Tema 1

DESCRIPCIÓN GRÁFICA Y NUMÉRICA DE UNA VARIABLE

1.INTRODUCCIÓN

La Tabla 1 recoge, parcialmente, el resultado de una encuesta[1] a la que fueron sometidas 250 personas con edad igual o superior a 15 años, tabla que aparece completa en el Anexo I. Las columnas de las respuestas están encabezadas por nombres, abreviados en algunos casos, que hacen referencia a la pregunta formulada. Su significado y el de la codificación correspondiente es el siguiente:

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
01	v	1	63	3	80,30	190	3	1	1	3
02	v	1	79	4	56,16	155	1	3	1	2
03	m	1	52	3	64,37	151	3	3	1	2
04	m	3	41	3	63,02	146	3	2	2	2
05	v	2	18	4	75,50	164	4	2	1	3
06	m	2	68	3	35,00	136	4	3	2	2
07	v	2	35	2	62,79	145	3	1	1	2
08	m	2	46	2	78,92	190	1	2	1	3
09	m	2	20	3	58,27	171	3	2	1	0
10	v	1	61	4	52,17	159	3	2	2	2
11	m	1	69	3	70,82	169	3	1	2	2
12	m	2	50	3	41,10	167	4	2	1	3
13	m	1	67	2	49,46	171	3	2	3	1

Tabla 1.- **Reproducción parcial de las 250 observaciones del Anexo I**

- La primera columna indica el número de caso
- **Sexo:** indica el sexo del entrevistado, **v** = varón, **m** = mujer
- **E_civil:** indica el estado civil, **1** = casado/a, **2** = soltero/a, **3** = viudo/a, **4** = div/sep
- **Edad:** edad expresada en años
- **Niv_ed:** nivel de educación, **1** = analfabeto/a, **2** = sin estudios, **3** = est. primarios, **4** = BUP o similares, **5** = est. universitarios
- **Peso:** peso expresado en kilogramos
- **Altura:** altura expresada en centímetros
- **Jueces:** opinión sobre los jueces "estrella", **1** = buena, **2** = indiferente, **3** = mala, **4** =ns/nc
- **E_penal:** opinión sobre el adelanto de la edad penal, **1** = a favor, **2** = en contra, **3** =ns/nc

[1] La encuesta es ficticia y las respuestas que en ella figuran han sido simuladas. Se trata tan solo de un ejemplo elaborado a los efectos de presentación y desarrollo del curso.

- **35horas:** opinión sobre la semana laboral de 35 horas **1** = a favor, **2** = en contra, **3** =ns/nc
- **C_alcohol:** consumo medio diario de alcohol medido en el equivalente a vasos de vino de 200cc, la escala va de 1 a 5, indicando esta última cifra 5 o más vasos diarios

Interpretar los datos que aparecen en la tabla presenta dificultades incluso para las personas con conocimientos de Estadística y, desde luego, prácticamente imposible para lo que podríamos denominar *gran público*. No por casualidad cuando se ofrece información de este tipo aparece resumida y transformada para hacerla fácilmente comprensible, resumen que pretende llamar nuestra atención sobre los aspectos más relevantes de los datos y que para conseguirlo utiliza las herramientas propias de la Estadística Descriptiva o Descripción de datos, a saber:

- distribuciones de frecuencia,
- gráficos,
- medidas de posición o centrales, y
- medidas de dispersión.

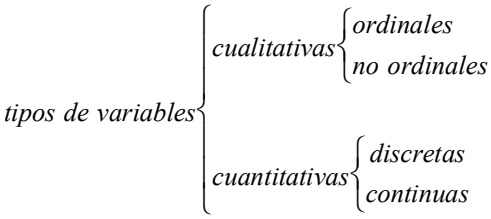
Antes de comenzar el resumen de los datos de nuestra tabla, introduciremos el lenguaje y las definiciones que nos permitan hacerlo.

2. MUESTRA Y VARIABLES

Los datos recogidos en la tabla reciben el nombre de **muestra**, que a su vez está constituida por las **observaciones muestrales** a cuyo número denominaremos **tamaño muestral**. En cada observación hay una o varias **variables observadas**. En el caso de nuestra tabla tenemos:

- una **muestra** que contiene las respuestas a determinada encuesta,
- el **tamaño muestral** es de 250 observaciones
- cada **observación muestral** se corresponde con las respuestas a la encuesta de una persona con edad igual o superior a 15 años,
- las **variables observadas** son: *sexo, estado civil, nivel de educación, peso, altura, opinión sobre los jueces, opinión sobre edad penal, opinión sobre semana laboral de 35 horas* y *consumo medio diario de alcohol.*

Las variables, lógicamente, han de centrar nuestra atención prioritariamente, razón por la cual conviene establecer una clasificación de las mismas:



variables cualitativas: son variables que describen categorías, razón por la cual se las denomina también **categóricas**. Cuando las categorías admiten algún tipo de ordenación se las denomina **ordinales** (por ejemplo, la variable *nivel de educación* de la tabla) y **no ordinales** en caso contrario (por ejemplo, las variables *sexo, estado civil, opinión sobre los jueces, opinión sobre edad penal, opinión sobre semana laboral de 35 horas)*

variables cuantitativas: son variables que expresan valores numéricos, **discretas** o **continuas** según la naturaleza de la observación. En la tabla, *consumo medio diario de alcohol* es un ejemplo de las primeras y *peso, altura* son ejemplos de las segundas.

La frontera entre variables discretas y continuas es en ocasiones difusa debido a la acción discretizadora que todo proceso de medida comporta. En efecto, si observamos la variable *edad* en la tabla nadie pondrá en duda su carácter continuo pues mide el *tiempo* transcurrido desde el nacimiento de una persona, pero, en general, las fracciones de año son irrelevantes razón por la cual viene medida en años y aparece expresada mediante valores enteros positivos.

3. DISTRIBUCIONES DE FRECUENCIAS

Una primera descripción resumida de los datos puede llevarse a cabo mediante la distribución de frecuencias de cada una de las variables. Como luego pondremos de manifiesto, el tipo de variables es determinante a la hora de analizar los datos con esta herramienta. Para variables categóricas y discretas con un rango pequeño de valores utilizaremos distribuciones de frecuencias no agrupadas de las que nos ocupamos a continuación:

Frecuencias no agrupadas Se trata simplemente de obtener y representar gráficamente el número de ocurrencias (**frecuencia absoluta)** de las distintas categorias o valores de la variable. En ocasiones es conveniente utilizar la **frecuencia relativa**, definida como:

$$frecuencia\ relativa = \frac{frecuencia}{n},$$

donde n es el tamaño muestral. La frecuencia relativa se suele expresar también en porcentaje.

Obtengamos la distribución de frecuencias asociada a alguna de las variables de la tabla.

Estado civil	Frec	Frec Rel
casado/a	131	0.524
soltero/a	98	0.392
viudo/a	16	0.064
div/sep	5	0.020
Total	250	1.00

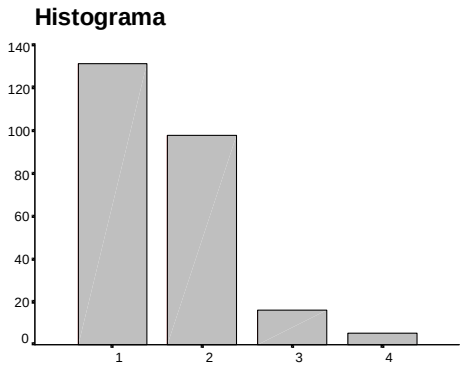


Figura 1.- **Tabla de frecuencias e Histograma de E_CIVIL**

Opinión	Frec	Frec Rel
buena	67	0.268
indiferente	49	0.196
mala	112	0.448
ns/nc	22	0.088
Total	250	1.00

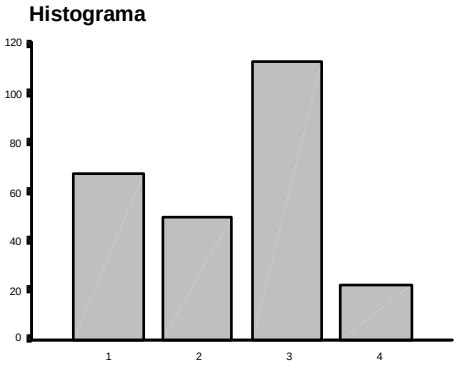


Figura 2.- **Tabla de frecuencias e Histograma de JUECES**

Vasos	Frec	Frec Rel
0	22	0.088
1	50	0.200
2	90	0.360
3	49	0.196
4	26	0.104
5	13	0.052
Total	250	1.00

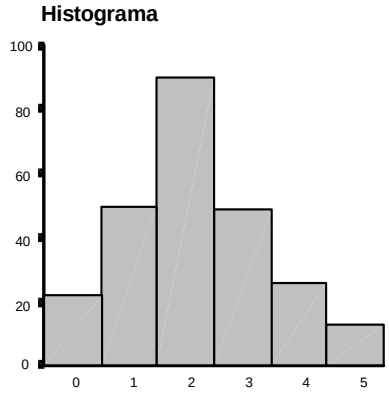


Figura 3.- **Tabla de frecuencias e Histograma de C_ALCOHOL**

La representación gráfica de las frecuencias, en los casos de variables categóricas o de variables discretas con pocos valores, puede también llevarse a cabo mediante **Diagramas de Sectores**, en los que cada valor o categoría de la variable se representa mediante un sector circular con área proprocional a su frecuencia. Las figura 4 y 5 son una muestra de estos diagramas para las variables *nivel estudios* y *35horas*

Estudios	Frec	Frec Rel
analf.	7	0.028
sin est.	57	0.228
primarios	112	0.448
BUP o sim.	61	0.244
univers.	13	0.052
Total	250	1.00

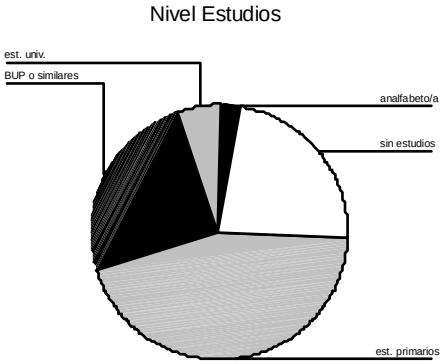


Figura 4.- **Tabla de frecuencias y Diagrama de Sectores de NIV_ED**

35 horas	Frec	Frec Rel
a favor	131	0.524
en contra	96	0.384
ns/nc	23	0.092
Total	250	1.00

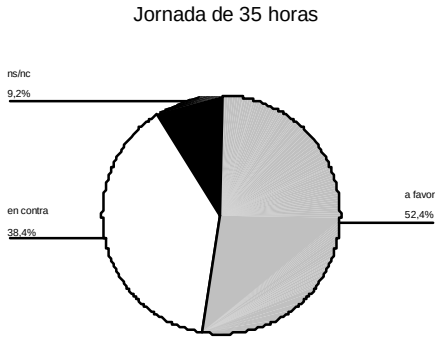


Figura 5.- **Tabla de frecuencias y Diagrama de Sectores de 35HORAS**

Frecuencias agrupadas Si pretendemos resumir la información de las variables *edad*, *altura* y *peso* tal como lo hemos hecho en las anteriores, es decir, considerando cada valor como una categoría obtendremos una tabla de frecuencias y un histograma que, al no condensar la información, nos servirán de poca ayuda. El motivo está en el carácter continuo de la variable.

El problema se resuelve agrupando los valores de la variable en **clases** y obteniendo la distribución de frecuencias para dichas clases.

Las clases son intervalos y están delimitadas por los **límites de clase**, y deben constituir una partición del conjunto de valores que toma la variable, es decir, las clases no se solapan y no deben excluir ningún valor de la variable, lo que permite clasificar a cualquier valor en una y solo una de las clases establecidas. La distancia entre los límites de la clase es la **amplitud de la clase**.

En la gráfica siguiente aparece la distribución de frecuencias de la variable *edad* que ha sido agrupada en los intervalos que se indica en la tabla, a saber, 8 clases de longitud 10, donde la clase i-ésima es el intervalo $[x_i, x_{i+1}[$, que al estar abierto en su límite superior no se solapa con la clase siguiente.

Edad	Frec	Frec Rel
15-25	48	0.192
25-35	43	0.172
35-45	50	0.200
45-55	34	0.136
55-65	23	0.092
65-75	30	0.120
75-85	16	0.064
85-95	6	0.024
Total	250	1.00

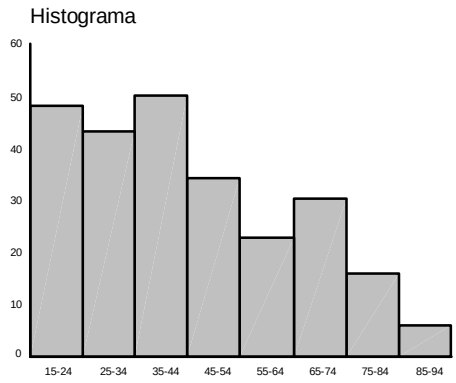
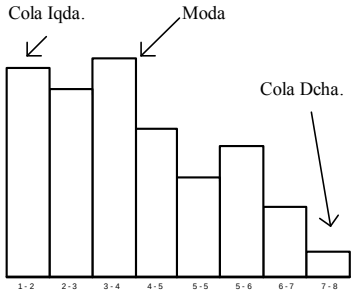


Figura 6.- **Tabla de frecuencias e Histograma de EDAD**

¿Qué información de interés nos proprociona el anterior histograma? Aunque más tarde estudiaremos con mayor detalle este problema, adelantemos ya algunos aspectos relevantes de la forma de la distribución de frecuencias. El pico, que representa la mayor frecuencia es la **moda**, valor alrededor del cual se distribuyen los valores que toma la variable, cuyas frecuencias van disminuyendo a derecha e izquierda para formar en los extremos las llamadas colas de la distribución. En nuestro caso, la **cola izqda.** es más *pesada* que la **derecha**, indicando con ello que hay mayor presencia de edades inferiores que de superiores y dando lugar a una distribución sin **simetría** y **sesgada** a la izquierda.



Número de clases a establecer. La pregunta que surge al observar la distribución de frecuencias anterior es ¿por qué 8 clases y no 14? No es difícil imaginar que un número de clases distinto producirá una gráfica de aspecto diferente, como puede observarse en los histogramas que aparecen a continuación; en ellos la variable edad ha sido representada con 3 y 30 clases, respectivamente.

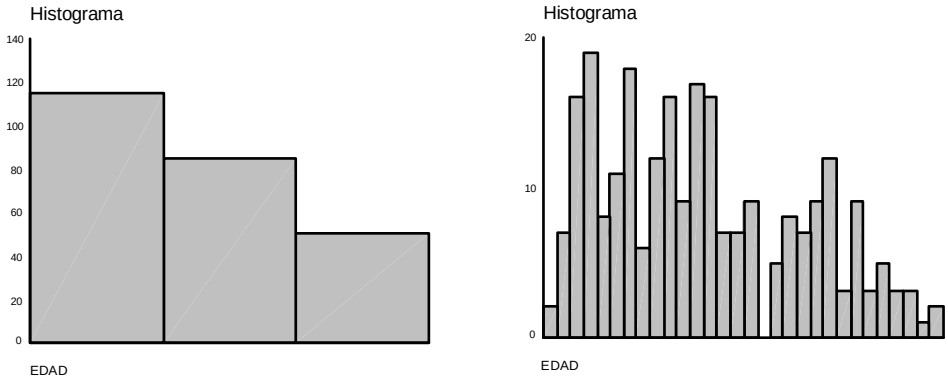


Figura 7.- **Histogramas de barras para EDAD con 3 y 30 clases**

No debemos olvidar que el objetivo de cualquier agrupación de datos es destacar los rasgos esenciales y eliminar los detalles irrelevantes, aún cuando esto se haga a expensas de perder una información que no consideramos esencial, de ahí la importancia de elegir adecuadamente el número y amplitud de las clases. Las

siguientes recomendaciones pueden ayudarnos, aunque puede ser conveniente llevar a cabo distintas elecciones y comparar los resultados:

- si el tamaño de la muestra es $n \leq 50$, un número de clases entre 5 y 15 suele ser apropiado; para muestras mayores este número puede superar las 20 clases,
- el rango de la variable, que es *rango = valor mayor - valor menor*, y la amplitud que deseamos para cada clase nos permitirán determinar su número. Por ejemplo, para la tabla de frecuencias e histograma de la variable *edad* que hemos representado en la figura 6, hemos calculado su rango = 92-15 = 77 y como deseábamos una amplitud de 10 años para cada clase, hemos obtenido un número de 7.7, que lógicamente se ha redondeado a 8, lo que supone que la última clase cubre el intervalo [85,95[,

Clases con amplitudes distintas Los histogramas que hemos utilizados hasta ahora provienen de distribuciones de frecuencias agrupadas cuyas clases tienen todas igual amplitud, razón por la cual su **altura** es directamente proporcional a su frecuencia.

Cuando las frecuencias de clases contiguas son bajas pueden agruparse en clases mayores cuya frecuencia será la suma de las frecuencias de las clases que constituyen la nueva clase. Por ejemplo, lo datos de la Tabla 2 son una muestra de 30 valores de la variable *peso*, extraídos de entre los 250 que constituyen los datos originales. La tabla de frecuencias muestra que la segunda clase, [35,45[, tiene una frecuencia 0.

81,72	52,44	69,24	58,34	81,43	52,35	28,60	92,78	87,82	59,44
86,39	68,26	57,29	83,62	26,14	68,47	56,00	96,97	57,79	65,10
78,37	56,74	45,41	65,85	48,95	81,84	74,82	91,93	71,48	68,34

Tabla 2.- **30 observaciones del peso**

Peso	**Frec**	**Frec Rel**
25-35	2	0.066
35-45	0	0.000
45-55	4	0.134
55-65	6	0.200
65-75	8	0.267
75-85	5	0.166
85-95	4	0.134
95-105	1	0.033
Total	30	1.00

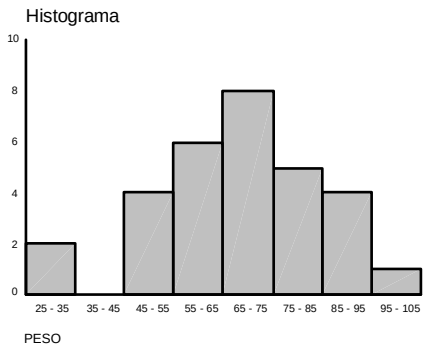


Figura 8.- **Tabla de frecuencias e Histograma de los 30 valores del *peso***

Puede pensarse en la conveniencia de agrupar las dos primeras clases en una sola, [25,45[, para conseguir una distribución de frecuencia más suavizada que evite la frecuencia 0. La consecuencia de esta agrupación es una distribución de frecuencias con clases de distinta amplitud, una de ellas el doble que las restantes, y debemos cambiar el método de representación del histograma para evitar distorsiones en su forma. En efecto, si, como hasta ahora, la altura de la barra correspondiente a cada clase es proporcional a su frecuencia, obtendremos el histograma B de la figura 9, que transmite visualmente la idea de una presencia de la primera clase mayor de la que en realidad le corresponde. Esto se evita haciendo que las **áreas** de las barras sean proporcionales a la frecuencia, como se ha hecho en el histograma A, lo que conduce en nuestro caso a una altura que es la mitad de la anterior puesto que la base del rectángulo es el doble.

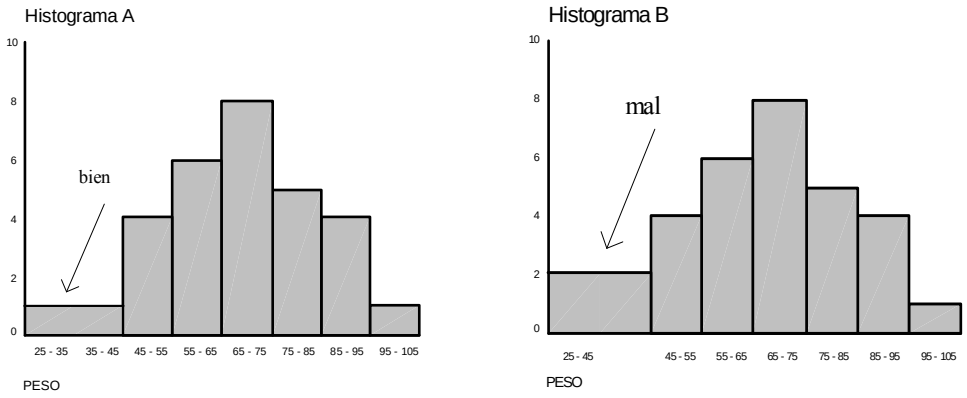


Figura 9.- **Histogramas de frecuencias para distribuciones con clases de distinta amplitud**

4. MEDIDAS DE POSICIÓN

Para las variables categóricas, las distribuciones de frecuencias y sus distintas representaciones gráficas nos proporcionan información concisa y completa, pero si las variables son cuantitativas es posible, y conveniente, completar aquella información con características numéricas asociadas a los datos. Estas características reciben el nombre de **estadísticos descriptivos** y los hay de dos tipos: de **posición** o **centrales** y de **dispersión**. Los primeros nos proporcionan información acerca de la posición de los datos si los representamos en una recta, mediante la obtención de lo que podríamos llamar *centro* de la distribución. Existen distintas formas de definir el centro de una distribución de datos, las más utilizadas son: *la media*, *la mediana*, *la moda y los percentiles*.

En adelante designaremos mediante las últimas letras mayúsculas del abecedario, **X**, **Y**, **Z**, …, a las variables observadas y con las minúsculas, **x**, **y**, **z**, …, las observaciones (datos), a las que cuando sea conveniente añadiremos un índice. Por ejemplo, si queremos designar las *n* observaciones de la variable **X** lo podemos hacer mediante $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$, …, $\mathbf{x_n}$.

La media Es sin duda la más conocida de las medidas de posición y es, sencillamente, la **media aritmética** de las observaciones correspondientes a la variable en estudio. Se le denomina **media muestral** y se le designa mediante el símbolo $\bar{\mathbf{x}}$. Su expresión es,

$$\bar{x} = \frac{suma\ de\ las\ x's}{n} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Retomemos los datos de las 30 observaciones de pesos contenidos en la Tabla 2, para calcular su media

$$\bar{x} = \frac{81{,}72+52{,}44+69{,}24\ +\cdots+91{,}93+71{,}48+68{,}34}{30} = \frac{2013{,}92}{30} = 67{,}13\ kgs.$$

La mediana Es aquel valor que, al ordenar las observaciones de menor a mayor, ocupa el lugar central, dividiendo el conjunto de observaciones en partes iguales. Es decir, que deja a su derecha y a su izquierda el 50% de las observaciones. Si el tamaño de la muestra, *n*, es **impar**, necesariamente existe una observación que ocupa el lugar central, concretamente la que al ordenar las observaciones está en la posición (n+1)/2; si, por contra, *n* es **par**, son dos las observaciones que ocupan el lugar central, las que están en las posiciones n/2 y (n/2)+1, definiéndose entonces la mediana como el punto medio entre ambas observaciones. Veamos algunos ejemplos:

- **Ejemplo 1:** Si ordenamos los 30 valores del *peso* de la Tabla 2 tendremos:

26,14	28,60	45,41	48,95	52,35	52,44	56,00	56,74	57,29	57,79
58,34	59,44	65,10	65,85	**68,26**	**68,34**	68,47	69,24	71,48	74,82
78,37	81,43	81,72	81,84	83,62	86,39	87,82	91,93	92,78	96,97

y siendo *n=30* par, la mediana será el valor medio de los valores que ocupan las posiciones 15 y 16, que aparecen en negrita en la ordenación. Así pues,

$$\text{mediana} = \frac{68{,}26 + 68{,}34}{2} = 68{,}30 \; kgs.,$$

valor que, como puede observarse, no coincide con el de la media antes calculada.

- **Ejemplo 2:** Las 13 primeras observaciones correspondientes al *consumo de alcohol* ordenadas de menor a mayor son: 0 1 2 2 2 2 **2** 2 2 3 3 3 3. La que ocupa la posición central, la séptima puesto que hay 13 valores, es la mediana y su valor es 2.

La moda Es aquel valor de la variable que tiene mayor frecuencia. En el caso de frecuencias agrupadas se toma la clase más frecuente como moda. Así, para la variable *consumo de alcohol* la moda es 2 (ver tabla de frecuencias de la Figura 3) y para la variable *edad* la moda es la clase 35-45 (ver Figura 6).

Los percentiles El percentil **p-ésimo** es aquel valor que verifica la condición de que un p% de las observaciones son menores o iguales que el. Así, el percentil 70-ésimo supone que el 70% de las observaciones son menores o iguales que el valor de dicho percentil.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
0	109	123	126	132	135	136	136	138	141	141	142	142	143	144	145	145	146	147	147	147
20	148	148	148	149	149	150	150	150	151	151	151	151	152	152	153	153	**154**	154	155	155
40	155	156	157	157	157	157	157	158	158	158	159	159	159	159	159	159	160	160	160	160
60	160	160	160	161	161	161	161	161	162	162	162	162	162	162	**162**	163	164	164	164	164
80	164	164	164	165	165	165	165	165	165	165	165	166	166	166	167	167	167	167	167	167
100	167	168	168	168	168	168	168	169	169	169	169	169	169	169	169	169	170	170	170	170
120	170	170	170	171	171	171	171	171	171	171	172	172	172	172	172	172	173	173	173	173
140	173	173	173	173	174	174	174	174	174	174	174	175	175	175	175	175	175	176	176	176
160	176	176	177	177	177	177	177	177	177	178	178	178	178	178	178	179	179	179	179	180
180	180	180	181	181	181	182	182	182	182	182	182	182	182	183	183	183	183	184	184	184
200	184	185	185	185	185	185	185	186	186	186	186	187	187	187	187	187	187	187	189	189
220	189	189	189	189	**190**	190	190	190	190	190	191	192	192	192	192	194	195	195	197	198
240	200	200	201	202	202	203	207	215	218	218										

Tabla 3.- **Las 250 observaciones de la variable *altura* ordenadas**

La Tabla 3 nos muestra, ordenadas de izquierda a derecha y de arriba a abajo, las 250 observaciones correspondientes a la variable *altura*. La primera fila y la primera columna, en negrita, han sido añadidas para mejor localizar las posiciones de cada valor en la ordenación. Así, si queremos conocer el percentil 30-ésimo, tendremos en cuenta que el 30% de 250 es 75 y buscaremos el valor

que ocupa esta posición en la tabla, el 162. El percentil 15-ésimo es 154 porque, aunque el 15% de 250 es 37.5, los valores correspondientes a las posiciones 37 y 38 son, ambos, 154. Si no hubiera sido así, hubiéramos tomado el valor correspondiente a la posición más cercana. De la misma manera calcularíamos el percentil 90 que es 190.

Los percentiles 25, 50, y 75-ésimo reciben el nombre de **primer cuartil**, **segundo cuartil** y **tercer cuartil**, respectivamente. El nombre les viene de dividir las observaciones en cuartos. Observemos que según la definición que hemos dado para la mediana, ésta coincide con el percentil 50-ésimo o segundo cuartil.

5. MEDIDAS DE DISPERSIÓN

La medidas de posición nos dan una información incompleta, por parcial, acerca de las observaciones. En efecto, supongamos que las notas de Matemáticas de los estudiantes pertenecientes a dos clases distintas, clase I y clase II con 10 estudiantes cada una, son las siguientes:

clase I: 4, 3, 5, 6, 4, 5, 5, 7, 5, 6

clase II: 1, 4, 3, 5, 6, 8, 2, 7, 5, 9

en ambos casos la media, como puede comprobarse con facilidad, es 5, pero sus histogramas de frecuencias son muy distintos.

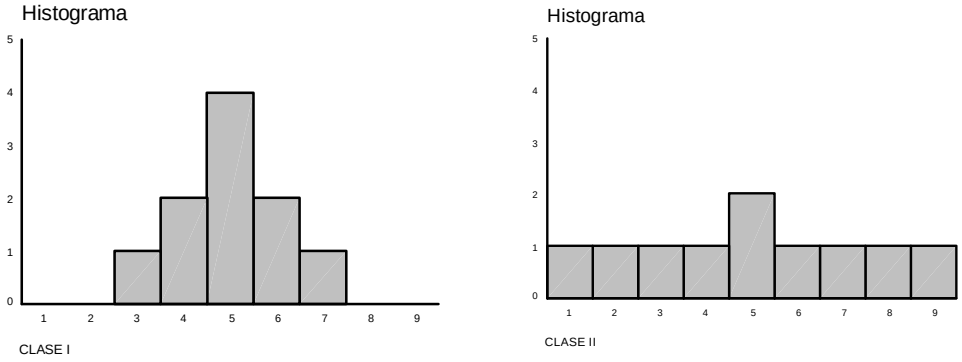


Figura 10.- **Histogramas de frecuencias para notas de las clases I y II**

La Figura 10 muestra que los valores se distribuyen simétricamente respecto de la nota 5, pero en la clase I existe una dispersión menor que en la clase II. ¿Cómo medir la distinta manera en que los valores se agrupan alrededor de la media? Las distintas

medidas de dispersión proporcionan esta información. Al igual que ocurre para la posición, existen diversas formas de medir la dispersión, de entre ellas vamos a ocuparnos de las siguientes: *rango, desviación tipica, varianza* y *rango intercuartílico*.

El rango Es la diferencia entre el máximo y el mínimo de las observaciones. Así, para los datos anteriores tendremos que rango de las notas en la clase I vale **4** y el rango en la clase II vale **8**, denotando la mayor dispersión de la variable en el segundo grupo de observaciones.

La varianza y la desviación típica Puesto que se trata de medir cómo se agrupan los valores alrededor de la media, podríamos utilizar como criterio las desviaciones de dichos valores respecto de aquella, es decir, la diferencias entre la media y los distintos valores y más concretamente la media de ellas. Aunque a primera vista la sugerencia pueda ser buena, vamos a aplicarla a los valores de las notas de clase para evidenciar el inconveniente insalvable que una medida de este tipo tiene.

En el cuadro aparecen las notas de cada clase y en columnas sucesivas sus desviaciones respecto de la media y el cuadrado de estas desviaciones, al que más tarde aludiremos. Al tratar de obtener la media de las diferencias, que recordemos es la suma de todas ellas dividida por su número, nos encontramos que dicha media será **0** en ambos casos, porque existiendo desviaciones positivas y negativas, unas anulan los efectos de las otras. En realidad eso nos ocurrirá con cualquier otro conjunto de datos, porque puede demostrarse que esa es una propiedad que tienen las desviaciones respecto de la media.

CLASE I			**CLASE II**		
nota	$\mathbf{d_i = x_i - \bar{x}}$	$\boldsymbol{d_i^2}$	**nota**	$\mathbf{d_i = x_i - \bar{x}}$	$\boldsymbol{d_i^2}$
4	1	1	1	4	16
3	2	4	4	1	1
5	0	0	3	2	4
6	-1	1	5	0	0
4	1	1	6	-1	1
5	0	0	8	-3	9
5	0	0	2	3	9
7	-2	4	7	-2	4
5	0	0	5	0	0
6	-1	1	9	-4	16
Suma	**0**	**12**	**Suma**	**0**	**60**

Tabla 4.- **Desviaciones respecto de la media y sus cuadrados para las notas de las clase I y II**

Puesto que el uso de las desviaciones respecto de la media parece razonable, ¿cómo soslayar el problema? Una manera sencilla de hacerlo es utilizar, no las desviaciones, sino sus cuadrados. Al ser estos cantidades positivas, su suma nunca podrá ser cero. Así, la media de los cuadrados de las desviaciones parece una medida adecuada, pero, por razones técnicas que están fuera del alcance y objetivos de este curso, la utilizaremos con una ligera modificación: en lugar de dividir por *n*, como se hace habitualmente para calcular una media, dividiremos por *n-1*. De acuerdo con esto, **la varianza** de un conjunto de observaciones se define mediante la fórmula:

$$s^2 = \frac{\textit{suma del cuadrado de las desviaciones}}{n-1} = \frac{\sum_{i=1}^{n} (\bar{x} - x_i)^2}{n-1}.$$

La **desviación típica** se define como la raiz cuadrada de la varianza y la designamos por *s*.

Para el caso de las clases I y II las sumas de los cuadrados de las desviaciones aparecen en la Tabla 4, sus varianzas y desviaciones típicas son:

clase I $\quad s^2 = \frac{12}{9} = 1{,}33 \qquad s = \sqrt{1{,}33} = 1{,}15$

clase II $\quad s^2 = \frac{60}{9} = 6{,}66 \qquad s = \sqrt{6{,}66} = 2{,}58$

que ponen de manifiesto la diferente distribución de los valores en un caso y otro. Para los 30 valores del *peso* de la Tabla 2,

peso $\quad s^2 = \frac{9040{,}76}{29} = 311{,}75 \text{ kgs}^2 \qquad s = \sqrt{311.75} = 17{,}65 \text{ kgs.}$

Obsérvese que las unidades de la varianza son el cuadrado de las unidades en las que venga expresada la variable, sin embargo la desviación no cambia de unidades.

Señalemos por último que si el tamaño de la muestra es grande, la diferencia entre dividir por *n* o por *n-1* es inapreciable y la varianza coincide, prácticamente, con la media de los cuadrados de las desviaciones.

Porcentajes típicos La desviación típica tiene una propiedad interesante, para distribuciones de frecuencias con una sola moda, de apariencia simétrica y con colas ni demasiado largas ni demasiado cortas, se suele verificar:

- aproximadamente el 68% de las observaciones distan como mucho **una** desviación típica de la media
- aproximadamente el 95% de las observaciones distan como mucho **dos** desviaciones típicas de la media

- aproximadamente más del 99% de las observaciones distan como mucho **tres** desviaciones típicas de la media

El rango intercuartílico Se define como la diferencia entre el tercer y el primer cuartil, IQR = Q_3 - Q_1. Directamente relacionado con él se define el **intervalo intercuartílico**, que es el intervalo definido por los cuartiles primero y tercero, [Q_1 , Q_3], cuya longitud es, precisamente, IQR. Contiene el 50% de las observaciones centrales. Para las 250 observaciones correspondientes a la *altura* estas medidas valen:

altura Q_1 = 160 cms Q_3 = 182 cms. IQR = 22 cms.

El coeficiente de variación Aún cuando no se trata, estrictamente, de una medida de dispersión este es el momento de definir esta nueva característica asociada a las observaciones. Para comprender mejor su interés tratemos de responder a la pregunta, ¿dónde hay mayor dispersión, en las observaciones del peso o en las notas de la clase I? La pregunta tiene difícil respuesta si, por ejemplo, pretendemos comparar directamente las correspondientes desviaciones típicas. En efecto, la del *peso* es mucho mayor que la de las *notas*, pero a nadie se le escapa que la magnitud de aquel es mucho mayor que las de éstas y, además, se trata de unidades diferentes, kilogramos en un caso y puntuación en el otro. Para resolver el problema se define el **coeficiente de variación** como el cociente entre la desviación típica y la media multiplicado por 100,

$$CV = \frac{s}{\bar{x}} 100 ,$$

que expresa la desviación típica en porcentaje de la media y que al no tener unidades permite comparaciones entre observaciones de distinta naturaleza. Volviendo a la pregunta inicial, para el peso, CV_{peso} = 100×(17,65/67,13) = 26,29%, y para las notas, $Cv_{notas\ I}$ = 100×(1,15/5) = 23%, lo que nos dice que en términos de porcentaje de sus medias, ambas distribuciones tienen dispersiones muy parecidas.

6. TRANSFORMACIÓN DE UNA VARIABLE: TIPIFICACIÓN

En ocasiones puede ser interesante transformar los valores observados mediante cambios sencillos. Por ejemplo, multiplicarlos por una constante y/o sumarles alguna cantidad fija. Una transformación de este tipo se denomina *lineal* y se expresa mediante la fórmula:

$$Y = aX + b ,$$

donde **Y** es la nueva variable que resulta de multiplicar la variable original, **X**, por *a* y añadirle *b* al resultado. Cuando $b = 0$, la transformación recibe el nombre de **homotecia** o **cambio de escala**, si $a = 1$ y $b \neq 0$, recibe el nombre de **traslación**. Por ejemplo, si decidimos cambiar la escala en las alturas observadas y expresarlas en metros, $Y = X/100$, con $a = 1/100$, puesto que 1m=100cms.

¿Cómo afecta una transformación lineal a las media, la varianza y la desviación típica? Con relativa sencillez puede comprobarse que éstas se ven afectadas de la manera que indica en el cuadro:

	X	**Y**
media	$\bar{x}$	$\bar{y} = a\bar{x} + b$
desviación típica	s_X	$s_Y = as_X$
varianza	s_X^2	$s_Y^2 = a^2 s_X^2$

Por ejemplo, si decidiéramos expresar los pesos en gramos, la tranformación sería de la forma $Y = 1000X$, y tendríamos $\bar{y} = 67130$ grs., $s_Y = 17656$ grs. y $s_Y^2 = 311750517$ grs^2.

Finalicemos con una transformación que tiene nombre propio y que es muy utilizada en estadística, se trata de la **tipificación de una variable**. Es una transformación lineal en la que $a = 1/s_X$ y $b = \bar{x} / s_X$ y que consiste en:

$$Y = \frac{X - \bar{x}}{s_X},$$

es decir, restarle a cada valor la media y dividirlo luego por la desviación típica. Si tenemos en cuenta el efecto de la transformación descrito en el cuadro anterior, $\bar{y} = 0$, $s_Y = 1$ y $s_Y^2 = 1$, y a la nueva variable se la conoce con el nombre de **variable tipificada**. Obsérvese que cualquiera que sea **X** inicialmente, la variable tipificada correspondiente tiene siempre media 0 y varianza y desviación típica 1.

Como comentario final, que se deja a la comprobación del lector, el *coeficiente de variación* no se altera cuando se lleva a cabo un cambio de escala.

7. MUESTRA Y POBLACIÓN: INFERENCIA ESTADÍSTICA

Solo en contadas ocasiones nuestro interés al analizar un conjunto de observaciones se limita a una simple descripción de las mismas mediante los distintos métodos explicados en párrafos anteriores. Casi siempre la descripción constituye un primer paso para conocer aquello que el conjunto de observaciones representa y que no nos es accesible.

Recordemos que al comenzar el tema aludíamos al conjunto de datos del Anexo I como a una **muestra** de personas con edad igual o superior a los 15 años. La pregunta que surge de inmediato es, *¿de dónde proviene esa muestra?* Para responder debemos introducir el concepto de **población**, que en nuestro caso podríamos definir como el total de individuos con edad igual o superior a 15 años. La pretensión habitual de los investigadores es conocer las características de la población a partir de lo observado en la muestra, en aquellas ocasiones en las que el estudio exhaustivo de la población es imposible en la práctica.

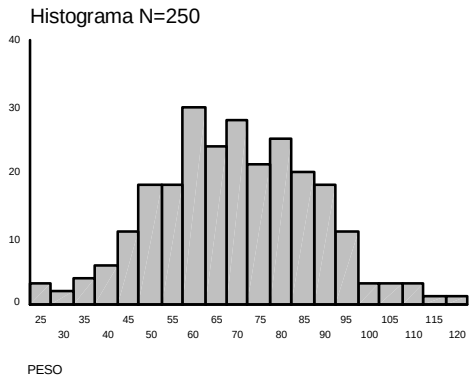
¿A qué nos referimos cuando hablamos de conocer la población a través de la muestra? Nos referimos a bajo qué condiciones, si la media de las 250 *alturas* de la muestra vale 170.68 cms., este valor puede tomarse como la altura media de las personas con edad igual o superior a 15 años. El proceso que estudia la manera de extraer conclusiones acerca de una población, partiendo de las observaciones contenidas en una muestra de aquella, se denomina **Inferencia Estadística**, y de él nos ocuparemos en temas posteriores. Podemos sin embargo adelantar la condición fundamental que toda muestra debe cumplir respecto de la población que pretende representar: la muestra debe ser *representativa*. Ello significa que ha de haber sido obtenida de tal manera que reproduzca los rasgos de aquella. Esto lo entenderemos mejor mediante algunos ejemplos de muestras que no serían representativas de las alturas de nuestra población original:

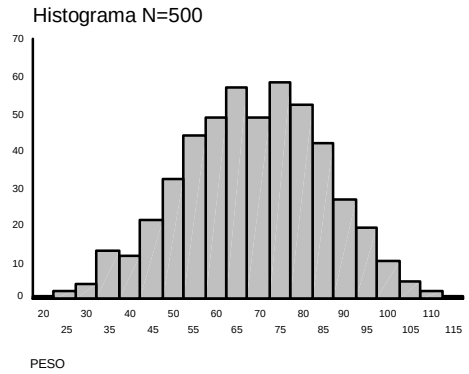
- que la muestra hubiera sido elegida sólo entre los hombres, en este caso podríamos extender nuestras conclusiones al conjunto de los hombres con edad igual o mayor a 15 años, pero de ninguna manera a todas las personas que cumplen la condición de edad,
- que la muestra hubiera sido elegida entre las personas con rentas altas, porque, si admitimos que mayor renta implica mejores condiciones de vida y alimentación, el resultado podría ofrecernos una altura media superior a la real.

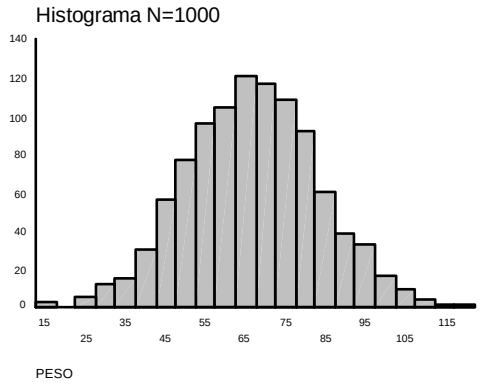
Bajo el nombre de **técnicas de muestreo** se conocen los distintos procedimientos que garantizan la representatividad de una muestra y estudian cómo el tamaño de la muestra influye en la calidad de nuestras conclusiones. Su importancia es obvia y su conocimiento primordial para llevar a cabo cualquier estudio en el que muestra y población estén implicados. El *tamaño de la muestra* es, no cabe duda, crucial en todo el proceso, pues a nadie se le escapa que a mayor tamaño, más acertados estaremos en nuestras conclusiones; pero, desgraciadamente, el aumento del tamaño encarece la obtención de la muestra, lo que impide que aquel crezca tanto como desearíamos.

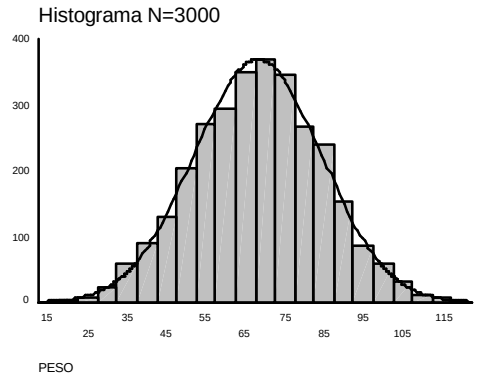
La curva de frecuencias Los histogramas que siguen representan las distribuciones de frecuencias de muestras de *pesos* cuyos tamaños hemos ido aumentando progresivamente. Observamos que a medida que N crece los histogramas evolucionan hacia una suavización del contorno superior de sus barras, evolución que nos permite intuir que, para una teórica muestra que contuviera toda la población, el histograma acabaría pareciéndose, si no coincidiendo, a la curva continua que hemos sobrepuesto a las gráficas correspondientes a los tamaños 3000 y 10000. Esta curva límite recibe el nombre

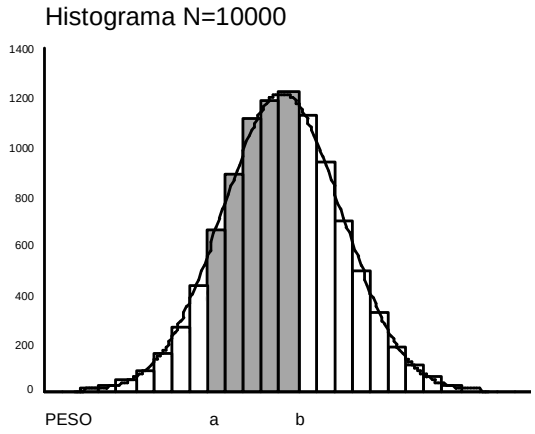
de **curva de frecuencias** o **curva de densidad** y tiene la propiedad de que las frecuencias relativas se representan en ella como área. Así, para cualesquiera dos pesos *a* y *b*, el área que hay bajo la curva entre *a* y *b* es la frecuencia relativa de los pesos que hay entre ambas cantidades.











ANEXO I

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
01	v	1	63	3	80,30	190	3	1	1	3
02	v	1	79	4	56,16	155	1	3	1	2
03	m	1	52	3	64,37	151	3	3	1	2
04	m	3	41	3	63,02	146	3	2	2	2
05	v	2	18	4	75,50	164	4	2	1	3
06	m	2	68	3	35,00	136	4	3	2	2
07	v	2	35	2	62,79	145	3	1	1	2
08	m	2	46	2	78,92	190	1	2	1	3
09	m	2	20	3	58,27	171	3	2	1	0
10	v	1	61	4	52,17	159	3	2	2	2
11	m	1	69	3	70,82	169	3	1	2	2
12	m	2	50	3	41,10	167	4	2	1	3
13	m	1	67	2	49,46	171	3	2	3	1
14	m	1	61	2	89,30	182	4	1	3	2
15	m	1	21	2	70,66	173	3	1	1	2
16	m	2	62	5	53,36	151	3	2	2	2
17	m	2	22	3	71,23	172	3	2	2	1
18	m	2	46	3	30,99	147	3	3	1	1
19	m	2	23	4	77,04	175	3	2	2	4
20	v	2	70	3	81,72	170	1	2	1	2
21	m	1	41	2	88,81	175	4	1	3	1
22	m	2	45	4	92,97	189	3	1	3	2
23	m	3	45	3	60,86	177	1	2	1	0
24	v	2	22	4	48,34	164	3	2	1	1
25	m	1	39	3	90,96	178	3	2	2	0
26	v	3	37	2	105,70	203	2	3	1	2
27	m	1	39	4	58,80	172	3	2	3	5
28	v	2	68	3	69,75	167	2	1	1	2
29	m	2	41	3	94,20	187	3	1	2	2
30	v	1	76	2	86,82	187	2	1	1	3
31	v	1	27	4	52,35	177	1	2	2	5
32	v	1	59	3	78,44	186	3	1	2	2
33	v	2	29	2	75,29	177	1	1	1	2
34	m	2	16	2	86,68	185	2	1	1	2
35	m	2	19	2	65,76	165	2	2	3	1
36	m	3	53	3	77,32	185	2	1	2	1
37	m	3	40	3	87,36	178	1	1	3	4
38	v	1	16	2	87,97	186	3	2	1	2
39	m	3	62	4	55,13	160	4	2	1	2
40	m	2	80	5	80,46	173	2	2	2	1
41	v	1	65	2	84,96	176	2	1	2	4

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
42	v	1	64	3	47,45	141	4	1	1	1
43	m	3	27	3	34,45	138	4	2	2	3
44	m	2	84	3	77,51	187	1	1	2	0
45	m	2	29	4	87,55	186	3	1	1	0
46	m	1	19	4	67,82	169	3	2	1	3
47	m	1	25	4	99,47	173	2	2	1	2
48	m	1	37	4	41,60	151	3	2	1	3
49	v	2	54	3	64,78	163	3	2	1	3
50	v	1	44	3	88,61	173	3	2	1	1
51	v	2	22	3	64,87	178	1	2	1	4
52	v	1	21	1	56,07	179	4	1	2	2
53	m	2	80	3	57,03	159	1	2	2	1
54	m	1	74	4	66,90	183	1	2	2	2
55	m	3	75	2	90,57	195	3	2	2	3
56	v	1	47	4	86,39	186	1	3	2	2
57	v	1	46	3	68,47	172	3	2	1	2
58	m	2	41	2	47,48	148	4	1	1	0
59	v	2	87	4	57,48	165	2	3	1	0
60	v	2	43	4	57,35	160	4	2	1	1
61	v	2	44	3	80,08	162	3	2	1	1
62	m	2	47	3	87,97	183	3	1	3	0
63	m	1	33	3	65,29	162	2	1	2	1
64	m	1	19	2	66,72	176	3	2	1	0
65	m	1	41	4	78,11	173	1	2	1	2
66	m	2	23	2	59,74	147	4	2	1	0
67	m	1	31	5	93,28	189	3	2	1	1
68	v	1	51	3	50,41	171	3	3	1	4
69	m	1	55	3	91,16	192	4	1	1	2
70	m	1	32	3	50,84	155	1	1	1	1
71	v	2	76	3	80,50	170	2	2	1	3
72	m	1	19	4	53,77	168	4	2	2	4
73	v	1	45	2	78,37	182	1	2	1	2
74	m	2	32	3	61,37	169	3	2	2	2
75	m	2	58	3	73,24	174	2	2	2	1
76	m	1	39	3	72,13	177	3	2	2	3
77	v	2	21	5	85,15	189	1	2	2	3
78	m	1	64	1	56,28	185	2	1	2	3
79	m	1	44	2	57,53	159	4	2	1	1
80	m	2	30	3	54,68	150	1	1	1	2
81	m	2	54	3	79,00	180	3	2	1	2
82	m	1	44	4	67,29	149	3	2	2	0
83	m	1	29	3	93,12	190	3	2	2	3
84	v	4	20	2	57,52	162	3	2	2	5
85	v	1	39	4	78,43	181	3	2	1	2

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
86	v	1	18	4	68,91	166	3	2	1	2
87	v	4	38	3	51,94	160	1	3	1	2
88	m	1	51	5	57,79	165	2	3	2	2
89	v	2	42	3	86,45	182	3	3	1	2
90	m	2	43	2	48,37	151	3	2	2	2
91	m	1	90	2	61,33	164	1	2	2	3
92	m	2	52	4	103,68	194	3	2	2	2
93	v	1	45	2	85,61	182	1	2	1	3
94	v	1	80	3	67,62	168	1	2	1	2
95	m	1	48	2	70,64	187	3	2	1	2
96	v	2	35	2	77,98	190	1	2	1	4
97	v	1	85	3	69,75	171	3	2	1	4
98	m	2	21	2	66,38	148	3	2	2	3
99	v	1	24	2	81,84	185	1	2	1	0
100	v	3	38	2	47,12	159	3	2	1	4
101	v	1	34	3	61,70	169	3	2	1	2
102	v	2	43	3	24,44	142	3	2	1	5
103	v	1	74	3	82,49	179	1	1	3	3
104	v	2	42	4	89,23	198	4	3	2	1
105	v	1	19	3	65,15	172	2	3	1	1
106	m	2	67	3	43,68	147	1	2	2	1
107	m	2	66	2	59,60	143	1	2	2	4
108	v	1	80	5	78,47	169	2	2	2	3
109	v	1	55	4	52,44	161	1	2	1	1
110	m	2	49	2	41,39	152	3	2	2	2
111	v	2	45	3	82,95	187	3	2	1	4
112	m	2	44	3	108,57	192	1	3	1	5
113	v	2	37	3	39,30	162	3	2	1	2
114	v	2	21	3	68,26	177	4	2	1	4
115	m	4	49	4	44,90	160	1	1	1	0
116	v	2	68	3	64,77	162	3	1	1	2
117	m	1	28	4	67,90	160	2	2	2	2
118	m	2	24	3	59,26	158	1	2	3	0
119	v	1	55	4	48,21	154	1	2	3	4
120	m	2	17	2	82,89	168	3	1	2	1
121	v	1	88	4	74,25	174	3	1	1	1
122	m	1	27	4	31,93	132	3	2	1	3
123	v	2	30	4	58,81	167	3	2	1	3
124	v	1	30	3	14,22	109	3	1	2	2
125	v	1	32	1	56,74	167	1	1	1	3
126	m	1	31	2	65,36	174	3	2	2	3
127	v	1	25	2	74,82	166	4	1	1	4
128	m	1	18	3	58,70	170	3	1	1	2
129	v	2	29	3	83,52	182	3	2	2	2

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
130	m	1	29	3	96,43	215	3	1	2	2
131	v	1	77	4	69,24	170	3	2	1	3
132	m	1	64	3	59,05	170	3	2	2	0
133	v	1	69	3	46,87	156	2	1	1	4
134	v	2	28	3	90,15	182	3	2	1	1
135	m	1	74	2	71,89	178	3	2	1	2
136	m	1	83	4	26,50	126	1	3	2	3
137	v	1	25	3	100,80	190	3	2	1	1
138	v	1	22	3	47,97	141	3	3	1	3
139	m	1	67	4	74,86	183	2	2	1	3
140	v	2	54	4	66,67	167	1	3	2	2
141	m	2	18	3	61,28	161	1	2	2	3
142	m	1	40	2	71,98	179	2	2	3	1
143	v	3	36	3	68,98	162	1	2	1	2
144	v	2	72	3	92,78	180	3	2	2	3
145	v	2	28	3	57,71	159	2	1	3	3
146	v	1	43	5	79,12	182	3	2	1	2
147	v	1	34	3	32,52	136	3	2	2	2
148	v	3	61	1	87,10	177	1	2	1	5
149	v	1	31	4	104,84	218	3	3	1	2
150	m	1	17	3	56,02	160	3	3	2	2
151	v	1	39	3	67,40	176	3	2	1	3
152	v	2	64	4	58,89	157	1	2	1	3
153	v	1	34	4	77,22	179	2	2	1	1
154	v	1	69	3	44,24	144	2	2	1	4
155	v	1	23	3	79,78	181	3	3	2	3
156	m	2	48	1	61,85	169	2	2	1	3
157	v	2	19	3	108,02	207	3	2	1	2
158	m	2	70	3	64,86	169	1	2	2	3
159	m	1	40	4	40,48	142	1	1	3	1
160	v	1	44	5	57,29	164	3	3	2	5
161	v	2	68	3	79,34	184	1	3	2	1
162	m	1	72	3	49,73	165	2	2	2	2
163	m	1	47	4	77,06	176	1	1	1	2
164	m	2	43	5	77,58	176	3	3	1	3
165	v	2	64	4	60,07	168	3	2	2	1
166	v	1	71	2	76,83	167	1	2	2	3
167	m	1	27	3	43,78	157	1	2	2	0
168	m	1	42	3	38,87	161	4	2	2	1
169	v	1	29	3	93,73	195	1	3	1	2
170	v	1	74	3	68,03	164	1	2	1	1
171	v	1	75	3	87,07	190	2	2	2	1
172	v	2	47	1	67,75	161	1	1	1	2
173	m	1	23	1	57,19	152	3	2	1	4

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
174	v	1	44	3	96,97	200	4	2	1	2
175	m	2	33	3	61,98	169	3	1	1	1
176	v	2	23	2	45,41	154	2	2	1	1
177	v	1	16	4	89,02	192	3	2	1	3
178	v	1	55	3	75,17	173	3	3	2	5
179	m	2	18	5	86,66	178	1	2	3	1
180	v	1	45	2	91,93	192	2	2	1	5
181	v	2	85	3	58,34	157	1	2	3	1
182	m	1	42	3	70,70	171	3	2	2	4
183	m	2	73	3	58,94	159	2	3	3	3
184	m	2	69	3	69,46	174	1	2	2	2
185	v	2	48	3	87,82	185	1	2	1	2
186	v	1	79	4	76,50	170	3	2	1	1
187	m	1	58	2	78,91	172	3	3	2	2
188	v	2	69	4	57,98	165	1	2	3	4
189	m	1	70	3	57,78	157	3	1	2	4
190	m	1	30	3	109,60	202	3	2	2	0
191	v	1	40	5	55,18	150	3	2	2	2
192	v	1	48	4	77,88	174	2	1	2	2
193	m	1	69	4	63,39	162	2	2	2	3
194	v	1	30	5	86,19	200	2	2	1	2
195	m	1	68	4	93,06	184	3	2	1	3
196	m	2	38	3	96,03	202	3	1	2	2
197	v	2	17	2	83,62	182	1	2	1	3
198	v	2	30	4	45,47	150	1	2	2	1
199	v	1	32	3	57,79	168	3	3	1	1
200	v	2	23	4	88,64	189	3	3	2	2
201	v	4	42	2	84,10	183	3	2	1	4
202	m	1	26	3	55,96	166	2	2	2	2
203	m	2	75	2	49,22	153	1	1	1	1
204	v	2	22	4	65,85	173	2	2	1	3
205	v	2	45	2	71,48	170	2	1	1	5
206	v	1	21	2	80,48	171	3	2	2	2
207	v	3	29	3	87,69	187	2	2	1	1
208	v	1	39	3	81,43	175	2	2	2	3
209	v	1	28	2	84,61	191	1	2	1	4
210	m	2	20	4	32,85	145	3	2	1	2
211	m	1	20	3	63,58	167	1	2	2	2
212	v	1	84	4	94,18	184	2	2	1	5
213	m	1	37	3	113,93	197	3	1	2	1
214	m	3	36	3	69,89	161	3	2	2	0
215	m	1	55	4	49,79	135	1	2	1	3
216	m	4	37	4	54,64	164	3	2	3	2
217	m	1	36	2	60,98	153	1	1	2	3

	SEXO	E_CIVIL	EDAD	NIV_ED	PESO	ALTURA	JUECES	E_PENAL	35HORAS	C_ALCOHOL
218	v	1	46	2	59,44	157	1	1	2	3
219	v	2	52	2	89,20	201	1	1	2	2
220	v	2	36	2	70,02	155	1	2	3	4
221	m	2	30	2	73,48	165	3	2	2	3
222	v	1	71	2	74,49	171	4	1	1	1
223	v	3	24	4	64,72	169	2	1	2	2
224	m	1	34	4	75,72	184	2	2	3	0
225	m	1	81	3	68,90	178	3	2	1	0
226	m	3	58	3	77,33	185	1	1	1	0
227	v	1	36	3	77,03	187	3	1	2	2
228	v	3	15	3	26,14	123	3	2	1	1
229	v	2	59	4	88,39	177	3	2	2	5
230	v	1	61	4	75,20	158	1	2	1	2
231	m	2	63	2	46,29	164	1	2	1	2
232	v	1	26	3	67,87	175	3	1	1	2
233	v	2	52	3	49,43	165	2	2	1	2
234	m	2	49	2	61,88	148	3	2	1	0
235	m	2	69	4	78,53	189	4	2	1	1
236	v	2	39	4	62,89	175	1	2	1	2
237	m	1	65	3	50,15	149	2	2	1	1
238	m	2	46	3	73,99	172	3	2	2	2
239	v	2	20	2	100,02	189	1	1	1	2
240	v	1	39	3	65,10	174	2	2	2	2
241	v	2	24	2	48,95	160	2	2	2	2
242	m	1	21	3	75,05	158	3	2	1	1
243	v	1	17	4	122,17	218	2	1	1	2
244	m	1	15	5	66,00	180	2	2	2	2
245	v	1	29	3	68,44	165	3	2	2	4
246	m	1	27	2	82,65	173	3	2	2	1
247	m	1	25	3	50,49	174	3	2	1	4
248	m	2	92	3	62,39	181	2	2	3	4
249	m	2	36	3	83,78	168	3	2	3	5
250	m	1	23	2	68,34	175	2	2	2	2