CAPÍTULO 7: ESTADÍSTICA. AZAR Y PROBABILIDAD

1. ESTADÍSTICA

1.1. Muestras. Estudios estadísticos

Si queremos hacer un estudio estadístico tenemos que:

- *Recoger los datos*
- *Describir esos datos con tablas y gráficas, cálculo de parámetros estadísticos….*
- *Extraer conclusiones.*

Para recoger los datos y determinar los valores de la variable se puede utilizar a toda la población, todo el universo sobre el que se realiza el estudio, o hacer una muestra. En muchas ocasiones no es conveniente recoger valores de toda la población, porque es complicado o demasiado costoso, o incluso porque es imposible como en el caso de un control de calidad en que se destruya el objeto a analizar. La parte de la Estadística que se ocupa de cómo seleccionar adecuadamente las muestras se denomina **Teoría de Muestras**.

Población o universo es todo el conjunto de individuos sobre el que se realiza el estudio. Una **muestra** es un subconjunto representativo de esa población. Cada uno de los elementos de la población es un **individuo**.

Las características de la población que se estudian se denominan **variables estadísticas**, que se clasifican en **cuantitativas** y **cualitativas** según que los valores que tomen sean o no numéricos. Las variables cuantitativas que toman valores aislados se denominan **variables discretas** y las que pueden tomar cualquier valor de un intervalo de la recta real, **variables continuas**.

La parte de la Estadística que ordena, analiza y representa un conjunto de datos para describir sus características se denomina **Estadística Descriptiva**.

Para extraer conclusiones se utilizan las probabilidades y la parte de la Estadística que se ocupa de ello es la **Inferencia Estadística**.

Ejemplos:

- *Si queremos conocer las preferencias en deportes del alumnado de 4º, es posible preguntar a toda la población (alumnado de 4º), aunque es adecuado elegir una muestra representativa, seleccionando a algunos estudiantes.*
- *En este estudio sobre preferencias deportivas, la variable utilizada es cualitativa.*
- *Para conocer la intención de voto ante unas elecciones europeas, municipales, autonómicas… se utilizan muestras, pues preguntar a toda la población sería muy costoso (y eso ya se hace en las elecciones). La variable en este caso también es cualitativa.*
- *Para estudiar lo que más preocupa a una población: paro, terrorismo, corrupción… también se utilizan muestras. En este caso sería muy costoso preguntar a toda la población, aunque sería factible. La variable en este caso también es cualitativa.*
- *Pero si una fábrica quiere conocer las horas de vida útil de una bombilla, una nevera, un camión… no puede poner a funcionar a toda la población, (todas las bombillas o neveras o camiones…) hasta que se estropeen pues se queda sin producción. En este caso es imprescindible seleccionar una muestra. La variable en este caso es cuantitativa, y el tiempo toma cualquier valor, es una variable cuantitativa continua.*
- *Si preguntamos por el número de hermanos es una variable cuantitativa discreta.*
- *En control de calidad se hacen estudios estadísticos y se toman muestras.*

Actividades propuestas

1. Queremos realizar un estudio estadístico sobre el tiempo dedicado al estudio por el alumnado de ESO de Madrid. Para ello se seleccionan adecuadamente 100 alumnos. Indica cuál es la población, cuál la muestra, qué tamaño tiene la muestra y quién sería un individuo.
2. Quieres pasar una encuesta para conocer, lo mismo que en el problema anterior, el tiempo dedicado al estudio, en este caso el de los compañeros y compañeras de tu centro escolar. ¿Se la pasarías sólo a las chicas? ¿Sólo a los chicos? ¿Preguntarías a los mejores de la clase? ¿A los de peores notas? Indica el criterio que seguirías para seleccionar la muestra a la que preguntar.

1.2. Variable discreta. Tablas y gráficos

Tablas

Al hacer un estudio estadístico o realizar un experimento aleatorio la información obtenida se resume en una tabla o distribución de frecuencias.

Posibles resultados	Frecuencia absoluta
Les gusta	28
No les gusta	12
Total	40

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Ejemplo:

- *Preguntamos a 40 estudiantes de 4º si les gusta, o no, el fútbol. En la tabla del margen reflejamos los resultados.*

Es una tabla de frecuencias absolutas. Al dividir la frecuencia absoluta entre el número total tenemos la frecuencia relativa, así la frecuencia relativa de los que les gusta el fútbol es 28/40 = 0,7, y la de los que no les gusta el futbol es 12/40 = 3/10 = 0,3.

Posibles resultados	Frecuencias relativas	Porcentaje
Les gusta	0,7	70
No les gusta	0,3	30
Suma total	1	100

La **frecuencia absoluta** es el número de veces que se ha obtenido ese resultado.

La **frecuencia relativa** se obtiene dividiendo la frecuencia absoluta entre el número total de datos.

La suma de las frecuencias relativas es siempre igual a 1.

Multiplicando por 100 se obtienen los porcentajes.

Actividad resuelta

- *Se han obtenido los datos sobre el número de visitas que se han hecho de los Textos Marea Verde de Matemáticas en los meses indicados, y se han reflejado en una tabla. Haz una tabla de frecuencias absolutas, relativas y porcentajes, de frecuencias acumuladas absolutas y de frecuencias relativas acumuladas.*

Marea verde	Frecuencias absolutas	Frecuencias relativas	Porcentajes	Frecuencias acumuladas absolutas	Frecuencias acumuladas relativas
Septiembre	1834	0,51	51	1834	0,52
Octubre	956	0,26	26	2790	0,77
Noviembre	432	0,12	12	3222	0,89
Diciembre	389	0,11	11	3611	1
TOTAL	3611	1	100		

Observa que las **frecuencias acumuladas** se obtienen sumando la frecuencia anterior e indica, en este ejemplo, el número de visitas hasta ese momento.

Actividades propuestas

3. Copia en tu cuaderno y completa la siguiente tabla de frecuencias absolutas de los valores obtenidos al tirar un dado con las frecuencias relativas y porcentajes, y con frecuencias acumuladas absolutas y frecuencias relativas acumuladas.

Resultados	Frecuencias absolutas
1	17
2	12
3	17
4	15
5	21
6	14

Gráficos estadísticos

Las representaciones gráficas ayudan a comprender el significado de los datos.

Dada una tabla de frecuencias (absolutas, relativas, porcentajes, acumuladas absolutas o acumuladas relativas) para representar un **diagrama de rectángulos o de barras** se traza para cada valor de la variable un rectángulo o barra de altura proporcional a la frecuencia que se esté representando.

Si se unen los puntos medios de los extremos superiores de las barras tenemos un **polígono de frecuencias o diagrama de líneas**.

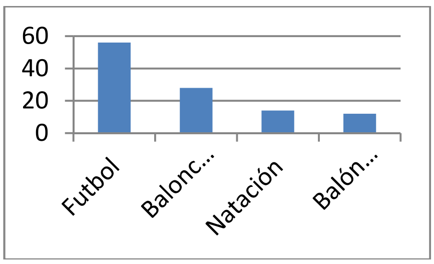
En un **diagrama de sectores** se dibuja un círculo que se divide en sectores de amplitudes proporcionales a las frecuencias.

Actividad resuelta

- *Tenemos un estudio estadístico sobre las preferencias deportivas del alumnado de 4º de un determinado centro escolar. Represéntalos en un diagrama de barras de frecuencias absolutas, en un polígono de frecuencias relativas y en un diagrama de sectores.*

Deportes	Frecuencia Absoluta
Futbol	56
Baloncesto	28
Natación	14
Balón volea	12

Diagrama de barras de frecuencias absolutas



Polígono de frecuencias relativas o diagrama de líneas

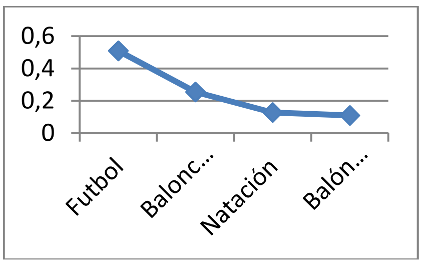
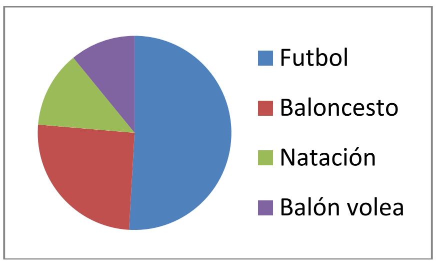


Diagrama de sectores



Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Actividades propuestas

4. Con la tabla de valores del ejercicio anterior, dibuja en tu cuaderno el diagrama de frecuencias relativas, el polígono de frecuencias absolutas acumuladas y el diagrama de sectores.
5. Haz un estudio estadístico preguntando a tus compañeros y compañeras de clase sobre el número de libros que leen al mes. Confecciona una tabla y represéntala en un diagrama de rectángulos, un polígono de frecuencias y un diagrama de sectores.
6. Selecciona una muestra entre tus compañeros y compañeras y realiza un estudio estadístico sobre el deporte que más le gusta a cada uno. Haz la representación que sea más sencilla de interpretar.

Utiliza el ordenador

Las hojas de cálculo son una herramienta muy útil para trabajar la Estadística. Suman, multiplican, y dibujan los gráficos con gran facilidad. Para la actividad resuelta anterior, copiamos la tabla con los datos en la hoja de cálculo a partir de la casilla A1. Calculamos la suma total en la casilla B6, simplemente apretando la tecla: $\sum$, o bien escribiendo =SUMA(B2:B5) que significa que queremos sumar lo que hay desde la casilla B2 a la B5.

Para calcular las frecuencias relativas escribimos en C1: Frecuencia relativa, y en C2, escribimos el signo igual, (con lo que estamos diciendo a la hoja que vamos a calcular algo), pinchamos en la casilla B2, escribimos: /, y pinchamos en B6: =B2/B6, nos sale 0,50909… La casilla B2 va a ir variando cuando calculemos C3, C4…, pero queremos que la casilla B6 se quede fija. Para decir eso, ponemos el símbolo $: =B2/$B$6. Y ahora arrastramos hasta la casilla C5. (Si arrastramos antes de poner el $ nos sale un error, pues está dividiendo por cero al ir modificando la casilla). Tenemos las frecuencias relativas calculadas.

	A	B	C
1	Deportes	Frecuencia absoluta	Frecuencia relativa
2	Futbol	56	0,509090909
3	Baloncesto	28	0,254545455
4	Natación	14	0,127272727
5	Balón volea	12	0,109090909
6	SUMA	110	
7			

Para dibujar los gráficos sólo tenemos que seleccionar las filas y columnas que nos interesen y en el menú de "Insertar" seleccionar el tipo de gráfico deseado: Columna, Línea, Circular…

1.3. Parámetros de centralización y dispersión

Parámetros de centralización

Ya sabes que los parámetros de centralización nos dan información sobre el "*centro*" de un conjunto de datos. Estudiamos la media aritmética, la moda y la mediana.

Actividad resuelta

- *Nieves ha tenido en Matemáticas las siguientes notas: 8, 4, 6, 10 y 10. Calcula su media, su moda y su mediana.*

Su nota media se calcula sumando todas las notas: 8 + 4 + 6 + 10 + 10 = 38, y dividiendo la suma entre el número total de notas que es 5: 38/5 = 7,6.

La moda es 10 pues es el valor más frecuente.

Una forma de calcular la mediana es ordenar los valores de menor a mayor, y si el número de datos es impar, el valor central es la mediana. Si el número de datos es par, la mediana es la media de los dos datos centrales.

En nuestro caso: $4 \leq 6 \leq 8 \leq 10 \leq 10$, por lo que la mediana es 8.

Para calcular la **media** (m) de $x_1, x_2, \ldots, x_n$, se suman todos y se divide por el número total de datos (n).

$$\text{Media} = m = (x_1 + x_2 + \ldots + x_n)/n$$

¿Qué es lo que está de moda? Lo que más se lleva.

La **moda** (mo) de una distribución de frecuencias es el valor más frecuente.

La **mediana** (me) es el valor central que deja por debajo el mismo número de valores de la variable que por encima.

Utiliza el ordenador

Para calcular la media, la mediana y la moda con la hoja de cálculo, copiamos en la casilla B2, B3… los datos: 8, 4, 6, 10 y 10. Escribimos en la casilla A7, Media, y para calcular la media escribimos un signo igual en B7. Buscamos, desplegando las posibles funciones, la función PROMEDIO, y escribimos

=PROMEDIO(B2:B6),

B7 =PROMEDIO(B2:B6)

	A	B	C	D	E
1		Datos			
2		8			
3		4			
4		6			
5		10			
6		10			
7	Media	7,6			
8	Mediana	8			
9	Moda	10			

que significa que calcule la media de los valores que hay en las casillas desde B2 hasta B6. Del mismo modo calculamos la mediana buscando en las funciones o escribiendo =MEDIANA(B2:B6) y la moda buscando en las funciones o escribiendo =MODA(B2,B6).

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Actividades propuestas

7. Dadas las temperatura en una ciudad a una hora determinada el día 1 de cada mes se tiene la siguiente tabla:

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Temperatura	–2	5	8	9	11	13	27	33	21	14	9	4

a) Calcula la temperatura media, la moda y la mediana.

b) Utiliza el ordenador para comprobar el resultado.

8. Calcula la media, la mediana y la moda de las distribuciones siguientes:

a) 2, 3, 4, 5, 7, 9, 9, 1000 b) 2, 3, 4, 5, 7, 9, 9, 10 c) 0, 0, 4, 5, 7, 9, 9, 100, 200

Utiliza el ordenador para comprobar los resultados.

Observa en cada caso cómo influyen los valores extremos. ¿Influyen en la moda? ¿Y en la mediana? ¿Y en la media?

Actividad resuelta

- *En una clase de 40 alumnos las calificaciones han sido:*

x_i	0	1	2	3	4	5	6	7	8	9	10	Suma
f_i	1	2	0	1	2	8	7	6	6	4	3	40

A cada nota la llamamos x_i y a la frecuencia absoluta de esa nota: f_i. Esto significa que ha habido un cero, dos unos, ningún 2… y 3 dieces.

Para calcular la media aritmética añadimos a la tabla una fila con los productos $x_i \cdot f_i$ y sumamos esa fila:

$x_i \cdot f_i$	0	2	0	3	8	40	42	42	48	36	30	251

Al ser 40 el número total de estudiantes la media es: **Media** = m = 251 / 40 = 6,275.

La **moda** es la nota más frecuente, que es mo = 5 pues es la de mayor frecuencia.

Para calcular la **mediana** añadimos una nueva fila, la de las frecuencias acumuladas:

x_i	0	1	2	3	4	5	6	7	8	9	10
Frecuencias acumuladas	1	3	3	4	6	14	21	27	33	37	40

La mitad de los datos es 40/2 = 20, y como 14 < 20 < 21, la mediana es 6.

Si la variable toma los valores $x_1, x_2, \ldots, x_n$, con una frecuencia absoluta $f_1, f_2, \ldots, f_n$, para calcular la **media** se multiplica cada valor por su frecuencia absoluta, se suman dichos productos y se divide por n el total de valores de la variable:

$$m = \text{Media} = (x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n) / (f_1 + f_2 + \ldots + f_n)$$

La **moda** es la frecuencia más alta.

Puede ocurrir que una distribución de frecuencias tenga más de una moda. Por ejemplo, la distribución:

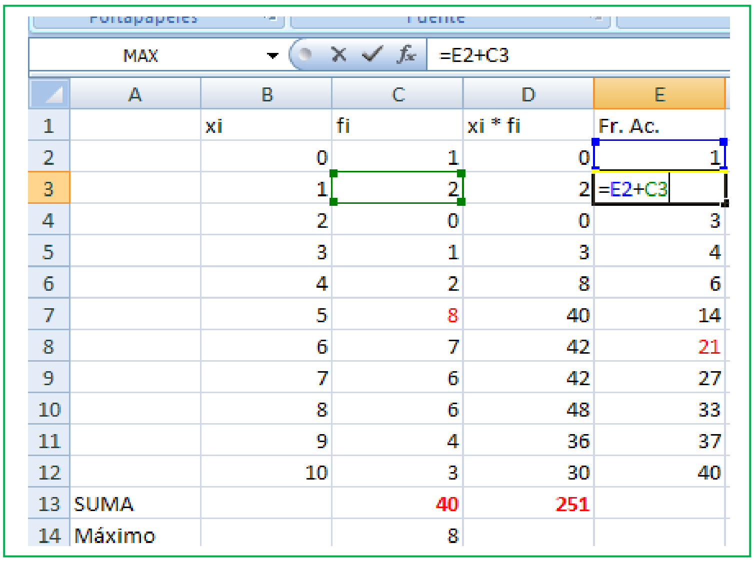
x_i	1	2	3	4	5	6
f_i	10	9	10	8	7	10

tiene 3 modas, 1, 3 y 6, ya que el valor más alto de la frecuencia absoluta es 10 en los tres casos. La moda permite clasificar los conjuntos de datos en *unimodales*, *bimodales* o *plurimodales*, según el número de modas que tengan.

Para obtener la **mediana** se calculan las frecuencias acumuladas y se busca el valor de la variable que ocupa el lugar central: n/2.

Utiliza el ordenador

Copiamos los datos de la actividad resuelta en una hoja de cálculo, escribiendo xi en la casilla B1, fi en la C1. En B2 escribimos 0, y en B3, 1. Seleccionamos estas dos casillas y arrastramos hasta la casilla B12. Copiamos las frecuencias en la columna C. En A13 escribimos SUMA. Calculamos la suma de las frecuencias con la tecla: $\sum$ y se obtiene 40 en la casilla C13. En la columna D1 escribimos $x_i \cdot f_i$. En D2 escribimos = y pinchamos en B2, escribimos * y pinchamos en C2 (=B2*C2). Seleccionamos D2 y arrastramos hasta D12. Calculamos la suma (251) y dividimos el valor de la casilla D12 entre el de la casilla C12.


MAX =E2+C3

	A	B	C	D	E
1		xi	fi	xi * fi	Fr. Ac.
2		0	1	0	1
3		1	2	2	=E2+C3
4		2	0	0	3
5		3	1	3	4
6		4	2	8	6
7		5	8	40	14
8		6	7	42	21
9		7	6	42	27
10		8	6	48	33
11		9	4	36	37
12		10	3	30	40
13	SUMA		40	251	
14	Máximo		8		

Podemos calcular el valor máximo de las frecuencias, que en este caso se ve a ojo, pero si hubiera muchos más valores, muchas más filas, se puede utilizar la función MAX.

Para calcular las frecuencias acumuladas utilizamos la columna E. En E2 escribimos =C2. En E3 escribimos =E2+C3. ¿Por qué? Y seleccionando E3 arrastramos hasta E12.

Actividades propuestas

9. Se ha lanzado un dado 100 veces y se ha confeccionado la siguiente tabla de frecuencias absolutas:

x_i	1	2	3	4	5	6
f_i	18	16	14	16	16	20

   a) Calcula la media, moda y mediana.
   b) Utiliza el ordenador para comprobar los resultados.

10. Lanzamos 2 dados y sumamos los valores obtenidos. Repetimos el experimento 1000 veces y obtenemos las siguiente tabla de frecuencias absolutas.

x_i	2	3	4	5	6	7	8	9	10	11	12
f_i	24	65	73	81	158	204	148	79	68	59	41

   a) Calcula la media, la mediana y la moda.
   b) Utiliza el ordenador para comprobar los resultados.
   c) Repite tú los lanzamientos, ahora sólo diez veces, y calcula de nuevo la media, mediana y moda.

11. Utiliza el ordenador para calcular la media, la mediana y la moda de la siguiente tabla de frecuencias absolutas, que indica el número de hijos que tienen 200 familias entrevistadas:

x_i	0	1	2	3	4	5	6	7	8	9	10
f_i	14	65	73	27	9	6	2	1	0	2	1

Parámetros de dispersión

Nos dan una medida de lo "*dispersos*" que están los datos.

La primera medida nos la da el **recorrido**, o el valor máximo menos el valor mínimo.

Las más utilizadas son la varianza y la desviación típica (o desviación estándar) que mide la distancia de los datos respecto de la media.

Ya sabes que la mediana nos indica el valor de la variable que ocupa el lugar central. Se denomina **primer cuartil** (Q1) al valor de la variable que deja menores o iguales que él a la cuarta parte de los datos, (o un 25 %), (siendo por tanto las tres cuartas partes mayores o iguales que él). La mediana es el segundo cuartil, que deja por debajo la mitad de los datos o un 50 %. El **tercer cuartil** (Q3) es el valor de la variable que deja menores o iguales que él las tres cuartas partes de los datos o un 75 % (y mayores o iguales la cuarta parte). Se llama **intervalo intercuartil** (o recorrido intercuartílico) a la distancia entre el tercer y el primer cuartil (Q3 – Q1). Por lo que hemos dicho, en ese intervalo están la mitad de los datos.

Actividad resuelta

Seguimos con la misma actividad anterior.

- *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.*

La mayor calificación ha sido un 10 y la menor un 6, luego el **recorrido** es 10 – 4 = 6.

Recorrido = Máximo – Mínimo.

La media ya la hemos calculado y es 7,6. Queremos analizar cómo las observaciones se separan de la media. Si a cada valor le restamos la media, unos salen positivos y otros negativos, y si sumamos todos, se compensan, por lo que sale 0. Es posible superar esa dificultad calculando esas diferencias en valor absoluto, o elevándolas al cuadrado. Si las elevamos al cuadrado, sumamos todo y dividimos por el número total de valores de la variable menos 1, obtenemos la varianza.

Se divide por *n* – 1 para mejorar las propiedades del estadístico: Varianza.

Si después calculamos la raíz cuadrada, se obtiene la desviación típica. Estamos evaluando la distancia de los valores de la variable a la media.

	x_i	x_i – media	(x_i – media)^2
1	8	0,4	0,16
2	4	-3,6	12,96
3	6	-1,6	2,56
4	10	2,4	5,76
5	10	2,4	5,76
Media = 7,6			Suma = 27,2

Si dividimos 27,2 entre 5 (*n*) se obtiene 5,44 que es la varianza.

Calculamos la raíz cuadrada: 2,33 que es la desviación típica.

Varianza = ((x_1 – media)^2 + (x_2 – media)^2 + … + (x_n – media)^2)/n = $\frac{\sum_{i=1}^{n}(x_i - m)^2}{n}$; S = Desviación típica = $\sqrt{\frac{\sum_{i=1}^{n}(x_i - m)^2}{n}}$

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Se puede demostrar, haciendo operaciones una fórmula más cómoda para calcular la varianza y la desviación típica:

Varianza = $\dfrac{\sum_{i=1}^{n} x_i^2}{n} - m^2$ S = $\sqrt{\dfrac{\sum_{i=1}^{n} x_i^2}{n} - m^2}$

$$\sum_{i=1}^{n}(x_i - m)^2 =$$
$$\sum_{i=1}^{n}(x_i^2 - 2x_i \cdot m + m^2) =$$
$$\sum_{i=1}^{n} x_i^2 - 2m\sum_{i=1}^{n} x_i + n \cdot m^2 =$$
$$\sum_{i=1}^{n} x_i^2 - 2m(n \cdot m) + n \cdot m^2 =$$
$$\sum_{i=1}^{n} x_i^2 - n \cdot m^2$$

	x_i	x_i^2
	8	64
	4	16
	6	36
	10	100
	10	100
m = 7,6	Suma = 38	Suma = 316

Varianza = (316/5) – (7,6)² = 63,2 – 57,76 = 5,44.
La desviación típica es la raíz cuadrada de la varianza, es decir, *s* = 2,33.
Para calcular los cuartiles debemos ordenar los datos; 4 ≤ 6 ≤ 8 ≤ 10 ≤ 10.

1	2	3	4	5
4	6	8	10	10

El primer cuartil deja por debajo la cuarta parte o el 25 % de los datos. Hay 5 datos y 5/4 = 1,25, como 1 < 1,25 < 2, el primer cuartil es 6. Q1 = 6. El tercer cuartil deja por debajo las tres cuartas partes o el 75 % de los datos: 3(5/4) = 3,75. Como 3 < 3,75 < 4, entonces Q3 = 10.

Intervalo intercuartil = Q3 – Q1.

En el ejemplo, el intervalo intercuartil = Q3 – Q1 = 10 – 6 = 4.

Utiliza el ordenador

Igual que hemos calculado la media, la mediana y la moda, la hoja de cálculo se puede utilizar para obtener:

- *El recorrido calculando MAX – MIN.*
- *La varianza utilizando VARP.*
- *La desviación típica usando DESVESTP.*
- *Los cuartiles, (CUARTIL), siendo el cuartil 0 el mínimo; el cuartil 1, Q1; el cuartil 2, la mediana; el cuartil 3, Q3; y el cuartil 4, el máximo.*

B10 =DESVESTP(B2

	A	B	C	D
1		xi		
2		8		
3		4		
4		6		
5		10		
6		10		
7	MAX	10	Recorrido =	6
8	MIN	4		
9	VARP	5,44		
10	DESVESTP	2,33		
11	CUARTIL 1	6	Intervalo Intercuartil =	

Actividades propuestas

12. Dadas las temperatura en una ciudad de un ejercicio anterior:

Meses	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Temperatura	–2	5	8	9	11	13	27	33	21	14	9	4

a. Calcula el recorrido, la varianza, la desviación típica, los cuartiles y el intervalo intercuartil.
b. Utiliza el ordenador para comprobar los resultados.

13. Calcula el recorrido, la varianza, la desviación típica, los cuartiles y el intervalo intercuartil. de las distribuciones siguientes:
a) 2, 3, 4, 5, 7, 9, 9, 1000   b) 2, 3, 4, 5, 7, 9, 9, 10   c) 0, 0, 4, 5, 7, 9, 9, 100, 200

Utiliza el ordenador para comprobar los resultados.

1.4. Diagrama de cajas

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 y donde se señala en el centro, la mediana. Se añaden dos brazos (o bigotes) donde se señalan los valores máximo y mínimo.

Se pueden calcular, además, unos límites superior e inferior. El inferior, L1; es Q1 menos 1,5 por el intervalo intercuartil, y el superior Ls es Q3 + 1,5 por el intervalo intercuartil.

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

En el ejemplo anterior, una vez ordenados los datos: 4 ≤ 6 ≤ 8 ≤ 10 ≤ 10, hemos calculado que:

Mediana = Me = 8.

Q1 = 6.

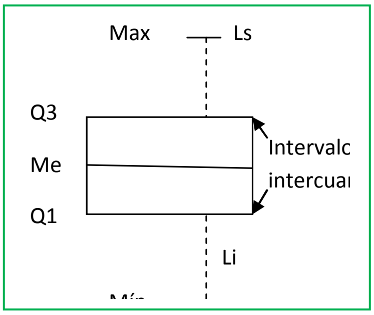
Q3 = 10.

Intervalo intercuartil = 4.

Los bigotes nos indican:

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Máx = 10.
Mín = 4.
Ls = Q3 + 4*1,5 = 16.
Li = Q1 – 4*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.



1.5. Variable continua: intervalos y marcas de clase. Histogramas

Recuerda que las variables pueden ser cualitativas, si no son numéricas, o cuantitativas, que a su vez pueden ser discretas o continuas.

- *Por ejemplo: Si se hace un estudio estadístico sobre la población de estudiantes, se puede preguntar sobre la profesión de sus padres y madres, que es una variable cualitativa, sobre el número de hermanos, que es una variable cuantitativa discreta (nadie tiene 3,7 hermanos), o sobre la edad, la estatura, la calificación media… que son variables cuantitativas continuas.*

Con las variables cuantitativas continuas tiene sentido agrupar los valores en intervalos.

Al valor central del intervalo se le denomina **marca de clase**.
La representación gráfica más adecuada es el **histograma** que es un diagrama de rectángulos en el que el área de cada rectángulo es proporcional a la frecuencia. Tiene la ventaja de que de esa forma la frecuencia de cada suceso viene representada por el área.

Actividad resuelta

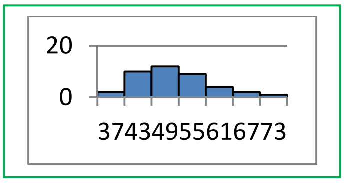
- *Realiza un estudio estadístico sabiendo que la tabla de frecuencias absolutas, con intervalos, de los pesos de 40 estudiantes de un centro escolar, es:*

Peso	[34, 40)	[40, 46)	[46, 52)	[52, 58)	[58, 64)	[64, 70)	[70, 76)
Estudiantes	2	10	12	9	4	2	1

La tabla nos dice que hay 2 estudiantes cuyo peso es mayor o igual a 34 y es menor que 40.
Calculamos las marcas de clase, buscando el punto medio de cada intervalo: (40 – 34)/2 = 3 y 34+3 =37. Todos los intervalos en este ejemplo tienen una longitud de 6. Reescribimos la tabla con las marcas de clase y las frecuencias absolutas:

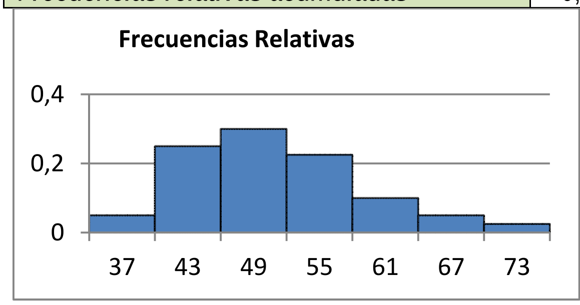
x_i	37	43	49	55	61	67	73
f_i	2	10	12	9	4	2	1

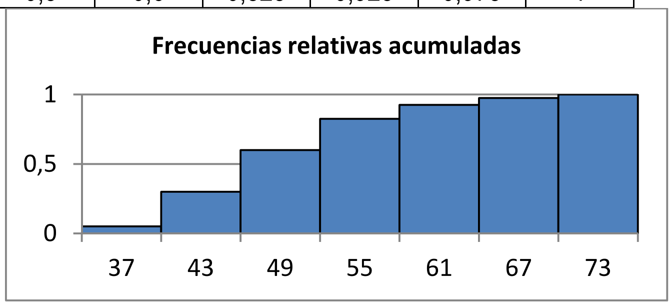
En este caso el histograma de las frecuencias absolutas es muy sencillo pues todos los intervalos tienen igual longitud. Si no fuera así, habría que calcular con cuidado las alturas de los rectángulos para que las áreas fueran proporcionales a las frecuencias.



Vamos a representar también el histograma de las frecuencias relativas y de las frecuencias relativas acumuladas:

x_i	*37*	*43*	*49*	*55*	*61*	*67*	*73*
Frecuencias relativas	0,05	0,25	0,3	0,225	0,1	0,05	0,025
Frecuencias relativas acumuladas	0,05	0,3	0,6	0,825	0,925	0,975	1





Cálculo de la media y la desviación típica:
Procedemos de la forma que ya conocemos, calculando el producto de las marcas de clase por las frecuencias:

x_i	37	43	49	55	61	67	73	Suma
f_i	2	10	12	9	4	2	1	40
$x_i \cdot f_i$	74	430	588	495	244	134	73	2038

La **media** es igual a 2038/40 = 50,95
Para calcular la **desviación típica** restamos a cada marca de clase, la media, elevamos al cuadrado y multiplicamos por la frecuencia relativa:

x_i	37	43	49	55	61	67	73	Suma
f_i	2	10	12	9	4	2	1	40
$x_i - m$	–13,95	–7,95	–1,95	4,05	10,05	16,05	22,05	
$(x_i - m)^2$	194,60	63,2025	3,8025	16,4025	101,0025	257,6025	486,2025	1122,8175
f_i	2	10	12	9	4	2	1	40
$(x_i - m)^2 \cdot f_i$	389,20	632,025	45,63	147,62	404,01	515,205	486,2025	2619,9

La suma de las diferencias de la media al cuadrado por las frecuencias relativas es 2619,9. Ahora dividimos entre *n* que en nuestro caso es 40, y se obtiene 65,5 que es la varianza. Calculamos la raíz cuadrada. La desviación típica es 8,09.

Actividad resuelta

- *Utilizamos la otra fórmula: Varianza =* $\dfrac{\sum_{i=1}^{n}(x_i^2 \cdot f_i)}{n} - m^2$

x_i	37	43	49	55	61	67	73	Suma
f_i	2	10	12	9	4	2	1	40
x_i^2	1369	1849	2401	3025	3721	4489	5329	22183
$x_i^2 \cdot f_i$	2738	18490	28812	27225	14884	8978	5329	106456

Varianza = (106456/40) –(50,95)^2 = 2661,4 – 2595,9 = 65,5 y *desviación típica* = *s* = 8,09.

- *Veamos otro ejemplo de cálculo de la media y la desviación típica utilizando la otra fórmula:*

$$\textit{Varianza} = \frac{\sum_{i=1}^{n}(x_i^2 \cdot f_i)}{n} - m^2$$

x_i	64	65	66	67	68	69	70	71	72	73	74	75	Suma
f_i	1	0	2	5	9	22	16	12	8	3	1	1	80
$x_i \cdot f_i$	64	0	132	335	612	1518	1120	852	576	219	74	75	5577
x_i^2	4096	4225	4356	4489	4624	4761	4900	5041	5184	5329	5476	5625	
$x_i^2 \cdot f_i$	4096	0	8712	22445	41616	104742	78400	60492	41472	15987	5476	5625	389063

n = 80.

La media es igual a *m* = 5577/80 = 69,7.

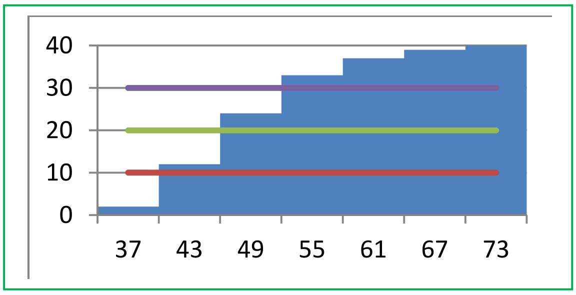
La varianza es igual a $\dfrac{\sum_{i=1}^{n}(x_i^2 \cdot f_i)}{n} - m^2 = \dfrac{389063}{80} - 69{,}7^2$ = 4863,2875 – 4858,09 = 5,1975

La *desviación típica* es igual a la raíz cuadrada de la varianza, *s* = 2,28.

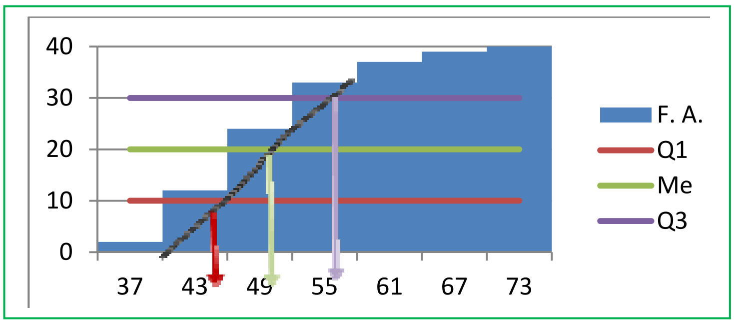
Cálculo de la mediana y los cuartiles.

Representamos el histograma de frecuencias absolutas acumuladas, y cortamos por las líneas *n*/2 para la mediana, *n*/4 para el primer cuartil, y 3*n*/4 para el segundo. En nuestro caso por 20, 10 y 30.

Observamos, viendo donde las rectas horizontales, *y* = 20, *y* = 10 e *y* = 30 cortan al histograma, que la mediana está en el intervalo [46, 52) cuya marca de clase es 49, el primer cuartil en el intervalo [40, 46) cuya marca de clase es 43, y el tercer cuartil en [52, 58) cuya marca de clase es 55.



x_i	[34, 40)	[40, 46)	[46, 52)	[52, 58)	[58, 64)	[64, 70)	[70, 76)
f_i	2	10	12	9	4	2	1
F_i	2	12	24	33	37	39	40



Podemos ajustarlo más haciendo una interpolación lineal, es decir, aproximando con una recta.

Para la mediana trazamos la recta que pasa por los puntos (46, 12) y (52, 24) (*y* = 2*x* – 80) y calculamos dónde corta a la recta *y* = 20. Corta en *x* = 50. Por tanto la mediana es Me = 50.

El tercer cuartil está en el intervalo [52, 58). Calculamos la ecuación de la recta que pasa por los puntos (52, 24) y (58, 33), que es *y* = (3/2)*x* – 54. Calculamos dónde corta a y = 30, que es en *x* = 56. Por tanto Q3 = 56.

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

El primer cuartil está en el intervalo [40, 46). La recta que pasa por los puntos:

(40, 2) y (46, 12)

tiene por ecuación $y = (5/3)x - 64{,}6666$, que corta a $y = 10$ en $x = 44{,}79999$…. Q1 = 44,8.

Utiliza el ordenador

Para dibujar histogramas con el ordenador utilizando una hoja de cálculo nos encontramos con la dificultad de que éste dibuja los rectángulos separados. Dibuja un diagrama de rectángulos. Para arreglarlo en el caso de que la longitud de todos los intervalos sea la misma, debes señalar uno de los rectángulos, entrar en "*dar formato a la serie de datos*" y, en "*Opciones de serie*" seleccionar en "*Ancho del intervalo*" un ancho del 0 %, es decir, "*sin intervalo*". Si las longitudes son distintas se debe calcular previamente las alturas de los rectángulos.

Actividades propuestas

14. Utiliza el ordenador para dibujar el histograma de la actividad 11.
15. Se conocen las cantidades de residuos sólidos recogidos en m^3/semana durante 12 semanas de una urbanización: 23, 27, 30, 34, 38, 21, 30, 33, 36, 39, 32, 24. Escribe en tu cuaderno una tabla de frecuencias absolutas con cuatro intervalos: [20, 25), [25, 30), [30, 35) y [35, 40). Calcula las marcas de clase. Dibuja el histograma de frecuencias absolutas. Calcula la media y la desviación típica. Calcula gráficamente la mediana y los cuartiles.
16. Haz un estudio estadístico preguntando a tus compañeros y compañeras de clase sobre el número de libros que leen al mes. Confecciona una tabla y represéntala en un diagrama de rectángulos, un polígono de frecuencias y un diagrama de sectores.

2. DATOS BIDIMENSIONALES

2.1. Ideas generales

Posiblemente, la aplicación más importante de la estadística no sea el estudio de una variable aislada sino el análisis de las relaciones entre variables. Si tenemos dos medidas que se dan juntas, es lógico querer saber en qué medida una influye en la otra. Veamos algunos ejemplos.

Ejemplos:

- *En una tienda de camisas, queremos saber cuántas venderemos (por término medio) en función del precio.*
- *Si sabemos la altura del padre de un niño, ¿cuál será la altura del hijo?*
- *Si a un grupo de alumnas le damos una paga y medimos sus calificaciones. ¿Las alumnas que reciben más dinero sacan mejores notas? ¿Cuánto más? Este mismo estudio puede hacerse con los trabajadores de una empresa. Si se les paga más ¿aumenta la producción?*
- *¿Son más inteligentes los hombres que las mujeres? ¿O viceversa?*

Puede parecerte que alguno de estos casos es elemental. Es obvio que los padres altos tienen hijos altos y que si bajo el precio, vendo más. Pero lo importante es CUÁNTO. Si yo tengo una tienda, lo que quiero es ganar dinero. Y por supuesto que si pongo las camisas a 0 € voy a vender mucho… pero no ganaré nada. Lo que quiero es una estimación de cuánto vendo a cada precio para poder saber el precio que me interesa poner.

2.2. Variables bidimensionales. Frecuencias conjuntas

Una **variable bidimensional** son dos variables que se miden conjuntamente. Si *X* e *Y* son las variables, la variable bidimensional es (*X*, *Y*).

Ejemplos:

- *El precio al que ponemos las camisas (X) y el precio anterior (Y).*
- *La altura de un padre (X) y la altura del hijo (Y)*
- *El color del pelo (X) y el color de los ojos (Y).*
- *El sexo de una persona (X) y su coeficiente de inteligencia (Y).*

Date cuenta que las variables bidimensionales pueden ser cualitativas o cuantitativas e incluso cada una de un tipo. Asimismo podríamos tener los datos agrupados, y entonces lo que habría sería parejas de intervalos.

La representación de forma de tabla de frecuencias es exactamente igual que en el caso unidimensional con la salvedad de que ahora tenemos parejas. Vamos primero con un ejemplo y luego introduciremos los conceptos.

Ejemplo:

- *Tenemos una muestra de 8 personas y miramos su color de ojos y pelo. Hay 4 morenos de ojos marrones, 1 moreno de ojos verdes, dos rubios de ojos azules y un rubio de ojos verdes.*

Aún no hemos definido las frecuencias pero creemos que lo puedes entender igual. La tabla es:

Individuo	Frecuencias absolutas	Frecuencias relativas
(Moreno, marrones)	4	0'5 = 4/8
(Moreno, verdes)	1	0'125 = 1/8
(Rubio, azules)	2	0'25 = 2/8
(Rubio, verdes)	1	0'125 = 1/10
TOTAL	8	1

Como puedes ver, para que dos elementos sean iguales, deben ser iguales las dos componentes. La variable *X* es el color del pelo y la variable *Y* el color de los ojos. Se tiene *X* = {"Moreno", "Rubio"} e Y={"Marrones", "Verdes", "Azules"}. No tiene por qué haber el mismo número de valores en cada variable.

Las definiciones son las mismas.

La **frecuencia absoluta** es el número de veces que se ha obtenido esa pareja de resultados (dos parejas son iguales si sus dos componentes son iguales).

La **frecuencia relativa** es la frecuencia absoluta dividida entre el número total de datos.

Tabla de frecuencias conjunta:

A veces, en vez de mostrar los datos en pares, se ponen en una **tabla de doble entrada o tabla de contingencia**. Se llama así porque la *X* está en vertical y la *Y* en horizontal. En los cruces se ponen las frecuencias, ya sean absolutas o relativas. Si se ponen las absolutas se dice **tabla de doble entrada de frecuencias absolutas** y si se ponen las relativas pues **tabla de doble entrada de frecuencias relativas.**

La tabla anterior, con (x_i, y_i) no tiene un nombre especial universalmente aceptado. Podemos llamarla **tabla de frecuencias de pares.**

Ejemplo:

- *Tenemos la misma muestra de antes: 4 morenos de ojos marrones, 1 moreno de ojos verdes, dos rubios de ojos azules y un rubio de ojos verdes. Vamos a colocarlos en tablas de doble entrada de frecuencias absolutas y luego relativas.*

Nos limitamos a poner en la primera columna los dos valores que tenemos de la *X*, que es el color de pelo ("Moreno" y "Rubio") y en la primera fila los de la *Y*, que es el color de los ojos ("Marrones", "Verdes" y "Azules").

Y / *X*	Marrones	Verdes	Azules
Moreno	4	1	0
Rubio	0	1	2

Observa que en esta tabla pueden aparecer ceros, que representan que no hay nadie con esa pareja de características.

Si dividimos las frecuencias absolutas por el número total de datos (que en este caso es 8) obtenemos la tabla de doble entrada de frecuencias relativas.

Y / *X*	Marrones	Verdes	Azules
Moreno	0'5 = 4/8	0'125=1/8	0
Rubio	0	0'125=1/8	0'25=2/8

Actividades propuestas

17. Con la tabla de valores del ejemplo, construye la tabla de frecuencias absolutas y relativas de la variable *X* ("Color de pelo") y la variable *Y* ("Color de ojos") por separado, como variables unidimensionales.
18. Completa la siguiente tabla y exprésala en forma de tabla de doble entrada, primero con frecuencias relativas y luego con frecuencias absolutas.

(x_i, y_i)	Frecuencia absoluta	Frecuencia relativa
(0, 1)	12	
(1,.2)	14	
(2,3)	14	

19. Completa la siguiente tabla de frecuencias conjunta y exprésala en frecuencias de pares (x_i, y_i), tanto con frecuencias relativas como absolutas.

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

2.3. Diagrama de dispersión y recta de regresión

Un **diagrama de dispersión**, también llamado **nube de puntos** por su apariencia, es un gráfico que se obtiene representando cada pareja como un punto del plano cartesiano. Se usa principalmente con variables cuantitativas y datos sin agrupar (si estuvieran agrupados tomaríamos las marcas de clase).

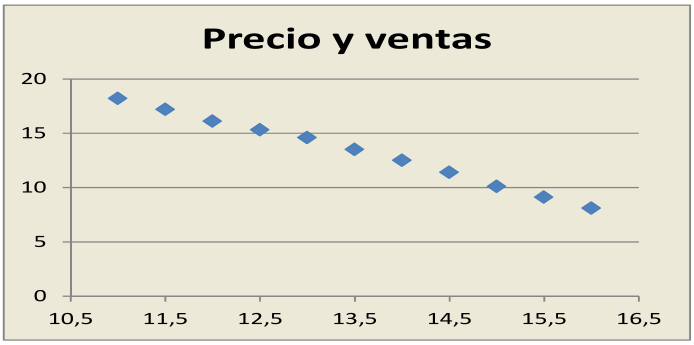
Es muy simple de dibujar. Basta con poner un punto en cada pareja. A veces si hay valores repetidos se ponen los puntos más gordos pero también es común ponerlos todos igual.

Ejemplo:

- *Tenemos una tienda y queremos estudiar las ventas de una camisa en función del precio. Para ello, probamos cada semana con un precio distinto y calculamos las ventas medias. Obtenemos así una tabla como la que sigue*

Precio	11	11'5	12	12'5	13	13'5	14	14'5	15	15'5	16
Ventas (medias)	18'2	17'2	16'1	15'3	14'6	13'5	12'5	11'4	10'1	9'1	8'1

Si copiamos los datos a una hoja de cálculo y le damos a dibujar un diagrama de dispersión, obtenemos algo como lo siguiente:

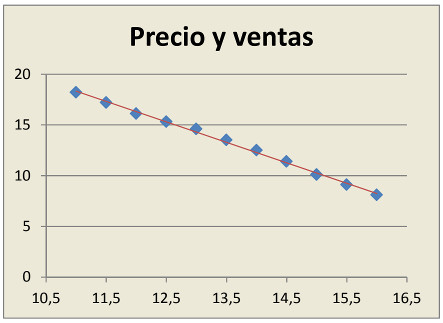


que es el típico gráfico que puede verse para hacer un estudio de resultados en cualquier empresa.

La recta de regresión

El problema con la nube de puntos es que simplemente describe lo que pasa. Esto ciertamente es importante en sí mismo, pero lo que es realmente interesante es PREDECIR qué pasará. En el ejemplo anterior, nuestros datos llegan a precios de 16 €. ¿Qué pasaría si subimos el precio a 17 €? ¿O lo bajamos a 9 €? ¿Y con los precios intermedios, como 12'25 €?

Como hay infinitos precios, no vamos a poder tener en cuenta infinitos precios. Lo interesante es tener un modelo matemático que nos diga, para un precio dado, cual es el valor esperado de las ventas. O, en general, para un valor de *X* cuál es el valor esperado de *Y*.



Lo más fácil es hacer una recta que se aproxime. Se puede dibujar prácticamente a mano, pero hay una fórmula matemática que la calcula. Esa fórmula es complicada y está fuera del alcance de este curso pero sí vamos a enseñarte cómo hacerla con ordenador.

Antes de nada, vamos a mostrarte en el ejemplo anterior la línea de tendencia.

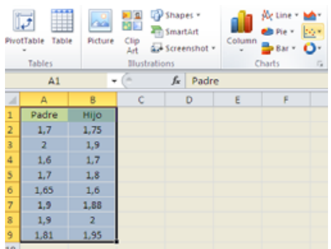
Observa que no pasa por todos los puntos, sino que unos quedan arriba y otros abajo. De hecho es imposible que una recta pase por todos y, en el mundo real, el ajuste nunca es exacto. La recta pasa por el medio de los puntos.

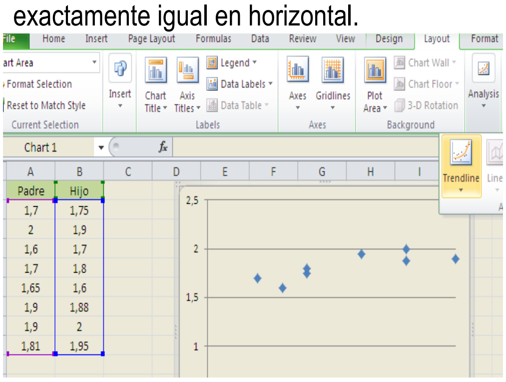
Utiliza el ordenador

- *Lo siguiente son datos de la altura de un padre y de la de su hijo con 15 años de edad. Las alturas están en metros.*

Padre	1'7	2	1'6	1'7	1'65	1'9	1'9	1'81
Hijo	1'75	1'9	1'7	1'8	1'6	1'88	2	1'95

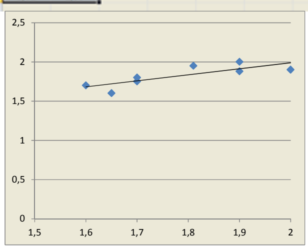
Lo primero, vamos a hacer el diagrama de dispersión. Copiamos los datos en una hoja de cálculo. Los vamos a poner en vertical para que se vea mejor, pero se podría hacer exactamente igual en horizontal.





Después, señalamos las dos series y le damos a *insertar gráfico de dispersión*.

Automáticamente nos aparece el diagrama de dispersión (nube de puntos). Si juegas un poco con las opciones puedes modificar el título, el formato, la escala de los ejes…



Más aún, la recta de regresión es muy fácil de dibujar. Basta con que selecciones el gráfico y le des a *análisis* y a *línea de tendencia*. Escogiendo una tendencia lineal, ya tienes la recta de regresión.

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Al final, si lo has hecho bien, el dibujo debe ser más o menos algo similar a esto:

Y fíjate, la recta tiene todos los valores posibles. Para ver qué valor correspondería a una altura del padre de 1,75 m, lo buscamos en la recta.

2.4. Interpretación de la recta. Introducción a la correlación

Una vez hemos dibujado la recta de regresión, podemos ver cómo es la relación entre las dos variables. En esencia el tipo de relación viene dada por la pendiente de la recta.

1. Si la recta de regresión tiene pendiente positiva (más informalmente, "si va hacia arriba") se dice que la relación entre las variables es **positiva**.
2. Si la recta de regresión tiene pendiente cero (más informalmente, "si queda horizontal") se dice que la relación entre las variables es **nula** o que **no hay relación lineal**.
3. Si la recta de regresión tiene pendiente negativa (más informalmente "si va hacia abajo") se dice que la relación entre las variables es **negativa**

La cuestión es, pues, sencilla. Basta dibujar la recta y ver hacia dónde va. Pero también nos interesa ver si los puntos están cerca de la recta o lejos. En otras palabras, mirar si la recta *ajusta bien* o *ajusta mal*.

Para calcular esto, se obtiene lo que se llama **coeficiente de correlación**. No vamos a entrar en cómo se calcula, es bastante complicado. Pero, como antes, basta con usar Excel o cualquier hoja de cálculo. La orden en Excel es COEF.DE.CORREL(serie1;serie2).

El coeficiente de correlación nos mide si la relación es positiva, negativa o nula. Y TAMBIÉN nos dice si el ajuste es bueno. Vamos a ver en un cuadro los detalles.

El **coeficiente de correlación**, ρ, mide la relación entre dos variables. Es un número entre -1 y 1 (puede ser exactamente -1 o exactamente 1).

Si el coeficiente de correlación es exactamente 1 la relación es **perfecta positiva**. La recta va hacia arriba y TODOS los puntos están sobre ella.

Si el coeficiente de correlación está en el intervalo (0, 1) la relación es **positiva**. La recta va hacia arriba pero no pasa por todos los puntos.

Si el coeficiente de correlación es exactamente 0, la relación es **nula** (no hay relación lineal).

Si el coeficiente de correlación está en $(-1, 0)$ la relación es **negativa**. La recta va hacia abajo pero no pasa por todos los puntos.

Si el coeficiente de correlación es exactamente -1 la relación es **perfecta negativa**. La recta va hacia abajo y TODOS los puntos están sobre ella.

Resumen

$\rho = 1 \to$ correlación perfecta positiva

$\rho = -1 \to$ correlación perfecta negativa

$\rho = 0 \to$ correlación nula

$\rho \in (0, 1) \to$ correlación positiva

$\rho \in (-1, 0) \to$ correlación negativa

Esto es lo que es objetivo. En algunas ocasiones, se habla de correlación positiva fuerte (si está cercana a 1) o positiva débil (si está entre 0 y 1 pero próxima a 0) y lo mismo negativa. Pero claro, eso depende de la interpretación de cada uno. Así, una correlación de 0'96 es positiva fuerte y una de -0'02 es negativa débil. Pero, ¿y 0'55? Pues depende de lo que consideres. Lo que sí es objetivo es si es perfecta o nula, positiva o negativa.

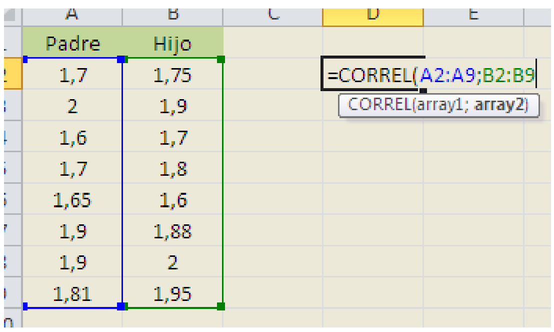
Utiliza el ordenador

- *Con los datos de la actividad anterior, vamos a calcular el coeficiente de correlación.*

Lo único que hay que hacer es poner, en la casilla correspondiente =COEF.DE.CORR. en nuestro ejemplo es la casilla D2.

Automáticamente nos da a escoger dos matrices y escogemos primero lo de la *X* y después lo de la *Y*.

Nos da el coeficiente de correlación, que en este caso resulta ser 0'81. Es una relación positiva fuerte como ya imaginábamos por la nube de puntos y la recta de regresión.

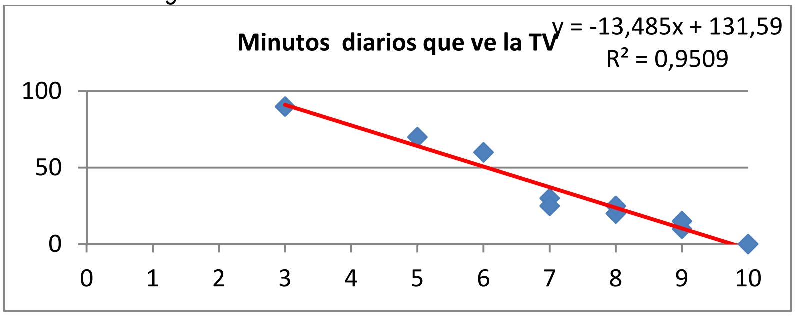


Utiliza el ordenador

- *Preguntamos a 10 alumnos de 4º ESO por sus calificaciones en Matemáticas, por el número de minutos diarios que ven la televisión, por el número de horas semanales que dedican al estudio, y por su estatura en centímetros. Los datos se recogen en la tabla adjunta. Queremos dibujar las nubes de puntos que los relacionan con las calificaciones de Matemáticas, el coeficiente de correlación y la recta de regresión.*

Calificaciones de Matemáticas	10	3	7	8	5	9	9	8	6	7
Minutos diarios que ve la TV	0	90	30	20	70	10	15	25	60	25
Horas semanales de estudio	15	2	9	12	7	14	13	11	7	8
Estatura (en cm)	177	168	157	159	163	179	180	175	169	170

Para hacerlo, entramos en Excel, y copiamos los datos. Seleccionamos la primera y la segunda fila, luego la primera y la tercera y por último la primera fila y la cuarta.
Con la primera y segunda filas seleccionadas, vamos a *Insertar*, *Dispersión* y elegimos la *nube de puntos*. Podemos conseguir que el eje de abscisas vaya de 0 a 10 en "*Dar formato al eje*". Pinchamos sobre un punto de la nube, y elegimos "*Agregar línea de tendencia*". Para que dibuje el ordenador la recta de regresión la línea de tendencia debe ser *Lineal*. En la pantalla que aparece marcamos la casilla que dice: "*Presentar ecuación en el gráfico*" y la casilla que dice "*Presentar el valor de R cuadrado en el gráfico*".



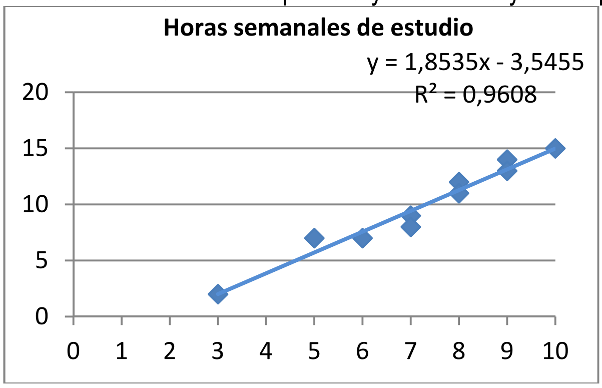
Observa, la recta de regresión, en color rojo, es decreciente y su ecuación es aproximadamente:

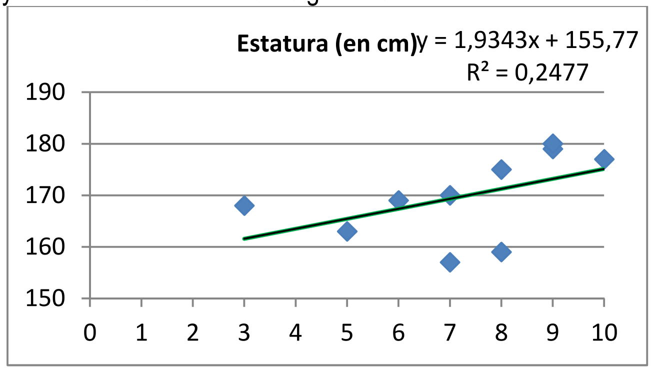
$y = -13'5\,x + 132$.

El cuadrado del coeficiente de correlación es $\rho^2 = 0'95$. La correlación es negativa y alta:

$$\rho = \sqrt{0'95} = -0{,}975$$

Hacemos lo mismo con la primera y tercera fila y con la primera y cuarta fila. Obtenemos los gráficos:





Observa que en ambos casos la pendiente de la recta de regresión es positiva pero en el primero el coeficiente de correlación, positivo, es próximo a 1, $\rho = \sqrt{0'96} = 0{,}98$. La correlación es alta y positiva.

En el segundo $\rho = \sqrt{0'25} = 0{,}5$.

Actividades resueltas

- *El propietario de una instalación mixta solar-eólica está realizando un estudio del volumen de energía que es capaz de producir la instalación. Para ello, mide dicha energía a lo largo de un total de* N = *16 días que considera suficientemente representativos. La energía (en kWh) producida en dichos días por las instalaciones solar y eólica se encuentra recogida en la siguiente tabla:*

Generación solar (x_i)	13'1	10'5	4'1	14'8	19'5	11'9	18	8'6	5'7	15'9	11'2	6'8	14'2	8'2	2'6	9'7
Generación eólica (y_i)	8'5	14'3	24'7	4	2'3	6'4	3'6	9'2	13'5	1'4	7'6	12'8	10'3	16'5	21'4	10'9

Vamos a realizar una actividad resuelta completa utilizando las fórmulas de la media, la desviación típica y de la correlación para que puedan servirte de modelos si necesitas alguna vez calcularlas sin ayuda del ordenador.
Vamos a denotar a la generación solar como variable *X* y la generación eólica como variable *Y*. Añadimos nuevas filas a nuestra tabla, los cuadrados de *x*, de *y* y los productos de ambas:

Generación solar (x_i)	13'	10'	4'1	14'	19'	11'	18	8'6	5'7	15'	11'	6'8	14'	8'2	2'6	9'7
Generación eólica (y_i)	8'5	14'	24'	4	2'3	6'4	3'6	9'2	13'	1'4	7'6	12'	10'	16'	21'	10'
x_i^2	17	11	16'	21	38	14	32	73'	32'	25	12	46'	20	67'	6'7	94'
y_i^2	72'	20	61	16	5'2	40'	12'	84'	18	1'9	57'	16	10	27	45	11
$x_i \cdot y_i$	11	15	10	59'	44'	76'	64'	79'	76'	22'	85'	87'	14	13	55'	10

Cálculo de las medias:
Sumando la primera fila y dividiendo por *N* = 16, obtenemos la media de la Generación Solar en Kwh.

Recuerda $\bar{x} = \dfrac{\sum_{i=1}^{N} x_i}{N}$:

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

$$\bar{x}=\frac{\sum_{i=1}^{N}x_i}{N}=\frac{13'1+10'5+4'1+14'8+19'5+11'9+18+8'6+5'7+15'9+11'2+6'8+14'2+8'2+2'6+9'7}{16}=10'925 \text{ Kwh}$$

Sumando la segunda fila y dividiendo por N = 16 obtenemos la media de la Generación Eólica en Kwh:

$$\bar{y}=\frac{\sum_{i=1}^{N}y_i}{N}=\frac{8'5+14'3+24'7+4+2'3+6'4+3'6+9'2+13'5+1'4+7'6+12'8+10'3+16'5+21'4+10'9}{16}=10'463 \text{ Kwh}$$

Las medias son: $\bar{x}=10'925$Kwh y $\bar{y}=10'463$ Kwh, Muy parecidas.

Cálculo de las desviaciones típicas:

En la tercera fila hemos calculado los cuadrados de los valores de la primera variable y los utilizamos para calcular la varianza: Recuerda $s_x{}^2=\frac{\sum_{i=1}^{N}x_i{}^2}{N}-\bar{x}^2$:

$$s_x{}^2=\frac{\sum_{i=1}^{N}x_i{}^2}{N}-\bar{x}^2=\frac{13'1^2+10'5^2+4'1^2+14'8^2+19'5^2+11'9^2+18^2+8'6^2+5'7^2+15'9^2+11'2^2+6'8^2+14'2^2+8'2^2+2'6^2+9'7^2}{16}-10'9^2=\frac{141'5}{16}-10'9^2=22'16$$

En la cuarta fila los cuadrados de los valores de la segunda variable y calculamos su varianza:

$$s_y{}^2=\frac{\sum_{i=1}^{N}y_i{}^2}{N}-\bar{y}^2=\frac{8'5^2+14'3^2+24'7^2+4^2+2'3^2+6'4^2+3'6^2+9'2^2+13'5^2+1'4^2+7'6^2+12'8^2+10'3^2+16'5^2+21'4^2+10'9^2}{16}-10'5^2=\frac{150'48}{16}-10'5^2=41'01$$

La desviación típica es la raíz cuadrada de la varianza, por tanto: $s_x=\sqrt{22'16}=4'71$ y $s_y=\sqrt{41'01}=6'4$

Cálculo del coeficiente de correlación:

Para calcular el coeficiente de correlación calculamos en la quinta fila los productos de la variable x por la variable y. Así, 13'1 · 8'5 = 111'4. Queremos calcula el término: $\frac{\sum_{i=1}^{N}(x_i\cdot y_i)}{N}$. Al sumar esa fila obtenemos 1401'2, que dividimos entre 16, le restamos el producto de las medias y dividimos por el producto de las desviaciones típicas:

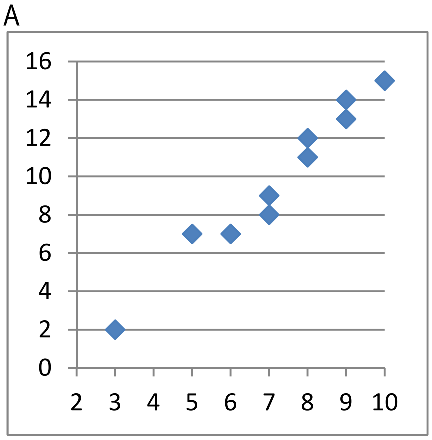
$$\rho=\frac{\frac{\sum_{i=1}^{N}(x_i\cdot y_i)}{N}-\bar{x}\cdot\bar{y}}{s_x\cdot s_y}=\frac{\frac{1401'2}{16}-(10'9\cdot 10'5)}{4'71\cdot 6'4}=\frac{-26'728}{4'71\cdot 6'4}=-0'887$$

Este coeficiente de correlación negativo y cercano a –1 nos indica que la relación entre las dos variables es negativa y bastante importante.

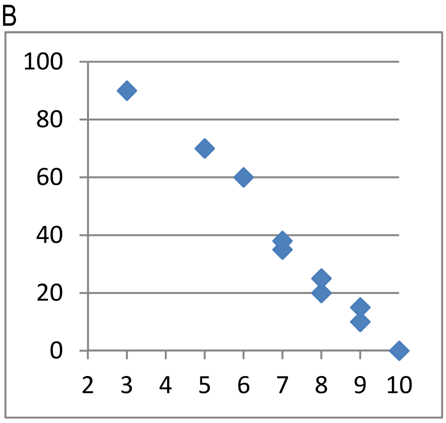
Actividades propuestas

20. María ha calculado los coeficientes de correlación de las tres nubes de puntos adjuntas, y ha obtenido: –0,05, 0,98 y –0,99, pero ahora no recuerda cuál es de cada una. ¿Puedes ayudarla a decidir qué coeficiente corresponde con cada nube?

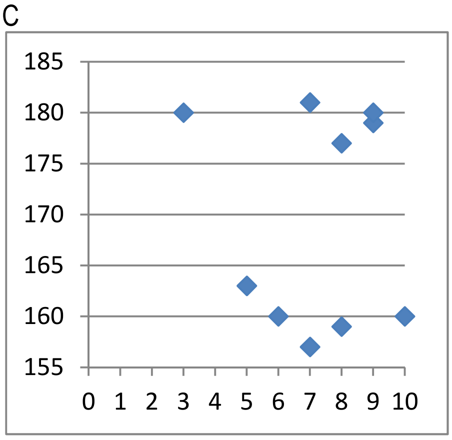
A



B



C



Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

21. Haz una encuesta entre tus compañeros de clase. Con ella vas a realizar un trabajo de investigación y presentar un informe. Elige con cuidado las preguntas. Vas a preguntar a cada uno de tus compañeros seleccionados, la muestra, dos preguntas, como por ejemplo lo que mide su mano y su nota en lengua, pero a ti pueden interesarte otras cuestiones muy distintas.
   a. Lo primero que vas a hacer es tabular las respuestas y confeccionar dos tablas de frecuencias absolutas. Luego completa esas mismas tablas con las frecuencias relativas y las frecuencias acumuladas. Haz representaciones gráficas de esas frecuencias: de barras, de líneas, de sectores.
   b. Calcula las medias, modas y medianas así como recorrido, desviación típica, cuartiles, intervalo intercuartílico… Representa los datos en una tabla de doble entrada y dibuja la nube de puntos. Calcula el coeficiente de correlación. Presenta un informe de este trabajo.

3. AZAR Y PROBABILIDAD

3.1. Experimento aleatorio y suceso

Un **fenómeno o experimento aleatorio** es aquel que, manteniendo las mismas condiciones en la experiencia, no se puede predecir el resultado.

- Son experimentos aleatorios:
  a) Lanzar una moneda y anotar si sale cara o cruz.
  b) Lanzar un dado y anotar el número de la cara superior.
  c) Lanzar dos dados o dos monedas.
  d) Si en una urna hay bolas blancas y rojas, sacar una al azar y anotar el color.
  e) Sacar una carta de una baraja.
  f) Sacar, sin reemplazamiento, dos cartas de la baraja.
  g) Abrir un libro y anotar la página por la que se ha abierto.

Sin embargo, calcular el coste de una mercancía, sabiendo el peso y el precio por kg, no es un experimento aleatorio. Tampoco lo es calcular el coste del recibo de la luz sabiendo el gasto.

- No son experimentos aleatorios
  a) Salir a la calle sin paraguas cuando llueve y ver si te mojas.
  b) El precio de medio kilo de rosquillas si las rosquillas cuestan a 3 € el kilo.
  c) Soltar un objeto y ver si cae.

Actividades propuestas

22. Indica si son, o no, fenómenos aleatorios:
- *La superficie de las comunidades autónomas españolas.*
- *Anotar el sexo del próximo bebé nacido en una clínica determinada.*
- *El área de un cuadrado del que se conoce el lado.*
- *Tirar tres dados y anotar la suma de los valores obtenidos.*
- *Saber si el próximo año es bisiesto.*

Al realizar un experimento aleatorio existen varios posibles resultados o **sucesos posibles**.
Al realizar un experimento aleatorio siempre se obtendrá uno de los **posibles resultados**.
Se llama **suceso elemental** a cada uno de los posibles resultados de un experimento aleatorio.
El conjunto de los posibles resultados de un experimento aleatorio se denomina **espacio muestral**.
Un **suceso** es un subconjunto del conjunto de posibles resultados, es decir, del espacio muestral.

Actividad resuelta

- Por ejemplo *los posibles resultados al tirar una moneda son que salga* cara *o salga* cruz*. El conjunto de sucesos elementales es {cara, cruz}.*
- El conjunto de posibles resultados de los experimentos aleatorios siguientes:
- *Extraer una bola de una bolsa con 9 bolas blancas y 7 negras es {*blanca, negra*}.*
- *Sacar una carta de una baraja española es {AO, 2O, 3O, …, SO, CO, RO, AC, …, RC, AB, …, RB, AE, …, RE }.*
- *Tirar dos monedas es: {(*cara, cara*), (*cara, cruz*), (*cruz, cara*), (*cruz, cruz*)}.*
- Al lanzar un dado, *el conjunto de posibles resultados es {1, 2, 3, 4, 5, 6}, el suceso obtener par es {2, 4, 6}, el suceso obtener impar es {1, 3, 5}, el suceso obtener múltiplo de 3 es {3, 6}, sacar un número menor que 3 es {1, 2}.*
- Al lanzar dos monedas *el conjunto de posibles resultados es {(C, C), (C, +), (+, C), (+, +)}. El suceso* sacar cero caras es *{(+, +)},* sacar una cara es *{(C, +), (+, C)} y* sacar dos caras *{(C, C)}.*

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Actividades propuestas

23. Escribe el conjunto de posibles resultados del experimento aleatorio: "*Escribir en cinco tarjetas cada una de las vocales y sacar una al azar*".
24. Escribe el conjunto de posibles resultados del experimento aleatorio: "*Tirar una chincheta y anotar si cae de punta o no*".
25. Inventa dos sucesos del experimento aleatorio: *Tirar dos monedas*.
26. En el juego de lotería, indica dos sucesos respecto a la cifra de las unidades del primer premio.
27. Escribe tres sucesos aleatorios del experimento aleatorio sacar una carta de una baraja española.

3.2. Frecuencia y probabilidad

No vamos a definir "probabilidad", pues existen varias definiciones posibles. Existe una axiomática debida a *Kolmogorov* relativamente reciente (1930), pero antes ya se había sido usado este concepto por ejemplo por *Fermat* y *Pascal* en el siglo XVII que se escribieron cartas reflexionando sobre lo que ocurría en los juegos de azar. Cuando no comprendían cómo asignar una determinada probabilidad, jugaban muchas veces al juego que fuese y veían a qué valor se aproximaban las frecuencias relativas. Así, la **probabilidad de un suceso** podría definirse como el **límite al que tienden las frecuencias relativas** de ese suceso cuando el número de experimentos es muy alto. Por tanto:

Para calcular probabilidades se usan dos técnicas, una experimental, analizando las frecuencias relativas de que ocurra el suceso, y la otra por simetría, cuando se sabe que los sucesos elementales son **equiprobables**, es decir, que **todos ellos tienen la misma probabilidad**, entonces **se divide el número de casos favorables por el número de casos posibles.**

Esto último, cuando se puede usar, simplifica la forma de asignar probabilidades y se conoce como **Regla de *Laplace*** que dice que: *"Si los sucesos elementales son equiprobables, la probabilidad de un suceso es el número de casos favorables dividido por el número de casos posibles"*.

Actividad resuelta

- La probabilidad de que salga cara al tirar una moneda es 1/2, pues sólo hay dos casos posibles {*cara*, *cruz*}, un único caso favorable, *cara*, y suponemos que la moneda no está trucada. Si sospecháramos que la moneda estuviera trucada para asignar esa probabilidad habría que tirar la moneda un montón de veces para observar hacia qué valor se acerca la frecuencia relativa de obtener cara.
- La probabilidad de sacar un 5 al tirar un dado es 1/6 pues hay seis casos posibles {1, 2, 3, 4, 5, 6}, un único caso favorable, 5, y suponemos que el dado no está trucado, luego todos ellos son equiprobables.
- La probabilidad de que al cruzar la calle te pille un coche NO es 1/2, aunque sólo hay dos casos posibles, que te pille el coche y que no te pille, pues ya te habría pillado un montón de veces. Para calcular esa probabilidad se recogen datos de peatones atropellados y se calcula utilizando las frecuencias relativas.
- La probabilidad de sacar una bola roja de una bolsa con 7 bolas rojas y 3 bolas blancas es 7/10.
- La probabilidad de que un bebé sea niña es aproximadamente 0,5, pero al hacer el estudio con las frecuencias relativas se ha visto que es 0,49.

- *Si consideramos una baraja española de 40 cartas y elegimos una carta, algunos de los sucesos que pueden ocurrir son "sacar un oro", o "sacar un as", o "sacar el caballo de copas"… Como de antemano no sabemos lo que va a ocurrir decimos que estos sucesos son aleatorios o de azar. Antes de sacar ninguna carta todas ellas son igualmente factibles, y como puede salir una cualquiera de las 40 cartas decimos que la probabilidad, de por ejemplo, sacar el caballo de copas es 1/40, la de sacar un oro es 10/40, y la de un as es 4/40.*
- *¿Cuál es la probabilidad de sacar el rey de copas? ¿Y de sacar un rey? ¿Y una copa?*

La probabilidad de sacar el *as de copas* es 1/40. Pero el suceso *sacar un as* se cumple si sale el as de oros, o de copas, o de bastos o de espadas. Es decir, no es un suceso simple, está formado, en este caso por 4 sucesos elementales, luego su probabilidad es 4/40 = 1/10. Lo mismo le ocurre a *sacar una copa*. Es un suceso compuesto, y como hay 10 copas su probabilidad es 10/40 = 1/4.

Actividades propuestas

28. Calcula la probabilidad de que al sacar una carta de la baraja sea una espada.
29. Para saber la probabilidad de que un recién nacido sea zurdo, ¿te basarías en el estudio de las frecuencias relativas o la asignarías por simetría?

3.3. Asignación de probabilidades

Suceso contrario

Actividades resueltas

- *¿Cuál es la probabilidad de sacar un as en la baraja de 40 cartas? ¿Y de* **no** *sacar un as? ¿Y de sacar una copa? ¿Y de* **no** *sacar una copa?*

El suceso *no sacar un as* es el suceso **contrario** al de *sacar un as*. Cartas que no son ases hay 36, luego la probabilidad de no sacar as es 36/40 = 9/10. Observa que se obtiene que p(as) + p(no as) = 1/10 + 9/10 = 10/10 = 1.

La probabilidad de *sacar copa* es 10/40, y hay 30 cartas que no son copas, luego la probabilidad de ***no** sacar copa* es 30/40, y 10/40 + 30/40 = 1.

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Si designamos por p(X) a la probabilidad de un suceso X y por p(noX) a la probabilidad de su **suceso contrario** resulta que:

$$p(X) + p(noX) = 1.$$

La probabilidad de un suceso más la probabilidad de su suceso contrario es igual a 1.

Actividades propuestas

30. ¿Cuál es la probabilidad de *no* sacar un 5 al tirar un dado? ¿Y de *no* sacar un múltiplo de 3? ¿Y de *no* sacar un número menor que 2?
31. Al tirar una moneda dos veces, ¿cuál es la probabilidad de no sacar ninguna cara? ¿Y de sacar al menos una cara? Observa que sacar al menos una cara es el suceso contrario de no sacar ninguna cara.

Sucesos dependientes e independientes

Ejemplo:

- *Tenemos una bolsa con 3 bolas rojas y 2 bolas negras. ¿Cuál es la probabilidad de sacar una bola roja? Si sacamos dos bolas, ¿cuál es la probabilidad de sacar dos bolas rojas?*

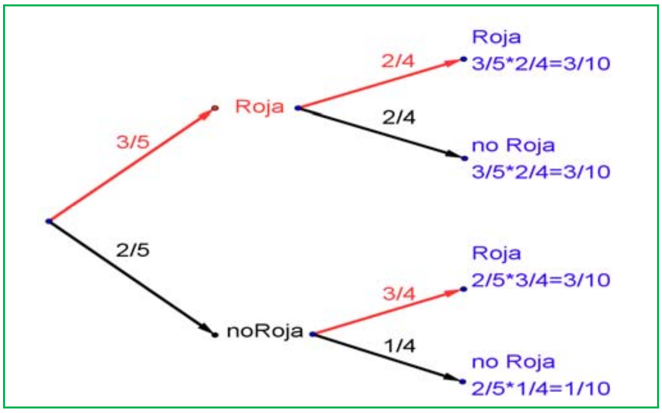
La probabilidad de sacar una bola roja es 3/5. Pero la de sacar dos bolas rojas, ¡depende!

Depende de si volvemos a meter en la bolsa la primera bola roja, o si la dejamos fuera.

En el primer caso decimos que es **con reemplazamiento** y en el segundo, **sin reemplazamiento**.

Si la volvemos a meter, la probabilidad de sacar bola roja volverá a ser 3/5, y la probabilidad de sacar dos bolas rojas es 3/5 · 3/5 = 9/25. La probabilidad de esta segunda bola *no depende* de lo que ya hayamos sacado, y en este caso la probabilidad se obtiene multiplicando.

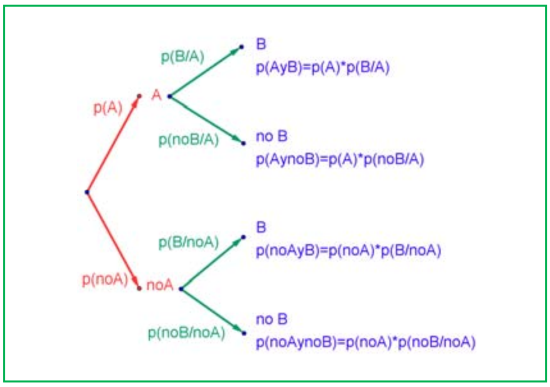
Si los sucesos A y B son **independientes: p(A y B) = p(A) · p(B).**



Pero si la dejamos fuera, ahora en la bolsa sólo hay 4 bolas y de ellas sólo quedan 2 bolas rojas, luego la probabilidad de que esa segunda bola sea roja es 2/4, y está condicionada por lo que antes hayamos sacado. Se escribe: p(Roja/Roja) y se lee *"probabilidad de roja condicionado a haber sacado roja"*. La probabilidad de sacar dos bolas rojas es ahora: 3/5 · 2/4 = 6/20 = 3/10.

Observa el diagrama de árbol y comprueba que la probabilidad de sacar primero una bola roja y luego una bola negra (no roja) es 3/5 · 2/4 = 3/10 pues después de sacar una bola roja en la bolsa quedan sólo 4 bolas y de ellas 2 son negras.

La probabilidad de sacar primero una bola negra y luego bola roja es 2/5·3/4 = 6/20 = 3/10, y la de sacar dos bolas negras es: 2/5·1/4 = 2/20 = 1/10.



Pero observa más cosas.

Por ejemplo, 3/5 + 2/5 = 1; 2/4 + 2/4 = 1; 3/4 + 1/4 = 1; 3/10 + 3/10 + 3/10 + 1/10 = 1.

Los sucesos no son independientes. El que ocurra A, o no ocurra A, afecta a la probabilidad de B. Por eso se dice que B **está condicionado** a A.

Si los sucesos A y B son **dependientes** entonces: **p(A y B) = p(A) · p(B/A)**

Actividades resueltas

- *Sacamos dos cartas de una baraja de 40 cartas sin reemplazamiento. ¿Cuál es la probabilidad de sacar dos ases?*

Si fuera con reemplazamiento la probabilidad sería 4/40 · 4/40, pero al ser sin reemplazamiento la probabilidad del segundo *as* viene condicionada por que hayamos sacado un as previamente. Ahora en la baraja ya no quedan 40 cartas sino 39, y no quedan 4 ases sino sólo 3, luego la probabilidad es:

$$4/40 \cdot 3/39 = 1/130.$$

Observa que:

Si dos sucesos son **dependientes** entonces: p(B/A) ≠ p(B).

Pero si dos sucesos son **independientes** entonces: p(B/A) = p(B/noA) = p(B).

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Actividades propuestas

32. En tu cuaderno haz un diagrama en árbol similar al anterior con los sucesos A y B: A = *sacar un as* en la primera extracción (noA = no sacarlo), y B = *sacar un as* en la segunda extracción (no B = no sacarlo). ¿Cuál es la probabilidad de *sacar as* en la segunda extracción condicionado a no haberlo sacado en la primera? ¿Y la de no sacar as en la segunda extracción condicionado a no haberlo sacado en la primera? ¿Cuál es la probabilidad de *sacar dos ases*? ¿Y la de sacar un solo as?
33. En el diagrama de árbol anterior indica cual es la probabilidad de "*no salen 2 ases*" y la de "*no sale ningún as*".
34. En el experimento "sacar tres cartas seguidas", ¿cuál es la probabilidad de *sacar tres ases*? Primero con reemplazo, y luego sin reemplazo.
35. Al tirar dos veces un dado calcula la probabilidad de que salga un seis doble.
36. Al tirar dos veces un dado calcula la probabilidad de sacar al menos un 6. *Ayuda*: Quizás te sea más fácil calcular la probabilidad de *no sacar ningún* 6, y utilizar el suceso contrario.

Sucesos compatibles e incompatibles

Ejemplo:

- *¿Cuál es la probabilidad de, en una baraja de 40 cartas, sacar una copa o un oro?*

Hay 10 copas y 10 oros, y ninguna carta es a la vez copa y oro, luego la probabilidad es 20/40.

- *¿Cuál es la probabilidad de, en una baraja de 40 cartas, sacar un as o un oro?*

Hay 4 ases y hay 10 oros, pero hay el *as de oros*, luego las cartas que son o bien un as o bien un oro son 13, luego la probabilidad es 13/40.

Llamamos sucesos incompatibles a los que, como copa y oro, no pueden realizarse a la vez, y sucesos compatibles a los que, como as y oro, pueden realizarse a la vez.

Designamos p(A o B) a la probabilidad del suceso "*se verifica A o bien se verifica B*". Hemos visto en el ejemplo que si los sucesos son incompatibles su probabilidad es igual a la suma de las probabilidades.

P(A o B) = p(A) + p(B), si A y B son incompatibles.

Pero si A y B si pueden verificarse a la vez habrá que restar esos casos, esas veces en que se verifican A y B a la vez.

P(A o B) = p(A) + p(B) – p(A y B), si A y B son compatibles.

Esta segunda expresión es más general que la primera, ya que en el caso en que A y B son incompatibles entonces p(A y B) = 0.

Actividades resueltas

- *Calcula la probabilidad de los sucesos siguientes: a) Sacar un rey o una figura; b) No sale un rey o sale un rey; c) Sacar un basto o una figura.*
- *Hay 4 reyes y hay 4 · 4 = 16 figuras (as, sota, caballo y rey), pero los cuatro reyes son figuras, por tanto p(Rey o Figura) = 4/40 + 16/40 – 4/40 = 16/40 = 0,4.*
- *Hay 40 – 4 = 36 cartas que no son reyes, y hay 4 reyes, luego p(no rey o rey) = 36/40 + 4/40 = 1. Esta conclusión es más general. Siempre:*

p(noA o A) = 1,

pues un suceso y su contrario ya vimos que verificaban que p(A) + p(noA) = 1.

- *Hay 10 bastos y hay 12 figuras, pero hay 4 figuras que son a la vez bastos (as, sota, caballo y rey), luego p(Basto o Figura) = 10/40 + 16/40 – 4/40 = 22/40 = 11/20.*

Resumen:
Suceso contrario: p(X) + p(noX) = 1.
Sucesos dependientes: p(A y B) = p(A) · p(B/A).
Sucesos compatibles: P(A o B) = p(A) + p(B) – p(A y B).

Actividades propuestas

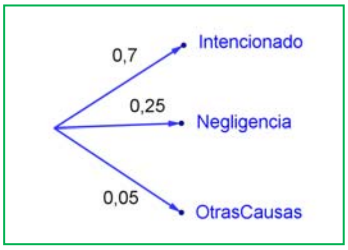
37. Lanzamos dos dados que no estén trucados y anotamos los números de su cara superior. Consideramos el suceso A que la suma de las dos caras sea 8, y el suceso B que esos números difieran en dos unidades. a) Comprueba que p(A) = 5/36 (2 + 6; 3 + 5; 4 + 4; 5 + 3; 6 + 2) y que p(B) = 8/36 ((1,3), (2, 4), …). b) Calcula las probabilidades de: p(A y B); p(A o B); p(A y noB); p(noA y B); p(noA y noB). c) Calcula p(A/B); p(A/noB); p(noA/B).

3.4. Experiencias compuestas: tablas de contingencia y diagramas de árbol

Diagramas de árbol

Ejemplo:

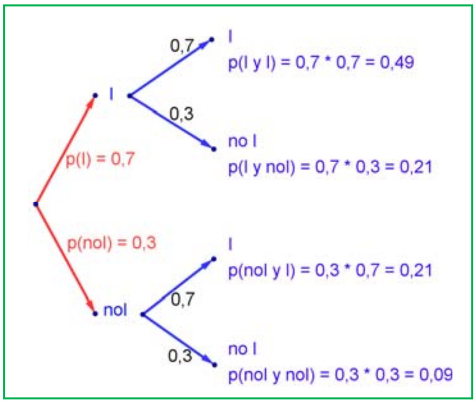
- *Se hace un estudio sobre los incendios y se comprueba que en una determinada zona el 70 % de los incendios son intencionados, un 25 % se deben a negligencias y 5 % a causas naturales como rayos o a otras causas. Representa esta situación con un diagrama de árbol.*



Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

Actividades resueltas

- *Si consideramos que la probabilidad de que un incendio sea intencionado es 0,7, ¿cuál es la probabilidad de que al considerar dos incendios, al menos uno haya sido intencionado?*



Llamamos I al suceso "ser intencionado" y noI al suceso "no ser intencionado". Representamos la situación en un diagrama de árbol. Como el que un incendio sea intencionado es independiente de cómo sea el segundo, tenemos que:

P(I y I) = 0,7 · 0,7 = 0,49

P(I y noI) = 0,7 · 0,3 = 0,21

ya que es la probabilidad de que el primer incendio sea intencionado y el segundo no.

P(noI y I) = 0,3 · 0,7 = 0,21

P(noI y noI) = 0,3 · 0,3 = 0,09

La probabilidad de que al menos uno haya sido intencionado la podemos calcular sumando las probabilidades de (I y I), (I y noI), y (noI y I) que es 0,49 + 0,21 + 0,21 = 0,91. Pero más sencillo es calcular la probabilidad del suceso contrario p(noI y noI) = 0,09 y restarla de 1:

p(al menos uno intencionado) = 1 – 0,09 = 0,91.

Actividades propuestas

38. Dibuja en tu cuaderno un diagrama en árbol para tres incendios, y calcula la probabilidad de que al menos uno haya sido intencionado siendo p(I) = 0,7.
39. En una aeronave se han instalado tres dispositivos de seguridad: A, B y C. Si falla A se pone B en funcionamiento, y si también falla B empieza a funcionar C. Las probabilidades de que funcione correctamente cada dispositivo son: p(A) = 0,95; p(B) = 0,97 y p(C) = 0,98. a) Calcula la probabilidad de que fallen los tres dispositivos. b) Calcula la probabilidad de que todo vaya bien.
40. Una fábrica de muñecas desecha normalmente el 0,5 % de su producción por fallos debidos al azar. Calcula la probabilidad de que: a) Al coger dos muñecas al azar haya que desechar ambas. b) Al coger dos muñecas al azar haya que desechar sólo una. c) Al coger dos muñecas al azar no haya que desechar ninguna d) Verificamos 4 muñecas, calcula la probabilidad de desechar únicamente la tercera muñeca elegida.
41. Lanzamos una moneda hasta que aparezca dos veces seguidas del mismo lado. Calcula las probabilidades de que: A) La experiencia termine al segundo lanzamiento. B) Termine al tercer lanzamiento. C) Termine en el cuarto. D) Termine a lo sumo en el cuarto lanzamiento (es decir, que termine en el segundo o en el tercero o en el cuarto lanzamiento).

Tablas de contingencia

Ejemplo:

- *Se han estudiado 500 enfermos del hígado analizando por un procedimiento nuevo si las lesiones son benignas o malignas. Luego se les volvió a analizar por el procedimiento usual determinando qué diagnósticos habían sido correctos y cuáles incorrectos. Los valores obtenidos se representan en la tabla:*

	Diagnóstico correcto	Diagnóstico incorrecto	Totales
Lesión maligna	206	12	218
Lesión benigna	268	14	282
Totales	474	26	500

Determinamos la tabla de frecuencias relativas:

	Diagnóstico correcto (C)	Diagnóstico incorrecto (I)	Totales
Lesión maligna (M)	0,412	0,024	0,436
Lesión benigna (B)	0,536	0,028	0,564
Totales	0,948	0,052	1

Actividades resueltas

- *Imagina que estas frecuencias relativas pudieran tomarse como probabilidades. Interpretamos entonces el significado de cada uno de estos valores.*

0,412 sería la probabilidad de que el diagnóstico de lesión maligna fuese correcto: p(M y C).

0,024 = p(M y I); 0,536 = p(B y C); 0,028 = p(B y I).

¿Y 0,436? El número de lesiones malignas es 218, luego 0,436 = p(M).

Del mismo modo: 0,564 = p(B); 0,948 = p(C); 0,052 = p(I).

Observa que p(M) + p(B) = 1 y que p(C) + p(I) = 1. Son sucesos contrarios.

Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF

¿Son dependientes o independientes los sucesos M y C?

Recuerda que p(M y C) = p(M) · p(C/M), por tanto: 0,412 = 0,436 · p(C/M), de donde p(C/M) = 0,412/0,436 = 0,945 que es distinto de 0,948 que es la probabilidad de C. Se puede afirmar que M y C son dependientes ya que p(C/M) ≠ p(C).

En general se denomina **tabla de contingencias** a:

	A	No A	
B	P(A y B)	P(noA y B)	P(B)
No B	P(A y noB)	P(noA y noB)	P(noB)
	P(A)	P(noA)	1

En una tabla de contingencia figuran todas las probabilidades o contingencias de los sucesos compuestos.

Observa que, como sabemos por la probabilidad del suceso contrario: p(A) + p(noA) = 1 y p(B) + p(noB) = 1.

Observa también que: p(A) = p(A y B) + p(A y no B), del mismo modo que p(B) = p(A y B) + p(noA y B) pues se obtienen sumando respectivamente la primera columna y la primera fila.

También: p(noA) = p(noA y B) + p(noA y no B) y p(noB) = p(A y noB) + p(noA y noB).

Actividades propuestas

42. Se ha hecho un estudio estadístico sobre accidentes de tráfico y se han determinado las siguientes probabilidades reflejadas en la tabla de contingencia:

	Accidente en carretera (C)	Accidente en zona urbana (U)	Totales
Accidente con víctimas (V)	0,27		0,56
Accidente con sólo daños materiales (M)			
Totales	0,58		1

a) Copia la tabla en tu cuaderno y complétala.

b) Determina las siguientes probabilidades: p(V y C); p(V y U); p(M y C); p(M y U); p(V); p(M); p(C) y p(U).

c) Calcula p(U/V); p(C/V); p(V/U); p(V/C). ¿Son dependientes o independientes los sucesos: accidente con víctimas y accidente en carretera?

43. Inventa una tabla de contingencia considerando que los accidentes puedan ser de carretera (C) o urbanos (U), pero que ahora los clasificamos en leves (L), graves (G) o mortales (M). *Observa que* lo fundamental para confeccionar la tabla es que los sucesos sean incompatibles dos a dos.

Diagramas de árbol y tablas de contingencia

Los diagramas de árbol y las tablas de contingencia están relacionados. Dado un árbol puedes obtener la tabla de contingencia, y viceversa. Tiene interés esta relación pues con los datos del problema a veces es más sencillo construir uno de ellos y dar la solución pasando al otro.

Actividades resueltas

Dada la tabla de contingencia, obtener el diagrama de árbol que comienza con A y noA.

	A	No A	
B	2/9	5/9	7/9
No B	1/9	1/9	2/9
	3/9 = 1/3	6/9 = 2/3	1

Conocemos la p(A) = 3/9 = 1/3, p(noA) = 6/9 = 2/3, p(B) = 7/9 y p(noB) = 2/9.

También conocemos p(A y B) = 2/9; p(A y noB) =1/9; p(noA y B) = 5/9 y p(noA y no B) = 1/9.

Nos falta conocer p(B/A) que podemos obtener dividiendo p(A y B) entre p(A): p(B/A) = p(A y B)/p(A) = 2/9 : 3/9 = 2/3.

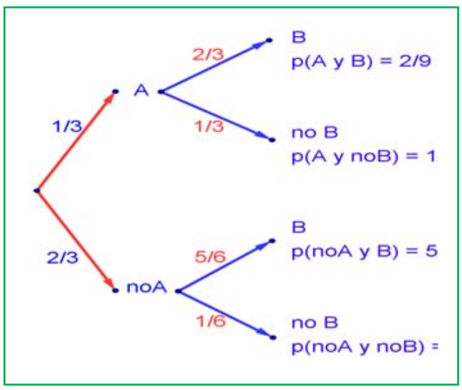
Del mismo modo calculamos:

p(noB/A) = p(A y noB)/p(A) = 1/9 : 3/9 = 1/3

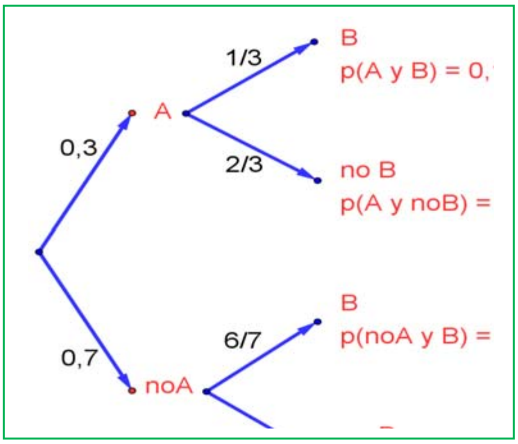
p(B/noA) = p(noA y B)/p(noA) = 5/9 : 6/9 = 5/6

p(noB/noA) = p(noA y noB)/p(noA) = 1/9 : 6/9 = 1/6.

El árbol es:



Autores: María Molero y Daniel García
Revisoras: Raquel Caro y Javier Rodrigo
Ilustraciones: Banco de Imágenes de INTEF



Actividades resueltas

- *Recíprocamente, dado el diagrama de árbol obtener el diagrama de contingencia:*

Ahora conocemos p(A) = 0,3 y p(noA) = 0,7. Además conocemos p(B/A) = 1/3; p(B/noA) = 6/7; p(noB/A) = 2/3 y p(noB/noA) = 1/7.

Calculamos, multiplicando: p(A y B) = 0,3·(1/3) = 0,1; p(A y noB) = 0,3·(2/3) = 0,2; p(noA y B) = 0,7·(6/7) = 0,6 y p(noA y noB) = 0,7·(1/7) = 0,1 que ponemos también en el árbol.

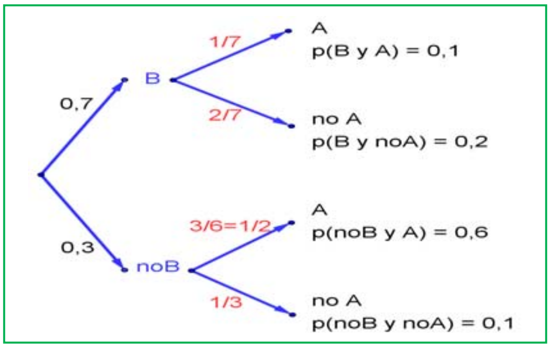
Rellenamos con estos datos, una tabla de contingencia:

	A	No A	
B	0,1	0,6	
No B	0,2	0,1	
	0,3	0,7	1

Calculamos, sumando, las casillas que nos faltan, p(B) = 0,1 + 0,6 = 0,7 y p(noB) = 0,2 + 0,1 = 0,3.

	A	No A	
B	0,1	0,6	0,7
No B	0,2	0,1	0,3
	0,3	0,7	1

Puede ser muy interesante pasar de un diagrama de árbol a la tabla de contingencia y de ésta, al otro diagrama de árbol, con el que podemos conocer p(A/B) = 0,1/0,7 = 1/7; p(noA/B) = 0,2/0,7 = 2/7; p(A/noB) = 0,3/0,6 = 3/6 = 1/2 y p(noA/noB) = 0,1/0,3 = 1/3.

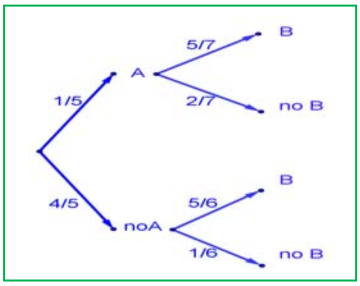


Actividades propuestas

44. Dada la tabla de contingencia, construye dos diagramas de árbol.

	A	No A	
B	0,4	0,2	0,6
No B	0,15	0,25	0,4
	0,55	0,45	1

45. Dado el diagrama de árbol, construye la tabla de contingencia, y después el otro diagrama de árbol.



46. Tenemos dos urnas, A y B. La primera con 8 bolas blancas y 2 bolas negras. La segunda con 4 bolas blancas y 6 bolas negras. Se saca una bola al azar, de una de las dos urnas, también al azar y resulta ser negra. ¿Cuál es la probabilidad de que proceda de la urna A?
47. Se está estudiando un tratamiento con un nuevo medicamento, para lo que se seleccionan 100 enfermos. A 60 se les trata con el medicamento y a 40 con un placebo. Los valores obtenidos se representan en la tabla adjunta

	Medicamento (M)	Placebo (no M)	
Curados (C)	50	30	80
No curados (no C)	10	10	20
	60	40	100

Se utilizan esos valores para asignar probabilidades. Calcula:

a) La probabilidad de que un enfermo curado haya sido tratado con el medicamento. *Ayuda*: p(M/C)

b) La probabilidad de que un enfermo curado haya sido tratado con el placebo. *Ayuda*: p(noM/C).

EJERCICIOS Y PROBLEMAS.

Estadística

1. En una clase se mira el color de los ojos de cada alumno y alumna y se obtiene lo siguiente:

N := negro; A := azul y V := verde.

N, N, A, V, N, V, A, N, A, N, V, A, A, N, N, N, V, A, N, N, A, N, V, N, N, A, N, A, N, N.

Haz una tabla de frecuencias absolutas, representa los valores en un diagrama de sectores y calcula la moda.

2. Las notas de un conjunto de alumnos de 4º son:

2, 10, 7, 8, 1, 0, 3, 5, 6, 9, 2, 4, 1, 6, 9, 10, 5, 6, 7, 8, 3, 1, 0, 1, 5, 9, 10, 9, 8, 7.

a) Haz una tabla de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas y frecuencias relativas acumuladas; b) Calcula la media, la mediana y la moda; c) Calcula la desviación típica y los cuartiles.

3. Se ha preguntado a 40 alumnos por el número de hermanos que tenía, y se ha obtenido

Número de hermanos	0	1	2	3	4	5	6 o más
Número de veces	5	15	7	6	4	2	1

a) Representa un diagrama de barras de frecuencias absolutas y un diagrama de líneas de frecuencias relativas; b) Calcula la media, la mediana y la moda.

4. Se han lanzado cuatro monedas 100 veces y anotado el número de veces que ha salido cara. Los resultados están reflejados en la tabla siguiente:

Número de caras	0	1	2	3	4
Número de veces	7	25	36	26	6

a) Escribe en tu cuaderno una tabla de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas y frecuencias relativas acumuladas; b) Representa un diagrama de barras de frecuencias absolutas acumuladas, un diagrama de líneas de frecuencias relativas y un diagrama de sectores de frecuencias absolutas; c) Calcula la media y la desviación típica; d) Calcula la mediana y los cuartiles.

5. Para conocer la distribución de un cierto país de las personas según su edad se ha recogido una muestra de diez mil personas y los valores obtenidos vienen reflejados en la tabla siguiente:

Edades	[0, 5)	[5, 10)	[10, 15)	[15, 25)	[25, 35)	[35, 45)	[45, 55)	[55, 65)	[65,100)
Número de personas	900	1000	900	1500	1300	1200	1300	900	1000

a) Utiliza las marcas de clase y escribe en tu cuaderno una tabla de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas y frecuencias relativas acumuladas; b) Representa un histograma de frecuencias absolutas. Cuidado: Los intervalos no son todos iguales. Recuerda: El área de los rectángulos debe ser proporcional a las frecuencias; c) Calcula la media y la desviación típica; d) Calcula la mediana y los cuartiles de forma gráfica usando un histograma de frecuencias absolutas acumuladas.

6. Con los datos del problema anterior calcula el intervalo [media – desviación típica, media + desviación típica]. ¿Cuántas personas están en dicho intervalo? ¿Qué porcentaje? Calcula también el intervalo [media – 2*desviación típica, media + 2*desviación típica] y [media – 3*desviación típica, media + 3*desviación típica]. Si la distribución fuera normal habría en el primer intervalo un 68 % de la muestra, en el segundo un 95 % y en el tercero más de un 99'7 %. Compara tus resultados con éstos.
7. Con los mismos datos calcula el intervalo intercuartil, e indica cuántas personas están en dicho intervalo y qué porcentaje.
8. Una compañía de seguros desea establecer una póliza de accidentes. Para ello, selecciona al azar a 200 propietarios y les pregunta cuántos euros han gastado en reparaciones del automóvil. Se han agrupado en intervalos los valores de la variable obtenidos:

Euros	[0, 100)	[100, 200)	[200, 400)	[400, 600)	[600, 800)	[800, 3000)
Número de personas	40	30	20	40	50	20

a) Calcula las marcas de clase y escribe en tu cuaderno una tabla de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas y frecuencias relativas acumuladas; b) Representa un histograma de frecuencias relativas. Cuidado: Los intervalos no son todos iguales; c) Calcula la media y la desviación típica; d) Calcula la mediana y los cuartiles de forma gráfica usando un histograma de frecuencias absolutas acumuladas.

9. Dos fabricantes de baterías de coches ofrecen su producto a una fábrica al mismo precio. La fábrica quiere elegir la mejor. Para ello escoge una muestra de 60 baterías de cada marca y obtiene de cada una los meses que ha funcionado sin estropearse. Obtiene la siguiente tabla:

Vida de la batería en meses	20	22	24	26	28	30	32
Marca A	2	7	13	16	12	8	2
Marca B	1	4	17	20	15	3	0

¿Qué marca crees que elegirá? Para tomar la decisión, calcula la media, la moda y la mediana para cada marca. Si aún no te decides, calcula el recorrido, la desviación típica, el intervalo [*m* – *s*, *m* + *s*] y el intervalo intercuartil.

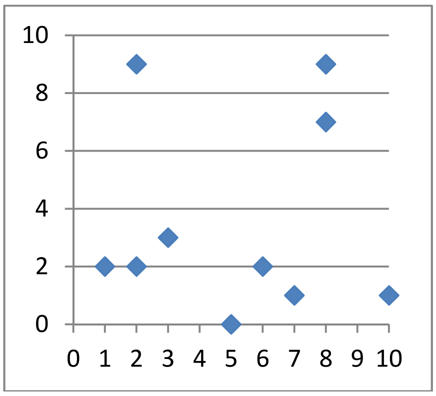
10. Haz un trabajo. Pasa una encuesta a tus compañeros y compañeras de clase. Hazles una pregunta con datos numéricos, como por ejemplo, cuánto mide su mano, qué número de zapato calzan, el número de libros que lee en un mes, el número de horas que ve la televisión a la semana, dinero que gasta al mes en comprar música… Representa los datos obtenidos en una tabla. Y haz un estudio completo. Puedes utilizar el ordenador:

a) Escribe en tu cuaderno una tabla de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas y frecuencias relativas acumuladas; b) Dibuja un diagrama de barras, un diagrama de líneas y un diagrama de sectores; c) Calcula la media, la mediana y la moda; d) Calcula la varianza y la desviación típica; e) Calcula los cuartiles y el intervalo intercuartil; f) Reflexiona sobre los resultados y escribe un informe.

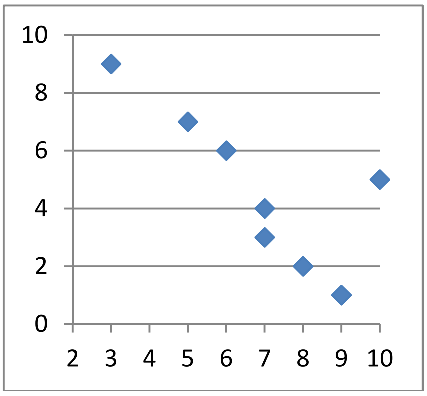
Coeficiente de correlación

11. Andrés ha calculado los coeficientes de correlación de las tres nubes de puntos adjuntas, y ha obtenido: –0,8, 0,85 y 0,03, pero ahora no recuerda cuál es de cada una. ¿Puedes ayudar a decidir qué coeficiente corresponde con cada nube?

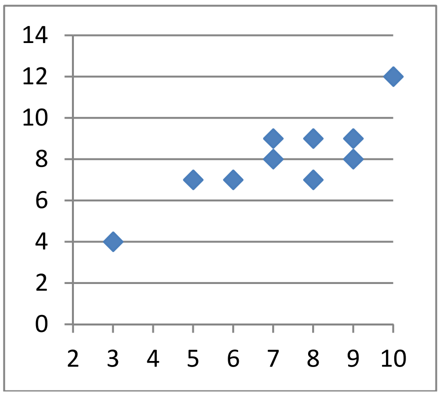
A



B



C



Probabilidad

12. En un colegio se selecciona un grupo de 200 estudiantes de los cuales todos estudian francés o inglés. De ellos 150 estudian inglés y 70 estudian francés. ¿Cuántos estudian francés e inglés? En otro centro escolar se estudian varios idiomas: francés, inglés, alemán, italiano. Se seleccionan también 200 estudiantes de los cuales, 150 estudian inglés, 70 francés y 40 ambos idiomas, ¿cuántos estudiantes de ese centro no estudian ni francés ni inglés?

13. Lanzamos un dado. Calcula la probabilidad de: a) Sacar un número impar. b) No sacar un 3. c) Sacar un número mayor que 3. d) Sacar un número mayor que 3 y que sea impar. e) Sacar un número mayor que 3 o bien que sea impar.

14. En una clase hay 24 alumnos y 14 alumnas. La mitad de las alumnas y la tercera parte de los alumnos tienen los ojos azules. Se elige un estudiante al azar. A) Calcula la probabilidad de que sea chico y tenga los ojos azules. B) Calcula la probabilidad de que sea chico o tenga los ojos azules.

15. Antonio, Juan y Jorge tienen una prueba de natación. Antonio y Juan tienen la misma probabilidad de ganar, y doble a la probabilidad de Jorge. Calcula la probabilidad de que gane Juan o Jorge.

16. Lanzamos dos monedas distintas, una de 50 céntimos y otra de un euro. Calcula la probabilidad de que: A) En la moneda de un euro salga cara. B) Salga una cara. C) Salga al menos una cara. D) No salga ninguna cara. E) Salga una cara y una cruz.

17. Lanzamos tres monedas. Calcula las probabilidades de: A) No salga ninguna cara. B) Salga al menos una cara. C) Salgan dos caras y una cruz.

18. Lanzamos dos dados y anotamos los valores de las caras superiores. Calcula las probabilidades de que la suma sea 1, sea 2, sea 3, …. sea 12.

19. ¿Qué es más probable al tirar tres dados, que la suma de sus caras superiores sea 9 o sea 10? Escribe el suceso "sea 9" y el suceso "sea 10" y calcula las probabilidades de sus sucesos elementales. ¡Sabes ya más que *Galileo*!

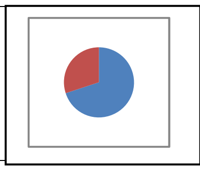
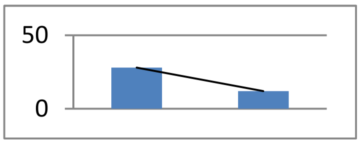
20. Lanzamos a la vez una moneda y un dado. Llama A al suceso "Salga cara y un número par". B al suceso "Salga cruz y un número primo" y C al suceso "salga un número primo". Calcula las probabilidades de A, B y C. ¿Cómo son estos sucesos? Indica cuáles de ellos son compatibles y cuáles son incompatibles.
21. Lanzamos una moneda 50 veces, ¿qué es más probable, obtener 50 caras seguidas o obtener en las primeras 25 tiradas cara y en las 25 siguientes cruz? Razona la respuesta.
22. Una moneda está trucada. La probabilidad de obtener cara es doble que la de obtener cruz. Calcula las probabilidades de los sucesos obtener cara y de obtener cruz al tirar la moneda.
23. Tres chicos y dos chicas juegan un torneo de ajedrez. Todos los chicos tienen idéntica probabilidad de ganar, y todas las chicas, también. Pero la probabilidad de ganar una chica es doble de la de ganar un chico. Calcula la probabilidad de que un chico gane el torneo.
24. Siete parejas de novios están en una habitación. Se seleccionan dos personas al azar. Calcula la probabilidad de: a) Sean un chico y una chica. b) Sean una pareja de novios. Ahora se escogen 4 personas al azar. Calcula la probabilidad de: c) Haya al menos una pareja de novios. d) No haya ninguna pareja de novios.
25. Tenemos un dado trucado de forma que los números impares tienen una probabilidad doble a la de los números pares. Calcula las probabilidades de: A) Salga un número impar. B) Salga un número primo. C) Salga un número primo impar. D) Salga un número que sea primo o sea impar.
26. En un grupo de 12 amigas hay 3 rubias. Se eligen dos chicas al azar. Calcula la probabilidad de que: A) Ambas sean rubias. B) Al menos una sea rubia. C) Ninguna sea rubia. D) Una sea rubia y la otra no.
27. Lanzamos dos dados y anotamos los valores de las caras superiores. Calcula las probabilidades de que: A) Los números obtenidos sean iguales. B) Los números obtenidos difieran en 3 unidades. C) Los números obtenidos sean pares.
28. Lanzamos una moneda hasta que salga cara. Calcula la probabilidad de que: A) Salga cara antes del cuarto lanzamiento. B) Salga cara después del octavo lanzamiento.
29. Un lote de 20 artículos tiene 2 defectuosos. Se sacan 4 al azar, ¿cuál es la probabilidad de que ninguno sea defectuoso?
30. Se lanzan dos dados y la suma de las caras superiores es 7. ¿Cuál es la probabilidad de que en uno de los dados haya salido un 3?
31. Se tienen 3 cajas, A, B y C. La caja A tiene 10 bolas de las cuales 4 son negras. La caja B tiene 6 bolas con una bola negra. La caja C tiene 8 bolas con 3 negras. Se coge una caja al azar y de esa caja se saca una bola, también al azar. Comprueba que la probabilidad de que la bola sea negra es 113/360.
32. Tenemos una moneda trucada cuya probabilidad de obtener cara es 3/5 y la de cruz es 2/5. Si sale cara se escoge al azar un número del 1 al 8, y si sale cruz, se escoge un número del 1 al 6. Calcula la probabilidad de que el número escogido sea impar.
33. En un proceso de fabricación de móviles se detecta que el 2 % salen defectuosos. Se utiliza un dispositivo para detectarlos que resulta que detecta el 90 % de los móviles defectuosos, pero señala como defectuosos un 1 % que no lo son. A) Calcula la probabilidad de que sea correcto un móvil que el dispositivo ha calificado como defectuoso. B) Calcula la probabilidad de que sea defectuoso un móvil que el dispositivo ha calificado como correcto. *Ayuda*: Utiliza primero un diagrama en árbol y luego una tabla de contingencia.

AUTOEVALUACIÓN

Con los datos siguientes, 1, 5, 2, 8, 9, 4, 7, 7, 5, 7, calcula:

1. La media: a) 5 b) 5'5 c) 6 d) 7
2. La mediana: a) 5 b) 5'5 c) 6 d) 7
3. La moda: a) 5 b) 5'5 c) 6 d) 7
4. La desviación típica: a) 2 b) 2,3 c) 2,5 d) 2,6
5. El intervalo intercuartil: a) 3 b) 2,75 c) 4 d) 2
6. Al tirar dos dados, la probabilidad de sacar al menos un 5 es:
   a) 5/6 b) 11/36 c) 25/36 d) 30/36
7. Al tirar 3 monedas, la probabilidad de sacar exactamente dos caras es:
   a) 1/2 b) 3/4 c) 3/8 d) 5/8
8. Al tirar 3 monedas, la probabilidad de sacar al menos dos caras es:
   a) 1/2 b) 3/4 c) 3/8 d) 5/8
9. Sacamos una carta de una baraja de 40 cartas, la probabilidad de que sea un oro o un múltiplo de 2 es:
   a) 22/40 b) 19/40 c) 36/40 d) 3/4
10. Indica cuál de las afirmaciones siguientes es **siempre** correcta:
    a) P(A) + P(noA) = 1 b) P(A y B) = P(A) · P(B) c) P(A o B) = P(A) + P(B)

RESUMEN

<table>
<tr><td></td><td></td><td>Ejemplos</td></tr>
<tr><td>Población y muestra</td><td>Población: Todo el conjunto de individuos sobre el que se hace el estudio. Muestra: Una parte de esa población.</td><td>Para conocer la intención de voto, la población es todo el país, y se selecciona una muestra</td></tr>
<tr><td>Frecuencia absoluta, relativa y acumulada</td><td>Frecuencia absoluta: Número de veces que se ha obtenido ese resultado. Frecuencia relativa: Se obtiene dividiendo la frecuencia absoluta por el número total. Frecuencia acumulada: Se obtiene sumando las frecuencias anteriores.</td><td><table><tr><td></td><td>Fr. Absoluta</td><td>Fr. Relativa</td><td>Fr. Acumulada Absoluta</td></tr><tr><td>A</td><td>28</td><td>0'7</td><td>28</td></tr><tr><td>B</td><td>12</td><td>0'3</td><td>40</td></tr></table></td></tr>
<tr><td>Gráficos estadísticos</td><td>Diagrama de barras
Diagrama de líneas
Diagrama de sectores
</td><td></td></tr>
<tr><td>Media</td><td>Media = m = $(x_1 + x_2 + \ldots + x_n)/n$</td><td>Con: 8, 4, 6, 10 y 10
Media = 38/5 = 7'6</td></tr>
<tr><td>Moda</td><td>Es el valor más frecuente</td><td>10</td></tr>
<tr><td>Mediana</td><td>Deja por debajo la mitad</td><td>4 < 6 < 8 < 10 = 10. Me = 8.</td></tr>
<tr><td>Varianza y Desviación típica</td><td>Varianza = $\frac{\sum_{i=1}^{n}(x_i-m)^2}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - m^2$. s = $\sqrt{\frac{\sum_{i=1}^{n}(x_i-m)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n} - m^2}$</td><td>Varianza = 5,4.
s = 2,33.</td></tr>
<tr><td>Cuartiles</td><td>Q1 deja por debajo la cuarta parte. Q3 deja por debajo las tres cuartas partes. Intervalo intercuartil = Q3 – Q1.</td><td>Q1 = 6; Q3 = 10; Intervalo intercuartil = Q3 – Q1 = 4.</td></tr>
<tr><td>Histograma</td><td>El área de cada rectángulo es proporcional a la frecuencia .</td><td></td></tr>
<tr><td>Correlación</td><td>El coeficiente de correlación, ρ, mide la relación entre dos variables. Es un número entre –1 y 1.</td><td>$\rho = 1 \rightarrow$ correlación perfecta positiva
$\rho = -1 \rightarrow$ correlación perfecta negativa
$\rho = 0 \rightarrow$ correlación nula
$\rho \in (0, 1) \rightarrow$ correlación positiva
$\rho \in (-1, 0) \rightarrow$ correlación negativa</td></tr>
<tr><td>Suceso</td><td>Al realizar un experimento aleatorio existen varios posibles resultados o sucesos posibles.
Un suceso es un subconjunto del conjunto de posibles resultados.</td><td>Tiramos un dado. Posibles resultados = {1, 2, 3, 4, 5, 6}
Suceso obtener múltiplo de 3 = {3, 6}</td></tr>
<tr><td>Probabilidad</td><td>Límite al que tienden las frecuencias relativas.
Si los sucesos elementales son equiprobables entonces:
p = casos favorables / casos posibles.</td><td>P(5) = 1/6.
P(sacar múltiplo de 3) = 2/6</td></tr>
<tr><td>Asignación de probabilidades</td><td>Suceso contrario: p(X) + p(noX) = 1.
Sucesos dependientes: p(A y B) = p(A) · p(B/A).
Sucesos compatibles: P(A o B) = p(A) + p(B) – p(A y B).</td><td>P(no 5) = 1 – 1/6 = 5/6.
P(5 o múl. 3) = 1/6 + 2/6 =3/6
P sacar primero un 5 y luego múltiplo de 3 =1/6·2/6 =2/36</td></tr>
</table>