Capítulo 1

Estadística descriptiva unidimensional

OBJETIVOS

- Conocer el origen de la estadística y entender las relaciones existentes entre estadística descriptiva, teoría de probabilidad e inferencia.
- Distinguir entre los diferentes tipos de datos que se pueden presentar en un estudio descriptivo: cualitativo y cuantitativo, discreto y continuo.
- Aprender a ordenar los datos creando tablas estadísticas de valores agrupados y sin agrupar, utilizando los distintos tipos de frecuencias.
- Realizar las representaciones gráficas adecuadas a cada tipo de distribución de frecuencias.
- Definir una serie de medidas que sinteticen la información contenida en una distribución de frecuencias unidimensional, tanto de valores agrupados como sin agrupar.
    - Aprender a calcular e interpretar las medidas usuales de posición central y no central.
    - Aprender a calcular e interpretar las medidas usuales de dispersión, para comparar la dispersión entre dos o más variables o distribuciones de frecuencias.
- Estudiar cómo se ven afectadas estas medidas al transformar los datos de una variable.

La palabra **estadística** procede del vocablo *estado*, pues era función principal de los gobiernos de los estados establecer registros de población, nacimientos, defunciones, cosechas, impuestos, etc.

Hoy en día, la mayoría de las personas entienden por estadísticas los conjuntos de datos distribuidos en tablas, gráficos publicados en los diarios, ... En la actualidad se entiende la estadística como un método de toma de decisiones. De ahí que se emplee en multitud de estudios científicos de todas las ramas del saber. Por ejemplo:

- ¿Cómo saber si un nuevo medicamento producirá efectos satisfactorios?

- ¿Cuáles serán las necesidades de puestos médicos para los próximos cinco años?

No quiere decir que gracias a la estadística se pueda contestar a estas preguntas con total exactitud, pero sí es cierto que, mediante procedimientos de inferencia estadística, se puede responder a las cuestiones planteadas con un margen de error prefijado.

La estadística se puede dividir en dos partes:

Estadística descriptiva: trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones. Se construyen tablas y se representan gráficos que permiten simplificar, en gran medida, la complejidad de todos los datos que intervienen en la distribución. Asimismo, se calculan parámetros estadísticos que caracterizan la distribución y se realizan deducciones directamente a partir de los datos y parámetros obtenidos.

Estadística inferencial: plantea y resuelve el problema de establecer previsiones y conclusiones generales sobre una población a partir de los resultados obtenidos de una muestra. Utiliza resultados obtenidos mediante la estadística descriptiva y se apoya en el cálculo de probabilidades.

1.1. Definiciones

Población: conjunto de elementos que cumplen una determinada característica. Los elementos de la población se llaman **individuos**.

Muestra: cualquier subconjunto de la población. El número de elementos de una muestra se llama **tamaño**.

El proceso mediante el cual se extrae una muestra representativa de la población se conoce con el nombre de **muestreo aleatorio**. En este muestreo, cada individuo de la población tiene la misma posibilidad de ser incluido en la muestra. Pero la composición de la muestra debe estar en proporción con la composición de la población. Por ejemplo, si se desea elegir una muestra formada por 1000 personas de una población en la que el 60 % son mujeres, deberemos elegir para la muestra 600 mujeres y 400 hombres.

Carácter estadístico: propiedad que permite clasificar a los individuos de la población. Por ejemplo, nivel de consumo o ahorro, edad, renta, peso, etc. Hay dos tipos de caracteres estadísticos: cuantitativo y cualitativo.

- **Carácter estadístico cuantitativo:** es aquel que se puede medir numéricamente. Este tipo de carácter estadístico determina una variable estadística. Se llaman valores de una variable estadística a los posibles resultados obtenidos al observar dicha variable estadística. Hay dos tipos de variables estadísticas: discreta y continua.

  - **Variable estadística discreta:** cuando puede tomar un número finito de valores. Ejemplos: número de proveedores, número de defectos en la fabricación de un producto, número de nacimientos registrados cada día en una ciudad, etc.
  - **Variable estadística continua:** cuando puede tomar, al menos teóricamente, todos los valores posibles dentro de un cierto intervalo de la recta real. Ejemplos: peso de un individuo, altura de los edificios de una ciudad, temperatura, etc.

  Una variable estadística se representa con letras mayúsculas, X, Y, Z, mientras que sus valores se suelen representar con letras minúsculas: $x_1, x_2, x_3, \ldots, x_n$ para la variable X; $y_1, y_2, y_3, \ldots, y_n$ para la variable Y; $z_1, z_2, z_3, \ldots, z_n$ para Z.

- **Carácter estadístico cualitativo:** es aquel que no se puede medir. Este tipo de carácter estadístico determina un atributo. Se llaman modalidades a los posibles resultados obtenidos al observar un atributo. Hay dos tipos de atributos: nominal y ordinal.

  - **Atributo nominal:** cuando no admite una ordenación natural de sus modalidades. Ejemplos: ciudad natal de una persona, color de los ojos, estado civil, etc.
  - **Atributo ordinal:** cuando admite una ordenación natural de sus modalidades. Ejemplos: nivel de idioma, cargo en una empresa, calificación de un alumno, etc.

  Por ejemplo, para el estado civil, las modalidades serán soltero, casado, separado, divorciado y viudo, mientras que para la calificación de un alumno tendremos suspenso, aprobado, notable, sobresaliente y matrícula de honor.

Por otra parte, se puede hablar de análisis de datos unidimensionales o multidimensionales, dependiendo del número de características a estudiar. En el caso multidimensional, la naturaleza de las características puede ser mixta. Por ejemplo, en distribuciones bidimensionales, puede trabajarse conjuntamente con características cuantitativas y cualitativas.

1.2. Distribución de frecuencias. Tablas estadísticas

Sea X una variable estadística que ha tomado los valores $x_1, \ldots, x_n$, que supondremos ordenados de menor a mayor, en un conjunto de N individuos.

- Se llama **frecuencia absoluta** del valor x_i, y la denotamos por n_i, al número de veces que se repite dicho valor. Se cumplirá, entonces, que $N = n_1 + \ldots + n_n$.

- Se llama **frecuencia absoluta acumulada** del valor x_i, y la denotamos por N_i, al número de individuos que toman un valor menor o igual que x_i:

$$N_i = n_1 + n_2 + \ldots + n_i$$

  Nótese que $N_n = N$.

- Se llama **frecuencia relativa** del valor x_i, y la denotamos por f_i, a la proporción de individuos que han tomado el valor x_i. Es, por tanto, el cociente entre la frecuencia absoluta de x_i y el número total de datos que intervienen en la distribución:

$$f_i = \frac{n_i}{N}$$

  Obsérvese que $f_1 + \ldots + f_n = \frac{n_1}{N} + \ldots + \frac{n_n}{N} = \frac{n_1 + \ldots + n_n}{N} = \frac{N}{N} = 1$.

- Se llama **frecuencia relativa acumulada** del valor x_i, y la denotamos por F_i, a la proporción de individuos que toman un valor menor o igual que x_i. Es, por tanto, el cociente entre la frecuencia absoluta acumulada de x_i y el número total de datos que intervienen en la distribución o, equivalentemente, la suma de las frecuencias relativas correspondientes a los valores menores o iguales que x_i:

$$F_i = \frac{N_i}{N} = f_1 + f_2 + \ldots + f_i$$

  Nótese que $F_n = 1$.

La distribución de frecuencias de una variable viene definida por los valores que toma la variable y sus respectivas frecuencias. Existen distribuciones de frecuencias de valores no agrupados y agrupados.

Veamos cómo proceder ordenadamente para analizar una muestra.

1. Recogida de datos. Consiste en la toma de datos numéricos procedentes de la muestra.

2. Ordenación de los datos. Una vez recogidos los datos, los colocaremos en orden creciente o decreciente, según el tipo de estudio que tengamos que hacer.

3. Recuento de frecuencias. Efectuaremos el recuento de los datos obtenidos.

4. Agrupación de los datos. En caso de que la variable sea continua o bien discreta pero con un número de datos muy grande, es muy aconsejable agrupar los datos en clases o intervalos. Pero, ¿cuál es el número idóneo de clases que debemos escoger a la hora de agrupar? No existe una contestación tajante a esta pregunta; existen incluso varios criterios para dar respuesta a esta cuestión.

   Con carácter muy general, podemos enunciar como uno de los criterios más sencillos el de Norcliffe, que establece que el número de clases debe ser aproximadamente igual a la raíz cuadrada positiva del número de datos.

   Una vez decidido el número de clases, es aconsejable escoger los límites de clase inferior y superior, de modo que se sitúen en números redondos. Los extremos del i-ésimo intervalo se denotan por L_{i-1} y L_i, donde L_{i-1} es su extremo inferior y L_i su extremo superior.

   Para el mejor tratamiento de la información, es más cómodo que los intervalos sean de la misma **amplitud**, la cual se denota por c_i, y se puede obtener como la diferencia entre el extremo superior e inferior del intervalo:

   $$c_i = L_i - L_{i-1}$$

   Si es así, se verificará que el número de clases o intervalos es igual al cociente entre el rango o recorrido y la amplitud. El rango o recorrido es la diferencia entre el mayor y el menor valor de la variable.

   Esta relación nos permitirá deducir el número de intervalos si fijamos la amplitud, o esta si fijamos el número de intervalos. Como ya hemos comentado, en el establecimiento del número de intervalos no existen reglas fijas, porque, en última instancia, dependerá de los objetivos de la investigación.

   Al punto medio de cada clase se le denomina **marca de clase**, y se podrá hallar como

   $$x_i = \frac{L_{i-1} + L_i}{2}$$

   Con el fin de que la clasificación esté bien hecha, los intervalos se deben construir de tal manera que el límite superior de una clase coincida con el límite inferior de la siguiente. Además, se debe adoptar el criterio de que los intervalos sean cerrados por la izquierda y abiertos por la derecha (es decir, de la forma $[L_{i-1}, L_i)$), a excepción del último, que será cerrado por ambos extremos (es decir, de la forma $[L_{n-1}, L_n]$). Puede tomarse como límite inferior del primer intervalo el mínimo valor de la variable o inferior a dicho valor. Fijado este y sumando la amplitud, se obtienen los demás intervalos. Debe comprobarse que el valor máximo de la variable está recogido en el último intervalo.

5. Construcción de la tabla estadística. Las tablas más simples son las que constan de una primera columna, donde se refleja los distintos valores o modalidades que presenta el carácter en estudio. Se añaden una o más columnas a su derecha que representan las respectivas frecuencias. En muchas ocasiones es interesante trabajar con porcentajes, que se obtienen multiplicando las frecuencias relativas por 100.

   Las tablas para organizar la información basada en datos no agrupados y agrupados en intervalos serán, respectivamente,

x_i	n_i	N_i	f_i	F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_i	N_i	f_i	F_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	n_n	N	f_n	1
	N		1	

Intervalos	n_i	N_i	f_i	F_i
$[L_0, L_1)$	n_1	N_1	f_1	F_1
$[L_1, L_2)$	n_2	N_2	f_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_{i-1}, L_i)$	n_i	N_i	f_i	F_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_{n-1}, L_n]$	n_n	N	f_n	1
	N		1	

   Obsérvese que en la última fila de cada tabla, se ha llevado a cabo la suma de las frecuencias (no acumuladas) por columnas, siendo la de las absolutas N y la de las relativas 1, como se comentó al principio de esta sección.

Ejemplo 1.1. *Un profesor tiene anotadas las calificaciones de 30 alumnos. Construir la tabla estadística sabiendo que son las siguientes:*

$$\begin{array}{cccccccccc} 5 & 3 & 4 & 1 & 2 & 8 & 9 & 8 & 7 & 6 \\ 6 & 7 & 9 & 8 & 7 & 7 & 1 & 0 & 1 & 5 \\ 9 & 9 & 8 & 0 & 8 & 8 & 8 & 9 & 5 & 7 \end{array}$$

Notas (x_i)	n_i	N_i	f_i	F_i
0	2	2	2/30	2/30
1	3	5	3/30	5/30
2	1	6	1/30	6/30
3	1	7	1/30	7/30
4	1	8	1/30	8/30
5	3	11	3/30	11/30
6	2	13	2/30	13/30
7	5	18	5/30	18/30
8	7	25	7/30	25/30
9	5	30	5/30	1
	30		1	

Ejemplo 1.2. *Agrupar los siguientes datos, correspondientes a las edades de las personas que acuden a un logopeda a lo largo de un mes, en intervalos de amplitud 5, considerando que el extremo inferior del primer intervalo es 0.*

3 2 11 13 4 3 2 4 5 6 7 3

4 5 3 2 5 6 27 15 4 21 12 4

3 6 29 13 6 17 6 13 6 5 12 26

Intervalos	n_i	N_i	f_i	F_i
[0,5)	13	13	13/36	13/36
[5,10)	11	24	11/36	24/36
[10,15)	6	30	6/36	30/36
[15,20)	2	32	2/36	32/36
[20,25)	1	33	1/36	33/36
[25,30]	3	36	3/36	1
	36		1	

Observemos que en el ejemplo anterior, si bien la variable estadística en cuestión es discreta, toma un número de valores distintos grande, por lo que interesa recopilar la información agrupando los datos en intervalos.

1.3. Representaciones gráficas

Aunque las tablas estadísticas contienen toda la información, a veces es conveniente expresarla mediante un gráfico, con el fin de hacerla más clara y evidente. Según como sea la naturaleza del carácter estudiado, utilizaremos un tipo u otro de representación gráfica.

1. Para fenómenos cuantitativos de tipo discreto.

   - **Diagrama de barras**. Son útiles cuando se desean comparar datos cualitativos o datos cuantitativos de tipo discreto. Para construirlo se llevan los valores de la variable sobre el eje de abscisas y sobre cada valor de la variable se levanta un segmento igual a la frecuencia (absoluta o relativa).

     Ejemplo 1.3. *El diagrama de barras para los datos del Ejemplo 1.1 se puede ver en la Figura 1.1 y el acumulativo en la Figura 1.2.*

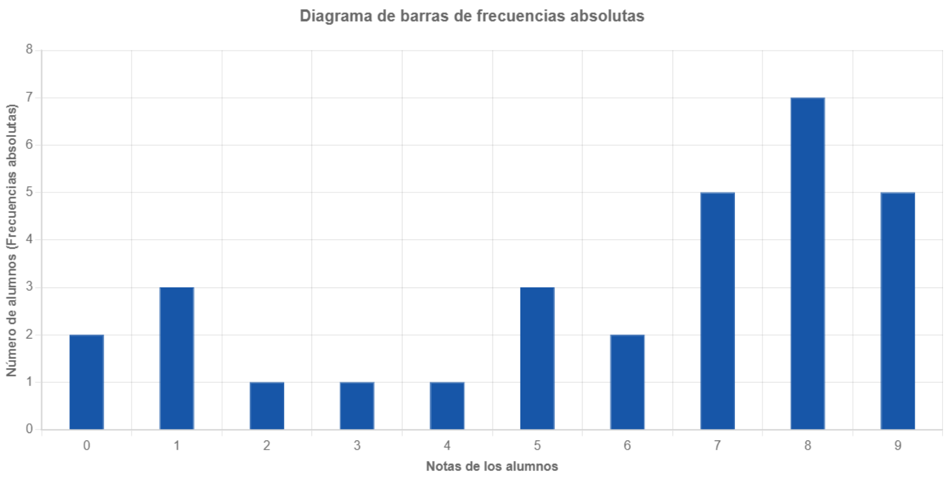

Figura 1.1: Diagrama de barras para los datos del Ejemplo 1.1

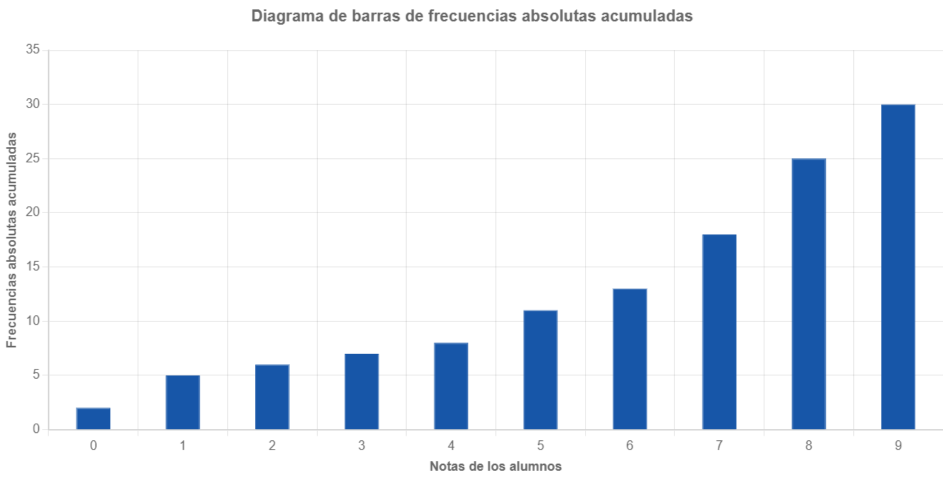

Figura 1.2: Diagrama de barras acumulativo para los datos del Ejemplo 1.1

- **Polígono de frecuencias** (absolutas o relativas). Se forma uniendo los extremos de las barras mediante una línea quebrada.

  Ejemplo 1.4. *El polígono de frecuencias absolutas para los datos del Ejemplo 1.1 se puede ver en la Figura 1.3 y el acumulativo en la Figura 1.4.*

Figura 1.3: Polígono de frecuencias absolutas para los datos del Ejemplo 1.1

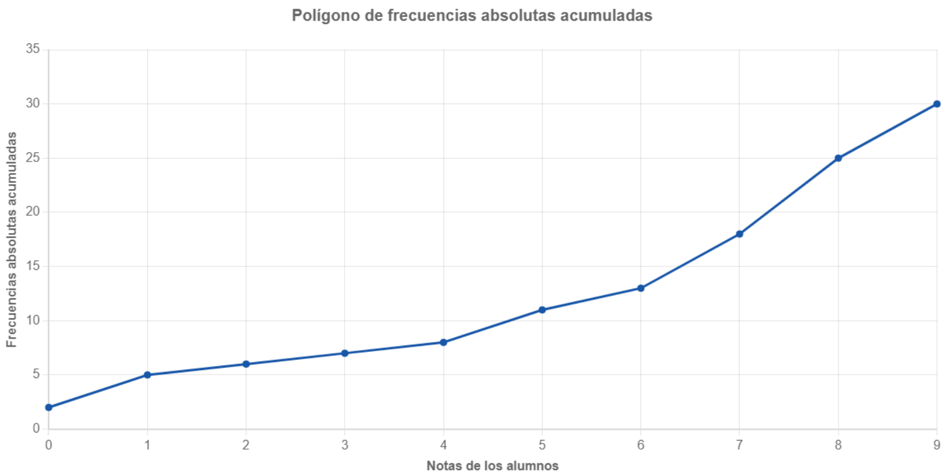

Figura 1.4: Polígono de frecuencias acumuladas para los datos del Ejemplo 1.1

2. Para fenómenos cuantitativos de tipo continuo o para fenómenos de tipo discreto con un gran número de valores distintos.

   - **Histograma**. Se utiliza generalmente para distribuciones de variable estadística continua, y para distribuciones de variable estadística discreta, con un gran número de datos, y que se han agrupado en clases. Para construir el histograma se representan sobre el eje de abscisas los límites de las clases. Sobre dicho eje se construyen unos rectángulos que tienen por base la amplitud del intervalo y por altura la frecuencia absoluta de cada intervalo, siempre que todos los intervalos tengan igual amplitud.

Ejemplo 1.5. *El histograma para los datos del Ejemplo 1.2 se puede ver en la Figura 1.5.*

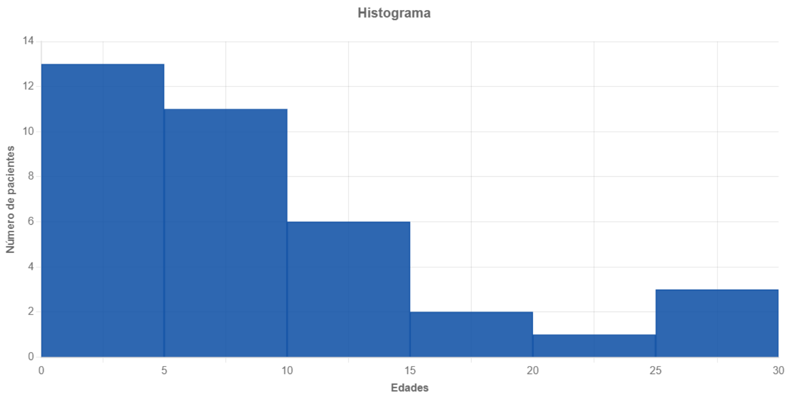

Figura 1.5: Histograma para los datos del Ejemplo 1.2

En caso de que los intervalos tengan distinta amplitud, las alturas de los rectángulos han de ser calculadas teniendo en cuenta que sus áreas deben ser proporcionales a las frecuencias de cada intervalo. De esta forma, las alturas de cada rectángulo se calcularían como el cociente entre la frecuencia absoluta y la amplitud de cada intervalo: $d_i = \frac{n_i}{c_i}$. Este cociente se llama **densidad de frecuencia**.

- **Polígono de frecuencias**. Se forma al unir los puntos medios de cada intervalo, a una altura proporcional a la frecuencia.

Ejemplo 1.6. *El polígono de frecuencias para los datos del Ejemplo 1.2 se puede ver en la Figura 1.6.*

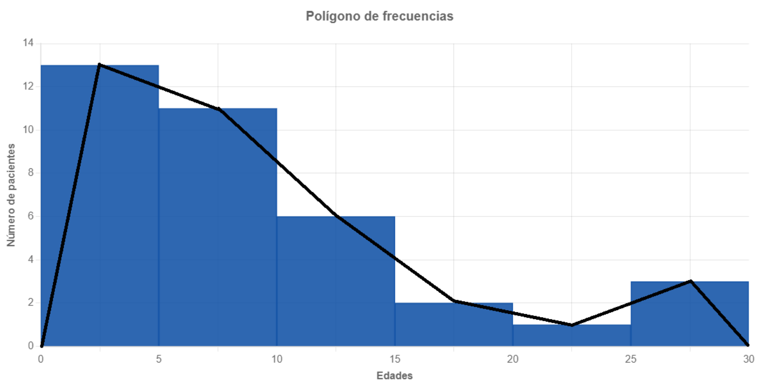

Figura 1.6: Polígono de frecuencias para los datos del Ejemplo 1.2

3. Para fenómenos cualitativos.

- **Diagrama de sectores**. Representa las distintas modalidades de un carácter mediante sectores circulares. El ángulo central de cada sector ha de ser proporcional a la frecuencia absoluta correspondiente. No es aconsejable representar un gran número de categorías, ni representar atributos ordinales.

  Ejemplo 1.7. *Se realiza una encuesta sobre el tipo de comida preferido por los individuos de una determinada población, resultando que el 60,31% prefiere verdura, el 15,06% pescado, el 11,77% carne y el resto otros productos. Pues bien, el diagrama de sectores correspondiente (ver Figura 1.7) se construye en base a los ángulos (en grados) que resultan de multiplicar cada frecuencia absoluta por 360. Así, por ejemplo, al sector en color rojo (verdura) le corresponde un ángulo de 0,6031·360=217,116°.*

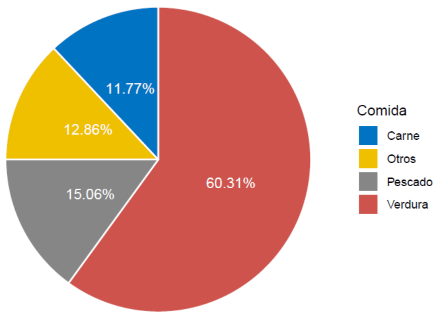

Figura 1.7: Diagrama de sectores

- **Pictograma**. Son dibujos alusivos a la distribución que se pretende estudiar y que mediante su forma, tamaño, etc., ofrecen una descripción lo más expresiva posible de la distribución estadística. Se comprende que este método tiene el grave inconveniente de la falta de precisión.

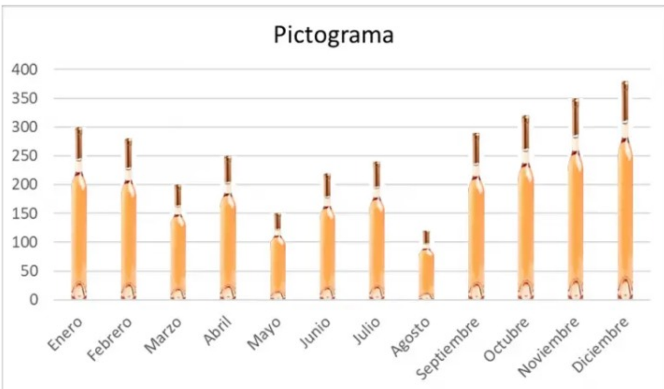

Figura 1.8: Pictograma

- **Cartograma**. Son los gráficos que se realizan sobre un mapa, señalando sobre determinadas zonas con distintos colores o rayados lo que se trata de poner de manifiesto. Se suelen utilizar estos tipos de diagramas para representar la densidad demográfica de una nación, la renta per cápita, los índices de lluvia de una nación, etc.

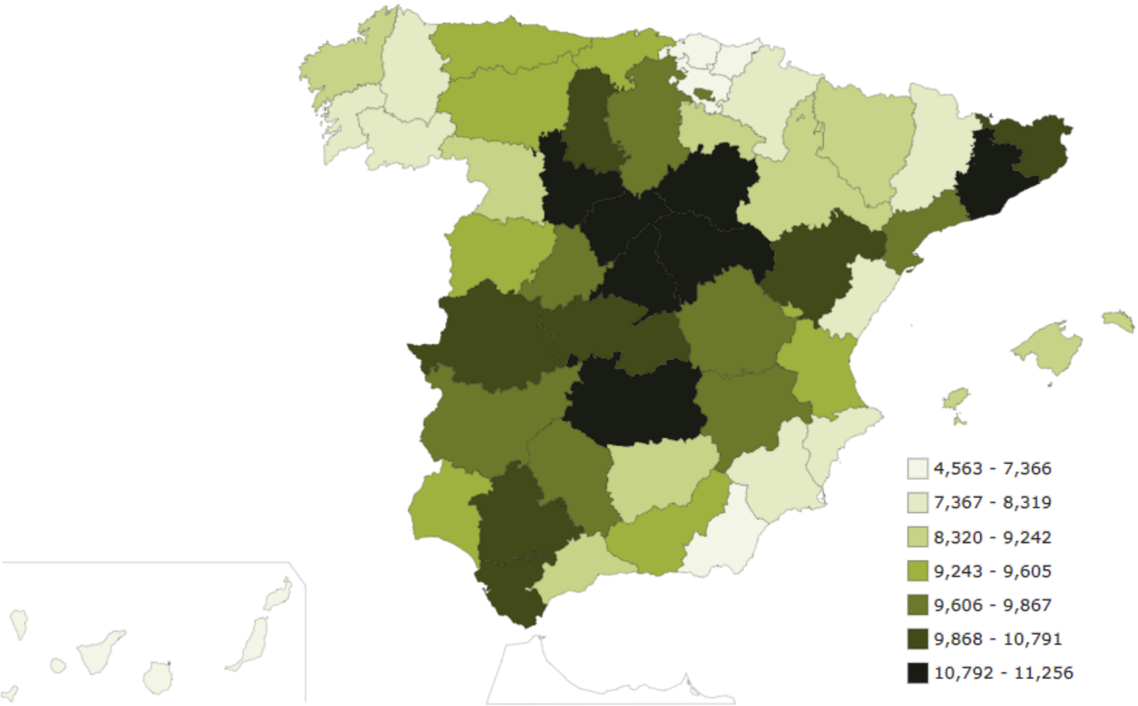

Figura 1.9: Cartograma sobre tanto por mil de personas que se llaman Cristina en España

1.4. Síntesis numérica de los datos

Aún cuando las tablas estadísticas y las representaciones gráficas permiten obtener, de una manera rápida, una idea aproximada del comportamiento de una distribución, lo que se intenta es resumirla en una serie de expresiones, que intentan representar el conjunto total de datos mediante un solo valor numérico. En este proceso de síntesis surgen distintas medidas.

1.4.1. Medidas de posición central (o de centralización)

Son medidas que tienden a situarse hacia el centro del conjunto de datos ordenados.

• La **media aritmética** (o **promedio**) de una variable estadística es la suma de todos los valores de dicha variable dividido por el número de valores. Se denota por $\bar{x}$.

Cálculo de la media aritmética:

Si la variable X toma los valores $x_1, \ldots, x_n$ con frecuencias absolutas respectivas $n_1, \ldots, n_n$, entonces

$$\bar{x} = \frac{x_1 n_1 + \ldots + x_n n_n}{N} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \sum_{i=1}^{n} x_i f_i$$

Si los datos están agrupados en intervalos, se toma como x_i la marca de clase.

Propiedades:

- Respeta los cambios de origen, ya que si a todos los valores de una variable les sumamos una constante, la media aritmética queda aumentada también en esa constante. Matemáticamente, si $Y = X + b$, entonces $\bar{y} = \bar{x} + b$ para todo $b \in \mathbb{R}$.
- Respeta los cambios de escala, es decir, si todos los valores de una variable se multiplican por una constante, la media aritmética queda multiplicada también por esa constante. Matemáticamente, si $Y = aX$, entonces $\bar{y} = a\bar{x}$ para todo $a \in \mathbb{R}$.
- Si $\bar{x}_i$ es la media de cada uno de k grupos $(i = 1, \ldots, k)$ de tamaño N_i, entonces
$$\bar{x} = \frac{\bar{x}_1 N_1 + \ldots + \bar{x}_k N_k}{N_1 + \ldots + N_k}$$
es la media del conjunto de los individuos de los k grupos.
- La suma de las desviaciones de los valores de la variable respecto a su media es cero:
$$\sum_{i=1}^{n} (x_i - \bar{x}) n_i = 0$$

Observaciones:

- La media aritmética es la medida de centralización que más se utiliza.
- Presenta la ventaja de tener en cuenta todos los datos de la distribución, además de resultar muy sencillo su cálculo.
- Tiene el grave el inconveniente de que si la distribución posee valores extremos, estos producen una distorsión sobre el valor de $\bar{x}$.
- No siempre es posible realizar el cálculo de la media: cuando los datos son cualitativos o cuando los datos se encuentran agrupados en clases, estando alguna de ellas abierta.

  En estos casos en los que no es posible calcular la media, se utilizan otras medidas, como la moda y la mediana.

Ejemplo 1.8. *La siguiente tabla recopila el número de veces que acuden al gimnasio, a la semana, los clientes que ingresaron el pasado mes:*

Veces que asiste al gimnasio	*1*	*2*	*3*	*4*
Número de individuos	*7*	*21*	*21*	*11*

Calcular el promedio de veces que van estos individuos al gimnasio por semana.

Solución. El promedio de veces por semana que acuden al gimnasio estos individuos se puede hallar como

$$\bar{x} = \frac{1 \cdot 7 + 2 \cdot 21 + 3 \cdot 21 + 4 \cdot 11}{60} = 2{,}6 \text{ veces}$$

■

Ejemplo 1.9. *Una encuesta realizada a los individuos que se apuntaron a cierto gimnasio el pasado mes recoge el tiempo que invirtieron, en minutos, en la última sesión de spinning realizada. Los resultados son los siguientes:*

Tiempo de spinning	*[0,10)*	*[10,24)*	*[24,50)*	*[50,60]*
Número de individuos	*6*	*21*	*26*	*7*

Calcular la media aritmética de los tiempos dedicados a la sesión de spinning.

Solución. La media aritmética de la variable estadística "minutos de la última sesión de spinning" se hallará teniendo en cuenta las marcas de clase. Estas vienen dadas por

$$x_1 = \frac{0+10}{2} = 5 \quad x_2 = \frac{10+24}{2} = 17 \quad x_3 = \frac{24+50}{2} = 37 \quad x_4 = \frac{50+60}{2} = 55$$

por lo que el promedio es

$$\bar{x} = \frac{5 \cdot 6 + 17 \cdot 21 + 37 \cdot 26 + 55 \cdot 7}{60} = 28{,}9 \text{ minutos}$$

■

Ejercicio 1.1. *El número de horas que dedica un sanitario a su consulta privada durante la semana es el siguiente: 3,5, 5,5, 4, 6, 5, 3. Calcular la media de horas dedicadas a su consulta. (Sol. 4,5 horas)*

Ejercicio 1.2. *Las calificaciones en la asignatura de bioestadística de los cuarenta alumnos de una clase vienen dadas por la siguiente tabla. Hallar la calificación media. (Sol. 5,3 puntos)*

Calificaciones	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Número de alumnos	*2*	*2*	*4*	*5*	*8*	*9*	*3*	*4*	*3*

Ejercicio 1.3. *Se ha aplicado un test sobre satisfacción en el trabajo a 88 sanitarios de un centro hospitalario, obteniéndose los siguientes resultados. Calcular la puntuación media. (Sol. 59,14 puntos)*

Puntuaciones	*38-44*	*44-50*	*50-56*	*56-62*	*62-68*	*68-74*	*74-80*
Número de sanitarios	*7*	*8*	*15*	*25*	*18*	*9*	*6*

La **moda** es el valor de la variable que presenta mayor frecuencia absoluta (o relativa). La moda no tiene por qué ser única; puede haber varios valores de la variable con la mayor frecuencia. En este caso se dirá que la distribución es bimodal, trimodal, ..., según sean dos, tres, ..., los valores de la variable que presentan mayor frecuencia.

Cálculo de la moda:

Como consecuencia de la definición, el cálculo de la moda resulta muy sencillo en los casos de datos simples y datos no agrupados en intervalos. En el caso de datos agrupados en intervalos, determinamos el intervalo modal, que es el de mayor densidad de frecuencia, y la moda se calcula como

$$Mo = L_{i-1} + c_i \frac{d_{i+1}}{d_{i-1} + d_{i+1}}$$

teniendo en cuenta que:

L_{i-1} es el extremo inferior de la clase modal.

c_i es la amplitud del intervalo modal.

d_{i+1} es la densidad de frecuencia del intervalo siguiente al modal.

d_{i-1} es la densidad de frecuencia del intervalo anterior al modal.

Se considera que $d_0 = 0$ y $d_{n+1} = 0$.

Si todos los intervalos tienen la misma amplitud, entonces $c_i = c$ para todo i, y la fórmula se puede escribir como

$$Mo = L_{i-1} + c \frac{n_{i+1}}{n_{i-1} + n_{i+1}}$$

donde

n_{i+1} es la frecuencia absoluta del intervalo siguiente al modal.

n_{i-1} es la frecuencia absoluta del intervalo anterior al modal.

Se considera que $n_0 = 0$ y $n_{n+1} = 0$.

Observaciones:

- Puede ocurrir que existan distribuciones que no tengan moda. Eso ocurre cuando las frecuencias de todos los datos son iguales.