Sesión 3

Variable cualitativa: Aquella cuyas modalidades son de tipo nominal.

Variable cuasicuantitativa: Modalidades de tipo nominal, en las que existe un orden.

Variable cuantitativa discreta: Sus modalidades son valores enteros.

Variable cuantitativa continua: Sus modalidades son valores reales.

1.4.2. Tablas estadísticas

Consideremos una población estadística de n individuos, descrita según un carácter o variable C cuyas modalidades han sido agrupadas en un número k de clases, que denotamos mediante c_1 , c_2, ..., c_k. Para cada una de las clases c_i, $i = 1, \ldots, k$, introducimos las siguientes magnitudes:

Frecuencia absoluta de la clase c_i es el número n_i, de observaciones que presentan una modalidad perteneciente a esa clase.

Frecuencia relativa de la clase c_i es el cociente f_i, entre las frecuencias absolutas de dicha clase y el número total de observaciones, es decir

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

Obsérvese que f_i es el *tanto por uno* de observaciones que están en la clase c_i. Multiplicado por $100\,\%$ representa el porcentaje de la población que comprende esa clase.

Frecuencia absoluta acumulada N_i, se calcula sobre variables cuantitativas o cuasicuantitativas, y es el número de elementos de la población cuya modalidad es inferior o equivalente a la modalidad c_i:

$$N_i = n_1 + n_2 + \ldots + n_i = \sum_{j=1}^{i} n_j$$

Frecuencia relativa acumulada , F_i, se calcula sobre variables cuantitativas o cuasicuantitativas, siendo el tanto por uno de los elementos de la población que están en alguna de las clases y que presentan una modalidad inferior o igual a la c_i, es decir,

$$F_i = \frac{N_i}{n} = \frac{n_1 + \ldots + n_i}{n} = f_1 + \ldots + f_i = \sum_{j=1}^{i} f_j$$

Llamaremos **distribución de frecuencias** al conjunto de clases junto a las frecuencias correspondientes a cada una de ellas. Una **tabla estadística**

sirve para presentar de forma ordenada las distribuciones de frecuencias. Su forma general es la siguiente:

Modali. C	Frec. Abs. n_i	Frec. Rel. f_i	Frec. Abs. Acumu. N_i	Frec. Rel. Acumu. F_i
c_1	n_1	$f_1 = \frac{n_1}{n}$	$N_1 = n_1$	$F_1 = \frac{N_1}{n} = f_1$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
c_j	n_j	$f_j = \frac{n_j}{n}$	$N_j = n_1 + \ldots + n_j$	$F_j = \frac{N_j}{n} = f_1 + \ldots + f_j$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
c_k	n_k	$f_k = \frac{n_k}{n}$	$N_k = n$	$F_k = 1$
	n	1		

Ejemplo de cálculo con frecuencias

Calcular los datos que faltan en la siguiente tabla:

$l_{i-1} — l_i$	n_i	f_i	N_i
$0 — 10$	60	f_1	60
$10 — 20$	n_2	$0{,}4$	N_2
$20 — 30$	30	f_3	170
$30 — 100$	n_4	$0{,}1$	N_4
$100 — 200$	n_5	f_5	200
	n		

Solución:

Sabemos que la última frecuencia acumulada es igual al total de observaciones, luego $n = 200$.

Como $N_3 = 170$ y $n_3 = 30$, entonces

$$N_2 = N_3 - n_3 = 170 - 30 = 140.$$

Además al ser $n_1 = 60$, tenemos que

$$n_2 = N_2 - n_1 = 140 - 60 = 80.$$

Por otro lado podemos calcular n_4 teniendo en cuenta que conocemos la frecuencia relativa correspondiente:

$$f_4 = \frac{n_4}{n} \qquad \Longrightarrow \qquad n_4 = f_4 \cdot n = 0,1 \times 200 = 20$$

Así:

$$N_4 = n_4 + N_3 = 20 + 170 = 190.$$

Este último cálculo nos permite obtener

$$n_5 = N_5 - N_4 = 200 - 190 = 10.$$

Al haber calculado todas las frecuencias absolutas, es inmediato obtener las relativas:

$$
\begin{aligned}
f_1 &= \frac{n_1}{n} = \frac{60}{200} = 0,3 \\
f_3 &= \frac{n_3}{n} = \frac{30}{200} = 0,15 \\
f_5 &= \frac{n_5}{n} = \frac{10}{200} = 0,05
\end{aligned}
$$

Escribimos entonces la tabla completa:

$l_{i-1} — l_i$	n_i	f_i	N_i
0 — 10	60	**0,3**	60
10 — 20	**80**	0,4	**140**
20 — 30	30	**0,15**	170
30 — 100	**20**	0,1	**190**
100 — 200	**10**	**0,05**	200
	200		

1.5. Representaciones Gráficas

Hemos visto que la tabla estadística resume los datos que disponemos de una población, de forma que ésta se puede analizar de una manera más sistemática y resumida . Para darnos cuenta *de un sólo vistazo* de las características de la población resulta aún más esclarecedor el uso de gráficos y diagramas, cuya construcción abordamos en esta sección.

1.5.1. Gráficos para variables cualitativas

Los gráficos más usuales para representar variables de tipo nominal son los siguientes:

Diagramas de barras: Siguiendo la figura 1.1, representamos en el eje de ordenadas las modalidades y en abscisas las frecuencias absolutas o bien, las frecuencias relativas. Si, mediante el gráfico, se intenta comparar varias poblaciones entre sí, existen otras modalidades, como las mostradas en la figura 1.2. Cuando los tamaños de las dos poblaciones son diferentes, es conveniente utilizar las frecuencias relativas, ya que en otro caso podrían resultar engañosas.

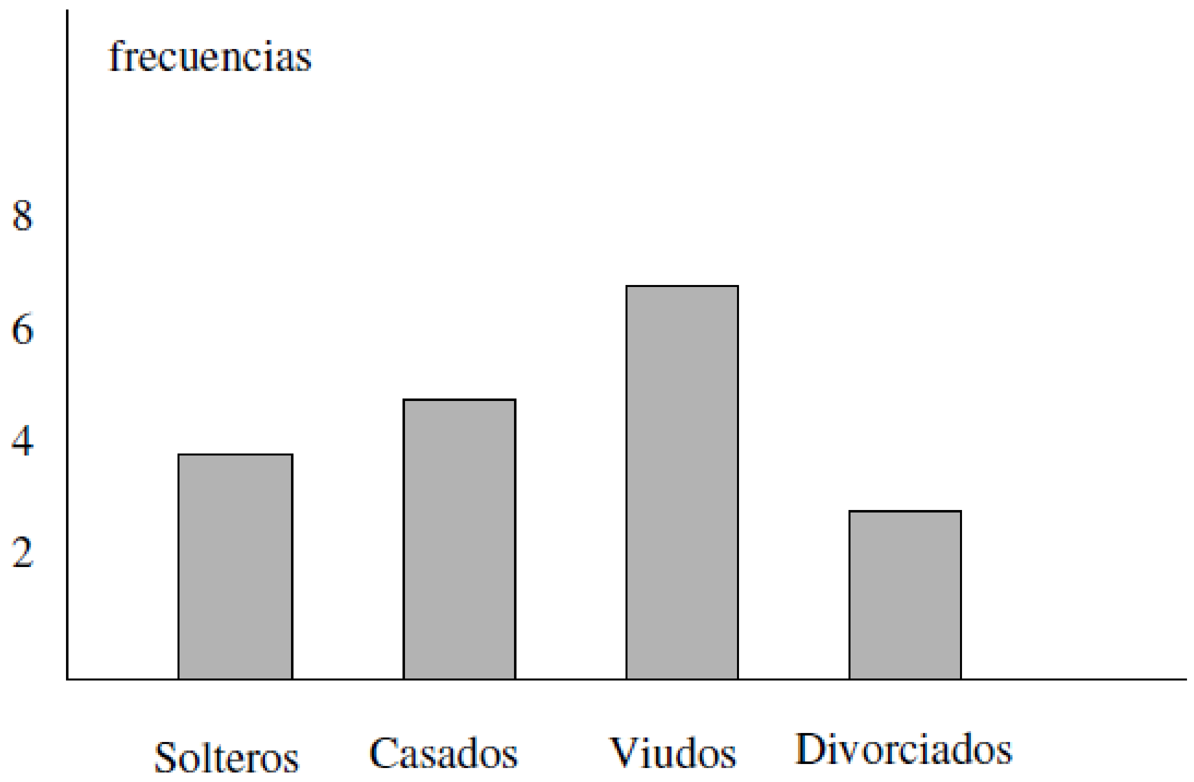


Figura 1.1: Diagrama de barras para una variable cualitativa.

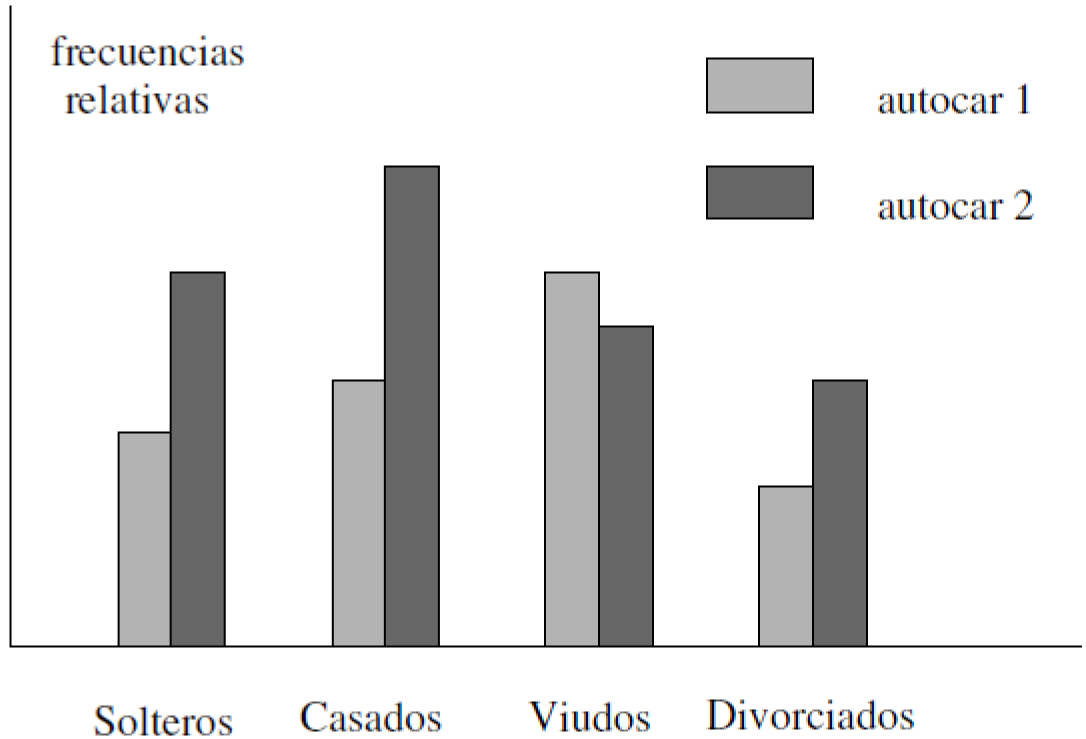


Figura 1.2: Diagramas de barras para comparar una variable cualitativa en diferentes poblaciones. Se ha de tener en cuenta que la altura de cada barra es *proporcional* al número de observaciones (frecuencias relativas).

Diagramas de sectores (también llamados *tartas*). Se divide un círculo en tantas porciones como clases existan, de modo que a cada clase le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa (figura 1.3).

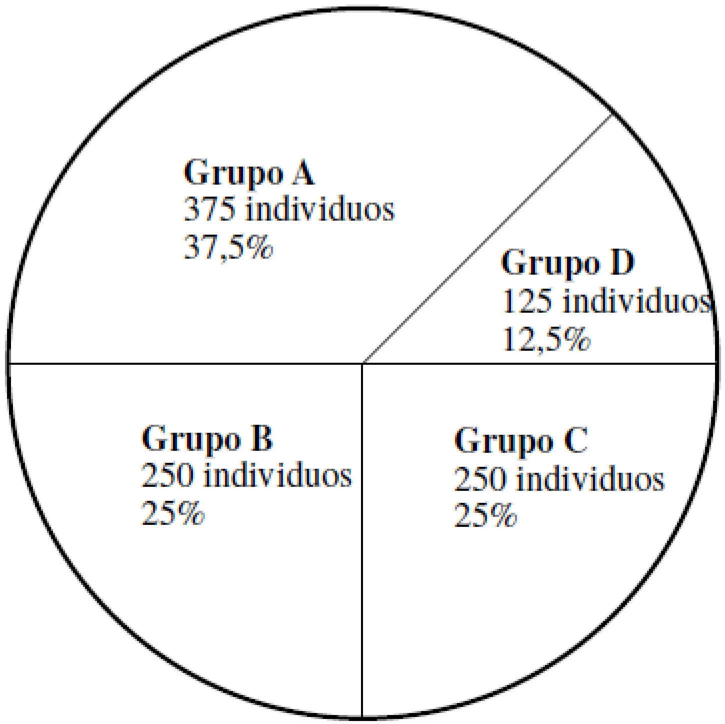


Figura 1.3: Diagrama de sectores.

El arco de cada porción se calcula usando la *regla de tres*:

$$\begin{array}{ccl} n & \longrightarrow & 360^\circ \\ n_i & \longrightarrow & x_i = \dfrac{360 \cdot n_i}{n} \end{array}$$

Como en la situación anterior, puede interesar comparar dos poblaciones. En este caso también es aconsejable el uso de las frecuencias relativas (porcentajes) de ambas sobre gráficos como los anteriores. Otra posibilidad es comparar las 2 poblaciones usando para cada una de ellas un diagrama semicircular, al igual que en la figura 1.4. Sean $n_1 \leq n_2$ los tamaños respectivos de las 2 poblaciones. La población más pequeña se representa con un semicírculo de radio r_1 y la mayor con otro de radio r_2.

La relación existente entre los radios, es la que se obtiene de suponer que la relación entre las areas de las circunferencias es igual a la de los tamaños de las poblaciones respectivas, es decir:

$$\frac{r_2^2}{r_1^2} = \frac{n_2}{n_1} \Longleftrightarrow r_2 = r_1 \cdot \sqrt{\frac{n_2}{n_1}}$$

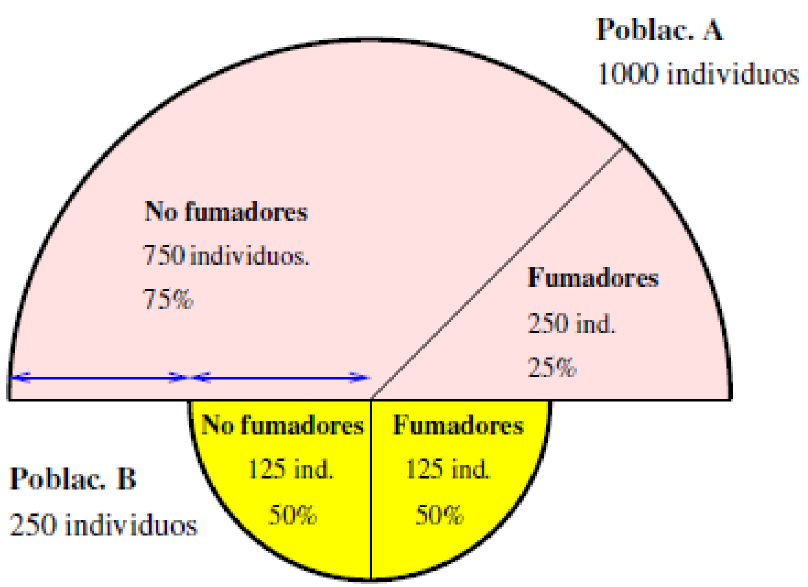


Figura 1.4: Diagrama de sectores para comparar dos poblaciones

Pictogramas Expresan con dibujos alusivo al tema de estudio las frecuencias de las modalidades de la variable. Estos gráficos se hacen representado a diferentes escalas un mismo dibujo, como vemos en la figura 1.5.

El escalamiento de los dibujos debe ser tal que el *área*[1] de cada uno de ellos sea proporcional a la frecuencia de la modalidad que representa. Este tipo de gráficos suele usarse en los medios de comunicación, para que sean comprendidos por el público no especializado, sin que sea necesaria una explicación compleja.

---

[1]Es un error hacer la representación con una escala tal que el *perímetro* del dibujo sea proporcional a la frecuencia, ya que a frecuencia doble, correspondería un dibujo de área cuadruple, lo que da un efecto visual engañoso.

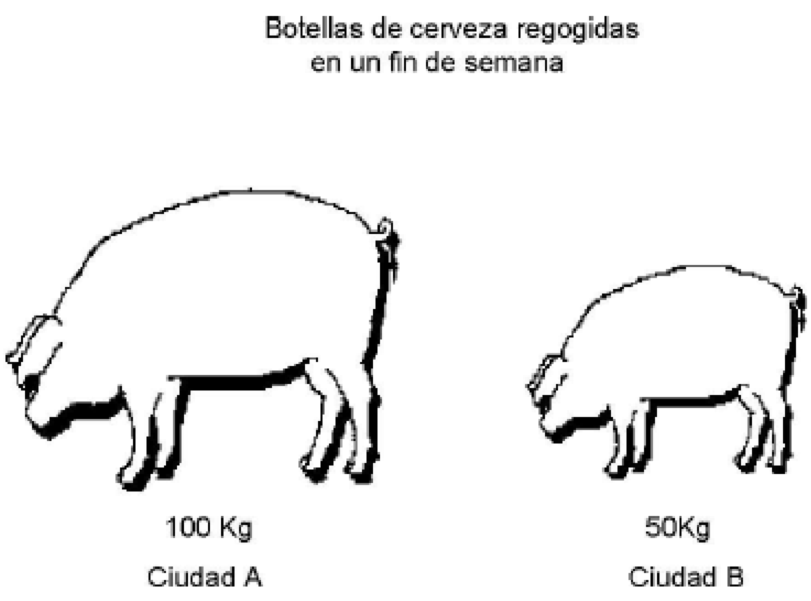


Figura 1.5: Pictograma. Las áreas son proporcionales a las frecuencias.

1.5.2. Gráficos para variables cuantitativas

Para las variables cuantitativas, consideraremos dos tipos de gráficos, en función de que para realizarlos se usen las frecuencias (absolutas o relativas) o las frecuencias acumuladas:

Diagramas diferenciales: Son aquellos en los que se representan frecuencias absolutas o relativas. En ellos se representa el número o porcentaje de elementos que presenta una modalidad dada.

Diagramas integrales: Son aquellos en los que se representan el número de elementos que presentan una modalidad inferior o igual a una dada. Se realizan a partir de las frecuencias acumuladas, lo que da lugar a gráficos crecientes, y es obvio que este tipo de gráficos no tiene sentido para variables cualitativas.

Según hemos visto existen dos tipos de variables cuantitativas: discretas y continuas. Vemos a continuación las diferentes representaciones gráficas que pueden realizarse para cada una de ellas así como los nombres específicos que reciben.

Gráficos para variables discretas

Cuando representamos una variable discreta, usamos el **diagrama de barras** cuando pretendemos hacer una gráfica diferencial. Las barras deben ser estrechas para representar el que los valores que toma la variable son discretos. El diagrama integral o acumulado tiene, por la naturaleza de la variable, forma de escalera. Un ejemplo de diagrama de barras así como su diagrama integral correspondiente están representados en la figura 1.6.

Ejemplo de variable discreta

Se lanzan tres monedas al aire en 8 ocasiones y se contabiliza el número de caras, X, obteniendose los siguientes resultados:

$$2,1,0,1,3,2,1,2$$

Representar gráficamente el resultado.

Solución: En primer lugar observamos que la variable X es cuantitativa discreta, presentando las modalidades: *0,1,2,3*

Ordenamos a continuación los datos en una tabla estadística, y se representa la misma en la figura 1.6.

x_i	n_i	f_i	N_i	F_i
0	1	$1/8$	1	$1/8$
1	3	$3/8$	4	$4/8$
2	3	$3/8$	7	$7/8$
3	1	$1/8$	8	$8/8$
	$n = 8$	1		

Ejemplo de regresentación gráfica

Clasificadas 12 familias por su número de hijos se obtuvo:

Número de hijos (x_i)	1	2	3	4
Frecuencias (n_i)	1	3	5	3

x_i	n_i
0	1
1	3
2	3
3	1

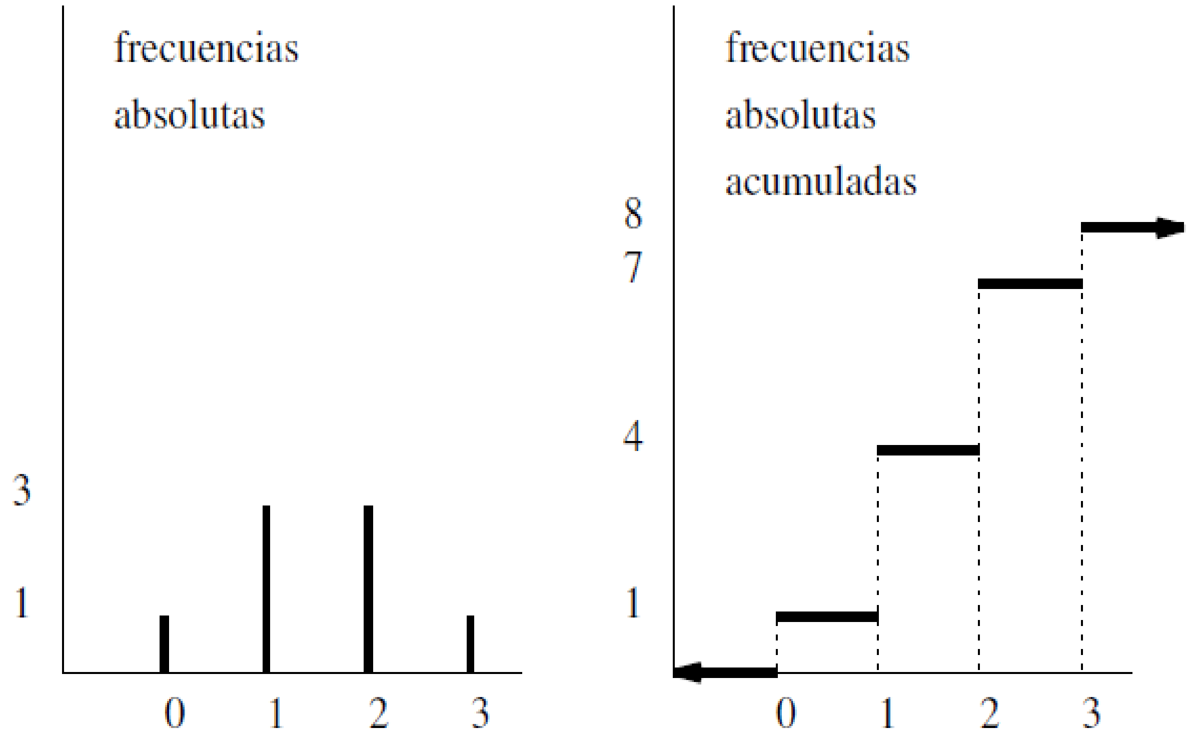


Figura 1.6: Diagrama diferencial (barras) e integral para una variable discreta. Obsérvese que el diagrama integral (creciente) contabiliza el número de observaciones de la variable inferiores o iguales a cada punto del eje de abcisas.

Comparar los diagramas de barras para frecuencias absolutas y relativas. Realizar el diagrama acumulativo creciente.

Variable x_i	F. Absolutas n_i	F. Relativas f_i	F. Acumuladas N_i
1	1	0,083	1
2	3	0,250	4
3	5	0,416	9
4	3	0,250	12
	12	1	

Con las columnas relativas a x_i y n_i realizamos el diagrama de barras para frecuencias absolutas, lo que se muestra en la figura 1.7. Como puede verse es identico (salvo un cambio de escala en el eje de ordenadas) al diagrama de barras para frecuencias relativas y que ha sido calculado usando las columnas de x_i y f_i. El diagrama escalonado (acumulado) se ha construido con la información procedente de las columnas x_i y N_i.

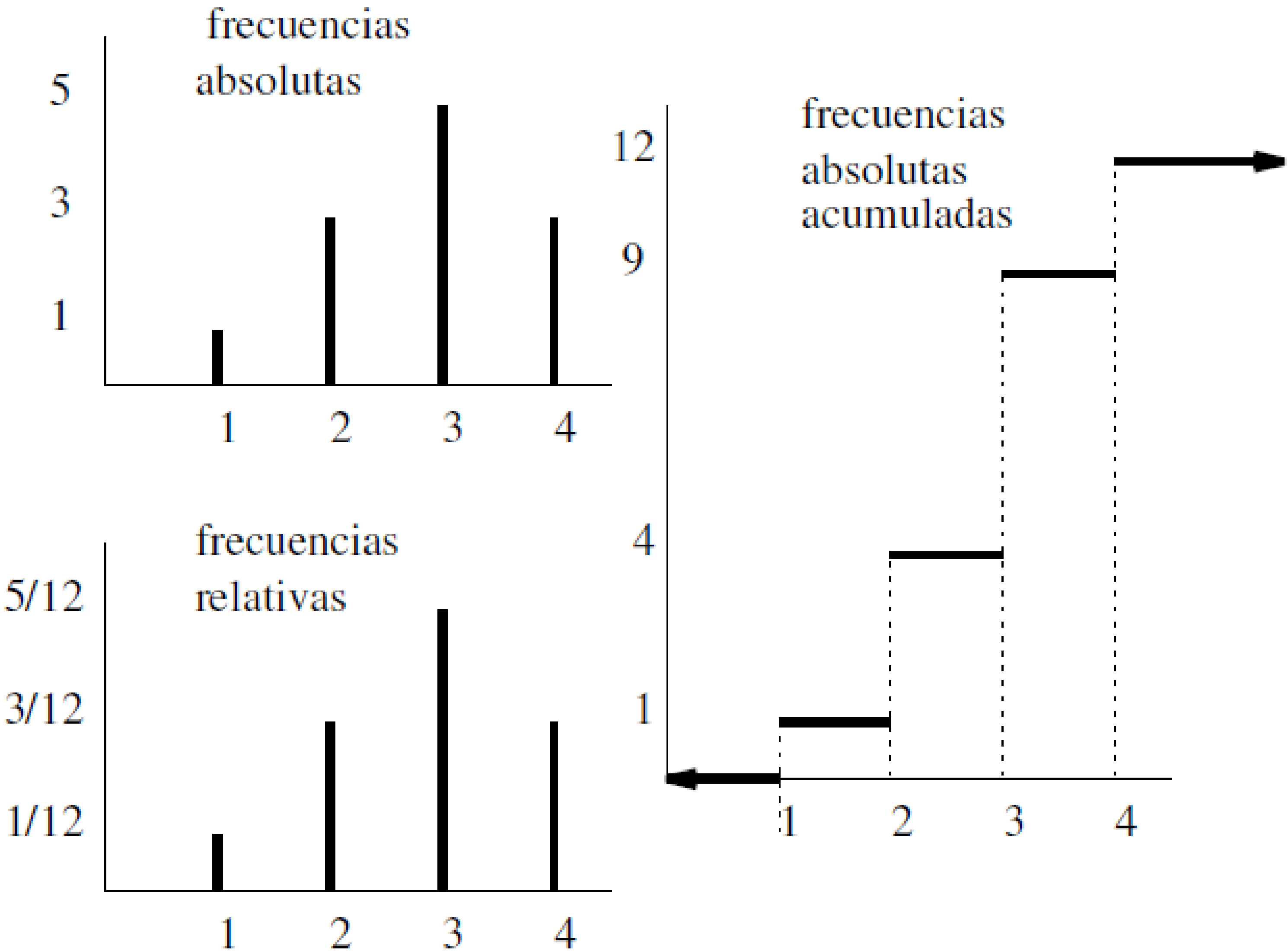


Figura 1.7: Diagramas de frecuencias para una variable discreta

Gráficos para variables continuas

Cuando las variables son continuas, utilizamos como diagramas diferenciales los *histogramas* y los *polígonos de frecuencias*.

Un *histograma* se construye a partir de la tabla estadística, representando sobre cada intervalo, un rectángulo que tiene a este segmento como base. El criterio para calcular la altura de cada rectángulo es el de mantener la proporcionalidad entre las frecuencias absolutas (o relativas) de cada intervalo y el área de los mismos. Véase la figura 1.8.

El *polígono de frecuencias* se construye fácilmente si tenemos representado previamente el histograma, ya que consiste en unir mediante lineas rectas los puntos del histograma que corresponden a las marcas de clase. Para representar el polígono de frecuencias en el primer y último interva-

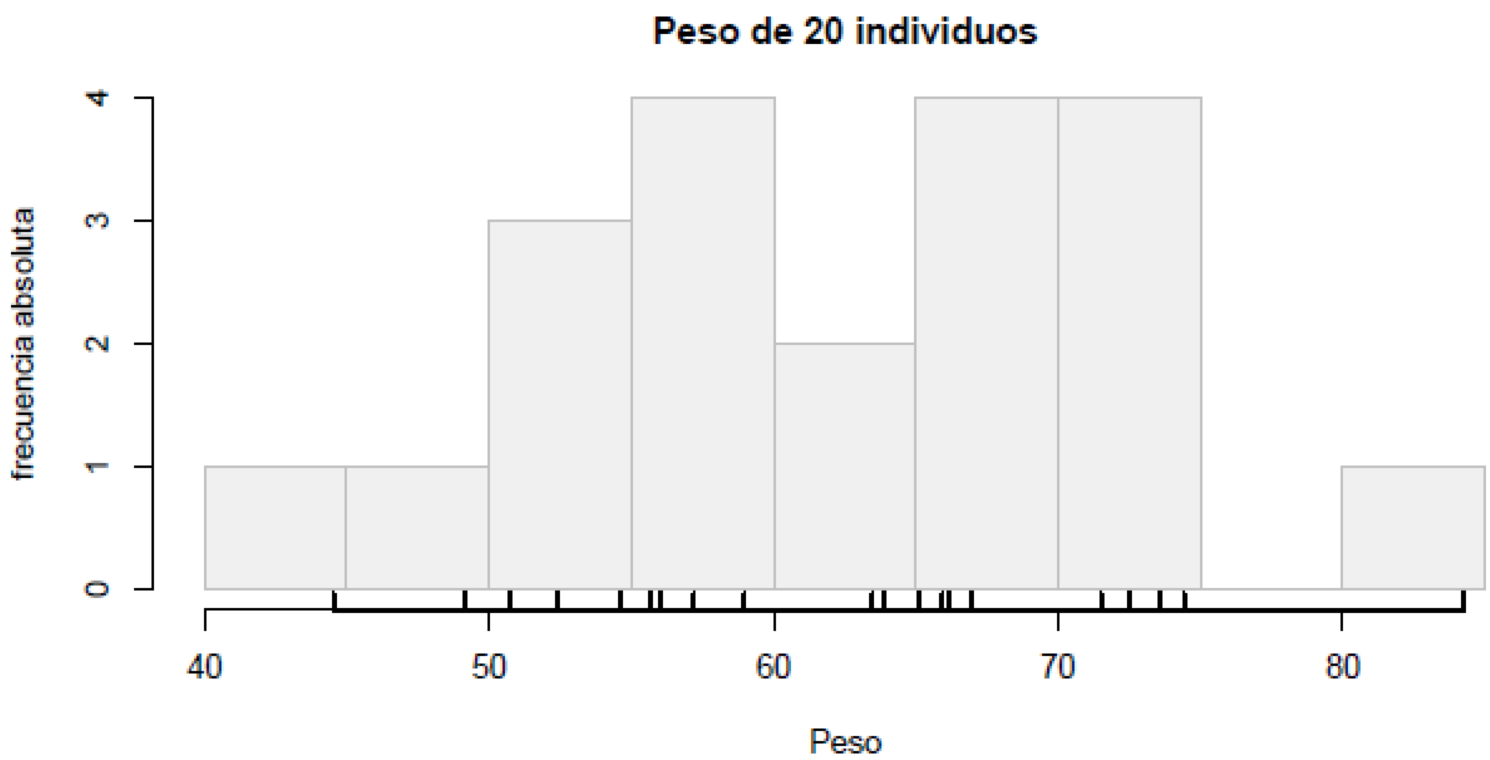
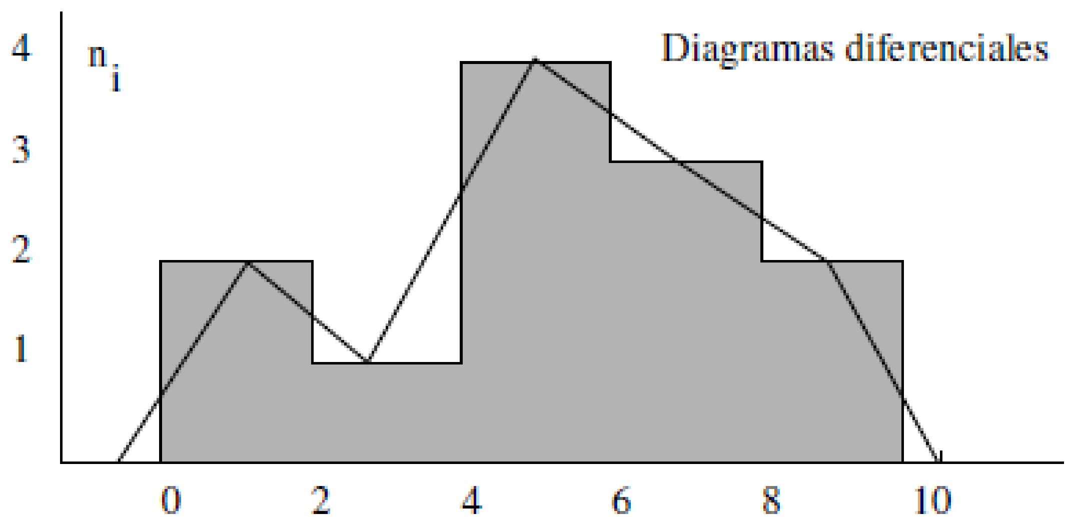


Figura 1.8: Histograma para una variable continua.

lo, suponemos que adyacentes a ellos existen otros intervalos de la misma amplitud y frecuencia nula, y se unen por una línea recta los puntos del histograma que corresponden a sus marcas de clase. Obsérvese que de este modo, el polígono de frecuencias tiene en común con el histograma el que las áreas de la gráficas sobre un intervalo son idénticas.

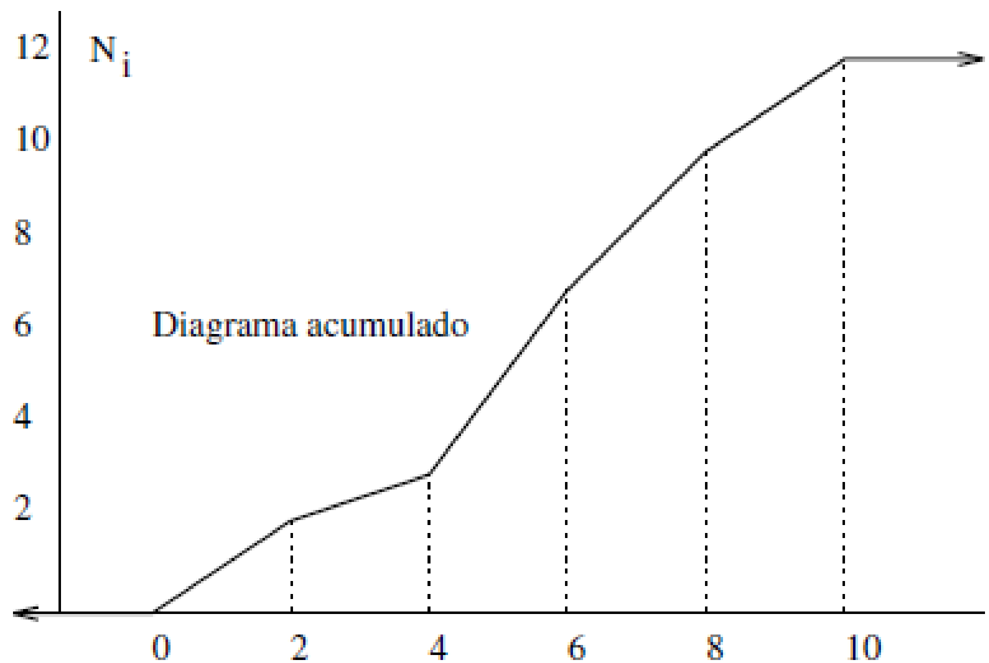
Intervalos	c_i	n_i	N_i
0 — 2	1	2	2
2 — 4	3	1	3
4 — 6	5	4	7
6 — 8	7	3	10
8 – 10	9	2	12
		12	



El diagrama integral para una variable continua se denomina también **polígono de frecuencias acumulado**, y se obtiene como la poligonal definida en abcisas a partir de los extremos de los intervalos en los que hemos organizado la tabla de la variable, y en ordenadas por alturas que son proporcionales a las frecuencias acumuladas. Dicho de otro modo, el polígono de frecuencias absolutas es una primitiva del histograma.

en la que se representa a modo de ilustración los diagramas correspondientes a la variable cuantitativa continua expresada en la tabla siguiente:

Intervalos	c_i	n_i	N_i
0 — 2	1	2	2
2 — 4	3	1	3
4 — 6	5	4	7
6 — 8	7	3	10
8 – 10	9	2	12
		12	



Ejemplo

La siguiente distribución se refiere a la duración en horas (completas) de un lote de 500 tubos:

Duración en horas	Número de tubos
300 — 500	50
500 — 700	150
700 — 1.100	275
más de 1.100	25
	Total 500

- Representar el histograma de frecuencias relativas y el polígono de frecuencias.
- Trazar la curva de frecuencias relativas acumuladas.
- Determinar el número mínimo de tubos que tienen una duración inferior a 900 horas.

Solución: En primer lugar observamos que la variable en estudio es discreta *(horas completas)*, pero al tener un rango tan amplio de valores resulta más conveniente agruparla en intervalos, como si de una variable continua se tratase. La consecuencia es una ligera perdida de precisión.

El último intervalo está abierto por el límite superior. Dado que en él hay 25 observaciones puede ser conveniente cerrarlo con una amplitud "razonable". Todos los intervalos excepto el tercero tienen una amplitud de 200 horas, luego podríamos cerrar el último intervalo en 1.300 horas[2].

Duración en horas	Número de tubos
300 — 500	50
500 — 700	150
700 — 1.100	275
más de 1.100	25
	Total 500

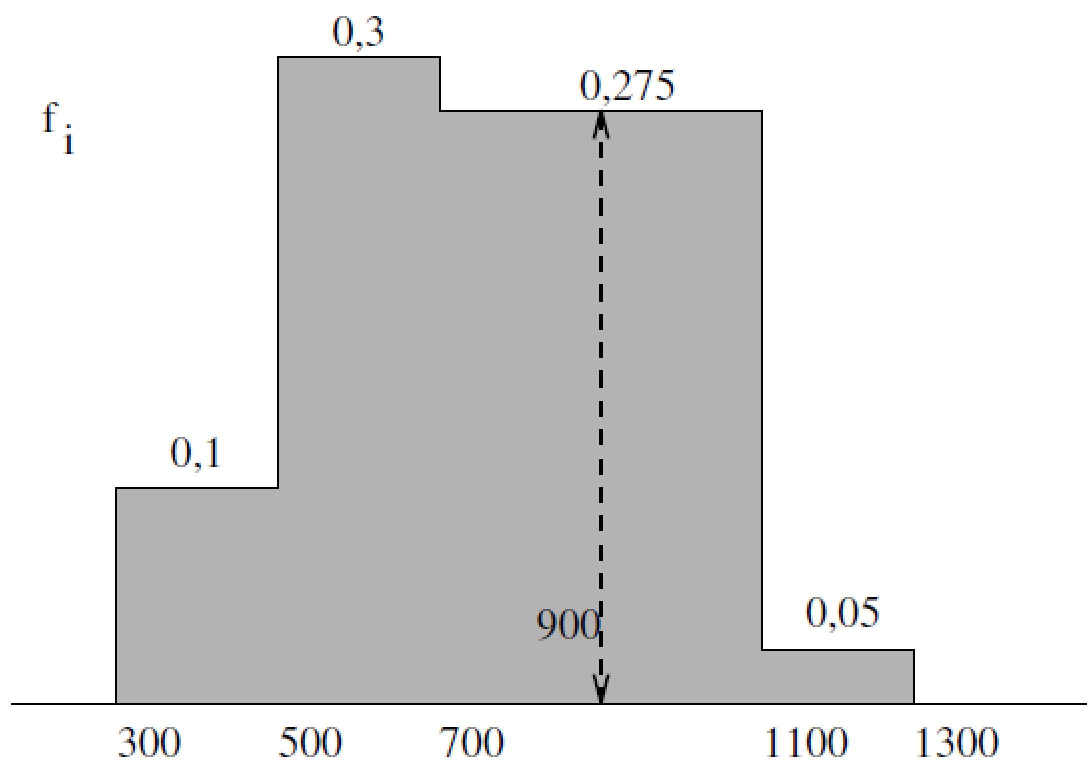
[2]Cualquier otra elección para el límite superior del intervalo que sea de "sentido comun" sería válida.

Antes de realizar el histograma conviene hacer una observación importante. El histograma representa las frecuencias de los intervalos mediante *áreas* y no mediante *alturas*. Sin embargo nos es mucho más fácil hacer representaciones gráficas teniendo en cuenta estas últimas. Si todos los intervalos tienen la misma amplitud no es necesario diferenciar entre los conceptos de área y altura, pero en este caso el tercer intervalo tiene una amplitud doble a los demás, y por tanto hay que repartir su área en un rectángulo de base doble (lo que reduce su áltura a la mitad).

Así será conveniente añadir a la habitual tabla de frecuencias una columna que represente a las amplitudes a_i de cada intervalo, y otra de frecuencias relativas rectificadas, f'_i, para representar la altura del histograma.

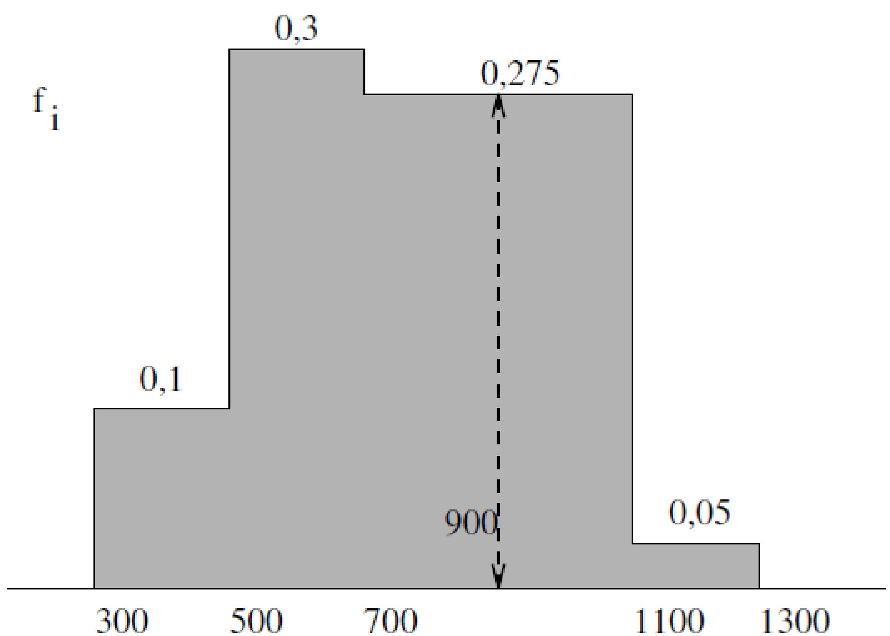
Intervalos	a_i	n_i	f_i	f'_i	F_i
300 — 500	200	50	0,10	0,10	0,10
500 — 700	200	150	0,30	0,30	0,40
700 — 1.100	400	275	0,55	0,275	0,95
1.100 — 1.300	200	25	0,05	0,05	1,00
		n=500			

Intervalos	a_i	n_i	f_i	f'_i	F_i
300 — 500	200	50	0,10	0,10	0,10
500 — 700	200	150	0,30	0,30	0,40
700 — 1.100	400	275	0,55	0,275	0,95
1.100 — 1.300	200	25	0,05	0,05	1,00
		n=500			



Histograma. Obsérvese que la altura del histograma en cada intervalo es f_i' que coincide en todos con f_i salvo en el intervalo 700 — 1.100 en el que ${f_i}' = 1/2\, f_i$ ya que la amplitud de ese intervalo es doble a la de los demás.

Intervalos	a_i
300 — 500	200
500 — 700	200
700 — 1.100	400
1.100 — 1.300	200



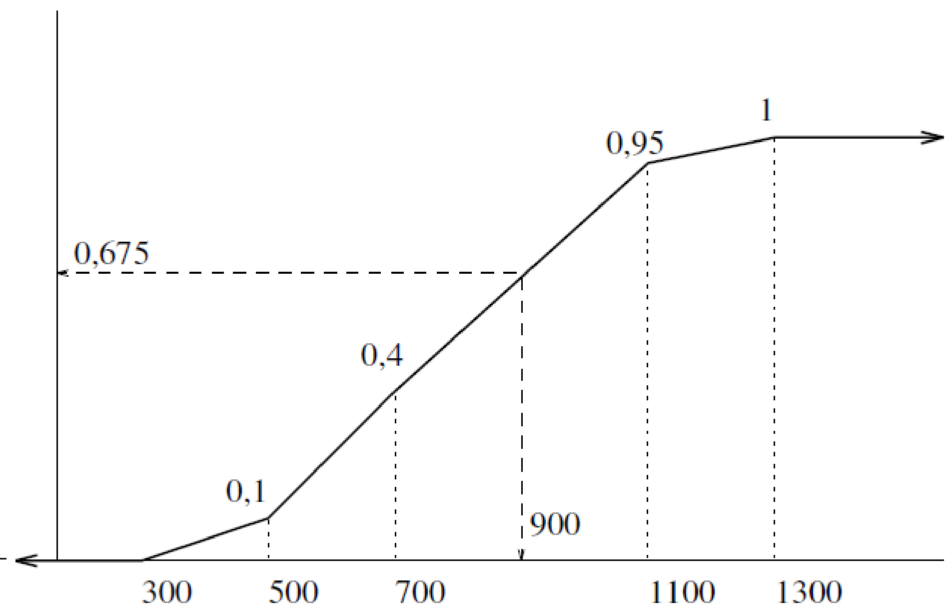


Diagrama acumulativo de frecuencias relativas

Por otro lado, se ve que sumando frecuencias relativas, hasta las 900 horas de duración hay

$$0{,}10 + 0{,}30 + 0{,}275 = 0{,}675 = 67{,}5\,\%\ \text{de los tubos.}$$

Esta cantidad se obtiene de modo más directo viendo a qué altura corresponde al valor 900 en el diagrama de frecuencias acumuladas

Como en total son 500 tubos, el número de tubos con una duración igual o menor que 900 horas es $0{,}675 \times 500 = 337{,}5$. Redondeando, 338 tubos.

Cuadro 1.1: Principales diagramas según el tipo de variable.

Tipo de variable	Diagrama
V. Cualitativa	Barras, sectores, pictogramas
V. Discreta	Diferencial (barras) Integral (en escalera)
V. Continua	Diferencial (histograma, polígono de frecuencias) Integral (diagramas acumulados)

1.6. Problemas

Ejercicio 1.1. Clasificar las siguientes variables:

1. Preferencias políticas (izquierda, derecha o centro).
2. Marcas de cerveza.
3. Velocidad en Km/h.
4. El peso en Kg.
5. Signo del zodiaco.
6. Nivel educativo (primario secundario, superior).

7. Años de estudios completados.
8. Tipo de enseñanza (privada o pública).
9. Número de empleados de una empresa.
10. La temperatura de un enfermo en grados Celsius.
11. La clase social (baja, media o alta).
12. La presión de un neumático en Nw/cm^2

Ejercicio 1.2. Clasifique las variables que aparecen en el siguiente cuestionario.

1. ¿Cuál es su edad?

2. Estado civil:

   a) Soltero
   b) Casado
   c) Separado
   d) Divorciado
   e) Viudo

3. ¿Cuanto tiempo emplea para desplazarse a su trabajo?

4. Tamaño de su municipio de residencia:

    a) Municipio pequeño (menos de 2.000 habitantes)
    b) Municipio mediano (de 2.000 a 10.000 hab.)
    c) Municipio grande (de 10.000 a 50.000 hab.)
    d) Ciudad pequeña (de 50.000 a 100.000 hab.)
    e) Ciudad grande (más de 100.000 hab.)

5. ¿Está afiliado a la seguridad social?

Ejercicio 1.3.

En el siguiente conjunto de datos, se proporcionan los pesos (redondeados a libras) de niños nacidos en cierto intervalo de tiempo:

4, 8, 4, 6, 8, 6, 7, 7, 7, 8, 10, 9, 7, 6, 10, 8, 5, 9, 6, 3, 7, 6, 4, 7, 6, 9, 7, 4, 7, 6, 8, 8, 9, 11, 8, 7, 10, 8, 5, 7, 7, 6, 5, 10, 8, 9, 7, 5, 6, 5.

1. Construir una distribución de frecuencia de estos pesos.
2. Encontrar las frecuencias relativas.
3. Encontrar las frecuencias acumuladas.
4. Encontrar las frecuencias relativas acumuladas.
5. Dibujar un histograma con los datos
6. ¿Por qué se ha utilizado un histograma para representar estos datos, en lugar de una gráfica de barras?

Medidas descriptivas

hemos visto cómo se pueden resumir los datos obtenidos del estudio de una muestra (o una población) en una tabla estadística o un gráfico. No obstante, tras la elaboración de la tabla y su representación gráfica, en la mayoría de las ocasiones resulta más eficaz “condensar” dicha información en algunos números que la expresen de forma clara y concisa.

Los fenómenos biológicos no suelen ser constantes, por lo que será necesario que junto a una medida que indique el valor alrededor del cual se agrupan los datos, se asocie una medida que haga referencia a la variabilidad que refleje dicha fluctuación.

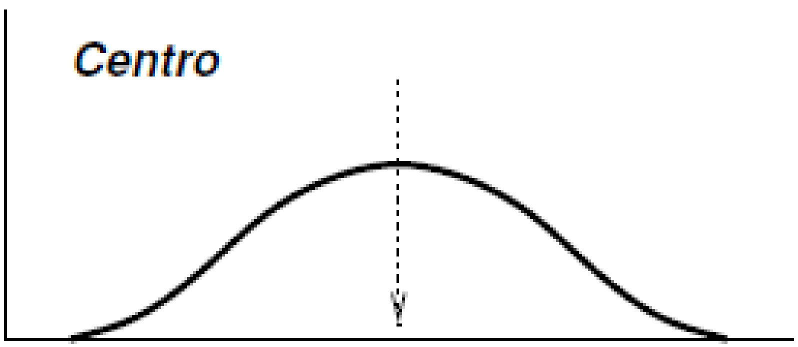
Por tanto el siguiente paso y objeto de este capítulo consistirá en definir algunos tipos de medidas (estadísticos o parámetros) que los sintetizan aún más.

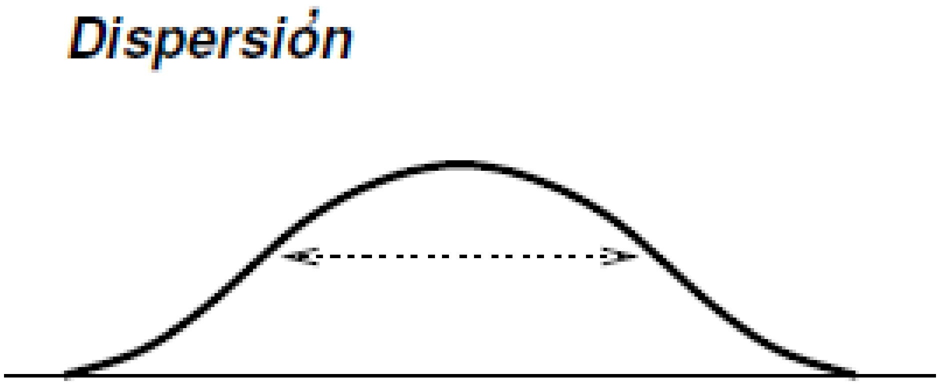
Es decir, dado un grupo de datos organizados en una distribución de frecuencias (o bien una serie de observaciones sin ordenar), pretendemos describirlos mediante dos o tres cantidades sintéticas.

En este sentido pueden examinarse varias características, siendo las más comunes:

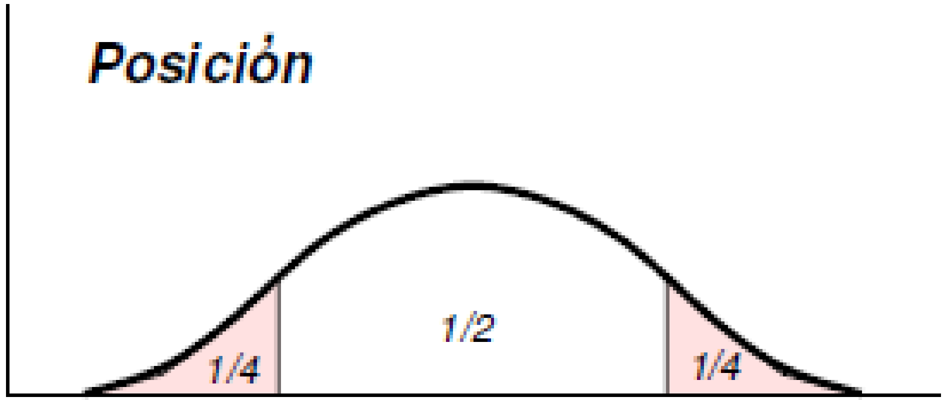
- La *tendencia central* de los datos;

- La *dispersión* o *variación* con respecto a este centro;

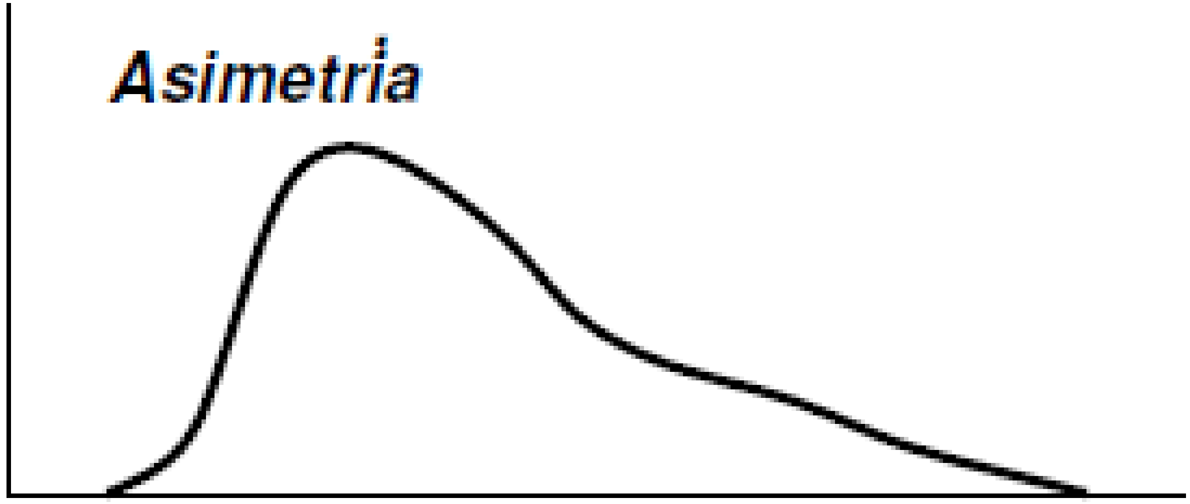




- Los datos que ocupan ciertas *posiciones*.



- La *simetría* de los datos.



Estadísticos de tendencia central

Las tres medidas más usuales de tendencia central son:

- la *media*,
- la *mediana*,
- la *moda*.

La media

La **media aritmética** de una variable estadística es la suma de todos sus posibles valores, ponderada por las frecuencias de los mismos. Es decir, si la tabla de valores de una variable X es

X	n_i	f_i
x_1	n_1	f_1
$\dots$	$\dots$	$\dots$
x_k	n_k	f_k

La media

la media es el valor que podemos escribir de las siguientes formas equivalentes:

$$\begin{aligned} \overline{x} &= x_1\, f_1 + \ldots + x_k\, f_k \\ &= \frac{1}{n}\,(x_1\, n_1 + \ldots x_k\, n_k) \\ &= \frac{1}{n}\sum_{i=1}^{k} x_i\, n_i \end{aligned}$$

La media

Si los datos no están ordenados en una tabla, entonces

$$\boxed{\overline{x} = \frac{x_1 + \ldots + x_n}{n}}$$

Algunos inconvenientes de la media

La media presenta inconvenientes en algunas situaciones:

- Uno de ellos es que es muy sensible a los valores extremos de la variable: ya que todas las observaciones intervienen en el cálculo de la media, la aparición de una observación extrema, hará que la media se desplace en esa dirección. En consecuencia,

- no es recomendable usar la media como medida central en las distribuciones muy asimétricas;

- Si consideramos una variable discreta, por ejemplo, *el número de hijos en las familias españolas* el valor de la media puede no pertenecer al conjunto de valores de la variable; Por ejemplo $\overline{x} = 1,2$ hijos.

Otras medias: Medias generalizadas

En función del tipo de problema varias generalizaciones de la media pueden ser consideradas. He aquí algunas de ellas aplicadas a unas observaciones $x_1, \ldots, x_n$:

La media geométrica $\overline{x}_g$, es la media de los logaritmos de los valores de la variable:

$$\log \overline{x}_g = \frac{\log x_1 + \ldots + \log x_n}{n}$$

Luego

$$\overline{x}_g = \sqrt[n]{x_1 \, x_2 \, \ldots \, x_n}$$

La media geométrica $\overline{x}_g$:

Si los datos están agrupados en una tabla, entonces se tiene:

$$\overline{x}_g = \sqrt[n]{x_1^{n_1}\, x_2^{n_2}\, \ldots\, x_k^{n_k}}$$

La media armónica $\overline{x}_a$, se define como el recíproco de la media aritmética de los recíprocos, es decir,

$$\frac{1}{\overline{x}_a} = \frac{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}{n}$$

Por tanto,

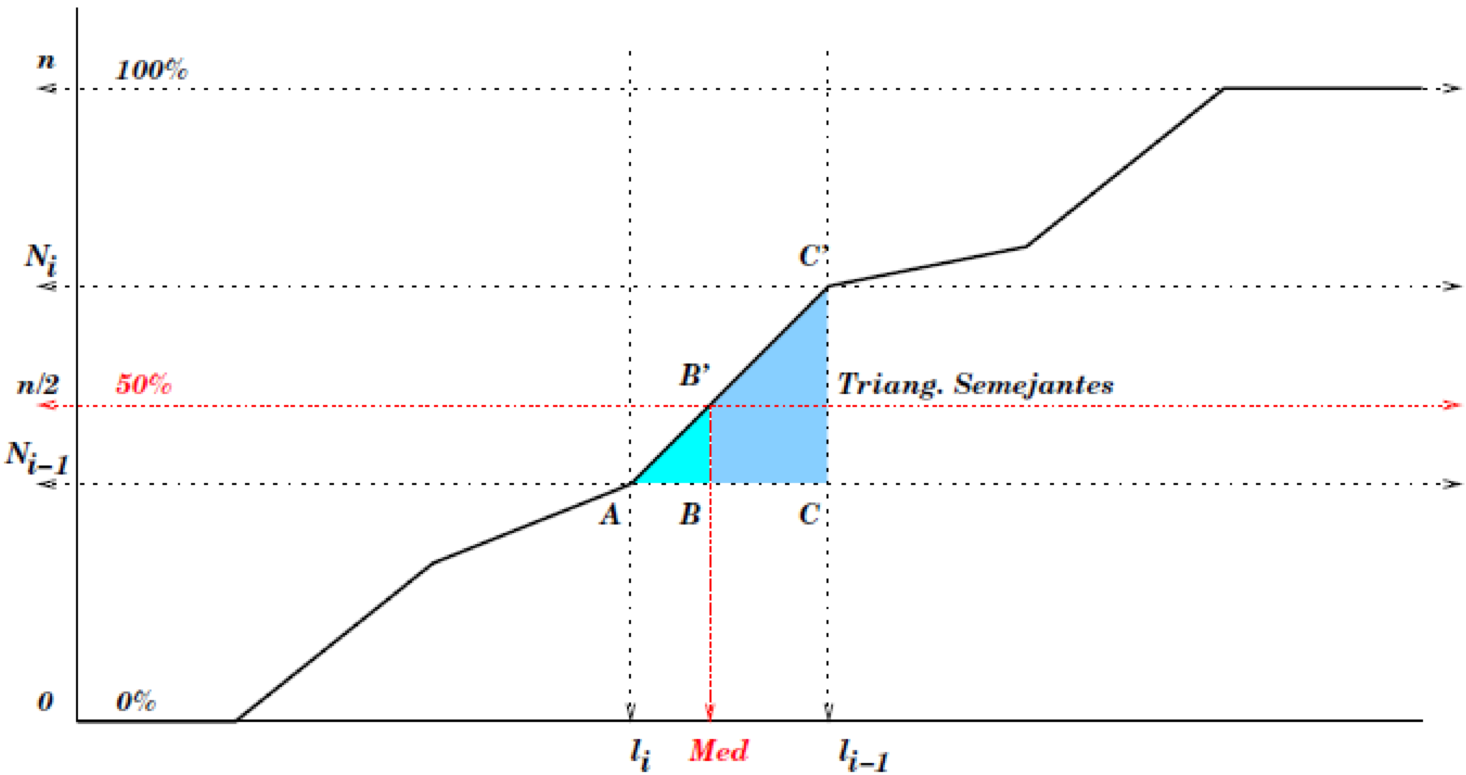
$$\overline{x}_a = \frac{n}{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}$$

La media cuadrática $\overline{x}_c$, es la raíz cuadrada de la media aritmética de los cuadrados:

$$\overline{x}_c = \sqrt{\frac{x_1^2 + \ldots + x_n^2}{n}}$$

La mediana

Consideramos una variable discreta X cuyas observaciones en una tabla estadística han sido ordenadas de menor a mayor. Llamaremos **mediana**, M_{ed} al primer valor de la variable que deja por debajo de sí al 50 % de las observaciones.



Cálculo geométrico de la mediana

En el caso de variables continuas, las clases vienen dadas por intervalos, y aquí la fórmula de la mediana se complica un poco más (pero no demasiado): Sea $(l_{i-1}, l_i]$ el intervalo donde hemos encontrado que por debajo están

el 50 % de las observaciones. Entonces se obtiene la mediana a partir de las frecuencias absolutas acumuladas, mediante interpolación lineal (teorema de Thales)

$$\frac{CC'}{AC} = \frac{BB'}{AB} \implies \frac{n_i}{a_i} = \frac{\frac{n}{2} - N_{i-1}}{M_{ed} - l_{i-1}}$$

$$\implies \boxed{M_{ed} = l_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} \cdot a_i}$$

Esto equivale a decir que *la mediana divide al histograma en dos partes de áreas iguales a* $\frac{1}{2}$.

Propiedades de la mediana

Entre las propiedades de la mediana, vamos a destacar las siguientes:

- Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas. Por ello es adecuado su uso en distribuciones asimétricas.

- Es de cálculo rápido y de interpretación sencilla.

- A diferencia de la media, la mediana de una variable discreta es siempre un valor de la variable que estudiamos (ej. La mediana de una variable *número de hijos* toma siempre valores enteros).

Un ejemplo de cálculo de mediana

Sea X una variable discreta que ha presentado sobre una muestra las modalidades

$$X \leadsto 2, 5, 7, 9, 12 \Longrightarrow \overline{x} = 7, \qquad M_{ed} = 7$$

Si cambiamos la última observación por otra anormalmente grande, esto no afecta a la mediana, pero si a la media:

$$X \leadsto 2, 5, 7, 9, 125 \Longrightarrow \overline{x} = 29, 6; \qquad M_{ed} = 7$$

En este caso la media no es un posible valor de la variable (discreta), y se ha visto muy afectada por la observación extrema. Este no ha sido el caso para la mediana.

Un ejemplo de cálculo de media y mediana

Obtener la media aritmética y la mediana en la distribución adjunta. Determinar gráficamente cuál de los dos promedios es más significativo.

$l_{i-1} - l_i$	n_i
$0 - 10$	60
$10 - 20$	80
$20 - 30$	30
$30 - 100$	20
$100 - 500$	10

$l_{i-1} - l_i$	n_i
0 – 10	60
10 – 20	80
20 – 30	30
30 – 100	20
100 – 500	10

$l_{i-1} - l_i$	n_i	a_i	x_i	$x_i n_i$	N_i	n_i'
0 – 10	60	10	5	300	60	60
10 – 20	80	10	15	1.200	140	80
20 – 30	30	10	25	750	170	30
30 – 100	20	70	65	1.300	190	2,9
100 – 500	10	400	300	3.000	200	0,25
	$n = 200$			$\sum x_i n_i = 6{,}550$		

La media aritmética es:

$$\overline{x} = \frac{1}{n}\sum x_i = \frac{6{,}550}{200} = 32,75$$

La primera frecuencia absoluta acumulada que supera el valor $n/2 = 100$ es $N_i = 140$. Por ello el intervalo mediano es $[10; 20)$. Así:

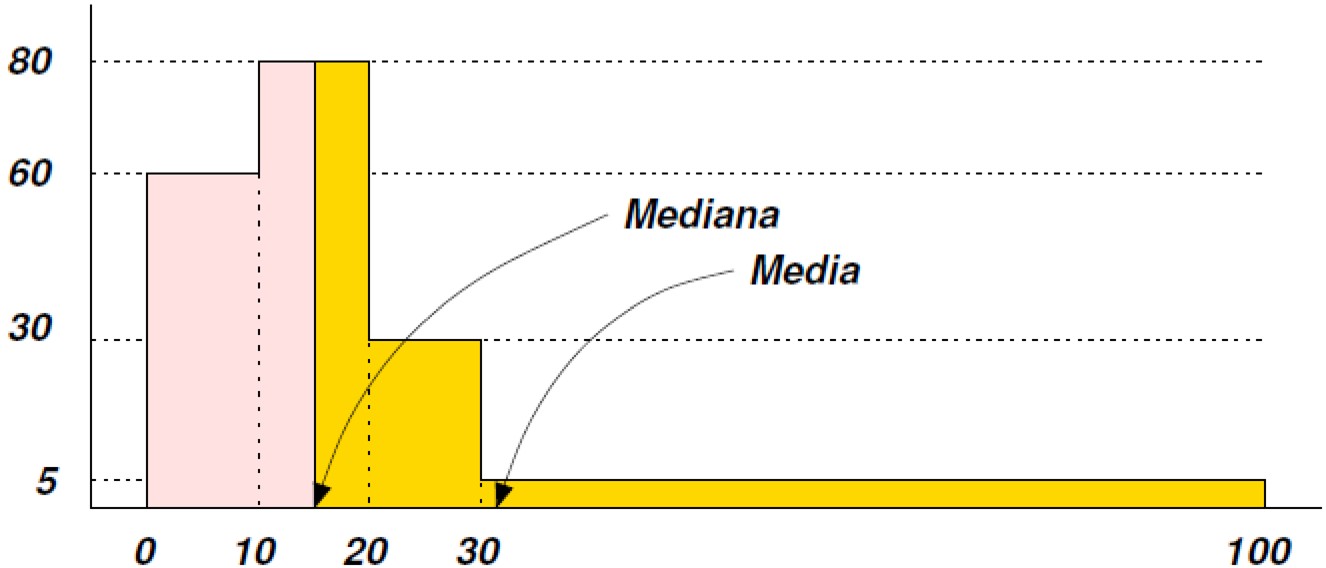
Un ejemplo de cálculo de media y mediana

$l_{i-1} - l_i$	n_i	a_i	x_i	$x_i n_i$	N_i	n_i'
0 – 10	60	10	5	300	60	60
10 – 20	80	10	15	1.200	140	80
20 – 30	30	10	25	750	170	30
30 – 100	20	70	65	1.300	190	2,9
100 – 500	10	400	300	3.000	200	0,25
	$n = 200$			$\sum x_i n_i = 6{,}550$		

$$M_{ed} = l_{i-1} + \frac{n/2 - N_{i-1}}{n_i} \cdot a_i = 10 + \frac{100 - 60}{80} \times 10 = 15$$

Para ver la representatividad de ambos promedios, realizamos el histograma y observamos que dada la forma de la distribución, la mediana es más representativa que la media.

Para ver la representatividad de ambos promedios, realizamos el histograma y observamos que dada la forma de la distribución, la mediana es más representativa que la media.



Para esta distribución de frecuencias es más representativo usar como estadístico de tendencia central la mediana que la media.

La moda

Llamaremos **moda** a cualquier máximo relativo de la distribución de frecuencias, es decir, cualquier valor de la variable que posea una frecuencia mayor que su anterior y su posterior.

Observación

De la moda destacamos las siguientes propiedades:

- Es muy fácil de calcular.
- Puede no ser única.

Resumen de las medidas de posición centrales.

	MEDIDAS DE TENDENCIA CENTRAL	
	DATOS SIN AGRUPAR	DATOS AGRUPADOS
	(ordenados) $x_1,\ x_2,\ \ldots,\ x_N$	$\begin{array}{cccc} \text{Interv.} & x_i & n_i & N_i \\ \hline l_0 — l_1 & x_1 & n_1 & N_1 \\ l_1 — l_2 & x_2 & n_2 & N_2 \\ \ldots & \ldots & \ldots & \ldots \\ l_{k-1} — l_k & x_k & n_k & N_k \end{array}$
MEDIA	$\overline{x} = \dfrac{x_1 + \cdots + x_n}{N}$	$\overline{x} = \dfrac{n_1\, x_1 + \cdots + n_k\, x_k}{N}$
MEDIANA	Primera observación que deja debajo de sí estrictamente a las $[N/2]$ observaciones menores: $x_{[N/2]+1}$	$M_{ed} = l_{i-1} + \dfrac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i$
MODA	$M_{oda} = x_i$ de mayor frecuencia	$M_{oda} == l_{i-1} + \frac{n'_i - n'_{i-1}}{(n'_i - n'_{i-1}) + (n'_i - n'_{i+1})}\, a_i$

Estadísticos de posición