

Matemáticas orientadas a las enseñanzas aplicadas:

3º A de ESO

Capítulo 11:

Estadística y probabilidad

Propiedad Intelectual

El presente documento se encuentra depositado en el registro de Propiedad Intelectual de Digital Media Rights con ID de obra AAA-0181-02-AAA-044030

Fecha y hora de registro: 2014-05-28 17:53:18.0

Licencia de distribución: CC by-nc-sa



Queda prohibido el uso del presente documento y sus contenidos para fines que excedan los límites establecidos por la licencia de distribución.

Más información en <http://www.drights.com>



LibrosMareaVerde.tk

www.apuntesmareaverde.org.es



Autor: Fernando Blasco

Revisor: David Hierro

Ilustraciones: Banco de Imágenes de INTEF

Índice

1. LA TOMA DE DATOS

- 1.1. UN EJEMPLO PARA REALIZAR UN ANÁLISIS
- 1.2. VARIABLES ESTADÍSTICAS
- 1.3. LAS FASES DE UN ESTUDIO ESTADÍSTICO

2. REPRESENTACIÓN DE LA INFORMACIÓN

- 2.1. UN EJEMPLO PARA TRABAJAR
- 2.2. DIAGRAMA DE BARRAS
- 2.3. HISTOGRAMA DE FRECUENCIAS
- 2.4. POLÍGONO DE FRECUENCIAS
- 2.5. DIAGRAMA DE SECTORES

3. PARÁMETROS ESTADÍSTICOS

- 3.1. INTRODUCCIÓN
- 3.2. MEDIDAS DE CENTRALIZACIÓN
- 3.3. MEDIDAS DE DISPERSIÓN
- 3.4. INTERPRETACIÓN CONJUNTA DE LA MEDIA Y LA DESVIACIÓN TÍPICA
- 3.5. CÁLCULO DETENIDO DE PARÁMETROS ESTADÍSTICOS
- 3.6. DIAGRAMA DE CAJAS

4. INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES

- 4.1. CONCEPTOS BÁSICOS DE PROBABILIDAD
- 4.2. CÁLCULO DE PROBABILIDADES
- 4.3. PROBABILIDAD Y FRECUENCIA RELATIVA

Resumen

La Estadística es una Ciencia que surgió para llevar la contabilidad del Estado. De ahí viene su nombre. En el siglo XX se desarrollaron sus técnicas y se separó de las Matemáticas, pasando a ser una ciencia con entidad propia. En los medios de comunicación encontramos frecuentes estadísticas. En medicina se necesitan métodos estadísticos para probar nuevos medicamentos. En todo experimento científico, tras la recogida de datos, se necesita utilizar pruebas estadísticas que permitan sacar información de esos datos.

El origen de la Probabilidad se encuentra en los juegos de azar. Cardano, Galileo, Pascal, Fermat son algunos de los matemáticos que se ocuparon en sus inicios.

1. LA TOMA DE DATOS

1.1. Un ejemplo para realizar un análisis

Ejemplo:

✚ La Casa de la Moneda quiere estudiar cuántas monedas debe emitir, teniendo en cuenta las que están en circulación y las que se quedan atesoradas (bien en casas particulares, o en máquinas de refrescos, o depositadas en un banco). Se ha hecho una encuesta a pie de calle a 60 personas y se ha apuntado cuántas monedas llevaba cada una de ellas en el bolsillo. Hemos obtenido estos datos:

12	7	11	8	8	9	6	12	7	7	13	0	10	9	13	18	7	6	11	12	16	0	10	10	8	8	9	11	10	8
16	8	5	2	12	8	14	14	16	6	2	0	18	10	10	12	14	6	7	3	12	11	10	18	9	7	12	1	15	8

El primer paso consiste en hacer un esquema para el recuento: usaremos una tabla y marcaremos palotes cada vez que aparezca ese número.

0	///
1	/
2	//
3	/
4	
5	/
6	////

7	//// /
8	//// ///
9	////
10	//// //
11	////
12	//// //
13	//

14	///
15	/
16	///
17	
18	///
19	
20	

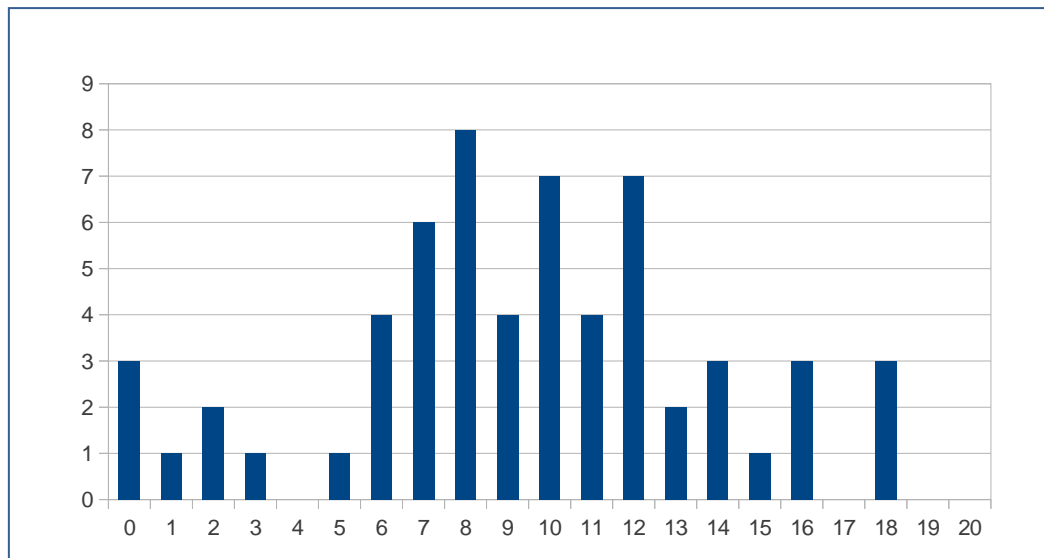
Pasar de ese recuento a una tabla de **frecuencias absolutas** es muy sencillo: solo hay que sustituir los palotes por el número que representan.

0	3
1	1
2	2
3	1
4	0
5	1
6	4

7	6
8	8
9	4
10	7
11	4
12	7
13	2

14	3
15	1
16	3
17	0
18	3
19	0
20	0

Es mucho mejor analizar los datos de modo visual. Estamos más acostumbrados a trabajar de esa manera. Podemos representar los datos de la tabla de frecuencias en un **diagrama de barras**, donde la altura de cada barra representa la frecuencia de aparición.



El procesamiento de datos estadísticos se utiliza mucho. Obviamente no se hacen las operaciones a mano, sino que se utilizan calculadoras u hojas de cálculo. Disponer de esos medios tecnológicos será un buen complemento para el capítulo, aunque recordamos que lo más importante es comprender qué se hace en cada momento.

Comenzaremos introduciendo algo de **nomenclatura**. Casi todos estos nombres los has escuchado puesto que los medios de comunicación los utilizan muchísimo

Población es el colectivo sobre el que se quiere hacer el estudio.

Muestra es un subconjunto de la población de modo que a partir de su estudio se pueden obtener características de la población completa.

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

Ejemplo:

✚ Se quiere hacer un estudio sobre hábitos alimenticios de los estudiantes de 3º de ESO de todo Madrid. Pero como es muy costoso entrevistar a todos los estudiantes se decide tomar un IES por cada distrito y entrevistar a los alumnos de 3º de ESO de esos colegios elegidos.

La **población** objeto del estudio serán todos los estudiantes madrileños matriculados en 3º de ESO.

La **muestra** son los estudiantes de 3º de ESO matriculados en los institutos elegidos.

Cada uno de los estudiantes de 3º de ESO es un **individuo** para este estudio estadístico.

Actividades propuestas

- Queremos hacer un estudio de la cantidad de monedas que llevan en el bolsillo los estudiantes de tu clase. Pero para no preguntar a todos elige 10 compañeros al azar y anota en tu cuaderno cuántas monedas lleva cada uno.
 - ¿Cuál es la población objeto del estudio?
 - ¿Cuál es la muestra elegida?
 - Especifica 5 individuos que pertenezcan a la población y no a la muestra.

1.2. Variables estadísticas

Ejemplo:

✚ En un estudio estadístico se puede preguntar cosas tan variopintas como

- ¿Qué frutas comes a lo largo de una semana?
- ¿Cuántas piezas de fruta comes al día?
- ¿Cuántas monedas llevas en el bolsillo?
- ¿Cuál es tu altura?
- ¿Cuántas marcas de chocolate recuerdas?
- ¿Cuáles son las marcas de chocolate que recuerdas?
- ¿Cuántos hermanos tienes?
- ¿Cuál es tu color favorito para un coche?
- ¿Cuánto tiempo pasas al día viendo la televisión?
- ¿Cuántos seguidores tienes en twitter?

Esas preguntas pueden corresponder a estudios de salud, económicos, publicitarios o socioeconómicos. Algunas se responden con un número y otras se responden con un nombre o un adjetivo. Incluso hay diferencias entre las que se responden con números: el número de monedas que llevas o el número de seguidores de *twitter* se contestan con números enteros, mientras que para hallar tu altura o las horas que pasas delante del televisor necesitamos utilizar números reales (normalmente con representación decimal).

Una variable se dice **cuantitativa** si sus valores se expresan con números.

Las variables cuantitativas pueden ser

- **discretas** si solo admiten valores aislados
- **continuas** si entre dos valores pueden darse también todos los intermedios

Una variable estadística es **cualitativa** cuando sus valores no se expresan mediante un número, sino con una cualidad.

Actividades propuestas

2. Clasifica en variables cualitativas y cuantitativas las que aparecen en el primer ejemplo de esta sección. Para las cuantitativas indica si son continuas o discretas.

1.3. Las fases de un estudio estadístico

En un estudio estadístico hay 6 fases fundamentales:

1. Determinación del objeto del estudio. Esto es, saber qué queremos estudiar.
2. Selección de las variables que se van a estudiar.
3. Recogida de los datos.
4. Organización de los datos.
5. Representación y tratamiento de los datos.
6. Interpretación y análisis.

En este libro empezaremos los ejemplos a partir del punto 4, con datos ya proporcionados en los enunciados.

2. REPRESENTACIÓN DE LA INFORMACIÓN

2.1. Ejemplos para trabajar

En la sección anterior lo comenzábamos analizando una variable discreta: el número de monedas que se llevan en el bolsillo. Puedes repasar qué hacíamos allí: cómo recontábamos los datos, cómo los llevábamos después a una tabla de frecuencias y cómo representábamos la información en un gráfico.

Haremos ahora el mismo proceso con una variable continua.

Ya sabes que:

Podemos distinguir entre **frecuencias absolutas**, si, como en este ejemplo, hacemos un recuento del número de veces que aparece cada dato. **Frecuencias relativas**, que estudiaremos con más detenimiento al final del capítulo, y que consiste en dividir cada frecuencia absoluta por el número total de observaciones. **Frecuencias acumuladas**, tanto frecuencias absolutas acumuladas como frecuencias relativas acumuladas si se calculan todos los valores menores o iguales a él.

Ejemplos:

- ✚ Se está realizando un control del peso de un grupo de niños. Para ello, se contabilizan el número de veces que comen al día una chocolatina 13 niños durante un mes, obteniendo los siguientes números: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2.

La información obtenida se puede resumir en una tabla de frecuencias **absolutas** y frecuencias **absolutas acumuladas**:

Valores	0	1	2	3	4	5	6	7
Frecuencia absoluta	2	2	4	1	2	1	0	1
Frecuencia absoluta acumulada	2	4	8	9	11	12	12	13

También se puede resumir en una tabla de frecuencias **relativas** y frecuencias **relativas acumuladas**:

Valores	0	1	2	3	4	5	6	7
Frecuencia relativa	0'154	0'154	0'307	0'077	0'154	0'077	0	0'077
Frecuencia relativa acumulada	0'154	0'308	0'615	0'692	0'846	0'923	0'923	1

- ✚ En una fábrica se realiza un estudio sobre el espesor, en *mm*, de un cierto tipo de latas de refresco. Con este fin, selecciona una muestra de tamaño $N = 25$, obteniendo los siguientes valores: 7'8, 8'2, 7'6, 10'5, 7'4, 8'3, 9'2, 11'3, 7'1, 8'5, 10'2, 9'3, 9'9, 8'7, 8'6, 7'2, 9'9, 8'6, 10'9, 7'9, 11'1, 8'8, 9'2, 8'1, 10'5.

Esta información se puede resumir haciendo cinco intervalos y haciendo una tabla de frecuencias absolutas, frecuencias absolutas acumuladas, frecuencias relativas y frecuencias relativas acumuladas

Intervalos de clase	(7, 8]	(8, 9]	(9, 10]	(10, 11]	(11, 12]
Marcas de clase	7'5	8'5	9'5	10'5	11'5
Frecuencia absoluta	6	8	5	4	2
Frecuencia relativa	0'24	0'32	0'2	0'16	0'08
Frecuencia relativa acumulada	0'24	0'56	0'76	0'92	1

Ejemplo:

- ✚ Las alturas de los 12 jugadores de la Selección Española de Baloncesto (en metros) que participaron en la Eurocopa 2013 se recogen en la siguiente tabla:

2'03	1'96	1'91	2'11	1'91	1'93	2'08	1'99	1'90	2'16	2'06	2'03
------	------	------	------	------	------	------	------	------	------	------	------

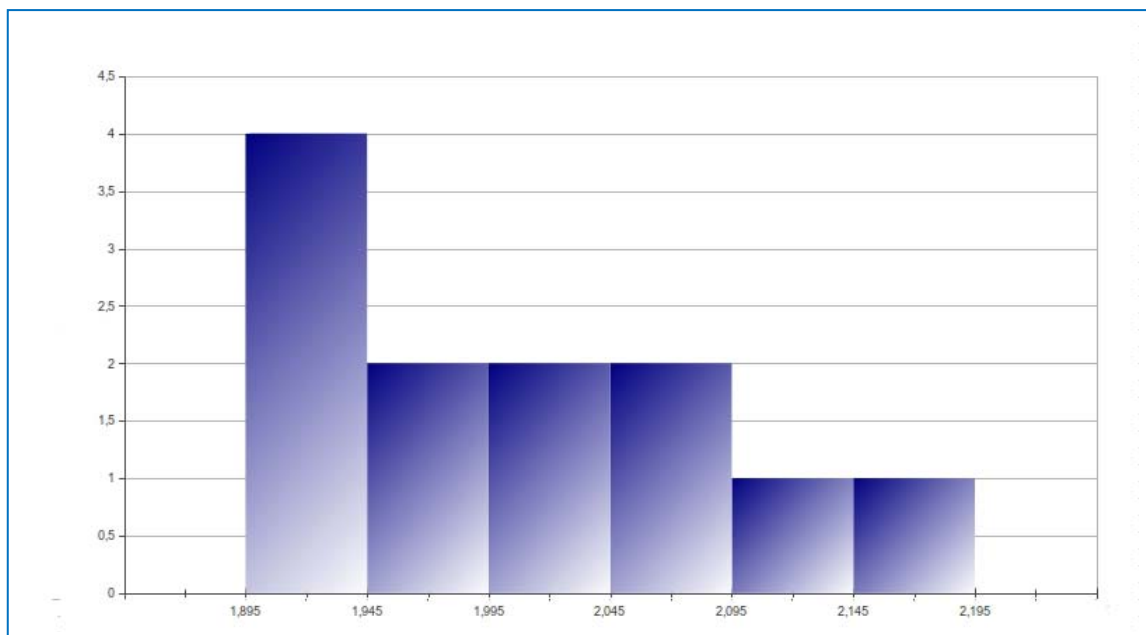
Como los datos son continuos, para hacer el recuento fijaremos **intervalos** de altura:

- entre 1'895 y 1'945 *////*
- entre 1'945 y 1'995 *//*
- entre 1'995 y 2'045 *//*
- entre 2'045 y 2'095 *//*
- entre 2'095 y 2'145 */*
- entre 2'145 y 2'195 */*

Ahora llevamos los datos del recuento a un diagrama de frecuencias:

entre 1'895 y 1'945	4
entre 1'945 y 1'995	2
entre 1'995 y 2'045	2
entre 2'045 y 2'095	2
entre 2'095 y 2'145	1
entre 2'145 y 2'195	1

En este caso la representación gráfica la hacemos con un **histograma de frecuencias**.



Observa la diferencia entre este gráfico (correspondiente a una variable continua) y el que hicimos para el recuento de monedas (que representaba una variable discreta). Este gráfico se denomina histograma de frecuencias y es similar a un diagrama de barras pero ahora representamos unas barras pegadas a otras, para recordar que se trata de intervalos de clase y no de valores aislados de las variables. En nuestro ejemplo todos los intervalos tienen la misma longitud, 0'05 cm. Si las longitudes de los intervalos fueran diferentes las alturas de los rectángulos deberían ser proporcionales al área.

2.2. Diagramas de barras

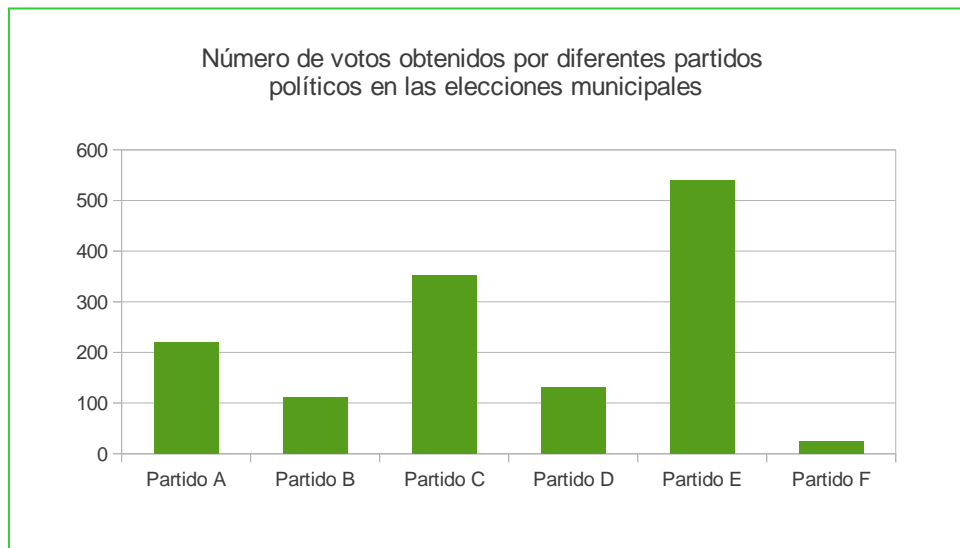
Se utiliza para representar datos de variables estadísticas discretas o variables estadísticas cualitativas.

Al principio del capítulo estudiando el número de monedas que se llevan en el bolsillo. Podemos utilizar este tipo de gráfico en otras situaciones.



El gráfico anterior representa el número de alumnos (de una clase de 35) que han aprobado todo, el de alumnos con 1 asignatura suspensa, con dos asignaturas suspensas, etc. Lo bueno de la representación gráfica es que **de un solo vistazo sabemos que 20 alumnos han aprobado todo y que hay un alumno que tiene 7 asignaturas suspensas.**

También podemos utilizar diagramas de barras para representar variables cualitativas, como la elección de la modalidad de bachillerato que cursan los alumnos de un IES o las preferencias políticas de los ciudadanos de un municipio.

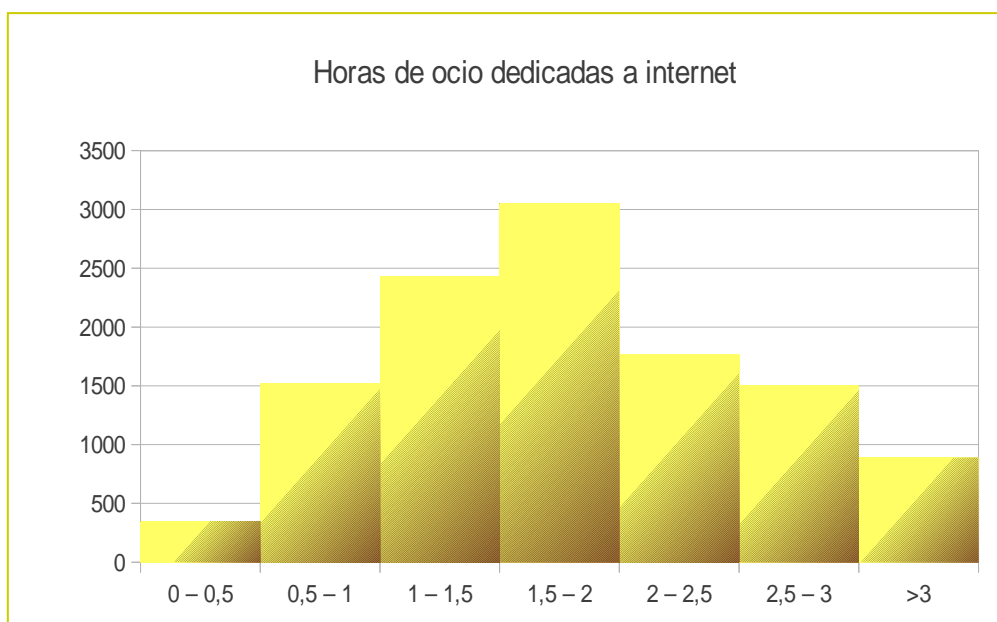


2.3. Histograma de frecuencias

Este tipo de gráfico lo hemos utilizado antes para representar las alturas de los jugadores de la Selección Española de Baloncesto.

Es similar a un diagrama de barras pero la altura de cada barra viene dada por el número de elementos que hay en cada clase.

Otras variables que podemos considerar como variables continuas son el número de horas que los jóvenes de una población dedican a internet en sus ratos de ocio o la cantidad de dinero que se lleva en el bolsillo (ojo, esto no es el número de monedas).



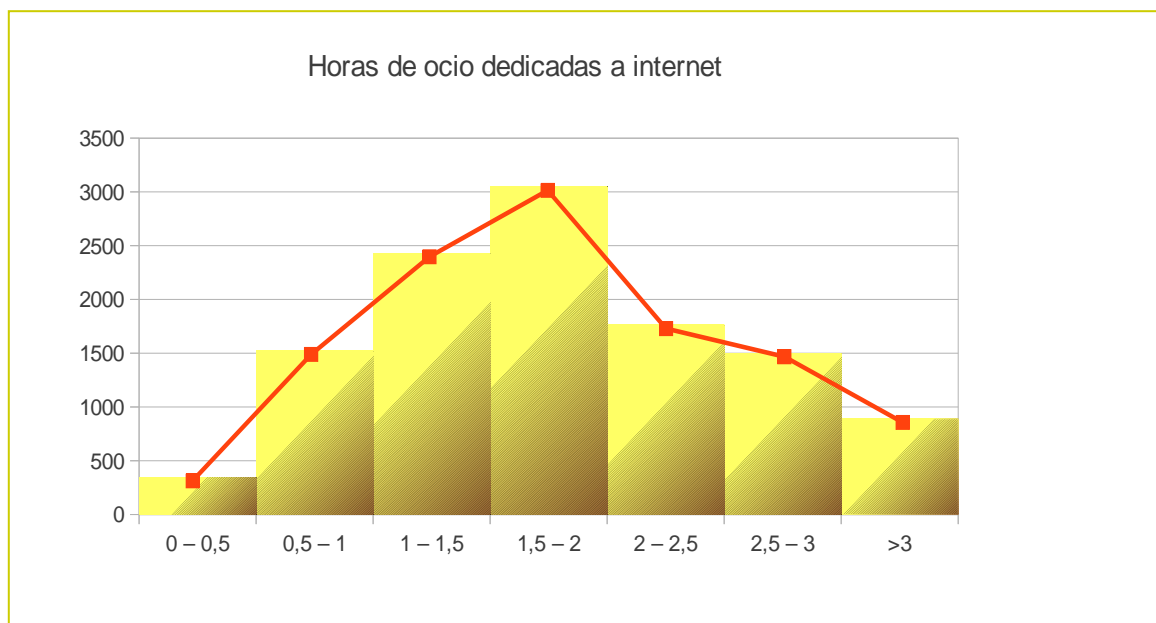
En el gráfico que incluimos a continuación las marcas del eje de las x se refieren a los tramos de dinero expresados de 5 en 5 euros. La altura del gráfico se corresponde con la cantidad de alumnos que llevan esa cantidad de dinero. De un simple vistazo se ve que hay algo más de 150 alumnos que llevan entre 5 € y 10 € al instituto y que poco más de 40 alumnos llevan entre 25 € y 30 €.



Las barras son más anchas y aparecen unas a continuación de otras para destacar que estamos representando una variable continua y que las alturas se corresponden con individuos dentro de un intervalo de datos. Pero recuerda, si los intervalos fueran distintos, las alturas de los rectángulos serían proporcionales al área.

2.4. Polígono de frecuencias

Se utiliza en los mismos casos que el histograma. Pero da idea de la variación de la tendencia. La línea poligonal se construye uniendo los puntos medios de los lados superiores de los rectángulos.

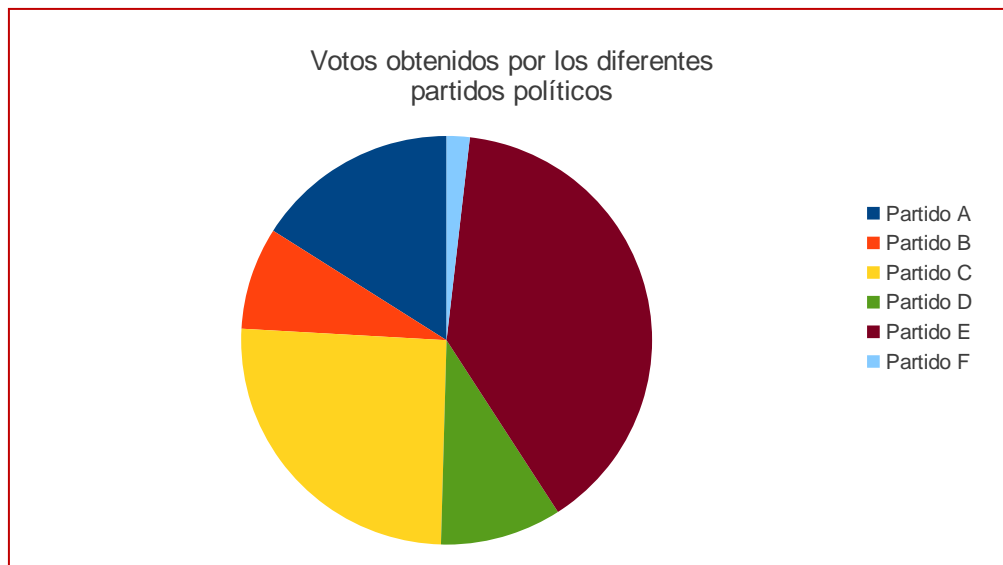


2.4. Diagrama de sectores

En algunas ocasiones nos interesa hacernos a la idea de la proporción que tiene cada resultado en relación con los demás. Se utiliza mucho con variables cualitativas. Por ejemplo, esta representación se utiliza para mostrar los resultados de unas las elecciones cuando queremos comparar los votos obtenidos por los diferentes partidos.

En un diagrama de sectores aparecen representados sectores circulares. El ángulo de estos sectores es proporcional a la frecuencia absoluta.

Retomando el ejemplo de los resultados obtenidos por diferentes partidos políticos vamos a representar esos mismos resultados mediante un diagrama de sectores:



Actividades propuestas

- Reúne a 10 amigos. Recuenta cuántas monedas de cada valor (1céntimo, 2 céntimos, 5 céntimos, ...) tenéis entre todos. Representa mediante un gráfico adecuado el número de monedas de cada clase que hay. ¿Hay algún otro diagrama que te permita ver qué tipos de monedas son más abundantes en la muestra que has tomado?
- En la clase de Educación Física el profesor ha medido el tiempo que tarda cada alumno en recorrer 100 metros. Los resultados están en esta tabla:

14'92	13'01	12'22	16'72	12'06	10'11	10'58	18'58	20'07	13'15	20'10	12'43	17'51	11'59	11'79
16'94	16'45	10'94	16'56	14'87	17'59	13'74	19'71	18'63	19'87	11'12	12'09	14'20	18'30	17'64

Agrupar estos resultados por clases' comenzando en 10 segundos y haciendo intervalos de longitud 1 segundo. Realiza una tabla de frecuencias y representa adecuadamente estos datos.

3. PARÁMETROS ESTADÍSTICOS

3.1. Introducción

Seguro que sabes qué es la media de dos números y probablemente sabes calcular la media de una serie de datos. Pero además de esa medida estadística hay otras medidas que pueden ser interesantes para conocer propiedades de los datos que tenemos.

Ahora estudiaremos las **medidas de centralización** (media, mediana y moda) que nos proporcionan un valor de referencia en torno al que se distribuyen los datos y las **medidas de dispersión** (recorrido, desviación media, varianza y desviación típica). Estas medidas nos indican cómo están de separados los datos en torno a la media.

Ejemplo:

✚ Imagina que en dos exámenes de matemáticas obtienes un 6 y un 5. La media es 5.5. Supón ahora que las notas que has tenido son 10 y 1. La media también es 5.5 pero deberás estudiarte la parte en la que has sacado 1 para recuperar. Las medidas de dispersión nos van a servir para detectar cuándo tenemos valores extremos, alejados de la media.

3.2. Medidas de centralización

La **media** se calcula sumando todos los valores y dividiendo entre el número de datos.

Si x_1, x_2, \dots, x_n son los valores que toma la variable estadística que estamos considerando, la media se representa por \bar{x} y se calcula mediante la fórmula:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Esa suma se puede escribir abreviadamente como $\bar{x} = \frac{\sum x_i}{n}$. El símbolo \sum se utiliza habitualmente para representar sumas de varios sumandos. Lo utilizarás mucho a partir de ahora.

Para calcular la **mediana** se ordenan todos los datos de menor a mayor y nos quedamos con el que ocupa la posición central. Si tenemos un número par de datos, tomamos como mediana la media de los dos números que ocupan las posiciones centrales. La representaremos por Me .

La **mediana** Me es un valor tal que el 50 % de las observaciones son inferiores a él.

Los **cuartiles** Q_1, Q_2 y Q_3 son los valores tales que el 25 %, 50 % y 75 % (respectivamente) de los valores de la variable son inferiores a él. Por tanto la mediana coincide con el segundo cuartil.

Usamos el término **moda** para referirnos al valor que más se repite. La denotamos por Mo .

Actividades resueltas

✚ Continuamos utilizando los datos de estatura correspondientes a los 12 jugadores de la Selección Española de Baloncesto (ver sección 2.1 de este capítulo).

La estatura **media** se calcula sumando todas las alturas y dividiendo entre el número de datos.

$$\sum x_i = 2'03 + 2'06 + 2'16 + 1'90 + 1'99 + 2'08 + 1'93 + 1'91 + 2'11 + 1'91 + 1'96 + 2'03 = 24'07$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{24.07}{12} = 2'0058.$$

En este ejemplo no podemos hablar de **moda**, puesto que no hay un único valor que sea el que más se repite.

La **mediana** en este caso es 2'01. Para calcularla ordenamos todos los datos de menor a mayor y nos quedamos con el que ocupa la posición central. Como en este caso tenemos un número impar de datos, tomamos como mediana la media aritmética de los 2 que ocupan las posiciones centrales.

Los datos, tras ordenarlos, quedarían así:

1'90	1'91	1'91	1'93	1'96	1'99	2'03	2'03	2'06	2'08	2'11	2'16
------	------	------	------	------	------	------	------	------	------	------	------

Media de ambos = 2'01

Para calcular los **cuartiles** tenemos que dividir el total de datos, en este ejemplo 12, entre 4, (o multiplicar por 0'25 que es lo mismo) y obtenemos 3. Luego el primer cuartil observamos que está entre 1'91 y 1'93, hacemos la media y obtenemos que $Q_1 = 1'92$. Para calcular el tercer cuartil multiplicamos por 3 y dividimos por 4, (o multiplicamos por 0'75) y en este caso se obtiene el valor que está entre 9, 2'06, y 10, 2'08, por lo que $Q_3 = 2'07$.

3.3. Medidas de dispersión

Recorrido es la diferencia entre el dato mayor y el dato menor. También se denomina **rango**.

Desviación media es la media de las distancias de los datos a la media de los datos de los que dispongamos.

$$DM = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{\sum |x_i - \bar{x}|}{n}$$

Varianza es la media de los cuadrados de las distancias de los datos a la media.

$$\text{Varianza} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n}$$

Equivalentemente (desarrollando los cuadrados que aparecen en la expresión) se puede calcular mediante esta otra expresión:

$$\text{Varianza} = \frac{\sum x_i^2}{n} - \bar{x}^2$$

Desviación típica es la raíz cuadrada de la varianza.

Se representa por σ .

$$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

Recorrido intercuartílico o **intervalo intercuartil** es la distancia entre el tercer y el primer cuartil:

$$R = \text{Recorrido intercuartílico} = Q_3 - Q_1.$$

Estas fórmulas provienen de diferentes modos de medir las distancias. Para el cálculo de la desviación media se usan valores absolutos, que es como se mide la distancia entre números en la recta real. La desviación típica tiene que ver con la forma de medir distancias en el plano (recordemos que la hipotenusa de un triángulo es la raíz cuadrada de la suma de los cuadrados de los catetos). No hace falta que comprendas ahora de dónde salen estas fórmulas pero sí es conveniente que sepas que no es por capricho de los matemáticos que lo inventaron. Cada cosa a su tiempo...

Actividades resueltas

✚ Volvemos a usar los datos del ejemplo de la Selección Española con los que venimos trabajando.

Recorrido: $2'16 - 1'90 = 0'26$ (metros). Esto es la diferencia de alturas entre el jugador más alto y el más bajo.

Para calcular la **desviación media** primero calcularemos la suma que aparece en el numerador. Después dividiremos entre el número de datos.

$$\begin{aligned} &|2'03 - 2'0058| + |2'06 - 2'0058| + |2'16 - 2'0058| + |1'90 - 2'0058| + |1'99 - 2'0058| + \\ &|2'08 - 2'0058| + |1'93 - 2'0058| + |1'91 - 2'0058| + |2'11 - 2'0058| + |1'91 - 2'0058| + \\ &|1'96 - 2'0058| + |2'03 - 2'0058| = 0'0242 + 0'0458 + 0'0958 + 0'1042 + 0'0958 + 0'0758 + 0'0742 + \\ &0'0158 + 0'1058 + 0'1542 + 0'9458 + 0'0242 = 0'87 \end{aligned}$$

Así la **desviación media** es $0'87/12 = 0'0725$

Para calcular la **varianza** primero calcularemos la suma que aparece en el numerador, de modo similar a como acabamos de hacer. Después terminaremos dividiendo entre el número de datos.

$$\begin{aligned} &(2'03 - 2'0058)^2 + (2'06 - 2'0058)^2 + (2'16 - 2'0058)^2 + (1'90 - 2'0058)^2 + (1'99 - 2'0058)^2 + \\ &(2'08 - 2'0058)^2 + (1'93 - 2'0058)^2 + (1'91 - 2'0058)^2 + (2'11 - 2'0058)^2 + (1'91 - 2'0058)^2 + \\ &(1'96 - 2'0058)^2 + (2'03 - 2'0058)^2 = 0'08934 \end{aligned}$$

Así la **varianza** es $0'08934/12 = 0'00744$

La **desviación típica** es la raíz cuadrada de la varianza: $\sigma = \sqrt{0'00744} = 0'08628$.

Recorrido intercuartílico o **intervalo intercuartil** se calcula restando $Q_3 - Q_1 = 2'07 - 1'92 = 0'15$.

Las medidas de posición nos permiten realizar otro tipo de gráfico estadístico que se llama el **gráfico de caja**.

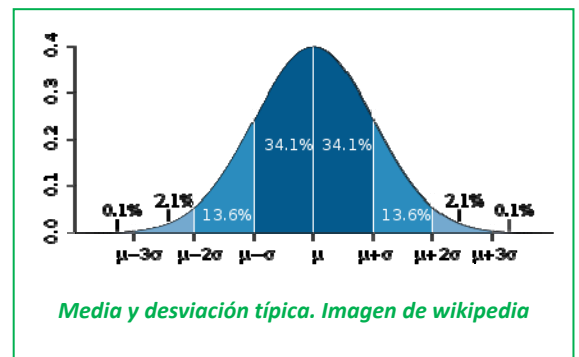
3.5. Interpretación conjunta de la media y la desviación típica

Hemos visto que la desviación típica nos mide la distancia de los datos respecto de la media. Nos da mucha información. Informa sobre cómo se agrupan los datos alrededor de la media.

La media y la desviación típica están relacionadas.

1. Aproximadamente el 68 % de los datos distan como mucho una desviación típica de la media.
2. Aproximadamente el 95 % de los datos distan como mucho dos desviaciones típicas de la media.
3. Aproximadamente más del 99 % de los datos distan como mucho tres desviaciones típicas de la media.

Si los datos que hemos recogido tuvieran una distribución normal (de momento no sabemos lo que esto significa exactamente dentro de la Estadística, pero puedes suponer que significa eso, que son normales, que no les pasa nada raro) resulta que en el intervalo entre la media menos una desviación típica y la media más una desviación típica están más del 68 % de los datos. En el intervalo entre la media menos 2 desviaciones típicas y la media más 2 desviaciones típicas están más del 95 % de los datos, y entre la media menos 3 desviaciones típicas y la media más 3 desviaciones típicas están más del 99,7 % de los datos.



Se podría decir que algo, por ejemplo la inteligencia de una persona, la altura de una planta o el peso de un animal... es normal si está dentro de ese intervalo $(\bar{x} - \sigma, \bar{x} + \sigma)$, que es inteligente, alto o pesado si está entre $(\bar{x} + \sigma, \bar{x} + 2\sigma)$, o que es un genio, gigante o muy pesado si está en el intervalo $(\bar{x} + 2\sigma, \bar{x} + 3\sigma)$.

Observa que estamos diciendo que prácticamente todos los datos distan de la media menos de tres desviaciones típicas y que más del 68 % distan menos de una desviación típica. Esto va a ser de gran utilidad pues conecta con otras ramas de la Estadística. Hasta ahora hemos estado describiendo lo que ocurre. Ahora vamos a poder tomar decisiones, inferir o predecir con una cierta probabilidad lo que va a ocurrir. Por eso vamos a estudiar a continuación las probabilidades.

3.5. Cálculo detenido de los parámetros estadísticos

Lo más cómodo para calcular parámetros estadísticos es utilizar una hoja de cálculo. Las calculadoras científicas también incorporan funciones para obtener los principales parámetros estadísticos. Para saber cómo usar tu calculadora puedes leer el manual que viene con ella.

Ahora veremos cómo se pueden utilizar las tablas de frecuencias para calcular la media y la varianza.

Cuando hay valores repetidos en vez de sumar ese valor varias veces podemos multiplicar el valor por su frecuencia absoluta. También, el número de datos es la suma de las frecuencias.

De este modo obtenemos la siguiente **fórmula para la media**

$$\bar{x} = \frac{\sum f_i \cdot x_i}{\sum f_i}$$

Análogamente, la **varianza** se puede calcular mediante

$$\text{Varianza} = \sigma^2 = \frac{\sum f_i \cdot (x_i - \bar{x})^2}{\sum f_i}$$

o, alternativamente, mediante la expresión

$$\sigma^2 = \frac{\sum f_i \cdot x_i^2}{\sum f_i} - \bar{x}^2$$

(Estas dos fórmulas son equivalentes. La segunda expresión se obtiene desarrollando los cuadrados de la primera y simplificando).

Por tanto la **desviación típica** se calcula:

$$\sigma = \sqrt{\frac{\sum f_i \cdot (x_i - \bar{x})^2}{\sum f_i}} = \sqrt{\frac{\sum f_i \cdot x_i^2}{\sum f_i} - \bar{x}^2}$$

Actividades resueltas

Las notas de 15 alumnos en un examen de matemáticas se reflejan en la siguiente tabla

7	7	6	6	10	1	4	5	5	3	9	5	5	8	6
---	---	---	---	----	---	---	---	---	---	---	---	---	---	---

Queremos calcular su media y su varianza.

En primer lugar, elaboramos una tabla de frecuencias con esos datos:

x_i	f_i
1	1
2	0
3	1
4	1
5	4
6	3
7	2
8	1
9	1
10	1

Añadimos una columna en la que escribiremos el resultado de multiplicaremos la frecuencia y el valor, esto es, $x_i \cdot f_i$.

x_i	f_i	$x_i \cdot f_i$
1	1	1
2	0	0
3	1	3
4	1	4
5	4	20
6	3	18
7	2	14
8	1	8
9	1	9
10	1	10
	$\sum f_i = n = 15$	$\sum x_i \cdot f_i = 87$

Sumando las frecuencias (columna central) obtenemos el número de datos.

Así la media es el cociente entre la suma de la columna de la derecha entre la suma de la columna central.

$$\bar{x} = \frac{87}{15} = 5,8$$

Para calcular la **varianza** añadiremos una columna más a la tabla anterior. En esa columna escribiremos el producto de la frecuencia por el cuadrado del valor.

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
1	1	1	1
2	0	0	0
3	1	3	9
4	1	4	16
5	4	20	100
6	3	18	108
7	2	14	98
8	1	8	64
9	1	9	81
10	1	10	100
$\sum f_i = n = 15$		$\sum x_i \cdot f_i = 87$	$\sum x_i^2 \cdot f_i = 577$

Así la **varianza** es $\sigma^2 = \frac{577}{12} - 5'8^2 = 14'4433$

Y la **desviación típica** es $\sigma = \sqrt{14'4433} = 3'8004$.

3.6. Diagrama de cajas o de bigotes

El **diagrama de cajas** es una representación gráfica en la que se utilizan los cuartiles, la mediana, los valores máximos y mínimos... intentando visualizar todo el conjunto de datos.

Se forma un rectángulo (o caja) cuyos lados son los cuartiles (Q_1 y Q_3) y donde se señala en el centro, la mediana (Me). Se añaden dos brazos (o *bigotes*) donde se señalan los valores máximo ($Máx$) y mínimo ($Mín$).

Se pueden calcular, además, unos límites superior e inferior. El inferior, L_i ; es $Q_1 - 1'5$ por el intervalo intercuartil, y el superior L_s es $Q_3 + 1'5$ por el intervalo intercuartil.

Ejemplo

- Nieves ha tenido en Matemáticas las siguientes notas: 8, 4, 6, 10 y 10. Calcula su recorrido, la varianza, la desviación típica, los cuartiles y el intervalo intercuartil.

Ordenamos los datos: $4 \leq 6 \leq 8 \leq 10 \leq 10$, y calculamos que:

Mediana = $Me = 8$.

$Q_1 = 6$.

$Q_3 = 10$.

Intervalo intercuartil = $10 - 6 = 4$.

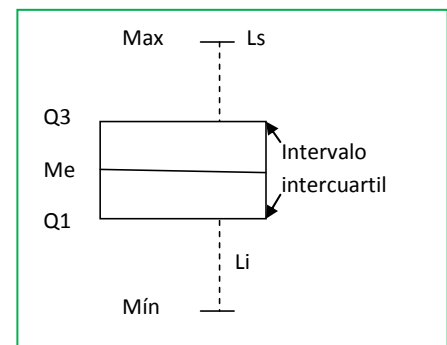
Los bigotes nos indican:

$Máx = 10$.

$Mín = 4$.

$L_s = Q_3 + 4 \cdot 1'5 = 16$.

$L_i = Q_1 - 4 \cdot 1'5 = 0$.



En este ejemplo el máximo es igual a 10, que es menor que el posible extremo superior, igual a 16. El mínimo es 4, mayor que el extremo inferior, luego no hay valores *atípicos* que sean mayores que el límite superior o menores que el límite inferior. Los extremos de los bigotes, en nuestro ejemplo son 10 y 4.

El diagrama de caja es el de la figura del margen.

4. INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES

4.1. Conceptos básicos en probabilidad

Todos los días aparecen en nuestra vida hechos que tienen que ver con la probabilidad. Si jugamos al parchís, intuimos que *más o menos* una de cada 6 veces saldrá un 5, con lo que podremos sacar una ficha a recorrer el tablero. En el 'Monopoly' sacar un doble tres veces seguidas nos manda a la cárcel ("sin pasar por la casilla de salida"). Esto no ocurre muchas veces; sin embargo, todos los que hemos jugado a esto hemos ido a la cárcel por ese motivo.

La **probabilidad** es una medida de lo factible que es que tenga lugar un determinado suceso.

Para estudiar la probabilidad, debemos introducir algunos nombres. Lo vamos a hacer con ayuda de un caso concreto.

Ejemplo

✚ Imaginemos que tenemos una bolsa con 5 bolas: 2 blancas, 2 rojas y una negra. Hacemos el siguiente **experimento aleatorio**: meter la mano en la bolsa y mirar el color de la bola que ha salido.

Hay 3 casos posibles: "que la bola sea blanca", "que la bola sea roja" o "que la bola sea negra". Abreviadamente los representaremos por *blanca*, *roja* o *negra* (también podremos representar los colores o escribir B, R o N; recuerda que en matemáticas siempre se debe simplificar, incluso la manera de escribir).

El **espacio muestral** es el conjunto de todos los casos posibles: {B, R, N}.

Los diferentes **sucesos** son los subconjuntos del espacio muestral. En nuestro ejemplo los sucesos posibles son {B}, {R}, {N}, {B,R}, {B,N}, {R,N}, {B,R,N}.

Es seguro que en nuestro experimento la bola que sacamos es "blanca", "negra" o "roja". Por eso al espacio muestral se le llama también **suceso seguro**.

Recuerda estos nombres:

Un **experimento aleatorio** es una acción (experimento) cuyo resultado depende del azar.

A cada uno de los resultados posibles de un experimento aleatorio le llamaremos **caso** o **suceso individual**.

El conjunto de todos los casos posibles se llama **espacio muestral** o **suceso seguro**.

Un **suceso** es un subconjunto del espacio muestral.

Ejemplos.

1. Baraja española de 40 cartas. Experimento: sacamos una carta al azar y miramos su palo.
Espacio muestral {oros, copas, espadas, bastos}
2. Experimento: lanzamos simultáneamente 1 moneda de euro y una de 2 euros al aire.
Espacio muestral:{Cara-Cara, Cara-Cruz, Cruz-Cara, Cruz-Cruz}
3. Experimento: lanzamos simultáneamente 2 monedas de 1 euro (indistinguibles)
Espacio muestral: {Salen 2 caras, Salen 2 cruces, Sale 1 cara y una cruz}
4. Experimento: lanzamos una moneda de 1 euro y apuntamos qué ha salido; la volvemos a lanzar y apuntamos el resultado.
Espacio muestral: {CC, CX, XC, XX}
5. Experimento: lanzamos simultáneamente dos dados y sumamos los números que se ven en las caras superiores.
Espacio muestral:{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}
6. Experimento: lanzamos un dado usual y sumamos los números que aparecen en la cara superior y la cara inferior (la que no se ve, que está sobre la mesa).
Espacio de sucesos: {7}

En los ejemplos anteriores, (2) y (4) son equivalentes: los posibles resultados del lanzamiento de 2 monedas que se distinguen son los mismos que los del lanzamiento de una misma moneda dos veces (por ejemplo, equiparamos el resultado del lanzamiento de la moneda de 1 euro del ejemplo 3 con el primer lanzamiento de la moneda del ejemplo 4 y el resultado del lanzamiento de la moneda de 2 euros con el segundo lanzamiento).

En el experimento 6 siempre sale el mismo resultado (por alguna razón los puntos en los dados usuales se distribuyen siempre de modo que las caras opuestas suman 7). Técnicamente éste no es un experimento aleatorio, puesto que el resultado no depende del azar.

Actividades propuestas

5. Para cada uno de los ejemplos 1 a 5 anteriores indica 3 sucesos diferentes que no sean sucesos individuales.
6. En una bolsa tenemos 10 bolas rojas numeradas del 1 al 10. Se hacen los dos experimentos siguientes:
EXPERIMENTO A: Se saca una bola de la bolsa y se mira su color.
EXPERIMENTO B: Se saca una bola de la bolsa y se mira su número.
¿Cuál de estos experimentos no es un experimento aleatorio? ¿Por qué?
Para el experimento que sí es un experimento aleatorio indica su espacio muestral.
7. Una baraja francesa tiene 52 cartas, distribuidas en 13 cartas de picas, 13 de corazones, 13 de tréboles y 13 de diamantes. Las picas y los tréboles son cartas negras mientras que los corazones y los diamantes son cartas rojas. Se mezcla la baraja, se corta y se hace el siguiente experimento: coger las dos cartas que han quedado arriba del todo y observar de qué color son.

Describe el espacio muestral.

4.2. Cálculo de probabilidades.

Ya hemos indicado que la probabilidad es una medida que nos indica el grado de confianza de que ocurra un determinado suceso.

La **probabilidad** se expresa mediante un número comprendido entre 0 y 1.

Si ese número está próximo a 0 diremos que es un suceso improbable (ojo, improbable no quiere decir que sea imposible), mientras que si está próximo a 1 diremos que ese suceso será mucho más probable.

Ejemplo

✚ En una bolsa que contiene 20 bolas blancas introducimos una bola negra (indistinguible al tacto). Mezclamos bien las bolas de la bolsa, y realizamos el experimento consistente en meter la mano en la bolsa y sacar una bola.

Sin que hayamos estudiado nada formalmente sobre probabilidad. ¿Qué piensas que es más probable, que la bola sacada es blanca o que es negra? Estaremos de acuerdo en que es más probable sacar una bola blanca.

Ahora ya sí que podemos plantearnos una pregunta: ¿En qué medida es más probable sacar una bola blanca?

No es difícil de calcular. Los datos que tenemos son los siguientes

- la bolsa tiene 21 bolas
- 1 bola es negra
- 20 bolas son blancas

La probabilidad de sacar la bola negra es 1 de entre 21. La probabilidad de sacar una bola blanca es de 20 entre 21.

Lo que acabamos de utilizar es conocido como **Ley de Laplace**. Si todos los casos de un espacio muestral son **equiprobables** (esto es, tienen la misma probabilidad de ocurrir), y S es un suceso de ese experimento aleatorio se tiene que

$$P(S) = \frac{\text{número de casos favorables al suceso } S}{\text{número de casos posibles}}$$

Ejemplo.

✚ Mezclamos una baraja española de 40 cartas (los palos son oros, copas, espadas y bastos y en cada palo hay cartas numeradas del 1 al 7 además de una sota, un caballo y un rey).

Se realiza el experimento consistente en *cortar la baraja y quedarnos con la carta superior*.

Consideraremos los siguientes sucesos:

- 1) Obtener una figura
- 2) Obtener una carta con un número impar
- 3) Obtener una carta de espadas
- 4) Obtener una carta de espadas o una figura
- 5) Obtener la sota de oros

En principio las cartas no van a estar marcadas, con lo que la probabilidad de que salga cada una de ellas es la misma. Esto es, estamos ante un experimento aleatorio con todos los casos equiprobables.

- 1) En la baraja hay 12 figuras (3 por cada palo). Así
 Casos favorables: 12
 Casos posibles: 40
 Probabilidad: $12/40 = 3/10$
- 2) Por cada palo hay 4 cartas con números impares: 1, 3, 5 y 7.
 Casos favorables: 16
 Casos posibles: 40
 Probabilidad: $16/40 = 2/5$
- 3) Hay 10 cartas de espadas en la baraja
 Casos favorables: 10
 Casos posibles: 40
 Probabilidad: $10/40 = 1/4$
- 4) Hay 10 cartas de espadas y además otras 9 figuras que no son de espadas (claro, las 3 figuras de espadas ya las hemos contado).
 Casos favorables: 19
 Casos posibles: 40
 Probabilidad: $19/40$
- 5) Solo hay una sota de oros
 Casos favorables: 1
 Casos posibles: 40
 Probabilidad: $1/40$

El que es capaz de calcular probabilidades rápidamente tiene ventaja en algunos juegos en los que se mezcla azar con estrategia. Por ejemplo, juegos de cartas o de dominó. Si sabemos qué cartas o fichas se han jugado podemos estimar la probabilidad de que otro jugador tenga una determinada jugada. Obviamente en esos casos no *cuantificamos* (no hacemos los cálculos exactos) pero sí que *estimamos* si tenemos la probabilidad a nuestro favor o en nuestra contra.

Para aprender más...

Jerónimo Cardano (1501-1576) fue un personaje inquieto y prolífico. Además de dedicarse a las matemáticas era médico, pero también era un jugador. De hecho él fue quien escribió el primer trabajo que se conoce sobre juegos de azar. Un siglo después el Caballero de Meré, un conocido jugador, planteó a Blas Pascal diversos problemas que le aparecían en sus partidas. Uno de los problemas que le planteó es el del reparto de las ganancias cuando una partida se tiene que interrumpir. Este problema ya había sido tratado con anterioridad por Luca Pacioli (el matemático que inventó la tabla de doble entrada para ayudar a los Medici a llevar la contabilidad de su Banca).

El problema enunciado y resuelto por Pacioli es éste:

✚ *Dos equipos juegan a la pelota de modo que gana el juego el primer equipo que gana 6 partidos. La apuesta es de 22 ducados, que se los llevará el ganador. Por algún motivo hay que interrumpir el juego cuando un equipo ha ganado 5 partidos y el otro 3. Se quiere saber cómo repartir los 22 ducados de la apuesta, de un modo justo.*

¡Piénsalo!

A pesar de haber pasado a la historia de las matemáticas, la solución que dio Pacioli a este problema hoy no se consideraría correcta por no tener en cuenta la probabilidad. ¿Qué propones tú? Este es un problema curioso, porque no tenemos todos los datos ni conocemos las probabilidades que intervienen en su resolución, pero es un bonito ejemplo para pensar en equipo y discutir sobre el tema. Decir qué es y qué no es justo es muy complicado.

Actividades resueltas

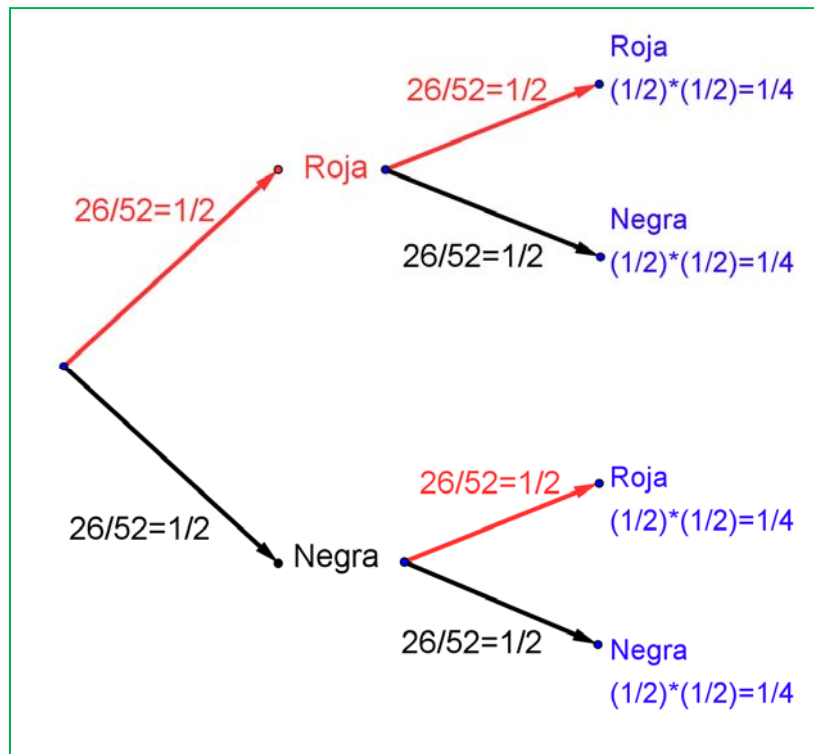
✚ *Una bolsa de bolas contiene 26 negras y 26 rojas. Se mezcla el contenido de la bolsa, se mete la mano y se saca una bola, se mira el color y se devuelve a la bolsa. A continuación se saca otra bola y se mira el color. ¿Cuál es la probabilidad de que hayan salido una bola roja y una bola negra?*

Antes de seguir leyendo, piénsalo. Si te equivocas no pasa nada: el sentido de probabilidad no lo tenemos demasiado desarrollado, pero este es el momento de hacerlo.

Este problema lo hemos planteado muchas veces a otros estudiantes. Algunos dicen que la probabilidad es $1/3$ porque hay 3 casos posibles: Roja-Roja, Negra-Negra y Roja-Negra. Esa respuesta no es correcta.

En realidad el suceso *sacar una bola de cada color* consta de 2 casos Roja-Negra y Negra-Roja. Dependiendo de cómo hubiésemos escrito el espacio muestral o de cómo hubiésemos planteado el problema ese detalle se podría ver con mayor o menor claridad.

Así, la probabilidad de sacar una bola de cada color es, en realidad $1/2$.



Si no te lo crees puedes hacer un experimento: será difícil que tengas 26 bolas negras y 26 bolas rojas, pero sí que es fácil que tengas una baraja francesa. Mézclala, corta y mira el color de la carta que ha quedado arriba en el montón. Apúntalo. Vuelve a dejar las cartas en el mazo, vuelve a mezclar, corta de nuevo y mira el color de la carta que ha quedado arriba ahora. Apunta los colores. Repite este experimento muchas veces: 20, 50 o 100.

Si tienes en cuenta los resultados verás que, aproximadamente, la mitad de las veces las dos cartas son del mismo color y la otra mitad las cartas son de colores diferentes. Con eso, hemos podido “comprobar” que la probabilidad de ese suceso era $1/2$.

Otra forma que te puede ayudar a razonar sobre este problema, y otros muchos de probabilidad, es confeccionar un **diagrama en árbol**. La primera bola que sacamos tiene una probabilidad de ser Roja igual a $26/52 = 1/2$. Ese número lo escribimos en la rama del árbol. Si devolvemos a la bolsa la bola y volvemos a sacar otra bola de la bolsa, la probabilidad de que sea Roja vuelve a ser $26/52 = 1/2$. Completamos con idéntico razonamiento el resto de las ramas.

La probabilidad de que las dos bolas que hayamos sacado sean rojas es el producto de sus ramas: $(1/2) \cdot (1/2) = 1/4$. Igual probabilidad obtenemos para los sucesos Negra-Negra, Negra-Roja y Roja-Negra. La probabilidad de Roja-Negra es por tanto $1/4$, igual a la de Negra-Roja. Como son sucesos elementales la probabilidad de que las dos bolas sean de distinto color es la suma: $1/4 + 1/4 = 1/2$.

4.3. Probabilidad y frecuencia relativa

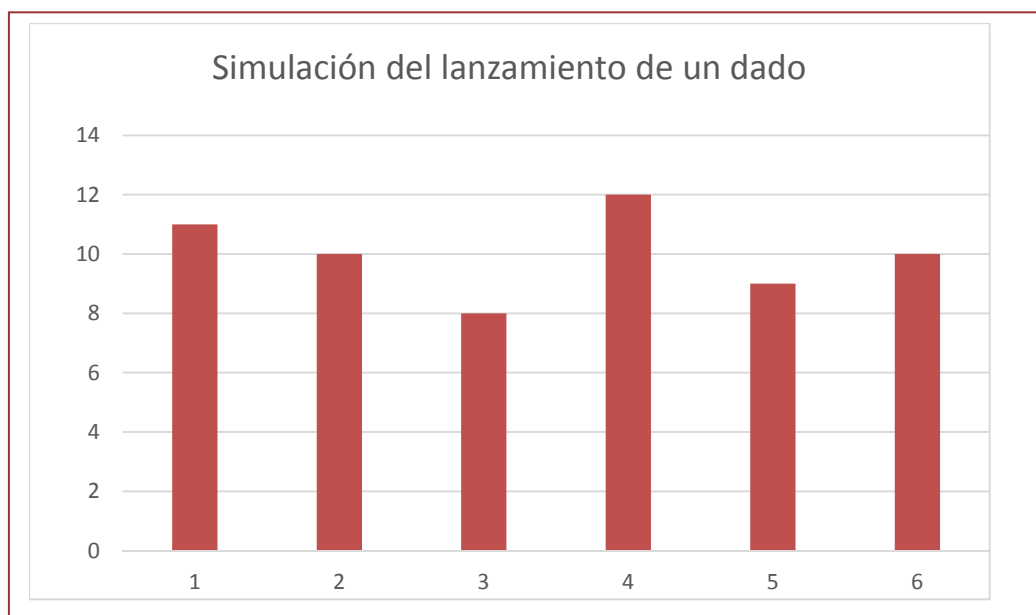
Al principio del capítulo, cuando introducíamos los principales conceptos estadísticos, hablábamos de la frecuencia. A esa frecuencia se le llama **frecuencia absoluta** para distinguirla de otro concepto, que es mucho más próximo a la probabilidad.

Llamaremos **frecuencia relativa** de un resultado de un experimento aleatorio a su frecuencia absoluta dividido entre el número de repeticiones del experimento.

Ejemplo

✚ Tira un dado 60 veces, copia esta tabla en tu cuaderno y apunta lo que sale:

Si dibujas un diagrama de barras con los resultados del experimento obtendrás algo parecido a esto:

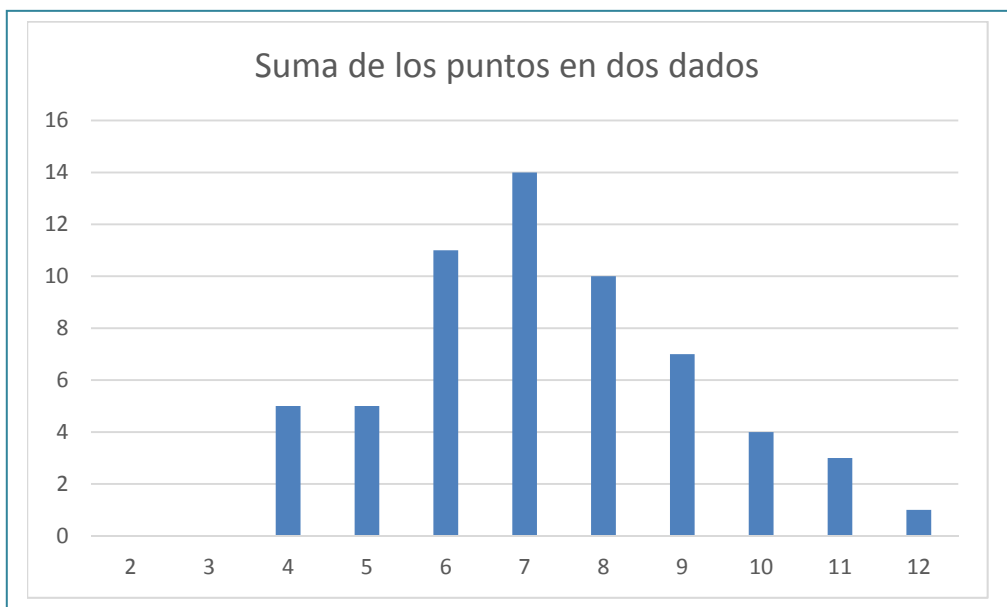


La frecuencia relativa de cada uno de los casos es bastante parecida a la probabilidad de ese caso (que es $1/6$).

Ejemplo.

✚ Haz ahora otro experimento: tira 2 dados 60 veces y apunta la suma de los valores de los dos dados en esta tabla.

Dibuja ahora un diagrama de barras. Lo que obtendrás será algo parecido a esto:



Si la probabilidad “se tiene que parecer” a las frecuencias relativas, en este caso vemos que el suceso *que la suma dé 7* es más probable que cualquiera de los demás. Y mucho más probable que *que la suma dé 2* o *que la suma dé 12*.

La **ley de los grandes números** nos dice que cuando se repite muchas veces un experimento aleatorio la frecuencia relativa de cada suceso S se aproxima a su probabilidad. Cuanto más grande sea el número de repeticiones, mejor va siendo la aproximación.

En este caso lo útil es utilizar las frecuencias relativas para estimar probabilidades cuando éstas no son conocidas.

CURIOSIDADES. REVISTA**Un problema resuelto: Las tres ruletas**

Disponemos de tres ruletas A, B y C cada una de ellas dividida en 32 sectores iguales con distintos puntos:

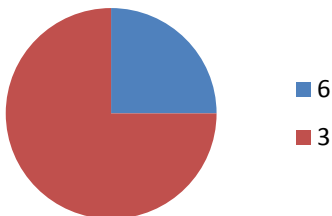
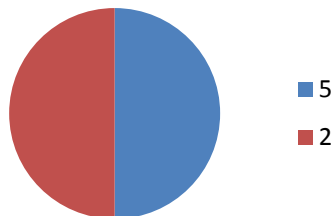
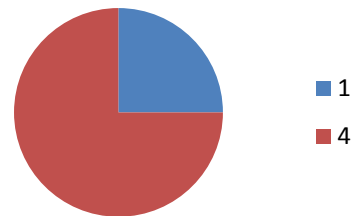
A: 8 sectores con la cifra 6 y 24 sectores con la cifra 3.

B: 16 sectores con la cifra 5 y 16 sectores con la cifra 2.

C: 8 sectores con la cifra 1 y 24 sectores con la cifra 4.

Dos jugadores seleccionan una ruleta cada uno. Gana quien obtenga mayor puntuación con la ruleta.

¿Quién tiene ventaja al elegir ruleta, la persona que elige primero o la que elige en segundo lugar?

Ruleta A**Ruleta B****Ruleta C****Solución: "Las tres ruletas"**

Haz un **diagrama de árbol** y comprueba que:

Jugando con la Ruleta A y la Ruleta B.

$$P(\text{ganar A}) = \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{2} = \frac{5}{8} \qquad P(\text{ganar B}) = \frac{25}{32} \cdot \frac{1}{2} = \frac{3}{8}$$

Gana el que juega con la Ruleta A.

Jugando con la Ruleta A y la Ruleta C.

$$P(\text{ganar A}) = \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{4} = \frac{7}{16} \qquad P(\text{ganar C}) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$$

Gana el que juega con la Ruleta C.

Jugando con la Ruleta B y la Ruleta C

$$P(\text{ganar B}) = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} = \frac{5}{8} \qquad P(\text{ganar C}) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$$

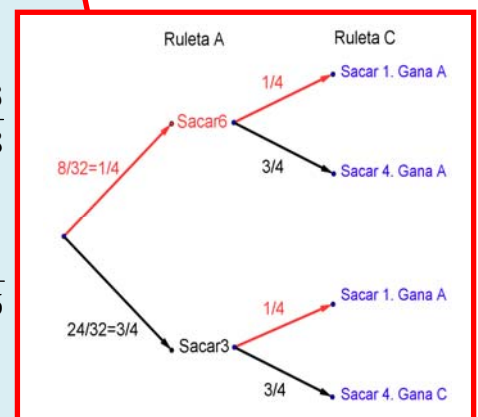
Gana el que juega con la Ruleta B.

Gana el jugador que elige en segundo lugar:

Si el primero elige la Ruleta A → El segundo elige la Ruleta C y gana.

Si el primero elige la Ruleta B → El segundo elige la Ruleta A y gana

Si el primero elige la Ruleta C → El segundo elige la Ruleta B y gana



Breve historia de la Probabilidad

Jerónimo Cardano (1501-1576) fue un personaje inquieto y prolífico. Además de dedicarse a las matemáticas era médico, pero también era un jugador. De hecho él fue quien escribió el primer trabajo que se conoce sobre juegos de azar.

Un siglo después el Caballero de Méré le planteó a Blaise Pascal algunos problemas sobre **juegos** como el siguiente:

✚ *Un jugador intenta obtener un 1 en 8 lanzamiento sucesivos de un dado, pero el juego se interrumpe después de 3 lanzamientos fallidos. ¿En qué proporción ha de ser compensado el jugador?*

Pascal escribió a Fermat sobre este problema y la correspondencia intercambiada se puede considerar como el inicio de la Teoría de Probabilidades, pero no publicaron por escrito sus conclusiones. Este problema ya había sido tratado con anterioridad por Luca Pacioli (el matemático que inventó la tabla de doble entrada para ayudar a los Medici a llevar la contabilidad de su Banca).

Huygens en 1657 publicó un breve escrito “*Los juegos de azar*” donde narra dicha correspondencia.

Pero el primer libro sobre Probabilidad es de 1713 de Jacques Bernoulli, “*El arte de la conjetura*”.

En él se enuncia la **ley de los grandes números** que viene a decir que *la probabilidad de un suceso se acerca a las frecuencias relativas cuando el número de experimentos es grande*. Conocer esto llevó a grandes jugadores a ganar en el Casino de Montecarlo, como se narra más abajo.

La Estadística y La Probabilidad se usaron en problemas sociales como defender la **vacunación de la viruela**, la educación pública... en la Ilustración Francesa.

Hasta aquí, ya sabes resolver todos los problemas históricos. Pero hay otros más difíciles, que requieren más conocimientos de Matemáticas, como el de la **aguja de Buffon**, que se ha utilizado para calcular cifras de π :

La ruleta

William Jagers llegó a Montecarlo con unos pocos francos en el bolsillo y, durante un mes anotó los números que salían en cada ruleta, y en cuatro días ganó dos millones cuatrocientos mil francos. *Jagers* consiguió quebrar a la banca en *Montecarlo* analizando las frecuencias relativas de cada número de la ruleta y observando que se había desgastado algo del mecanismo de una de ellas, con lo que todos los valores no tenían igual probabilidad. Apostó a los números más probables y ganó.



Luca Pacioli

Luca Pacioli (1445 – 1517), de nombre completo **Fray Luca Bartolomeo de Pacioli** o **Luca di Borgo San Sepolcro**, cuyo apellido también aparece escrito como **Paccioli** y **Paciolo** fue un fraile franciscano y matemático italiano, precursor del cálculo de probabilidades. Ya hemos hablado de él en estas revistas por sus trabajos sobre la proporción áurea o divina proporción como él la llamó.



Escribió un libro con 36 capítulos sobre **contabilidad** donde utiliza la partida doble o tabla de doble entrada para ayudar a los Medici a llevar la contabilidad de su Banca, define sus reglas, tales como no hay deudor sin acreedor, o que la suma de lo que se adeuda debe ser igual a lo que se abona. No fue su inventor, pero sí su divulgador.



Ducado

El problema enunciado y resuelto por Pacioli es éste:

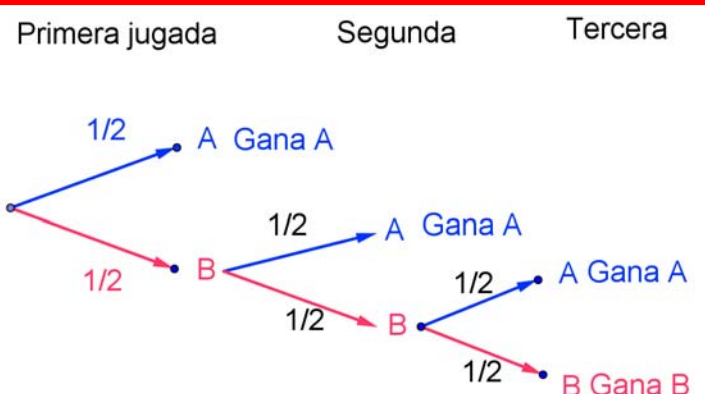
✚ Dos equipos juegan a la pelota de modo que gana el juego el primer equipo que gana 6 partidos. La apuesta es de 22 ducados, que se los llevará el ganador. Por algún motivo hay que interrumpir el juego cuando un equipo ha ganado 5 partidos y el otro 3. Se quiere saber cómo repartir los 22 ducados de la apuesta, de un modo justo.

Luca sabía de proporciones, y la solución que dio hoy no se considera válida. ¡No sabía probabilidades! Pero tú, sí.

Partimos de la hipótesis de que cada uno de los jugadores tiene la misma probabilidad de ganar: $1/2$. Llamamos A al jugador que ya ha ganado 5 partidos y B al que lleva ganadas 3.

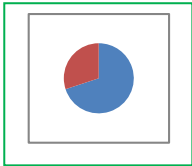
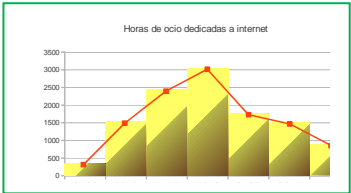
Si hicieran una nueva partida podría ganar A con probabilidad $1/2$ o B con igual probabilidad. Si gana A ya se lleva la bolsa. Si gana B entonces B llevaría 4 jugadas ganadas y A 5. Se continúa el juego. Puede ganar A o B. Observa el diagrama de árbol.

La probabilidad de que gane B es $(1/2) \cdot (1/2) \cdot (1/2) = 1/8$, y la de que gane A es $7/8$.



¿Cómo repartirías los 22 ducados?

RESUMEN

Población	Colectivo sobre el que se hace el estudio	Estudiantes de todo Madrid
Muestra	Subconjunto de la población que permita obtener características de la población complete.	Alumnos se 3º de ESO seleccionados
Individuo	Cada uno de los elementos de la población o muestra	Juan Pérez
Variabes estadística	Cuantitativa discreta Cuantitativa continua Cualitativa	Número de pie que calza Estatura Deporte que practica
Gráficos estadísticos	Diagrama de barras Histograma de frecuencias Polígono de frecuencias Diagrama de sectores	 
Media	$\bar{x} = \frac{\sum x_i}{n} = (x_1 + x_2 + \dots + x_n)/n$	Con los datos: 8, 2, 5, 10 y 10 $Media = 35/5 = 7$
Moda	Es el valor más frecuente	$Mo = 10$
Mediana	Deja por debajo la mitad	$4 < 6 < 8 < 10 = 10$. $Me = 8$.
Rango o recorrido	Es la diferencia entre el dato mayor y el dato menor.	$10 - 2 = 8$
Desviación media	Es la media de las distancias de los datos a la media de los datos de los que dispongamos.	$(8-7 + 2-7 + 5-7 + 10-7 + 10-7)/5$ $= (1+5+2+3+3)/5 = 14/5 = DM$
Varianza	Es la media de los cuadrados de las distancias de los datos a la media: $\frac{\sum_{i=1}^n (x_i - m)^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - m^2$	$V = (1 + 25 + 4 + 9 + 9)/5 = 47/5 = 9,4$
Desviación típica	Es la raíz cuadrada de la varianza= $\sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - m^2}$	$\sigma = \sqrt{47/5} = 3,06$
Probabilidad	Valor entre 0 y 1 que nos da una medida de lo factible que sea que se verifique un determinado suceso.	$P(3) = 1/6$ al tirar un dado
Espacio muestral	El conjunto de todos los casos posibles	{1, 2, 3, 4, 5, 6}
Suceso	Subconjunto del espacio muestral	Sacar par: {2, 4, 6}
Ley de Laplace.	$P(S) = \frac{\text{número de casos favorables al suceso } S}{\text{número de casos posibles}}$	$P(\text{par}) = 3/6 = 1/2$.

EJERCICIOS Y PROBLEMAS**Estadística**

1. Se han recogido los datos sobre el número de hijos que tienen 20 matrimonios. ¿Cómo es la variable utilizada? Escribe una tabla de frecuencias de los datos recogidos y representa los datos en un diagrama de sectores:

3, 1, 1, 2, 0, 2, 3, 1, 1, 1, 1, 0, 3, 2, 1, 2, 1, 2, 2, 3.

2. Con los datos del problema anterior calcula la media, la mediana, la moda y los cuartiles.
 3. Con los datos del problema anterior calcula el rango, la desviación media, la varianza, la desviación típica y el intervalo intercuartílico.
 4. Representa esos datos en un diagrama de cajas.
 5. La siguiente tabla expresa las estaturas, en metros, de 1000 soldados:

Talla	1,50 – 1,56	1,56 – 1,62	1,62 – 1,68	1,68 - 1,74	1,74 - 1,80	1,80-1,92
Nº de soldados	10	140	210	340	210	90

- a) Representa los datos en un histograma.
 b) Calcula la media y la desviación típica.
 c) Determina el intervalo donde se encuentran la mediana.
 6. Se pregunta a un grupo de personas por el número de televisores que hay en su hogar y los resultados son:

Número de televisores	0	1	2	3	4	5
Número de hogares	2	27	15	4	2	1

¿Qué tipo de variables es? Representa los datos en la representación que te parezca más adecuada.

Calcula la media y la desviación típica-

7. Con los datos del problema anterior calcula la mediana y el intervalo intercuartílico.
 8. En un centro escolar se ha recogido información sobre el número de ordenadores en las casas de 100 familias y se han obtenido los siguientes resultados:

Número ordenadores	0	1	2	3	4
Número de familias:	24	60	14	1	1

Representa los datos en un diagrama de barras y calcula la media, la mediana y la moda.

9. Con los datos del problema anterior calcula el rango, la desviación media, la varianza y la desviación típica. Haz un diagrama de cajas.
 10. Se pregunta a un grupo de personas por el número de veces que han visitado al dentista en el último año. Las respuestas obtenidas se recogen en la siguiente tabla:

Número de visitas:	1	2	3	4	5
Número de personas:	13	18	7	5	7

Representa los datos en un diagrama de sectores y calcula la media, la mediana y la moda.

11. Se pregunta a un grupo de personas por el número de veces que han visitado al dentista en el último año. Las respuestas obtenidas se recogen en la siguiente tabla:

Número de visitas:	1	2	3	4	5
Número de personas:	13	18	7	5	7

Calcula el rango, la desviación media, la varianza y la desviación típica.

12. En las elecciones de 2014 al Parlamento Europeo se obtuvieron los siguientes escaños por grupo parlamentario (DM: demócrata – cristianos; S: socialistas; L: Liberales; V: verdes; C: conservadores; I: izquierda unitaria; LD: Libertad y democracia; NI: No inscritos; Otros).

Partidos	DM	S	L	V	C	I	LD	NI	Otros	Total
Escaños	213	190	64	52	46	42	38	41	65	751

¿Qué representación de los datos te parece más adecuada? ¿Puedes calcular la media o el rango? ¿Qué tipo de variables es la de la tabla?

13. En las elecciones de 2014 al Parlamento Europeo se obtuvieron los siguientes escaños por alguno de los estados miembro:

Estado	Alemania	España	Francia	Italia	Polonia	Reino Unido	Portugal	Grecia	Otros	Total
Escaños	96	54	74	73	51	73	21	21		751

¿Qué representación de los datos te parece más adecuada? ¿Puedes calcular la media o el rango? ¿Qué tipo de variables es la de la tabla? Determina el número de escaños de los otros países miembros de la Unión Europea.

14. En las elecciones de 2004, 2009, 2014 al Parlamento Europeo se obtuvieron los siguientes porcentajes de votos por algunos de los estados miembros:

Estado	Alemania	España	Francia	Italia	Reino Unido	Portugal	Grecia	Bélgica	% total
2004	43	45'14	42'76	71'72	38'52	38'6	63'22	90'81	45'47
2009	43'27	44'87	40'63	65'05	34'7	36'77	52'61	90'39	43
2014	47'6	45'9	43'5	60	36	34'5	58'2	90	43'09

¿Qué representación de los datos te parece más adecuada? ¿Puedes calcular la media o el rango? ¿Qué tipo de variables es la de la tabla? Ordena a los países de mayor a menos porcentaje de votantes en las elecciones de 2014.

15. Con los datos del problema anterior sobre las elecciones de 2004' 2009' 2014 al Parlamento Europeo se obtuvieron los siguientes porcentajes de votos por algunos de los estados miembros:

Estado	Alemania	España	Francia	Italia	Reino Unido	Portugal	Grecia	Bélgica	% total
2004	43	45'14	42'76	71'72	38'52	38'6	63'22	90'81	45'47
2009	43'27	44'87	40'63	65'05	34'7	36'77	52'61	90'39	43
2014	47'6	45'9	43'5	60	36	34'5	58'2	90	43'09

Representa en un polígono de frecuencias los porcentajes de participación del total de los estados miembros.

16. Con los datos del problema anterior sobre las elecciones de 2004, 2009, 2014 al Parlamento Europeo se obtuvieron los siguientes porcentajes de votos por algunos de los estados miembros:

Estado	Alemania	España	Francia	Italia	Reino Unido	Portugal	Grecia	Bélgica	% total
2004	43	45'14	42'76	71'72	38'52	38'6	63'22	90'81	45'47
2009	43'27	44'87	40'63	65'05	34'7	36'77	52'61	90'39	43
2014	47'6	45'9	43'5	60	36	34'5	58'2	90	43'09

Separa los Estados Miembros en dos grupos, los que tuvieron un porcentaje superior al porcentaje medio y los que lo tuvieron menor en 2004. Haz lo mismo para 2014. ¿Son los mismos? Analiza el resultado.

17. Con los datos del problema anterior sobre las elecciones de 2004, 2009, 2014 al Parlamento Europeo se obtuvieron los siguientes porcentajes de votos por algunos de los estados miembros:

Estado	Alemania	España	Francia	Italia	Reino Unido	Portugal	Grecia	Bélgica	% total
2004	43	45'14	42'76	71'72	38'52	38'6	63'22	90'81	45'47
2009	43'27	44'87	40'63	65'05	34'7	36'77	52'61	90'39	43
2014	47'6	45'9	43'5	60	36	34'5	58'2	90	43'09

Calcula el porcentaje de participación medio para Alemania en esas tres convocatorias y la desviación típica. Lo mismo para España, para Bélgica y para Portugal.

18. En las elecciones de 2014 al Parlamento Europeo los resultados de España han sido:

Censo	Total de votantes	Abstención	Votos nulos	Votos en blanco
35.379.097	15.920.815	19.458.282	290.189	357.339

Representa en un diagrama de sectores estos datos. Haz una tabla de porcentajes: el censo es el 100 %. Determina los otros porcentajes. ¿Consideras que ha ganado la abstención?

19. En las elecciones de 2014 al Parlamento Europeo los resultados de España han sido:

PP	PSOE	Izquierda plural	Podemos	UPyD	Otros	Total de votantes
4.074.363	8.001.754	1.562.567	1.245.948	1.015.994		15.920.815

Determina el número de votos de los otros partidos. Representa en un diagrama de barras estos datos. Haz una tabla de porcentajes para cada partido. Tienes que distribuir 54 escaños, ¿cómo los distribuirías por partidos?

Probabilidad

20. Se considera el experimento aleatorio de tirar un dado dos veces. Calcula las probabilidades siguientes:
- Sacar algún 1.
 - La suma de los dígitos es 8.
 - No sacar ningún 2.
 - Sacar algún 1 o bien no sacar ningún 2.
21. Se considera el experimento aleatorio sacar dos cartas de la baraja española. Calcula la probabilidad de:
- Sacar algún rey.
 - Obtener al menos un basto.
 - No obtener ningún basto.
 - No obtener el rey de bastos.
 - Sacar alguna figura: sota, caballo, rey o as.
 - No sacar ninguna figura.
22. Se considera el experimento aleatorio de tirar una moneda tres veces. Calcula las probabilidades siguientes:
- Sacar cara en la primera tirada.
 - Sacar cara en la segunda tirada.
 - Sacar cara en la tercera tirada.
 - Sacar alguna cara.
 - No sacar ninguna cara.
 - Sacar tres caras.
23. Con una baraja española se hace el experimento de sacar tres cartas, con reemplazo, ¿cuál es la probabilidad de sacar tres reyes? Y si el experimento se hace sin reemplazo, ¿cuál es ahora la probabilidad de tener 3 reyes?
24. En una urna hay 6 bolas blancas y 14 bolas negras. Se sacan dos bolas con reemplazo. Determina la probabilidad de que:
- Las dos sean negras.
 - Haya al menos una negra.
 - Ninguna sea negra.
25. En una urna hay 6 bolas blancas y 14 bolas negras. Se sacan dos bolas sin reemplazo. Determina la probabilidad de que:
- Las dos sean negras.
 - Haya al menos una negra.
 - Ninguna sea negra.
 - Compara los resultados con los de la actividad anterior.

- 26.** Al lanzar cuatro monedas al aire,
- ¿cuál es la probabilidad de que las cuatro sean caras?
 - ¿Cuál es la probabilidad de obtener a lo sumo tres caras?
 - ¿Cuál es la probabilidad de tener exactamente 3 caras?
- 27.** Dos tiradores al plato tienen unas marcas ya conocidas. El primero acierta con una probabilidad de 0,7 y el segundo de 0,5. Se lanza un plato y ambos disparan. Expresa mediante un diagrama de árbol y las distintas posibilidades: a) ¿Qué probabilidad hay de que uno de los tiradores dé en el plato? b) Calcula la probabilidad de que ninguno acierte. c) Calcula la probabilidad de que los dos acierten.
- 28.** Se lanza una moneda hasta que aparezca cara dos veces seguidas. a) Calcula la probabilidad de que la experiencia termine en el segundo lanzamiento. b) Calcula la probabilidad de que termine en el tercer lanzamiento.
- 29.** En el lanzamiento de naves espaciales se han instalado tres dispositivos de seguridad A, B y C. Si falla A se pone automáticamente en marcha el dispositivo B, y si falla este, se pone en marcha C. Se sabe que la probabilidad de que falle A es 0,1, la probabilidad de que B funcione es 0,98 y la probabilidad de que falle C es 0,05. Calcula la probabilidad de que todo funcione bien.
- 30.** Se hace un estudio sobre los incendios forestales de una zona y se comprueba que el 40 % son intencionados, el 50 % se deben a negligencias y el 10 % a causas naturales. Se han producido tres incendios, a) ¿cuál es la probabilidad de que al menos uno haya sido intencionado? b) Probabilidad de que los tres incendios se deban a causas naturales. c) Probabilidad de que ningún incendio sea por negligencias.
- 31.** Se lanza dos veces un dado equilibrado con seis caras. Hallar la probabilidad de que la suma de los valores que aparecen en la cara superior sea múltiplo de tres.
- 32.** Se sabe que se han eliminado varias cartas de una baraja española que tiene cuarenta. La probabilidad de extraer un as entre las que quedan 0,12, la probabilidad de que salga una copa es 0,08 y la probabilidad de que no sea ni as ni copa es 0,84.
- Calcular la probabilidad de que la carta sea el as de copas. ¿Se puede afirmar que entre las cartas que no se han eliminado está el as de copas?
- 33.** Una persona despistada tiene ocho calcetines negros, seis azules y cuatro rojos, todos ellos sueltos. Un día con mucha prisa, elige dos calcetines al azar. Hallar la probabilidad de:
- que los calcetines sean negros.
 - que los dos calcetines sean del mismo color.
 - que al menos uno de ellos sea rojo.
 - que uno sea negro y el otro no.
- 34.** Tres personas viajan en un coche. Si se supone que la probabilidad de nacer en cualquier día del año es la misma y sabemos que ninguno ha nacido en un año bisiesto,
- hallar la probabilidad de que solamente una de ellas celebre su cumpleaños ese día.
 - calcular la probabilidad de que al menos dos cumplan años ese día.

AUTOEVALUACIÓN

- Se hace un estudio sobre el color que prefieren los habitantes de un país para un coche. La variable utilizada es:
 - cuantitativa
 - cualitativa
 - cuantitativa discreta
 - cuantitativa continua
- En un histograma de frecuencias la altura de los rectángulos es:
 - proporcional al área
 - igual a la frecuencia absoluta
 - proporcional a la frecuencia relativa
 - proporcional a la frecuencia acumulada
- Ana ha obtenido en Matemáticas las siguientes notas: 7, 8, 5, 10, 8, 10, 9 y 7. Su nota media es de:
 - 7,6
 - 8,2
 - 8
 - 9
- En las notas anteriores de Ana la mediana es:
 - 9
 - 8
 - 7,5
 - 8,5
- En las notas anteriores de Ana la moda es:
 - 10
 - 8
 - 7
 - 7, 8 y 10
- El espacio muestral de sucesos elementales equiprobables del experimento “tirar dos monedas y contar el número de caras” es:
 - {2C, 1C, 0C}
 - {CC, CX, XC, XX}
 - {XX, XC, CC}
 - {CC, CX, XC, CC}
- Tiramos dos dados y contamos los puntos de las caras superiores. La probabilidad de que la suma sea 7 es:
 - 1/6
 - 7/36
 - 5/36
 - 3/36
- Al sacar una carta de una baraja española (de 40 cartas), la probabilidad de que sea un oro o bien un rey es:
 - 14/40
 - 13/40
 - 12/40
 - 15/40
- En una bolsa hay 7 bolas rojas, 2 negras y 1 bola blanca. Se sacan 2 bolas. La probabilidad de que las dos sean rojas es:
 - 49/100
 - 42/100
 - 49/90
 - 7/15
- Tiramos tres monedas al aire. La probabilidad de que las tres al caer sean caras es:
 - 1/5
 - 1/7
 - 1/8
 - 1/6