



# **APUNTES DE ESTUDIO BASES DE LA FORMACIÓN CIENTÍFICA**

*Eje Formación Remedial*

Unidad N°2: Representación de datos y  
medidas de resumen

# Índice

<b>1. Tipos de datos y variables</b>	<b>2</b>
1.1. Definiciones . . . . .	2
1.2. Tipos de variables . . . . .	2
1.3. Medida de las variables . . . . .	3
1.3.1. Escalas nominales . . . . .	3
1.3.2. Escalas ordinales . . . . .	3
1.3.3. Escalas de intervalo . . . . .	4
1.3.4. Escalas de razón . . . . .	4
1.4. Variables cualitativas . . . . .	4
1.5. Variables cuantitativas . . . . .	4
1.6. Transformación de variables . . . . .	5
1.7. Necesidad de una definición clara de las variables . . . . .	6
1.8. Dominio de una variable . . . . .	6
1.9. Variables aleatorias y controladas . . . . .	6
1.10. Medida de una variable continua . . . . .	6
<b>2. Representación de la información mediante tablas y gráficos</b>	<b>7</b>
2.1. Tablas . . . . .	7
2.2. Forma de tabular . . . . .	7
2.2.1. Variable cualitativas . . . . .	7
2.2.2. Variables cuantitativas . . . . .	8
2.3. Representaciones gráficas . . . . .	8
2.3.1. Componentes . . . . .	9
2.3.2. Tipos de gráficos . . . . .	9
<b>3. Medidas simples de resumen de datos</b>	<b>16</b>
3.1. Definición de parámetros e índices . . . . .	16
3.2. Parámetros de tendencia central . . . . .	16
3.2.1. Media aritmética . . . . .	16
3.2.2. Media aritmética ponderada . . . . .	17
3.3. Mediana . . . . .	17

Tipo	Datos	Expresión	Variantes	Ejemplos
Cualitativas o Atributos	Catagóricos	Modalidades o Categorías	2 ó más modalidades	Sexo (H o M), caras de un dado: 1, 2, 3, 4, 5, 6
Cuantitativas	Métricos	Valores	Continuas o Discretas	Talla: 170 cm

Cuadro 1: Ejemplos de variables y sus características

## 1. Tipos de datos y variables

### 1.1. Definiciones

Por característica se entiende una propiedad o condición claramente reconocible en diversos individuos. El individuo es la unidad estadística y puede ser una persona, un animal, una planta, un objeto o una acción. Las características pueden ser constantes o variables. Las constantes no varían, siempre ocurren de la misma forma, como las constantes físicas o la certeza de la muerte en los seres vivos. Siguen el llamado modelo determinista de los fenómenos naturales. Tienen un resultado fijo, que se puede resumir por una fórmula matemática. Al lanzar una bola es posible saber con exactitud la velocidad y la aceleración que va a tener en un determinado momento.

Las variables presentan una gama de variaciones (al menos dos) en los diversos individuos, como el sexo o la talla de las personas. Siguen el modelo indeterminista (= probabilístico, casual o estocástico). No tienen un resultado fijo. Hay un conjunto de posibles resultados, conocidos de antemano, de los que sólo se producirá uno. Los factores que influyen en que se produzca ese resultado u otro son múltiples, complejos, incontrolables y en parte desconocidos, de forma que el resultado ocurre de forma aparentemente casual, al azar. El azar no es ciego, tiene sus modelos de comportamiento, predecibles con un margen de variación mediante fórmulas matemáticas, basadas en el cálculo de probabilidades. Son las llamadas distribuciones fundamentales de probabilidad (Distribución normal, de Poisson, binomial, hipergeométrica, etc.). Los fenómenos biológicos siguen uno u otro modelo, que una vez conocido nos permite calcular las probabilidades de que ocurra tal o cual resultado<sup>1</sup>.

### 1.2. Tipos de variables

Hay variables: **cualitativas (CL)** y **cuantitativas (CT)**, sus características se resumen en el cuadro 1.

<sup>1</sup>Pág 35. Spiegel M. Estadística. Mc Graw Hill 2002.

### 1.3. Medida de las variables

Se hace según las llamadas escalas. Básicamente hay 4 escalas de medidas:

- nominales
- ordinales
- de intervalo
- de razón

Las variables ordinales son una variante de las nominales y las de razón de las de intervalo.

#### 1.3.1. Escalas nominales

Se utilizan para medir atributos, es decir, variables cualitativas. Se da un nombre a cada una de las modalidades, se asignan los individuos a ellas y se cuentan los individuos de cada modalidad (frecuencia). El orden en que se designan las modalidades es indiferente, p.e. alto y bajo o bajo y alto.

Ejemplo: la variable sexo tiene dos modalidades, hombre y mujer. Medimos este atributo en 100 personas y encontramos 52 hombres y 48 mujeres.

En vez de dar un nombre convencional a las modalidades se las puede designar con un número, lo que facilita sobre todo el tratamiento informático. Estos números son realmente un nombre y por tanto no pueden hacerse con ellos operaciones matemáticas. Así podríamos llamar a los hombres “1” y a las mujeres “2” ( ó 7 y 8...)

#### 1.3.2. Escalas ordinales

Una escala ordinal es una escala nominal en la que las diversas modalidades guardan entre sí una relación de orden o jerarquía, que debe ser respetada, siendo indiferente que el orden sea de mayor a menor o viceversa. Ese orden viene marcado por el sentido común y también por la costumbre.

Un ejemplo clásico son las notas académicas tradicionales : sobresaliente-notable-aprobado- suspenso o suspenso-aprobado-notable-sobresaliente. En la variable “evolución de la enfermedad” podríamos distinguir las siguientes modalidades : muerto-peor-igual-mejor-curado , o bien, curado-mejor-igual-peor-muerto.

También pueden emplearse números como nombre de modalidades, pero respetando el orden. Podríamos hacer muerto=1, peor=2, igual=3, mejor=4, curado=5 . O bien, curado=1, mejor=2, igual=3, peor=4 , muerto=5 .

### **1.3.3. Escalas de intervalo**

Se utilizan para medir variables cuantitativas cuando no hay cero absoluto en la zona de medición, lo que permite valores negativos. El cero se asigna arbitrariamente así como la unidad de medida. La escala ha sido diseñada de tal manera que sus números permiten valorar exactamente la diferencia que hay entre dos medidas (= intervalo). Ejemplo típico es la temperatura medida de la forma habitual, lo que puede hacerse de diversas maneras. En Europa se mide en grados centígrados o Celsius (C). El "0" se asigna a la temperatura de congelación del agua destilada y el "100" a su temperatura de ebullición. Ese intervalo se divide en 100 partes y así se obtienen los grados centígrados. En USA se mide en grados Fahrenheit (F). 0° C equivalen a 32° F y 0° F equivalen a -17,78° C. Por tanto 32° C no representa el doble de calor que 16° C, simplemente el doble de grados C. Esas temperaturas medidas en grados Fahrenheit serían 0° F y -8,9° F. Un niño con un proceso febril en Castellón puede tener 40° C de fiebre; en USA tendría 104° F. Por la Física sabemos que hay un mínimo infranqueable de temperatura, el llamado "cero absoluto", que en grados centígrados corresponde a -273,15°. Este cero no significa la ausencia de temperatura, sino el mínimo de temperatura posible. La escala de Kelvin asigna su 0 a esta temperatura.

### **1.3.4. Escalas de razón**

Se utilizan para medir variables cuantitativas cuando hay un cero absoluto, siendo la unidad de medida lo único arbitrario. Una longitud puede ser medida en cm., Km., yardas, varas, etc. pero el cero es el mismo para todos. El tiempo de reacción a un estímulo siempre empieza en cero cualquiera que sea el sistema que utilicemos para medir el tiempo. Aquí sí puede decirse que una persona que pesa 50 Kg. pesa el doble que un niño que pesa 25. Y que la diferencia de peso entre una persona que pese 80 Kg. y otra que pese 50 Kg. es la misma que la existente entre dos piedras de 35 y 5 Kg., respectivamente. No hay valores negativos.

## **1.4. Variables cualitativas**

Las variables cualitativas (CL) o atributos se miden por escalas nominales u ordinales según corresponda. Cuando sólo tienen dos modalidades se llaman dicotómicas. Ejemplos: cara-cruz, varón-hembra, vivo-muerto. Todos los atributos, con independencia del número de modalidades que tengan, pueden ser siempre reducidos a dicotómicos si así se desea. Los 4 palos de la baraja española (oros, copas, espadas y bastos) pueden ser reducidos a oros-no oros, bastos-no bastos, etc. ; las marcas de autos a Hyundai-no Hyundai. ; el estado civil a casado-no casado.

## **1.5. Variables cuantitativas**

Las variables cuantitativas (CT) se miden por escalas de intervalo o de razón, según su naturaleza. Pueden ser continuas o discretas.

Una variable CT es continua cuando puede tomar cualquier valor en su zona de variabilidad. Son continuas la talla, el peso, la tensión arterial, el contenido de un frasco, la glucemia, etc.

Las variables CT discretas no pueden adoptar cualquier valor, sino solamente ciertos valores. Una familia puede tener 0, 1, 2, 3, ... hijos, pero no 3,1416 hijos. El nº de pacientes que ingresa en un hospital,, el nº de ataques que sufre un paciente en un mes, el nº de cápsulas de un envase medicamentoso son discretas.

Una variable CT continua se mide a menudo, porque resulta más práctico, de forma “discretizada”. La edad suele expresarse en años enteros, o en meses en los niños pequeños, pero no por eso deja de ser continua.

## **1.6. Transformación de variables**

Las variables cuantitativas pueden ser transformadas en cualitativas, dicotómicas o no, con una pérdida en la calidad de la medida, que a veces se asume si mejora la información. La talla podemos medirla en alta-normal-baja. Los valores de colesterol en mayor de 200 mg/dl - igual o menor de 200 mg/dl. Como la variable CT proporciona más información que la CL debe ser usada siempre que no sea más conveniente hacerlo de forma cualitativa. Las variables CL en cambio no pueden ser transformadas en CT. Veamos dos ejemplos al respecto:

### **Ejemplo 1.**

Variable: Ingestión de Alcohol

- 1) abstemio – bebedor *Variable CL con dos modalidades, nominal.*
- 2) abstemio – bebedor – alcohólico *Variable CL con tres modalidades, ordinal.*
- 3) nº de copas o vasos bebidos en una semana *Variable CT discreta.*
- 4) gramos de alcohol tomados en una semana *Variable CT continua.*

### **Ejemplo 2.**

Variable: Estudio de tres tratamientos de la isquemia coronaria.

- 1) sexo : hombre – mujer *CL con 2 modalidades, nominal*
- 2) medicamento: A – B – C *CL con 3 modalidades, nominal*
- 3) nº ataques del día anterior *CT discreta*
- 4) distancia caminada sin disnea *CT continua*

## **1.7. Necesidad de una definición clara de las variables**

Es esencial que todo el mundo sepa qué se está midiendo y cómo. Está claro lo que es medir el peso en Kg. o la talla en cm. Pero, ¿que es ser “fumador”? ¿El que fuma un pitillo, aunque sea una vez al año? ¿O el que fuma cada día o al menos cada tres?. Hay que concretar y decir por ejemplo: “en este estudio se considera fumador a quien fuma al menos un cigarrillo cada semana” o “se considera desnutridos a los niños que en los gráficos peso/talla de Tanner están por debajo del percentil 3”, etc., etc.

## **1.8. Dominio de una variable**

Es el conjunto de valores o modalidades que puede adoptar. El dominio de la variable CL “puntuación de la cara de un dado” es (1, 2, 3, 4, 5 y 6). El de la variable sexo: (hombre, mujer). El de la “longitud de las hojas de la planta P” cualquier valor entre 1 y 8 cm. o  $e(1/8)$ , etc.

## **1.9. Variables aleatorias y controladas**

Una variable es controlada o independiente cuando su valor o la modalidad elegida en cada uno de los individuos depende únicamente del investigador. En un estudio podemos seleccionar sólo individuos del sexo masculino. O fijar la dosis de medicamento que se da a los ratoncillos, etc. Una variable es aleatoria o dependiente cuando su valor en cada uno de los individuos no depende del investigador, sino de la naturaleza o reacción del propio individuo. Por ejemplo la talla de los alumnos de una clase, la tensión arterial de un grupo de pacientes, etc.<sup>2</sup>

## **1.10. Medida de una variable continua**

Debido a la imperfección de los instrumentos de medida, aún de los más sofisticados, el valor exacto o real de una medida ( $X_e$ ) es realmente desconocido y sólo podemos expresarlo de una forma aproximada mediante el valor medido ( $X$ ). Supongamos que estamos midiendo una longitud con una regla graduada. Cuando la medida no se corresponde con un valor marcado en la regla, hay que aproximar (=redondear) a la marca más cercana. Si hay equidistancia se aproxima al valor par<sup>3</sup>.

---

<sup>2</sup>Pág 2-4. Spiegel M. Estadística. Mc Graw Hill 2002.

<sup>3</sup>Pág 2-4. Spiegel M. Estadística. Mc Graw Hill 2002.

## 2. Representación de la información mediante tablas y gráficos

La tabulación consiste en presentar los datos estadísticos en forma de tablas o cuadros.

### 2.1. Tablas

- TITULO de la tabla, que debe ser preciso y conciso
- CONTENIDO, con
  - la fila de encabezamiento o cabecera (títulos de las columnas)
  - la columna matriz, con las modalidades o clases de la variable
  - columnas de parámetros
- NOTAS EXPLICATIVAS (opcional), como fuente de los datos, abreviaturas, etc.

### 2.2. Forma de tabular

#### 2.2.1. Variable cualitativas

- la frecuencia absoluta (símbolo :  $f$  ó  $n$ ), que es el nº de veces que aparece cada modalidad (resultado del recuento). La frecuencia total, de todas las modalidades juntas, se representa por  $N$ .
- la frecuencia relativa (  $fr$  ) o proporción se obtiene dividiendo la frecuencia de cada modalidad entre el total de datos.  $fr = f / N$  . Los valores posibles oscilan entre 0 y 1. Suele expresarse con 3 decimales. La suma de todas las  $fr$  tiene que dar 1 ó un número muy cercano al 1, si ha habido redondeos.
- el porcentaje ( $P$  o %), que es la frecuencia relativa multiplicada por 100.  $P = fr * 100$  ó  $\% = (f*100)/N$  . Suele expresarse con 3 dígitos. La suma de todos los porcentajes debe dar 100 o un número muy próximo, si ha habido redondeos.
- las frecuencia acumuladas ( $\sum f$  ó  $\sum n$  ) que se obtienen sumando la frecuencia de cada modalidad a las frecuencias ya acumuladas anteriormente. En la primera modalidad no hay nada acumulado de antes y por tanto su frecuencia acumulada será su misma frecuencia. La última modalidad tiene que dar una frecuencia acumulada igual a  $N$ .
- las frecuencias relativas acumuladas y los porcentajes acumulados se obtienen de forma similar.
- En las variables nominales las modalidades pueden ponerse en el orden que se quiera, pero en las ordinales hay que respetar el orden lógico.



Ejemplo:

Sección	f	fr	%	$\sum f$	$\sum fr$	$\sum \%$
Neonatología	25	0,125	12,5	25	0,125	12,5
Lactantes	95	0,475	47,5	120	0,6	60
Preescolares	80	0,400	40	200	1	100
Total	200	1	100			

### 2.2.2. Variables cuantitativas

Los datos se agrupan según la frecuencia de los valores. Es lo que se denomina **Distribución de frecuencias**. La forma de tabular depende del nº de datos.

Si son pocos (la mayoría de autores pone el tope en 30) , se hace una tabla simple de forma similar a lo visto para las variables CL. Cada dato equivale a una modalidad. Al final nos quedaremos con la f de cada número y si se prefiere también con el%. Los números se ordenan de menor a mayor o de mayor a menor. La tabla puede hacerse en sentido vertical u horizontal. Ejemplo: Si  $x = (4, 1, 7, 2, 2, 9, 7, 2, 2, 9, 7, 1, 4)$ , ejemplos de representaciones tablas pueden ser:

x	1	2	4	7	9
f	2	4	2	3	2

x	f
1	2
2	4
4	2
7	3
9	2

Si son muchos se agrupan en clases, que son intervalos sucesivos de valores. Los datos se asignan a la clase que les corresponde y se cuentan los datos de cada clase, que está representada por el punto medio o centro de clase (pm ó c).

Esta agrupación es arbitraria con dos condiciones esenciales: que las clases sean mutuamente excluyentes y que todos los datos puedan ser asignados a una clase. Ahora bien, la experiencia ha ido introduciendo una serie de normas, que permiten hacer esta agrupación de la forma más racional posible.

### 2.3. Representaciones gráficas

Los datos estadísticos pueden ser también representados por medio de gráficos. Un viejo proverbio chino dice que una imagen vale más que mil palabras (o que mil números, aplicado a la Estadística). Los gráficos son una simplificación y un complemento de una tabla estadística. Son más sencillos, más llamativos y a menudo más inteligibles, aunque se pierde información.

### 2.3.1. Componentes

Como en las tablas estadísticas se pueden distinguir:

- el título
- el gráfico en sí (casi siempre complementado con números)
- notas explicativas , si procede

### 2.3.2. Tipos de gráficos

Si bien existen varios tipos de gráficos, existen los tipos que se listan a continuación son los típicamente utilizados durante la presentación de datos clínicos, especialmente a nivel de publicación de la información científica. Estos tipos son:

- Diagramas
  - de barras
  - histogramas
  - polígonos de frecuencias
- Gráficos sectoriales
- Pictogramas
- Otros

Los **diagramas** utilizan un sistema de coordenadas cartesianas. En el eje de abscisas (x) se representa la variable. En el de ordenadas (y) las frecuencias o porcentajes. Si la variable es CL se marcan en el eje de abscisas las modalidades y sobre ellas se dibujan líneas o barras de altura proporcional al parámetro representado. Si la variable es CT se marcan los valores y clases correspondientes al recorrido de la variable. La escala de y debe de empezar siempre en 0 para evitar manipulaciones y engaños ópticos. Habitualmente se trata de una escala aritmética, pero cuando hay frecuencias o valores muy dispares el gráfico es apenas legible y es mejor utilizar escalas logarítmicas o semilogarítmicas. Una alternativa, algo chapucera, es quebrar claramente la escala y las barras. Todo antes que violar la norma del comienzo de y en 0. En un buen diagrama la longitud de x debe de estar entre 1 y 2 veces la de y. Ambas escalas deben de estar claramente rotuladas, directamente o por medio de una nota explicativa. Son preferibles números cortos (redondeados) y hay que evitar dar excesivos datos, sobre todo en presentaciones, ya que el gráfico se muestra un corto espacio de tiempo. Otra cosa es un gráfico impreso al que el lector puede dedicarle el tiempo que quiera. Los ordenadores permiten fácilmente dibujar los gráficos en 3D. Las barras pasan a ser prismas o incluso cilindros o conos, a gusto del usuario <sup>4</sup>.

---

<sup>4</sup>Extraído de de [www.eduardobuesa.es/Tema1-11.pdf](http://www.eduardobuesa.es/Tema1-11.pdf)

El diagrama de barras o columnas es propio de variables discretas (todas las CL y las CT discretas). Cada barra corresponde a una modalidad o valor de la variable. La altura de la barra es proporcional a la frecuencia a representar. Todas las barras deben de tener la misma anchura y la distancia entre ellas debe de ser como máximo la anchura de las barras. Se pueden distinguir tres tipos de diagramas de barras: a) simples (Figura 1 y 2) b) de barras adosadas o parcialmente superpuestas, cuando se presentan de forma paralela dos conceptos que interesa comparar, p.e. hombres y mujeres (Figura 3 y 4) c) de barras mixtas, apiladas, una variante del anterior (Figura 5).

El histograma es propio de variables CT continuas agrupadas en clases. Las barras están unas al lado de otras sin separación, a no ser que alguna clase tenga una frecuencia de 0. Cada barra empieza en el límite real inferior de la clase que representa y termina en el límite superior, que a su vez es el comienzo de la clase siguiente. El punto medio de la clase coincide con el centro de la base. La superficie de cada barra es proporcional a la frecuencia de la clase. Si todas las clases tienen la misma amplitud, como en principio debe ser, la altura es la frecuencia de la clase. Si hay clases con distinta amplitud no puede ponerse la etiqueta de frecuencia ( $f$  ó  $n$ ) en el eje vertical, ya que sería engañoso. Debe figurar la de “densidad de frecuencias” ( $fd$ ), que se calcula  $fd = f/i$ . (Figura 6)

Se pueden distinguir tres tipos de histogramas:

- 1) el H. simple, que es el que acabamos de ver (Figura 7)
- 2) el H. de frecuencias acumuladas, en el que cada barra representa las frecuencias acumuladas en cada clase. El gráfico tiene forma de escalera más o menos irregular. (Figura 8)
- 3) el H. doble, cuyo paradigma es la pirámide de población. Este gráfico nos informa de la distribución por edades de un grupo poblacional, separando hombres y mujeres y rotando el gráfico de tal forma que las edades de las personas, agrupadas en clases, están en el eje vertical y la frecuencia de cada clase en el eje horizontal <sup>5</sup>. (Figura 9).

Un **polígono de frecuencias** se obtiene uniendo los puntos medios de los techos de un hipotético histograma, que se corresponden, al ser la barra un rectángulo, con los puntos medios o centros de cada clase. La línea debe comenzar y terminar en el eje de abscisas, precisamente en el sitio que correspondería al punto medio de dos clases inexistentes, la que precedería a la primera y la que seguiría a la última. Si se superponen un histograma y el correspondiente polígono de frecuencias se ve que la superficie del histograma y el área que incluye el polígono es la misma <sup>6</sup>. Por tanto ambos representan igualmente a la distribución. Los hay también simples y de frecuencias acumuladas (Figuras 10 y 11).

<sup>5</sup>Extraído de [www.eduardobuesa.es/Tema1-11.pdf](http://www.eduardobuesa.es/Tema1-11.pdf)

<sup>6</sup>Tomado de [www.eduardobuesa.es/Tema1-11.pdf](http://www.eduardobuesa.es/Tema1-11.pdf)



Figura 1: Diagrama de barras simple



Figura 2: Diagrama de barras simple, rotado



Figura 3: Diagrama de barras adosadas

Los **gráficos sectoriales** o de **tarta** equivalen a un diagrama de barras y por tanto sirven para representar variables discretas. Se utilizan círculos o semicírculos y a cada modalidad o valor se le adjudica un sector circular, cuya superficie sea

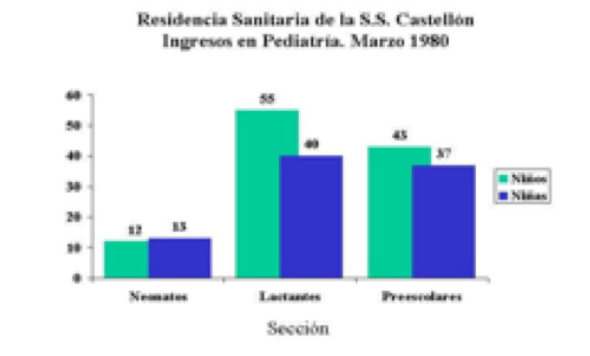


Figura 4: Diagrama de barras parcialmente superpuestas

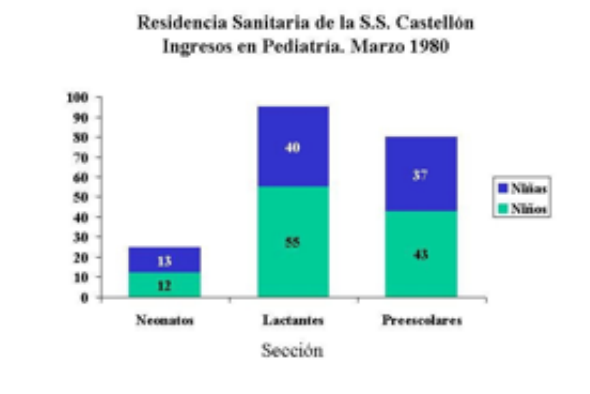


Figura 5: Diagrama de barras mixtas o apiladas

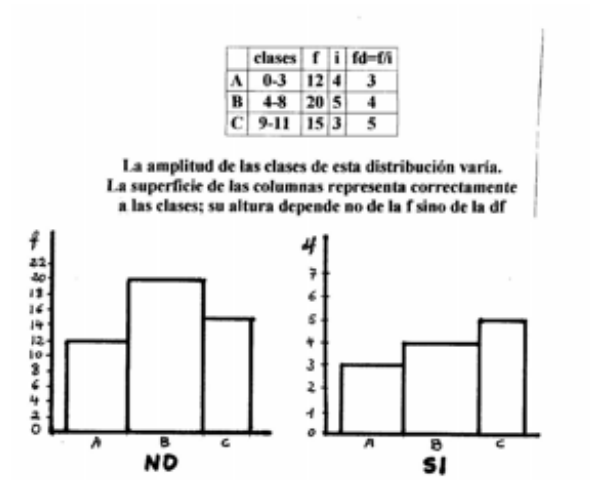


Figura 6: Si no son iguales todas las cla- ses, hay una regla especial

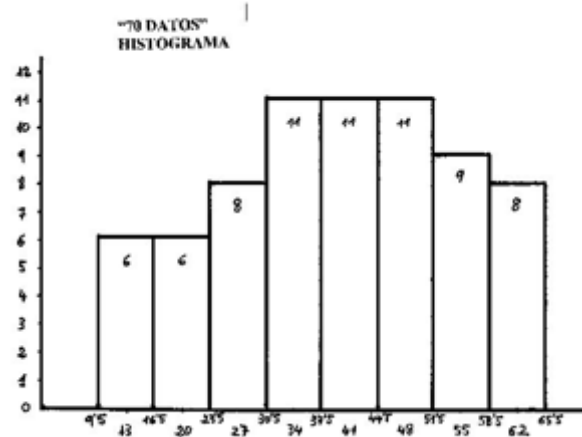


Figura 7: Histograma simple "70 Datos" de la tabla del tema anterior

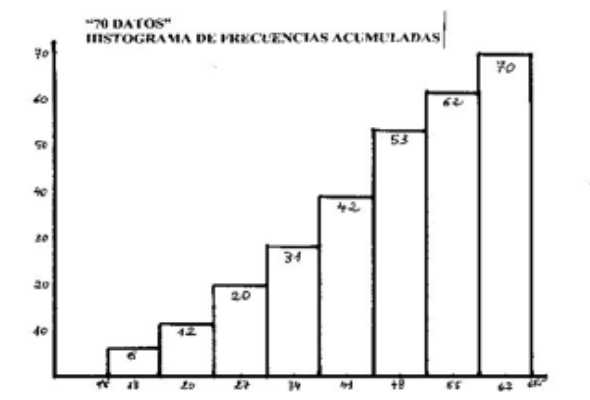


Figura 8: Histograma de frecuencias acumuladas

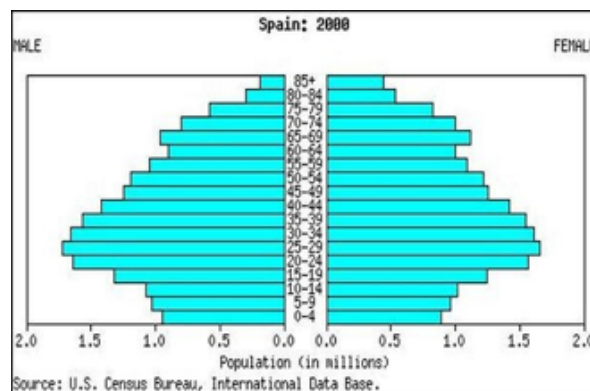


Figura 9: Pirámide de población de España en 2002

proporcional a la frecuencia relativa o porcentaje. Para ello se calcula el ángulo que le corresponde mediante una simple regla de tres. A todo el círculo le corresponden 360° y si es un semicírculo 180°. En el ejemplo de los ingresos en Pediatría:



Figura 10: Pirámide de población de España en 2002

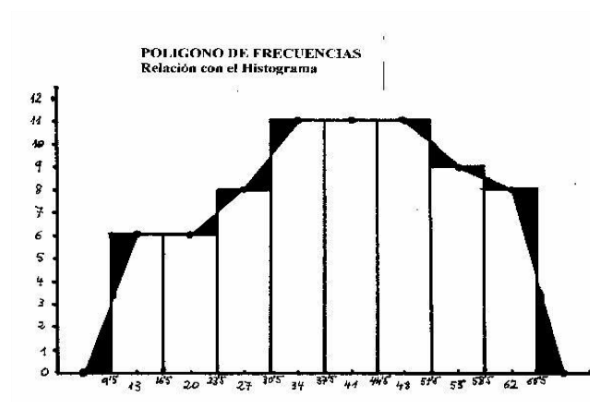


Figura 11: Pirámide de población de España en 2002

$$\frac{\text{al } 100\% \text{ (todos)}}{\text{al } 12,5\%} = \frac{\text{corresponden } 360^{\circ}}{\text{corresponden } x^{\circ}}$$

Por lo tanto  $x^{\circ}=45^{\circ}$ , y así para las otras Secciones se obtiene: Lactantes  $171^{\circ}$  y Preescolares  $144^{\circ}$  Luego mediante un transportador se trazan en el círculo las líneas correspondientes. Los sectores circulares se pueden desgajar del conjunto para que resalten más. (Figura 12 y 13)



Figura 12: Pirámide de población de España en 2002

Residencia Sanitaria de la S.S. Castellón  
Ingresos Pediatría. Marzo 1980

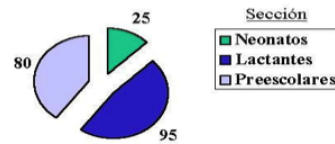


Figura 13: Pirámide de población de España en 2002

Los PICTOGRAMAS utilizan figuras e imágenes de todo tipo, como pilas de monedas, balanzas, coches, muñequitos, mapas distorsionados, etc. Siempre deben respetar el espíritu del gráfico básico. (Figura 14)

La fantasía y la inspiración pueden sugerir OTROS tipos de gráficos. Pero lo esencial no es que sean bonitos, sino que informen bien. Pero si son buenos, bonitos y sencillos, mejor que mejor.

Los gráficos se prestan mucho a la manipulación (no respetando las normas básicas que se han citado) y pueden ofrecer por tanto una información falsa (Figura 15). En este caso se podría decir que una imagen puede mentir más que mil palabras.

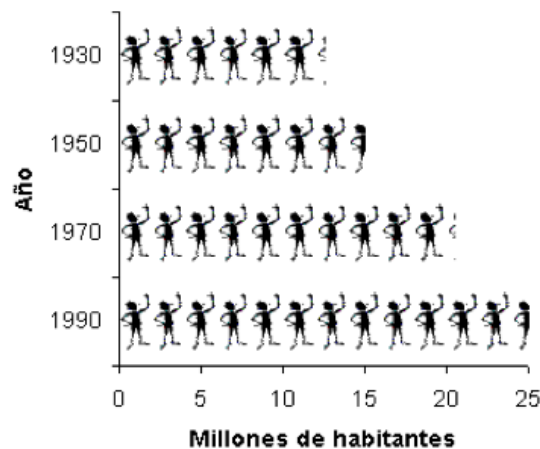


Figura 14: Pirámide de población de España en 2002



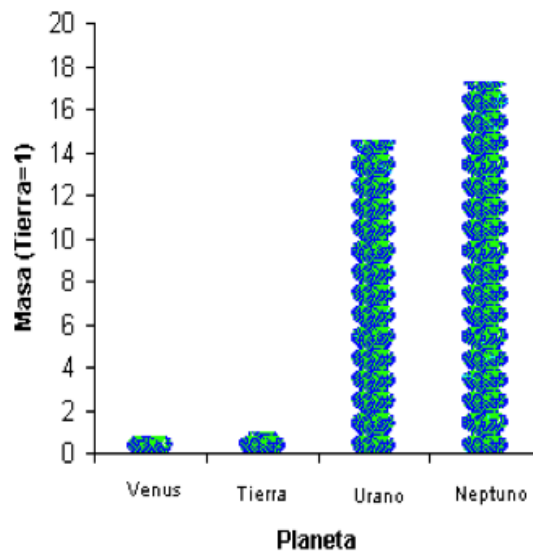


Figura 15: Pirámide de población de España en 2002

### 3. Medidas simples de resumen de datos

#### 3.1. Definición de parámetros e índices

Los parámetros o índices (ya vimos en el tema 3 que consideramos ambos conceptos como equivalentes) son otra forma de presentar resumidos los datos estadísticos. Hay que distinguir:

- parámetros de tendencia central, que informan del centro de la distribución
- parámetros de dispersión, que informan de la dispersión de los datos
- parámetros de posición, que sitúan a los datos en el conjunto de la distribución ordenada. Los más utilizados en Bioestadística son los percentiles. Algunos de ellos pueden ser considerados también como parámetros de tendencia central y otros como de dispersión.
- parámetros de forma, que precisan la forma de la distribución. Podría decirse que expresan numéricamente la forma del histograma.

En este apunte revisaremos únicamente los parámetros de tendencia central, pues los otros son materia de los cursos de bioestadística.

#### 3.2. Parámetros de tendencia central

##### 3.2.1. Media aritmética

La media es la suma de todos los valores dividida por el número de ellos. Se simboliza mediante  $\bar{x}$ . Supongamos que tenemos los datos numéricos 8, 1, 4, 8,

8, 5, 1. Calcularemos la media aritmética. En primer lugar contamos el número de datos que tenemos, para este ejemplo particular son siete, por lo que decimos que le número de datos  $N = 7$ . A continuación calculamos la media:

$$\bar{x} = \frac{\sum x}{N}$$

$$\bar{x} = \frac{8+1+4+8+8+5+1}{7} = \frac{35}{7} = 5$$

Por lo tanto la media aritmética es igual a 5.

### Propiedades de la media

- 1) si a cada valor de  $x$  le sumamos, restamos, multiplicamos o dividimos por una constante, la media queda sumada, restada, multiplicada o dividida por esa constante.
- 2) la media es sensible a la variación de cada valor de  $x$ .
- 3) la media se expresa en la misma unidad de medida que los datos originales.
- 4) si la media tiene decimales es habitual expresarla con uno más que los datos originales.

### 3.2.2. Media aritmética ponderada

Se puede usar una media ponderada para calcular la nota final de un curso, en donde se asigna distinta importancia (peso) a los distintos exámenes que se realicen. Por ejemplo, si las dos primeras pruebas tienen un peso de 30 % y 20 % respectivamente, mientras que la última prueba tiene un peso de 50 % y las calificaciones respectivas son de 6.4, 3.2, 4.0 entonces la nota final corresponde a la siguiente media ponderada <sup>7</sup>:

$$\bar{x} = \frac{6,4*0,30+3,2*0,20+4,0*0,50}{0,30+0,20+0,50} = \frac{1,92+0,64+2,00}{1} = 4,56$$

Es importante destacar que en este ejemplo los porcentajes se expresaron como un número decimal. El 30 % es igual a 0.3, el 20 % es igual a 0.2 y el 50 % es igual a 0.5.

### 3.3. Mediana

La **mediana** es el valor que ocupa el centro de la distribución una vez ordenados los datos. El símbolo es  $M$ .

Sean  $x_1, x_2, x_3, \dots, x_n$  los datos de una muestra ordenada en orden creciente y designando la mediana como  $M_e$ , distinguimos dos casos:

<sup>7</sup>Pág 58-64. Spiegel M. Estadística. Mc Graw Hill 2002.

a) Si  $n$  es impar, la mediana es el valor que ocupa la posición  $(n+1)/2$  una vez que los datos han sido ordenados (en orden creciente o decreciente), porque éste es el valor central. Es decir:  $M_e = x_{(n+1)/2}$ . Por ejemplo, si tenemos 5 datos, que ordenados son:  $x_1 = 3, x_2 = 6, x_3 = 7, x_4 = 8, x_5 = 9 \Rightarrow$  El valor central es el tercero:  $x_{(5+1)/2} = x_3 = 7$ . Este valor, que es la mediana de ese conjunto de datos, deja dos datos por debajo ( $x_1, x_2$ ) y otros dos por encima de él ( $x_4, x_5$ ).

b) Si  $n$  es par, la mediana es la media aritmética de los dos valores centrales. Cuando  $n$  es par, los dos datos que están en el centro de la muestra ocupan las posiciones  $n/2$  y  $n/2+1$ . Es decir:  $M_e = (x_{n/2} + x_{n/2+1})/2$ . Por ejemplo, si tenemos 6 datos, que ordenados son:  $x_1 = 3, x_2 = 6, x_3 = 7, x_4 = 8, x_5 = 9, x_6 = 10 \Rightarrow$  Hay dos valores que están por debajo del  $x_{n/2} = x_3 = 7$  y otros dos que quedan por encima del siguiente dato  $x_{n/2+1} = x_4 = 8$ . Por tanto, la mediana de este grupo de datos es la media aritmética de estos dos datos:  $M_e = \frac{x_3+x_4}{2} = \frac{7+8}{2} = 7,5$ .<sup>8</sup>

---

<sup>8</sup>Pág 58-64. Spiegel M. Estadística. Mc Graw Hill 2002.