TEMA 5 ESTADÍSTICA

Estadística obtención, estudio e interpretación de grandes masas de datos

Población es el conjunto de todos los elementos que cumplen una determinada característica.

Muestra es cualquier parte de la población, cuyo estudio sirve para inferir características de toda la población.

Individuo es cada uno de los elementos que forma la población o la muestra.

Carácter estadístico es un aspecto de la población que se puede observar. Las variantes que puede tomar un carácter son las modalidades del carácter. Un carácter será **cualitativo** si sus modalidades no se pueden expresar con números, y será **cuantitativo** cuando sí que se pueden expresar. Los caracteres cualitativos se llaman variables estadísticas

Variable estadística es el conjunto de valores que toma un carácter estadístico cuantitativo (que se puede medir) y pueden ser de dos tipos:

_ **Variable estadística discreta**: La que puede tomar un número finito de valores numéricos, o infinito numerable.

_ **Variable estadística continua**: La que puede tomar, por lo menos teóricamente, todos los valores dentro de un intervalo de la recta real.

Ejemplos

- Caracteres estadísticos cuantitativos:

– La altura de un individuo.
– El diámetro de una pieza de precisión.
– El cociente intelectual de un individuo.
– La renta per cápita de una comunidad autónoma.

- Caracteres estadísticos cualitativos:

– La profesión de una persona.
– El color de los ojos.
– La lengua que habla un individuo.

- Variables estadísticas discretas:

– Numero de empleados de una fábrica.
– Número de hijos de una familia.
– Número de goles marcados por la selección de fútbol.
– Numero de periódicos vendidos en un día.

- Variables estadísticas continuas:

– Presión sanguínea de un paciente.
– Diámetro de una rueda.
– Medida del cráneo de un bebé.
– Horas dormidas en una noche.
– Altura de un individuo.

Actividad

1. De cada uno de los siguientes estudios estadísticos, indica cuál es la población a la que se refiere, si consideras necesario elegir una muestra, y el carácter estadístico y su tipo.

a) Horas diarias de sueño de los habitantes de una provincia.
b) Preferencias literarias de las personas mayores de edad que viven en un edificio.
c) Color de ojos de la población de Vigo

2. Indica la población, la variable y el tipo (cualitativa, cuantitativa discreta o continua) de:

_ Peso al nacer de los bebés que nacieron en Barcelona en 2009.
_ Profesiones que quieren estudiar los estudiantes de un centro escolar.
_ Número de tarjetas amarillas mostradas en los partidos de fútbol de la liga del año pasado.
_ Un fabricante de tornillos desea hacer un control de calidad. Para ello recoge 1 de cada 100 tornillos producidos los analiza.

3. ¿Cómo debe de ser una muestra para ser correcta?

4. En una gasolinera se pretende hacer un estudio de su clientela. Para ello se observan y se anotan ciertas características de algunos coches, elegidos al azar, como el nº de ocupantes, el tipo de carburante y el coste del producto repostado.
- ¿Cuál es la población, la muestra, el individuo?
- ¿Qué caracteres estadísticos se estudian?¿Cómo son?

RECOGIDA DE DATOS
La información estadística llega a nosotros mediante gráficas o tablas muy bien construidas, con las que resulta fácil entender la información dada. Pero para llegar a ellas, es necesario realizar un largo proceso, que iniciamos ahora.
Frecuencia absoluta, f_i, de un valor de la variable, x_i, es el número de veces que se repite dicho valor
Frecuencia relativa de un valor de la variable, x_i, es el cociente entre la frecuencia absoluta del valor y el número total de datos.

La frecuencia relativa del valor x_i la representaremos por h_i : $h_i = \frac{f_i}{N}$ donde N es el número total de datos, es decir: $N = f_1 + f_2 + \ldots + f_n$

Frecuencia absoluta acumulada de un valor de la variable, x_i, es la suma de las frecuencias absolutas de los valores menores o iguales a x_i. La representamos por F_i
$F_i = f_1 + f_2 + f_3 + \ldots + f_i$

Frecuencia relativa acumulada de un valor de la variable, x_i, es el cociente entre la frecuencia absoluta acumulada del valor x_i y el número total de datos. La representaremos F_{ri} :

$$F_{ri} = \frac{F_i}{N} = \frac{f_1 + f_2 + f_3 + \ldots + f_i}{N} = \frac{f_1}{N} + \frac{f_2}{N} + \frac{f_3}{N} + \ldots + \frac{f_i}{N} = f_{r1} + f_{r2} + f_{r3} + \ldots + f_{ri}$$

<u>TABLAS</u>
Para ordenar los datos elaboramos tablas con los datos de la muestra. ¿Cómo se elabora una tabla estadística?

1. Recogida de datos: consiste en la toma de datos procedente de la muestra.
2. Ordenación de los datos: se colocarán en orden creciente o decreciente.
3. Recuento de frecuencias: se hace el recuento de datos.
4. Agrupación de datos: si la variable es continua (toma todos los infinitos valores de un intervalo), o discreta (toma sólo valores aislados) con un número de datos muy grande, resulta aconsejable agrupar los datos en intervalos (**clases**).
   Todas las clases deben tener la misma amplitud.
   Al punto medio de cada clase, se le llama **marca de clase**. Los intervalos se deben construir de tal manera que el extremo superior de una clase coincida con el extremo inferior de la siguiente. Así, en el intervalo [40 – 45) se contabilizan todos los pesos desde 40 kg (incluido) hasta 45 kg (excluido).

5. Elaboración de la tabla: Figurarán los valores de la variable (si están agrupados en clases, los extremos inferior y superior, así como la marca de clase), y las frecuencias absolutas y relativas. A veces, es conveniente incluir las frecuencias absolutas acumuladas, las frecuencias relativas acumuladas y los porcentajes.

Ejemplo 1. En una clase con 30 alumnos, hay 3 que no tienen hermanos, 9 con un hermano, 13 con dos hermanos, 2 con 3 hermanos, 1 con 4 hermanos, 1 con 5 hermanos y 1 con 8 hermanos. Elabora la tabla estadística de la variable "nº de hermanos"

Ejemplo 2. El peso de los alumnos anteriores:

x_i	f_i	F_i	h_i	H_i
0	3	3	$\frac{3}{30}$	$\frac{3}{30}$
1	9	12	$\frac{9}{30}$	$\frac{12}{30}$
2	13	25	$\frac{13}{30}$	$\frac{25}{30}$
3	2	27	$\frac{2}{30}$	$\frac{27}{30}$
4	1	28	$\frac{1}{30}$	$\frac{28}{30}$
5	1	29	$\frac{1}{30}$	$\frac{29}{30}$
8	1	30	$\frac{1}{30}$	$\frac{30}{30}$
	N=30		1	

ejemplo 1

Peso (kg)	Marca de clase x_i	Nº de alumnos f_i	F_i	f_{ri}	F_{ri}
[40-45)	42,5	1	1	$\frac{1}{30}$	$\frac{1}{30}$
[45-50)	47,5	3	4	$\frac{3}{30}$	$\frac{4}{30}$
[50-55)	52,5	10	14	$\frac{10}{30}$	$\frac{14}{30}$
[55-60)	57,5	9	23	$\frac{9}{30}$	$\frac{23}{30}$
[60-65)	62,5	4	27	$\frac{4}{30}$	$\frac{27}{30}$
[65-70)	67,5	2	29	$\frac{2}{30}$	$\frac{29}{30}$
[70-75)	72,5	1	30	$\frac{1}{30}$	$\frac{30}{30}$

ejemplo 2

PARÁMETROS ESTADÍSTICOS. CÁLCULO Y SIGNIFICADO

Después de obtener los datos de una distribución, necesitamos sintetizar la información para su posterior análisis. Para eso, obtendremos los parámetros estadísticos que serán de dos tipos: de centralización y de dispersión.

_ **Parámetros de centralización** Nos indican en torno a qué valor central se distribuyen los datos.

_ **Parámetros de dispersión** Nos informan sobre cuánto se alejan del centro los valores de la distribución.

MEDIDAS DE CENTRALIZACIÓN

Media aritmética, $\bar{x}$, de una variable estadística es el cociente entre la suma de todos los valores de dicha variable y el número de estos.

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{f_1 + f_2 + \dots + f_n}$$

En caso de datos agrupados, se toman para x_i las marcas de clase.

Moda de una variable estadística es el valor de dicha variable que tiene mayor frecuencia absoluta. La moda se representa **M_o**.

En caso de datos agrupados en clases, se toma como valor aproximado de la moda, la marca de clase que presenta mayor frecuencia absoluta. Esta clase se llama **clase modal**.

Mediana de una variable estadística, es un valor de la variable tal que el número de valores menores que él es igual al número de observaciones mayores que él. La mediana se representa por **M**.

Si los datos están agrupados, el **intervalo** o **clase mediana** es el primer intervalo cuya frecuencia absoluta acumulada es mayor que la mitad del número de datos.

Cuartiles de una variable estadística son Q1, Q2, y Q3, de tal modo que:

– Q_1 deja a su izquierda el 25 % de los datos.

– Q_2 deja a su izquierda el 50 % de los datos y coincide con la mediana.

– Q_3 deja a su izquierda el 75 % de los datos.

Ejercicio 5. Juan fue anotando las temperaturas de su pueblo durante los siete días de una semana:

19 Cº; 21 Cº; 19 Cº; 18Cº; 18 Cº; 20 Cº; 18Cº

Calcula la media, moda y la mediana. También Q_1 y Q_3

MEDIDAS DE DISPERSIÓN

Se desea saber si los datos numéricos están agrupados o no alrededor de los valores centrales. A esto es a lo que se le llama **dispersión**, y los parámetros que miden la desviación respecto de la media, se llaman **parámetros de dispersión**. Son: rango o recorrido, varianza y desviación típica.

Rango o **recorrido** de una distribución es la diferencia entre el mayor y el menor valor de la variable estadística.

Se llaman valores extremos, al mayor y al menor valor de los datos estadísticos.

Varianza de una variable estadística es la media aritmética de los cuadrados de las desviaciones respecto a la media. Se representa por σ^2.

$$\sigma^2 = \frac{f_1(x_1-\bar{x})^2 + f_2(x_2-\bar{x})^2 + \ldots + f_n(x_n-\bar{x})^2}{N} = \frac{f_1x_1^2 + f_2x_2^2 + \ldots + f_nx_n^2}{N} - \bar{x}^2$$

Desviación típica es la raíz cuadrada positiva de la varianza. Se representa por σ.

Ejercicio 6. Calcula las medidas de centralización y de dispersión del ejemplo 1

USO DE LA CALCULADORA

Cuando se trata de un número importante de datos, el proceso se simplifica con la ayuda de una **calculadora científica**.

Ejercicio: Hallar la media aritmética y la desviación típica de la distribución que estudia el número de goles por partido marcados en la Liga de Fútbol 86-87.

nº de goles	0	1	2	3	4	5	6	7	8
nº de partidos	32	71	80	62	36	15	6	2	2

1º Se selecciona el MODE SD (Statistics Descriptive).

2º Se borra la memoria, no sea que haya datos anteriores almacenados.

3º Se introducen los datos: 0x32 ENTER 1X71 ENTER ... 8X2 ENTER

En algunas calculadoras, la tecla ENTER se sustituye por la tecla DATA

4º Se pulsa la tecla $\bar{x}$ y se obtiene la media: $\bar{x} = 2{,}29$

5º Se pulsa la tecla σ_n y se obtiene la desviación típica: $\sigma_n = 1{,}55$

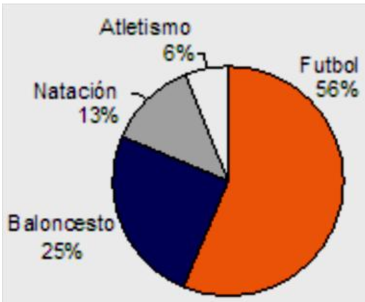
GRÁFICAS

Es conveniente expresar la información mediante gráficas, con el fin de hacerla más clara y evidente. Los principales tipos de gráficos son: diagrama de sectores, diagrama de barras-polígono de frecuencias, histograma-polígono de frecuencias, diagrama lineal.

Diagrama de sectores, consiste en un círculo dividido en tantos sectores circulares como modalidades tiene el carácter. El ángulo central de cada sector ha de ser proporcional a la frecuencia absoluta correspondiente.

	f_i	h_i	Graos
Fútbol	2 304	0,5625	202,5º
Baloncesto	1 024	0,25	90º
Natación	512	0,125	45º
Atletismo	256	0,0625	22,5º
	4 096		

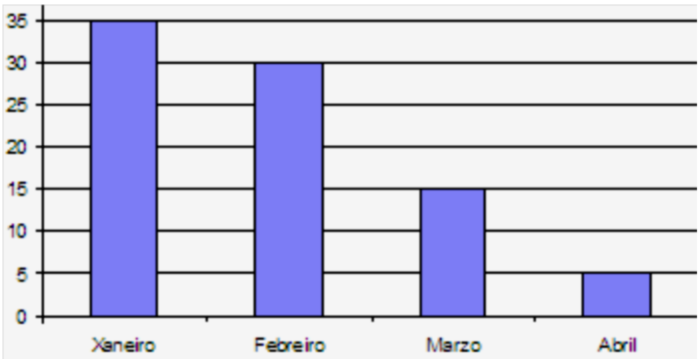
X 360º



Ejercicio 7. Representa mediante un diagrama de sectores la distribución estadística que clasifica a 30 alumnos según la Autonomía de nacimiento: 19 de Andalucía, 7 de castilla-La Mancha, 2 de Cataluña, 1 de Galicia y 1 de País Vasco.

Diagrama de barras, consiste en representar sobre el eje de abscisas los datos y en esos puntos, se levantan barras con altura proporcional a las frecuencias absolutas o a las frecuencias absolutas acumuladas. Si unimos los extremos de las barras, obtenemos el **polígono de frecuencias**.

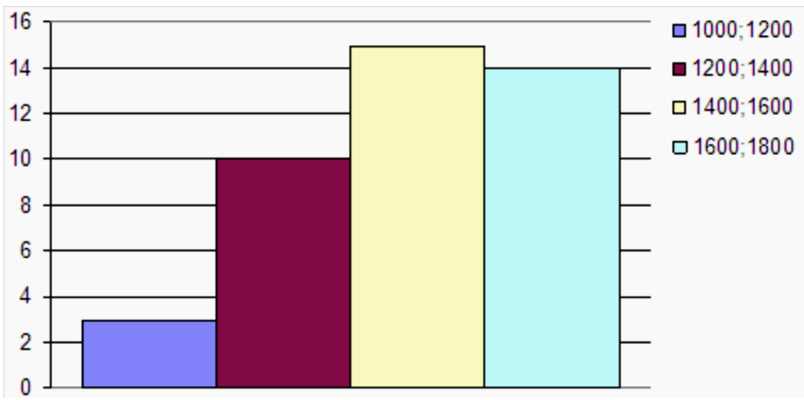
Mes	Nº vedas
▪ Xaneiro	35
▪ Febreiro	30
▪ Marzo	15
▪ Abril	5



Ejercicio 8. Representa mediante diagramas de barras las frecuencias absolutas y las frecuencias absolutas acumuladas, los datos del ejercicio anterior. Obtén los polígonos de frecuencias correspondientes.

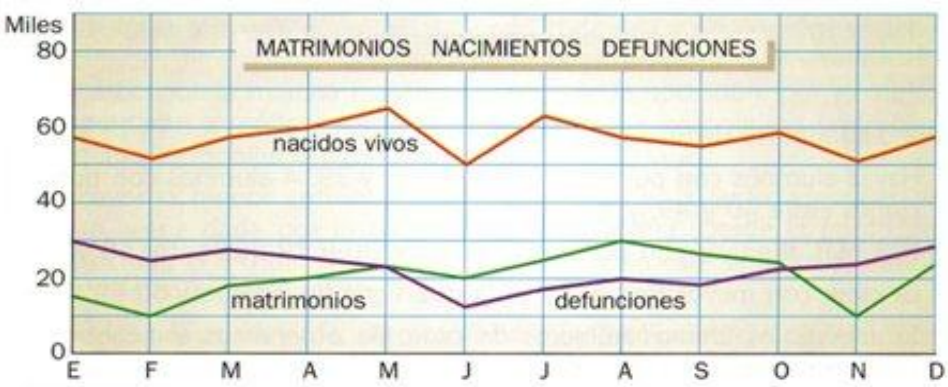
Histograma, se utiliza para variables continuas o discretas con datos agrupados en clases. Se representan sobre el eje de abscisas los extremos de las clases y sobre ellos se levantan rectángulos con altura proporcional a las frecuencias absolutas. El polígono de frecuencias se obtiene al unir los puntos medios de los lados superiores de cada rectángulo.

Peso (g)	Nº de polos
$1\,000 \leq x < 1\,200$	3
$1\,200 \leq x < 1\,400$	10
$1\,400 \leq x < 1\,600$	15
$1\,600 \leq x < 1\,800$	14



Ejercicio 9. Obtén el histograma de la distribución que clasifica a 30 alumnos según su peso en kg (Ejemplo 2)

Diagrama lineal, se utiliza para mostrar las fluctuaciones de uno o varios caracteres estadísticos con el paso del tiempo. El gráfico siguiente, expresa en miles, los matrimonios, nacimientos y defunciones producidos un determinado año:



TEMA 6 SUCESOS ALEATORIOS y PROBABILIDAD

Un *experimento es aleatorio* cuando no se puede predecir el resultado que se va a obtener al realizarlo.

Ejemplo: Estudiemos la experiencia aleatoria consistente en lanzar un dado y observar lo que sale:

- Los posibles resultados son

Espacio muestral es el conjunto formado por todos los resultados posibles de un experimento. Se designa por **E** . E= {1, 2, 3, 4, 5, 6}

- **Suceso aleatorio** es cualquier conjunto formado por uno o más elementos del espacio muestral.
  A={1}, B={1, 3, 5}, C= {los múltiplos de 3}={3, 6}
- **Suceso elemental** es el que está formado por un solo resultado.
- **Suceso compuesto** está formado por más de un resultado.
- **Suceso seguro** es el que siempre se realiza. Se designa por E. *El suceso sacar un nº menor que 7 es un suceso seguro*
- **Suceso imposible** es el que nunca se realiza. Se designa por Φ . p.e*.: El suceso sacer un nº mayor que 7*
- **Suceso contrario** de A es el que se realiza cuando no se realiza A. Se designa por $\overline{A}$.
  ={2, 4, 6}

Ejercicio 9.- Se lanzan tres monedas. Formar el espacio muestral.

Ejercicio 10.- Se lanzan una moneda y un dado. Establecer el espacio muestral.

Ejercicio 11 .Determinar si los siguientes experimentos son o no aleatorios.

_ Lanzar una moneda al aire.

_ Meter una botella en un cubo de agua y ver qué cantidad vierte.

_ Extraer una carta de una baraja.

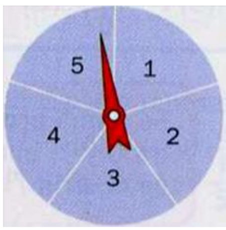
_ Observar el número de días con lluvia de un mes.

_ Medir una circunferencia de 2 cm de radio.

_ Tirar una piedra y medir su aceleración.

Ejercicio 12

¿Cuál es el espacio muestral en casa caso? Escriba los sucesos elementales y un suceso compuesto. Ponga un ejemplo de suceso seguro para cada caso. Ponga un ejemplo de suceso imposible para cada caso.



La probabilidad de un suceso indica el grado de confianza que podemos tener de que acontezca. Lo expresaremos mediante un número comprendido entre 0 y 1. Para designar la probabilidad de un suceso S, pondremos P[S].

_ Cuando la probabilidad sea un número próximo a cero, el suceso será poco probable.

_ Siempre que la probabilidad sea un número próximo a uno, será muy probable.

Ley de Laplace: Si todos los resultados de un experimento son equiprobables, es decir, tienen la misma probabilidad de salir, se tiene que

$$P = \text{probabilidad de un suceso} = \frac{n^{o}\ de\ casos\ \ favorables\ al\ suceso}{n^{o}\ de\ casos\ \ posibles} = \frac{C_F}{C_P}$$

Ejemplo: En una bolsa tenemos 90 bolas de colores, todas del mismo tamaño: Si sacamos una al azar, calcular las posibilidades de que sea de uno u otro color

Rojas	40
verdes	25
azules	15
negras	10

Solución: P(roja)=——————— — –

P(verde) = — — P(azul) = — – P(negra) = — –

Problema 8.- Lanzamos un dado. Halla la probabilidad de obtener:

a) Un número impar b) Un múltiplo de 3

Problema 9.- Lanzamos dos monedas. Hallar las siguientes probabilidades:

a) Obtener dos caras b)Obtener al menos una cruz

DECISIÓN y PROBABILIDAD. VALORACIÓN de la participación en JUEGOS de AZAR.

El número 77777 es muy llamativo, pero como nos parece "raro", casi nadie quiere jugarlo en la Lotería de Navidad, lo mismo que el número 00001 o el 99999. Todos ellos tienen exactamente las mismas probabilidades de salir premiados.

Ejercicio 12. De una rifa se han vendido 1 000 papeletas numeradas del 1 al 1 000. ¿Cuál es la probabilidad de que me toque si he comprado una papeleta? ¿Y si compro siete?

Dependencias y LUDOPATÍAS. Mucha gente sueña con que le toque la Lotería Primitiva o algún otro juego de azar. Todos podemos soñar, pero la teoría de probabilidades nos demuestra que nuestras opciones son pocas. Claro que cuántos más boletos distintos rellenes, más posibilidades tienes. ¿Sabes cuántas apuestas necesitarías para estar seguro de acertar un pleno de la Primitiva? 13.983.816 apuestas distintas. Imagínate cuánto dinero necesitarías y, lo que es peor, cuánto tiempo para escribir todas esas apuestas.

Si fueras capaz de rellenar un boleto cada 15 segundos, necesitarías trabajar sin parar durante seis años y medio para tenerlos todos.

Si prefieres jugar a las quinielas de fútbol y quieres asegurarte un premio de quince, tendrías que hacer 14.348.907 columnas.

Se llama **dependencia** al hecho de estar vinculado a "algo" y no poder prescindir de ello. La **ludopatía** es la dependencia de los juegos de azar.