

Estadística

Descriptiva

Ángel Mediano Morando

Institut Joan Ramon Benaprès-Sitges

30 de abril de 2015



En la realización de cualquier experimento se pueden distinguir dos fases:

Primera: consiste en la observación y análisis de los hechos que acontecen.

Segunda: interpretación y obtención de conclusiones.

Definición (Estadística descriptiva)

*La **ESTADÍSTICA DESCRIPTIVA** es la parte de la estadística que se encarga de dar una descripción numérica, ordenar y simplificar la información recogida en al primera fase.*



Definición (Población estadística o colectivo)

Llamamos población estadística al conjunto de referencia sobre el cual van a recaer las observaciones.

Este conjunto, normalmente, viene definido por comprensión, es decir, citando la propiedad que caracteriza a sus elementos.

Las poblaciones pueden ser **finitas o infinitas**, dependiendo del número de elementos que la forman.

Definición (Unidad estadística o individuo)

Se llama unidad estadística o individuo a cada uno de los elementos que componen la población estadística.

Definición (Muestra)

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



Definición (Censo)

*Decimos que realizamos un censo cuando se observan **todos** los elementos de una población estadística.*

Definición (Caracteres)

La observación del individuo la describimos mediante uno o más caracteres. El carácter, por tanto, es una cualidad o propiedad inherente en el individuo.

Hay caracteres que son medibles, es decir, se pueden cuantificar, por ejemplo, la edad de las personas, el peso, la estatura. Pero hay otros que no se pueden medir, por ejemplo, el sexo, el color del cabello, etc.



Tipos de caracteres

Cualitativos: Caracteres que **NO** se pueden determinar numéricamente. Por ejemplo: el sexo, color de los ojos, profesión laboral.

NOMINALES: Son los caracteres cualitativos **NO** ordenables.

- Sexo
- Color del cabello.

ORDINALES: Son los caracteres cualitativos que se pueden ordenar. Las notas I, S, B, N, E pueden ser ordenadas.

Cuantitativos: Caracteres que **SÍ** se pueden medir numéricamente.

Definición (Variable estadística)

*A los valores numéricos de las distintas modalidades que adoptan un carácter **cuantitativo** se les llama variable estadística y se representa con la letra X .*



Clasificación de las variables estadísticas

Variables estadísticas discretas: Son aquellas que toman un número finito de valores o infinito numerable, y que no pueden tomar ningún valor entre dos consecutivos fijados.

Ejemplos

- 1 Número de monedas que lleva una persona en el bolsillo.
- 2 Número de pulsaciones de una persona.
- 3 Número de varones de una familia.

Variables estadísticas continuas: Son aquellas que pueden tomar infinitos valores en un intervalo dado. Pueden tomar valores intermedios entre dos dados, por muy próximos que estén.

Ejemplos

- 1 La medida del tiempo.
- 2 La altura de una persona.



CARACTERES

Cuantitativos: **determinan una v. estadística: X**

Discreta

Continua

Cualitativos: determinan un atributo. Modalidades

Nominales

Ordinales



Frecuencias absolutas

Sean $x_1, x_2, \dots, x_k$ los distintos valores que toma la variable X :

Definición (Frecuencia absoluta (f_i))

Llamamos frecuencia absoluta de un valor x_i de la variable estadística X , al número de veces que aparece repetido dicho valor en el conjunto de las observaciones realizadas.

Definición (Frecuencia absoluta acumulada (F_i))

Llamamos frecuencia absoluta acumulada en el valor x_i de la variable estadística X , a la suma de las frecuencias absolutas de los valores inferiores o iguales a él.



Definición (Frecuencia relativa (h_i))

Llamamos frecuencia relativa de un valor x_i de la variable estadística X al cociente entre la frecuencia absoluta y el número de observaciones realizadas (N):

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

ATENCIÓN

Los valores de x_i deben estar ordenados en **forma creciente** y la frecuencia absoluta acumulada del último valor será N . Es decir, $F_k = N$.

Definición (Frecuencia relativa acumulada(H_i))

Llamamos frecuencia relativa acumulada en el valor x_i de la variable estadística X al cociente entre la frecuencia acumulada y el número de observaciones realizadas (N).

$$H_i = \frac{F_i}{N}$$



Propiedades de las frecuencias

- $f_1 + f_2 + \dots + f_k = N \iff \sum_{i=1}^k f_i = N$
- $h_1 + h_2 + \dots + h_k = 1 \iff \sum_{i=1}^k h_i = 1$

Tabla de frecuencias de una variable estadística DISCRETA

x_i	f_i	F_i	h_i	H_i
x_1	f_1	F_1	h_1	H_1
x_2	f_2	F_2	h_2	H_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	f_k	N	h_k	1
	N		1	



Ejemplo

Se lanza un dado 50 veces y los resultados obtenidos se recogen en la

siguiente tabla:

5	5	4	6	6	1	3	2	5	6
1	1	2	5	2	2	6	6	6	5
6	3	3	5	2	6	1	2	6	3
2	1	5	4	5	1	2	3	4	3
5	1	2	1	6	6	4	3	1	2

- Variable estadística DISCRETA, $X = \{1, 2, 3, 4, 5, 6\}$.
- Tabla estadística asociada:

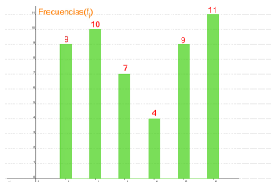
x_i	f_i	F_i	h_i	H_i
1	9	9	0,18	0,18
2	10	19	0,2	0,38
3	7	26	0,14	0,52
4	4	30	0,08	0,6
5	9	39	0,18	0,78
6	11	50	0,22	1
	50		1	



Definición (Diagramas de barras o bastones)

Son útiles cuando se comparan datos cualitativos o cuantitativos de tipo discreto.

- Eje de abscisas: valores x_i de la v.e.
- Eje de ordenadas: frecuencias absolutas o relativas.



x_i	f_i	F_i
1	9	9
2	10	19
3	7	26
4	4	30
5	9	39
6	11	50
	50	

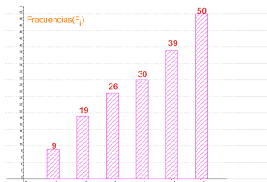


Tabla de frecuencias de una variable estadística

CONTINUA

- Si la v.e. es continua los datos se agrupan en **clases** (intervalos).

Definición (Marca de clase)

Se llama marca de clase (x_i) al punto medio de cada clase:

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

Definición (Longitud de una clase)

Se llama longitud (amplitud) de una clase a la diferencia entre el extremo superior (L_{i+1}) y el inferior (L_i).

$$L = L_{i+1} - L_i, \forall i$$



Variable estadística **discreta** con un número elevado de datos

Se define como una v. e. continua siguiendo los siguientes criterios:

- Número de clases: $k \sim \sqrt{n}$ donde n es el número de datos.
- Amplitud de las clases: $L \sim \text{rango}(R)/k$ redondeando por exceso.
(El exceso $kL - R$ se reparte entre la primera y la última clase)
- Las clases son intervalos cerrados por la izquierda y abiertos por la derecha: $[L_i, L_{i+1})$.

Tabla de frecuencias

Clases	Marca de clase(x_i)	f_i	h_i	F_i	H_i
$[L_1, L_2)$	x_1	f_1	h_1	F_1	H_1
$[L_2, L_3)$	x_2	f_2	h_2	F_2	H_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_{k-1}, L_k)$	x_k	f_k	h_k	N	1
		N	1		



Ejemplo

Los puntos obtenidos por el equipo de rugby en los 40 partidos jugados en la temporada 2010-11 se recogen en la siguiente tabla:

Puntos obtenidos	$[0, 5)$	$[5, 10)$	$[10, 15)$	$[15, 20)$	$[20, 25)$
Nº de partidos	3	9	15	10	3

Sea $X = \{\text{puntos obtenidos}\}$ la v. e. continua, efectuar la tabla de frecuencias asociada.

Clases	Marca de clase(x_i)	f_i	F_i	h_i	H_i
$[0, 5)$	2,5	3	3	0,075	0,075
$[5, 10)$	7,5	9	12	0,225	0,3
$[10, 15)$	12,5	15	27	0,375	0,675
$[15, 20)$	17,5	10	37	0,25	0,925
$[20, 25)$	22,5	3	40	0,075	1
		40		1	



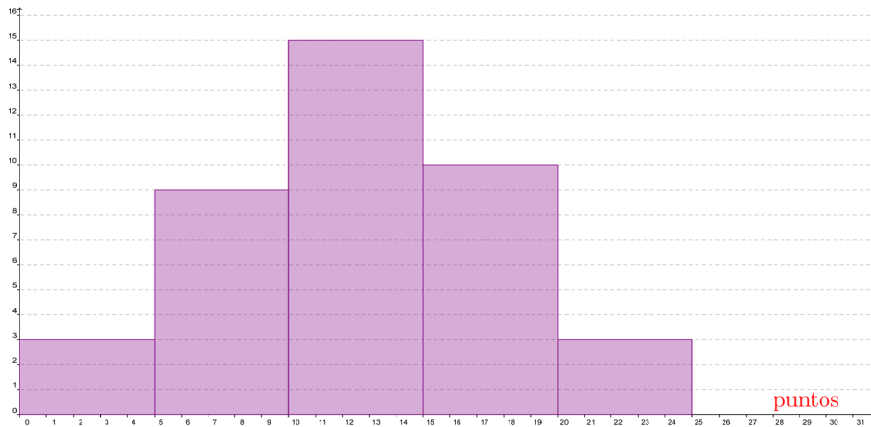
Definición (Histograma)

Se utilizan, generalmente, para distribuciones de v.e. continua o v.e. discreta con un gran número de datos que se han agrupado en clases.

- Eje de abscisas: las clases de la v.e.
- **El área de los rectángulos que se forman debe ser igual a la frecuencia absoluta de cada una de las clases**, para ello, la altura de cada uno de los rectángulos se calcula dividiendo la frecuencia absoluta por la amplitud del intervalo, que es la base del rectángulo.
- **Caso particular:** si las clases tienen todas la misma amplitud puede tomarse como altura de cada uno de los rectángulos la frecuencia de la clase correspondiente.



Histogramas rugby



Polígonos de frecuencias

Variable estadística discreta

El polígono de frecuencias es la línea poligonal que une los extremos de las barras.

Variable estadística continua

Frecuencias absolutas: se forma el polígono uniendo los puntos medios de las bases superiores de cada rectángulo del histograma.

Frecuencias absolutas acumuladas: se forma el polígono uniendo los puntos extremos de las bases superiores de cada rectángulo del histograma.



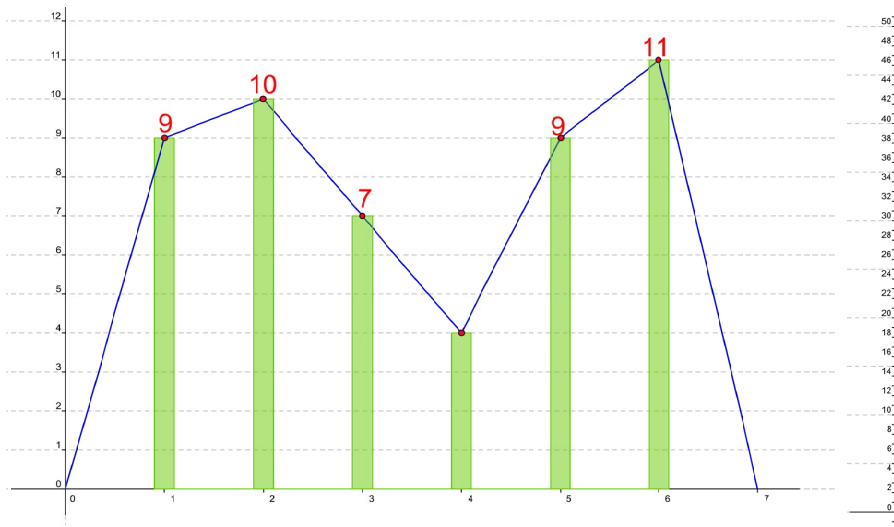
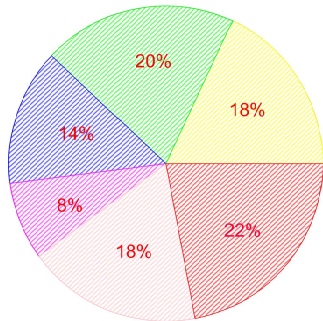


Diagrama de sectores (pie-chart)

Se emplean tanto para caracteres cualitativos como cuantitativos. El ángulo central de cada sector ha de ser proporcional a la frecuencia absoluta correspondiente, por lo tanto el área del sector es proporcional a la frecuencia absoluta (f_i).



x_i	f_i	h_i	$h_i \cdot 100$
1	9	0,18	18
2	10	0,20	20
3	7	0,14	14
4	4	0,8	8
5	9	0,18	18
6	11	0,22	22
	50	1	



Para caracteres **cualitativos** se emplean pictogramas, cartogramas, que representan mediante dibujos a escala, las frecuencias absolutas de cada una de las modalidades del carácter.

