′1	65′5	62′0
63′3	72′4	69′4	71′3	70′2	70′4	64′1	69′1	66′4	65′2

a) Agrupar los datos en intervalos de la misma amplitud aplicando la regla de Sturges. Hay que especificar los extremos de los intervalos, la marca de clase, frecuencia absoluta, frecuencia relativa, frecuencia acumulada absoluta y frecuencia acumulada relativa. Calcular la media y la desviación típica, directamente y mediante codificación.

b) Agrupar de nuevo los datos sabiendo que la amplitud de los intervalos es 1.

E.8.2. Los siguientes datos corresponden al nivel de glucosa en sangre de diez niños:

56	62	63	65	65	65	65	68	70	72

a) Calcular la media, mediana y moda.

b) Calcular la desviación media e interpretar el resultado.

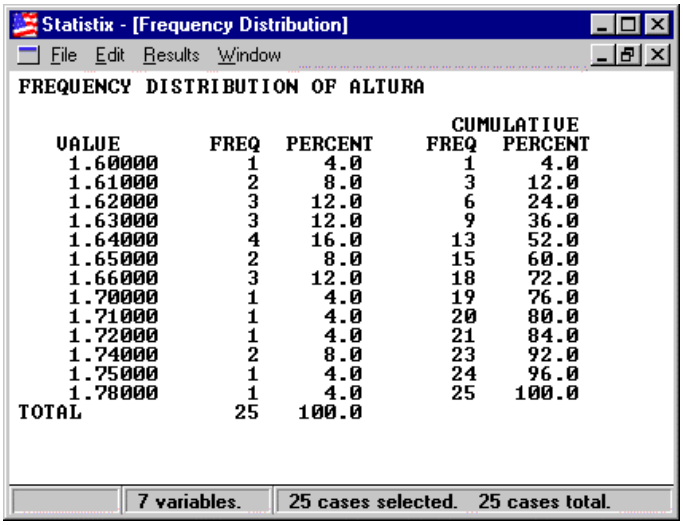


Figura 8.27: Pantalla del programa que muestra la distribución de frecuencias de la variable ALTURA.

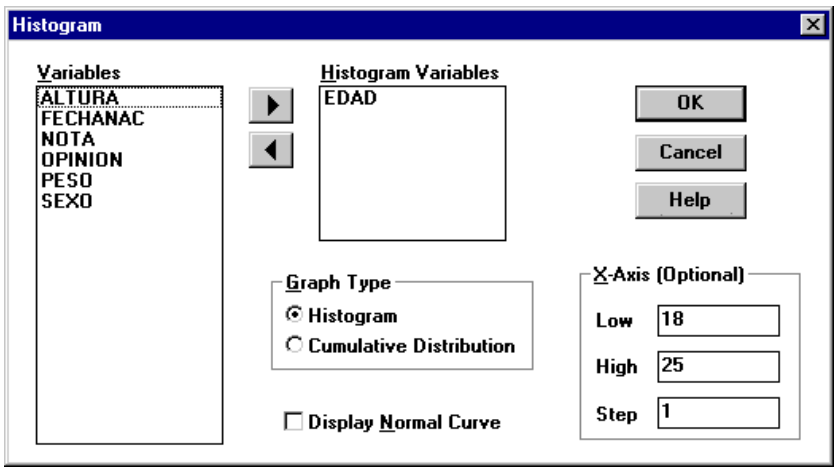


Figura 8.28: Pantalla del programa que permite seleccionar la variable para la que vamos a representar su histograma.

c) Calcular la desviación mediana e interpretar el resultado.

E.8.3. En un mismo examen de una asignatura, pasado a dos clases de 40 alumnos, se han registrado las siguientes calificaciones:

clase A	2	3	7	5	4	6	7	5	6	7
	8	4	3	7	5	6	9	5	7	2
	7	4	8	5	6	5	7	8	5	3
	6	5	8	4	2	5	4	5	6	7
clase B	2	5	3	4	7	9	5	2	7	4
	8	3	5	8	7	9	3	2	4	1
	10	9	4	8	6	9	3	3	7	1
	2	8	6	7	3	6	9	7	4	8

Realizar un estudio estadístico de las calificaciones en cada clase y decir qué conclusiones pueden sacarse de dicho estudio. ¿Son ambas clases muy similares por lo que respecta a la variable calificación, o por el contrario son muy diferentes?

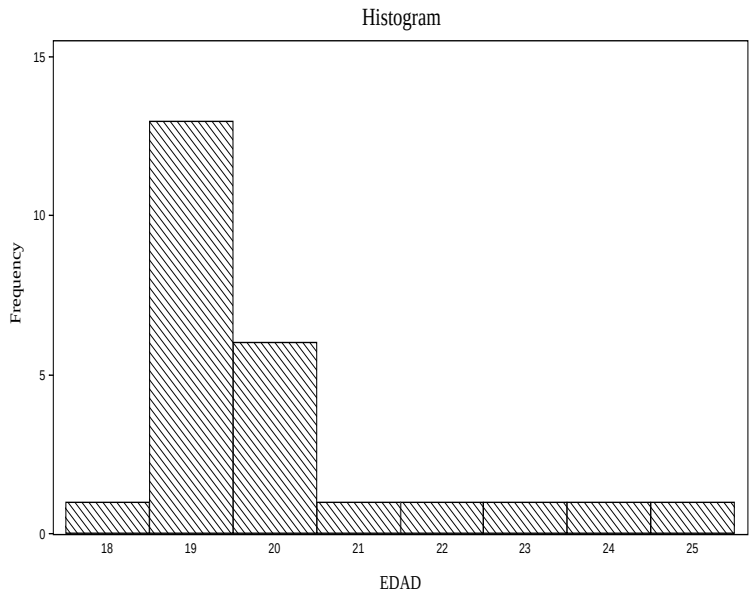


Figura 8.29: Pantalla del programa que muestra el histograma de frecuencias de la variable EDAD.

E.8.4. Los siguientes datos corresponden a los precios (en cientos de pesetas) de una colección de 30 libros.

27	40	12	3	30	16	20	21	30	12
45	18	25	22	35	24	37	12	21	7
35	17	21	27	14	15	25	45	12	24

a) Agrupar los datos en intervalos de la misma amplitud (mediante la regla de Sturges) y clasificarlos en una nueva tabla que contenga la marca de clase, la frecuencia absoluta, la frecuencia relativa y la frecuencia acumulada. Dibujar el histograma de frecuencias absolutas.

b) Calcular el séptimo decil. Compararlo con el valor que resultaría si lo hallásemos a partir de los datos agrupados en intervalos.

c) Calcular un coeficiente de asimetría e interpretarlo.

E.8.5. El número de veces que fueron consultados 30 artículos de investigación archivados en una hemeroteca (durante el año pasado) viene dado por la siguiente tabla.

18	25	20	24	19	23	21	22	20	22
23	19	21	24	21	22	20	21	22	21
20	24	21	19	23	21	22	23	22	23

a) Calcular la mediana y la mediana de las desviaciones respecto de la mediana. Deducir si los datos están cerca de la mediana.

b) Calcular la media y la desviación típica. Deducir si los datos están cerca de la media.

E.8.6. **a)** Agrupar los datos del ejercicio anterior en intervalos de la misma amplitud y clasificarlos en una nueva tabla en la que aparezca la marca de clase, la frecuencia absoluta y la frecuencia acumulada.

b) A partir de la agrupación en intervalos anterior, calcular de nuevo la media y la mediana. Comparar estos dos nuevos valores con los resultados obtenidos en el ejercicio anterior (datos no agrupados en intervalos).

E.8.7. El número de personas que visitan diariamente un centro de información fue observado durante 74 días elegidos al azar, y los resultados fueron:

nº de personas	47	59	62	64	71	76	78	80
nº de días	4	6	10	17	16	10	7	4

a) Hacer el gráfico de frecuencias acumuladas absolutas.

b) Hallar la moda, la mediana y el recorrido intercuartílico. ¿Están los datos cerca de la mediana?

c) Calcular el coeficiente de curtosis e interpretar el resultado.

E.8.8. Las personas que aprobaron una determinada oposición tienen la siguiente distribución por edades:

edad	[20,25]	(25,30]	(30,35]	(35,40]	(40,50]	(50,60]
frecuencia	41	123	44	13	7	3

a) Calcular la edad media. ¿Es representativa de todos los datos?

b) Calcular la mediana. ¿Es representativa de todos los datos?

c) Hacer el polígono de frecuencias acumuladas absolutas.

ANOTACIONES