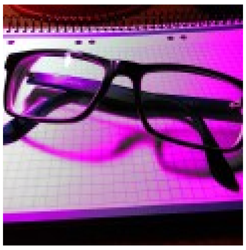
ESTADÍSTICA Y PROBABILIDAD

 Ángel Vázquez Hernández 2024



https://cienciamorada.es

Sumario

Este documento puede abrirse y editarse usando Libre Office.

Estatua de Florence Nightingale, Waterloo Place, Londres (Dominio público)

En 1854 Florence Nightingale llegó a un hospital sobrecargado y la mortalidad era altísima. Nightingale, como casi todo el personal sanitario de mediados del siglo XIX, nada sabía de microbios. Pero era muy buena en estadística, y sus informes sobre la influencia de las condiciones sanitarias en la mortalidad de los soldados heridos durante la Guerra de Crimea influyeron de forma decisiva en el diseño, construcción y gestión de hospitales.

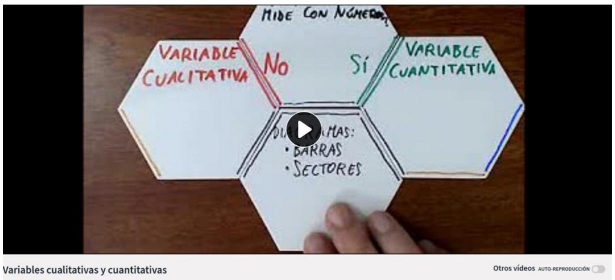
ESTADÍSTICA DESCRIPTIVA: TABLAS Y GRÁFICOS

Población, muestra y tipos de variables

Población y muestra. Representatividad de una muestra

- **Población:** sistema del que se está midiendo una variable.
- **Muestra:** parte del sistema en el que realmente se está midiendo una variable. Se dice que es representativa si podemos suponer que los resultados obtenidos en la muestra pueden ser extrapolados a la población.

Variables cualitativas, cuantitativas discretas y cuantitativas continuas.

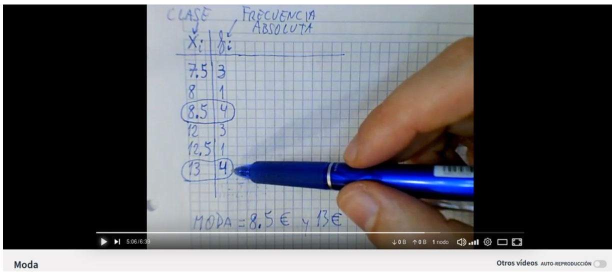


Las variables estadísticas pueden ser cualitativas o cuantitativas.

- **Variables cualitativas:** sin valor numérico. Ejemplos: color, ideología, gustos, sexo, género, etc.
- **Variables cuantitativas:** valores numéricos. Pueden ser discretas o continuas:
    - Discretas: números enteros. Ejemplo: número de hijos.
    - Continuas: números reales. Ejemplos: altura, masa, etc.

Cálculo de los parámetros de localización o posición de una variable estadística

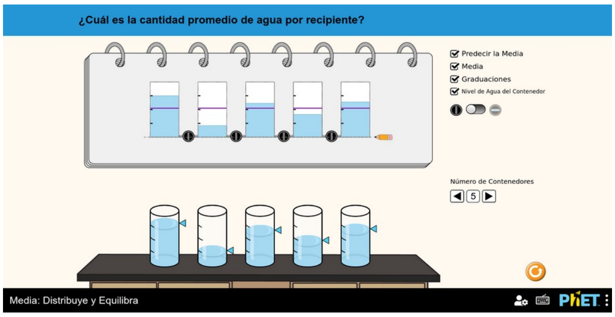
- **Frecuencia absoluta (f_i):** número de veces que aparece un valor en un conjunto de datos.
- **Número de datos (N):** $N=\sum f_i$
- **Moda:** valor de mayor frecuencia. Si hay varios de igual frecuencia se habla de distribución bimodal, trimodal, etc.



La moda es el valor (o los valores) que más se repite.

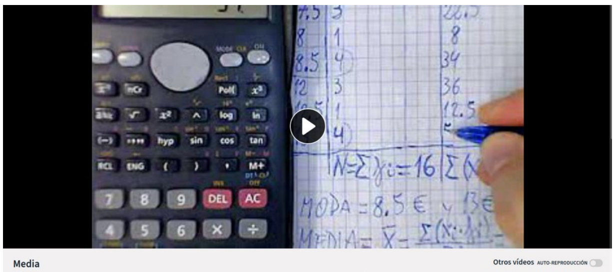
- **Media:** media aritmética de los datos.

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



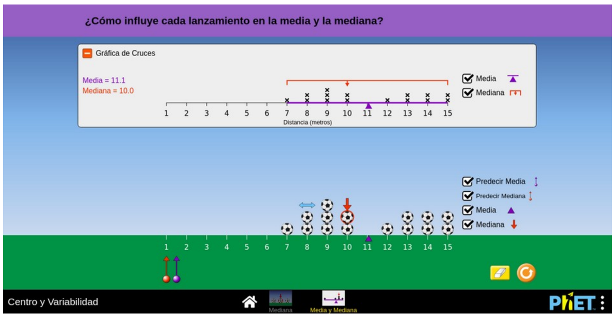
Media del agua contenida en un conjunto de recipientes

(Imagen: [Distribuye y Equilibra](https://phet.colorado.edu), [CC-By](https://phet.colorado.edu) [PhET Interactive Simulations University of Colorado Boulder](https://phet.colorado.edu) https://phet.colorado.edu)



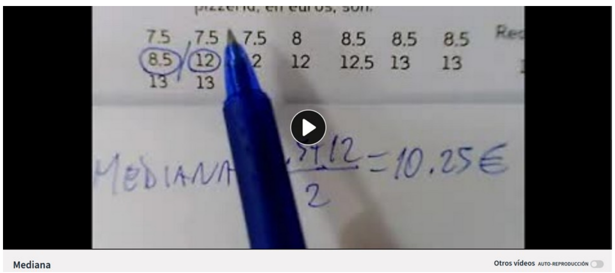
[En el cálculo de la media artimética es necesario tener en cuenta la frecuencia.]()

- **Mediana:** valor intermedio de una serie de datos ordenados de menor a mayor. Si el número de datos es par la mediana es la media aritmética de los dos datos intermedios.



Media y mediana son, junto a la moda, las formas más habituales de representar con pocos valores toda una serie de datos.

(Imagen: [Centro y Variabilidad](https://phet.colorado.edu), [CC-By](https://phet.colorado.edu) [PhET Interactive Simulations University of Colorado Boulder](https://phet.colorado.edu) https://phet.colorado.edu)



[La mediana es el valor intermedio de una serie ordenada de datos.]()

Representaciones gráficas de una variable estadística

Tablas de frecuencias.

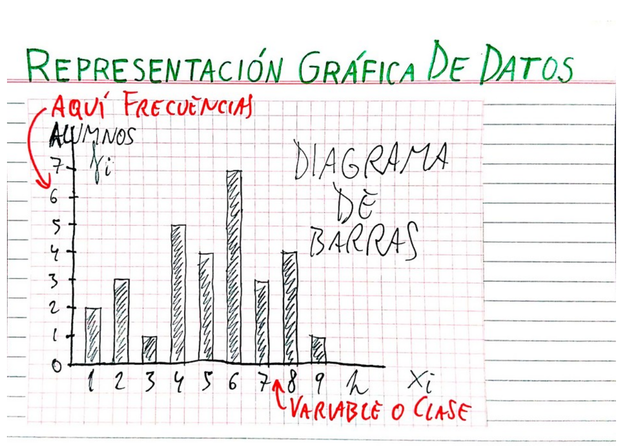
- **Frecuencia relativa (f_r):** $f_r = \frac{f_i}{N}$
- **Porcentaje:** $f_r \cdot 100$

Frecuencia relativa y porcentaje son dos formas de expresar la misma información.

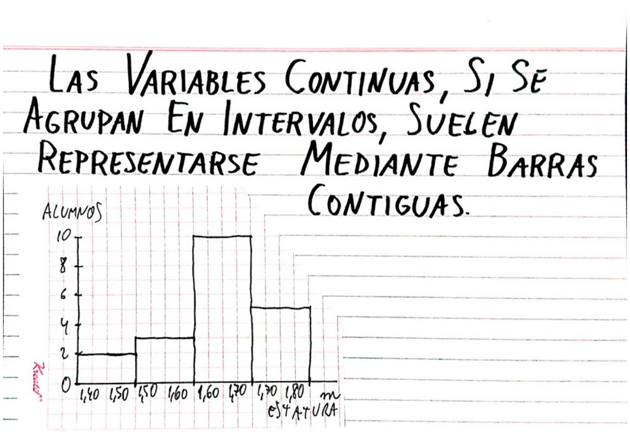
Gráficos estadísticos

Se utiliza una multitud de gráficos estadísticos, pero los más habituales suelen ser variantes de los diagramas de barras y de los diagramas de sectores:

En un diagrama de barras el tamaño de cada barra es directamente proporcional a la frecuencia de la clase que representa.



Ejemplo de diagrama de barras sencillo. En el eje vertical se ha representado la frecuencia absoluta, y en el eje horizontal la variable o clase.



Las variables cuantitativas continuas suelen ser agrupadas en intervalos:

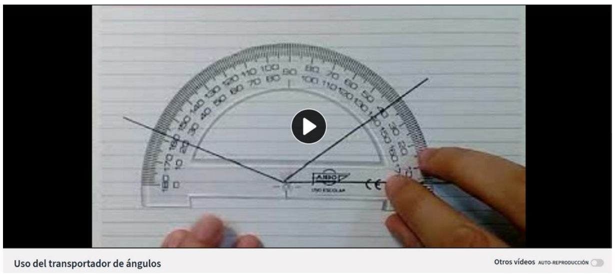
[a,b] desde **a** hasta **b**, ambos inclusive

[a,b) desde **a** hasta **b**, excepto **b**

(a,b] desde **a** hasta **b**, excepto **a**

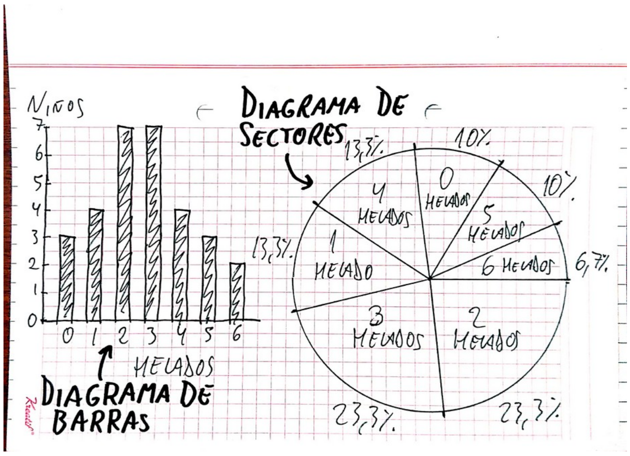
(a,b) desde **a** hasta **b**, excepto **a** y **b**

Marca de clase: $x_i = \frac{a+b}{2}$

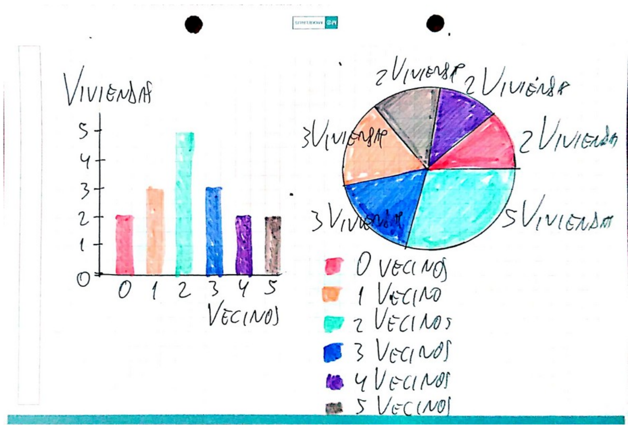


En el SI los ángulos se miden en **radianes**, pero la unidad más habitual es el **grado sexagesimal (símbolo º)**. [Se pueden medir con un transportador](#).

[En un diagrama de sectores el tamaño de cada sector es directamente proporcional al tamaño de la clase que representa](#).

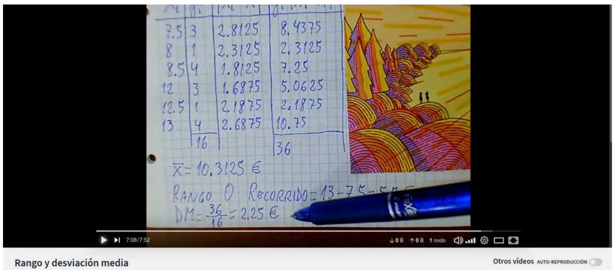


Ejemplo de diagrama de barras y diagrama de sectores. En el diagrama de sectores, en este caso, se han representado las frecuencias absolutas y los porcentajes (es imprescincible indicar a que clase corresponde cada sector).



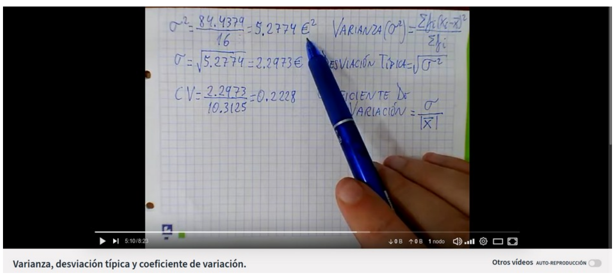
El uso de una leyenda de colores puede facilitar mucho la interpretación de los gráficos.

Cálculo de los parámetros de dispersión de una variable estadística



- **Rango o recorrido:** diferencia entre el valor máximo y el mínimo de la variable.
- **Desviación media:** media aritmética de los valores absolutos de las desviaciones.

$$DM=\frac{\sum (f_i \cdot |x_i-\bar{x}|)}{N}$$



- **Varianza:** media aritmética de los cuadrados de las desviaciones.

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

- **Desviación típica:** raíz cuadrada de la varianza.

$$\sigma=\sqrt{\sigma^2}$$

- **Coeficiente de variación:** cociente entre la desviación típica y el valor absoluto de la media.

$$CV=\frac{\sigma}{|\bar{x}|}$$

Actividades

1. En la baraja *Científicos y científicas* ¿Cuántas mujeres y cuántos hombres hay? Representa los datos en un diagrama de barras y uno de sectores.
2. En la baraja *Científicos y científicas* ¿Cuántas personas hay en cada área de conocimiento? Representa los datos en un diagrama de barras y uno de sectores.
3. En la baraja *Científicos y científicas* ¿Cómo se distribuyen por antigüedad? Representa los datos en un diagrama de barras y uno de sectores. Calcula moda, media y mediana.
4. Las edades de un grupo de víctimas de delitos sexuales se repartían en los siguientes intervalos: 18-20 años (1), 21-30 años (3), 31-40 años (11), 41-50 años (10), 51-64 años (4), 65-74 años (3). Representa los datos en un diagrama de barras y uno de sectores. Calcula moda, media y mediana.
5. Los precios de la carta de una pizzería, en euros, son:

7.5 7.5 7.5 8
8.5 8.5 8.5
8.5 12 12 12
12.5 13 13
13 13

Realiza un estudio estadístico completo.

6. Los precios, en euros, de las raciones de un bar son:

3 5 5.5 8
8 10 10

Realiza un estudio estadístico completo.

7. El número de piezas de fruta en unos tarros de conserva es:

6 6 7 8 6 6 6
7 7 7 8 8 7 6
8 7 7 6 6 6 7
8

Realiza un estudio estadístico completo.

8. El número de hijos de una serie de familias es:

0 0 3 2 1 1 3
2 2 2 3 0 1 1
2 3 2 0 1 1 2
3

Realiza un estudio estadístico completo.

9. El número de huevos que hay en un conjunto de nidos es:

1 1 3 2 3 2 2
1 2 3 2 2 1 3
2 1 2

Realiza un estudio estadístico completo.

10. En una serie de cajas hay las siguientes gominolas:

6 6 6 7 8 7 7
6 6 8 6 6 8 7
7 7 8 8 8 6 6
6 9 9 9

Realiza un estudio estadístico completo.

11. Los precios, en euros, de los "durum" de un restaurante turco son:

4 4 4.5 4.5 5 5 4.5
6.5 6.5 7 7 7 7

Realiza un estudio estadístico completo.

12. Los precios, en euros, de la carta de una pizzería son:

8 9 8 8
8.5 8.5 8.5
8 9 9
12.5 13.5
12.5 12.5 1313 13 12.5

Realiza un estudio estadístico completo.

13. Los precios de unos juguetes, en euros, son:

2 2 5 8
7 5 6
6 4 4 2
8 6 6
5 4 3 2
6 7 8

Realiza un estudio estadístico completo.

14. Los precios, en euros, de los complementos de un restaurante de comida rápida son:

2.7 2.7 2.7 2.7 2.7 2.7 1.95
4.95 1.75 3.25 1.95 3.5 1.95

Realiza un estudio estadístico completo.

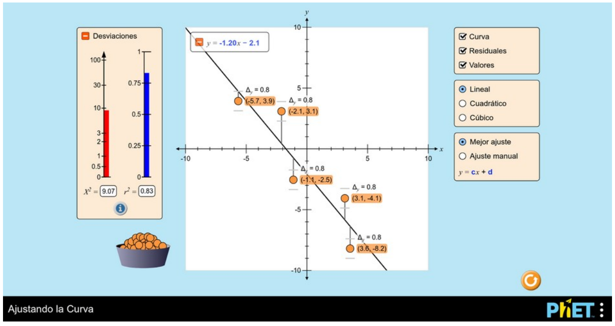
15. Las calificaciones de un grupo de alumnos y alumnas son las siguientes:

2 4 4 8
7 6 4
5 5 4 6 9 1 0
7 4 4 6 5 5 3
4 1

Realiza un estudio estadístico completo.

En las secciones de Ciencia Morada ¡A comer! y De viaje puedes encontrar más ejemplos de datos para practicar.

Representaciones gráficas de una variable bidimensional

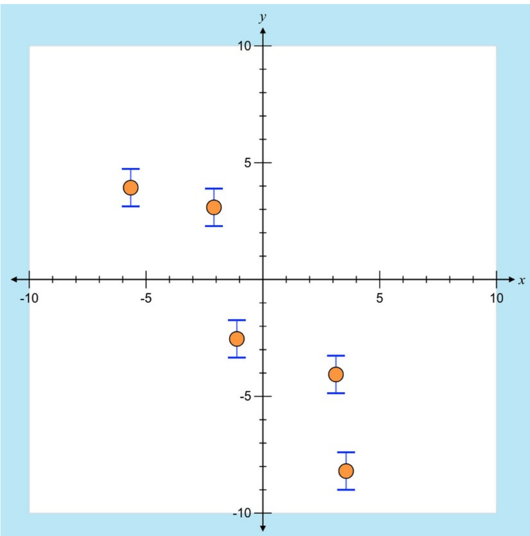


(Imagen: Ajustando la Curva. CC-By PhET Interactive Simulations University of Colorado Boulder https://phet.colorado.edu)

Si tenemos dos series de datos (**x** e **y**, por ejemplo) supuestamente relacionados entre ellos por una función desconocida, podemos buscar esa función intentando que las desviaciones respecto a la función teórica sean lo más pequeñas posibles. Un método muy utilizado para lograrlo es el llamado **método de mínimos cuadrados.**

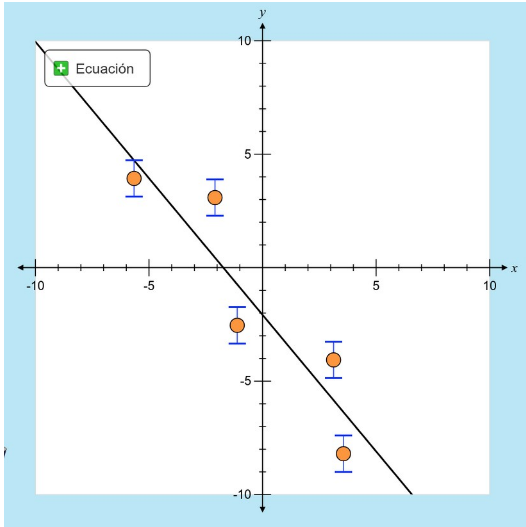
Pasos:

1. **Recopilar los datos y representarlos graficamente**: utilizar los pares de datos para crear una nube de **N** puntos en un sistema de dos ejes coordenados.



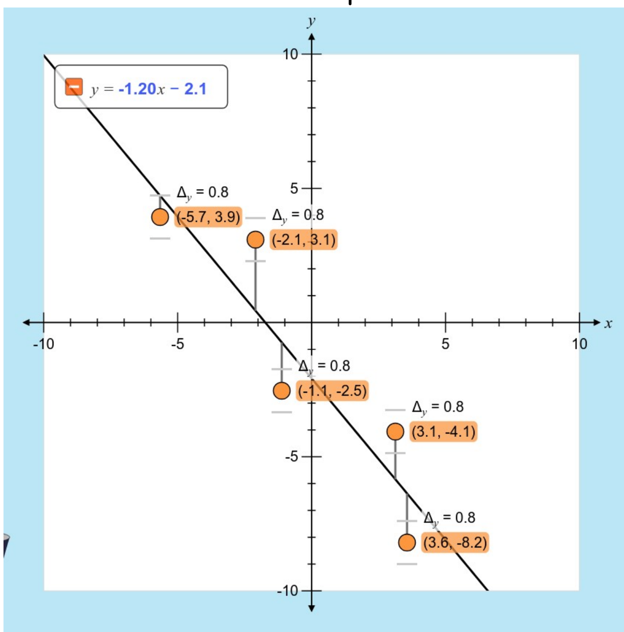
Nube de puntos obtenida al emparejar dos series de valores.

2. **Definir el tipo de modelo a utilizar**: representar los datos gráficamente en forma de una nube de puntos, y establecer el tipo de función que más se puede aproximar (lineal, cuadrática, etc).



En este caso optamos por una función lineal

3. **Ajustar la línea de regresión**: se busca la función lineal que menos se desvíe de los puntos.



Supongamos que la recta que estamos buscando tiene la forma **y=mx+n**. Para hallar los valores **m** y **n** debemos hacer lo siguiente:

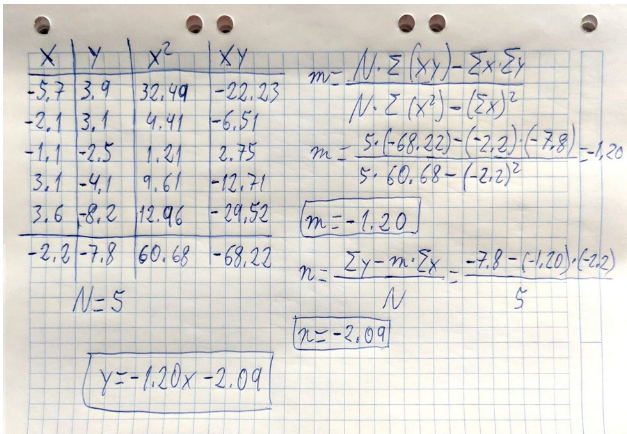
a) Calcular x^2 y xy para cada punto.
b) Calcular calcular Σx, Σy, Σx^2 y Σxy. Si estamos trabajando en una tabla será la suma de cada columna.
c) Calcular la pendiente m:

$$m=\frac{N\,\Sigma(xy)-\Sigma x\,\Sigma y}{N\,\Sigma(x^2)-(\Sigma x)^2}$$

d) Calcular la ordenada en el origen n:

$$n=\frac{\Sigma y-m\,\Sigma x}{N}$$

e) Construir la función lineal.



X	Y	X^2	XY
-5,7	3,9	32,49	-22,23
-2,1	3,1	4,41	-6,51
-1,1	-2,5	1,21	2,75
3,1	-4,1	9,61	-12,71
3,6	-8,2	12,96	-29,52
-2,2	-7,8	60,68	-68,22

N=5

$$m=\frac{N\cdot\Sigma(xy)-\Sigma x\cdot\Sigma y}{N\cdot\Sigma(x^2)-(\Sigma x)^2}$$

$$m=\frac{5\cdot(-68{,}22)-(-2{,}2)\cdot(-7{,}8)}{5\cdot 60{,}68-(-2{,}2)^2}=-1{,}20$$

$$\boxed{m=-1.20}$$

$$n=\frac{\Sigma y-m\cdot\Sigma x}{N}=\frac{-7{,}8-(-1{,}20)\cdot(-2{,}2)}{5}$$

$$\boxed{n=-2{,}09}$$

$$\boxed{y=-1{,}20x-2{,}09}$$

Cálculos de los que se deduce la ecuación de la recta que mejor define a la nube de puntos de este problema.

¡CUIDADO! Algunas veces hay que eliminar del estudio uno o varios puntos. Esto ocurre cuando, debido a alguna circunstancia atípica o a algún error de medida, uno de los puntos "se aleja" de la nube. No siempre es posible determinar cuando hay que hacerlo, pero a veces eliminar esos puntos anómalos dan lugar a un resultado más fiel a la realidad.

Una vez obtenida la función (una función lineal en este caso, aunque es posible ajustar la nube a otros tipos de funciones) podemos utilizar dicha función para realizar una estimación de valores que desconocemos.

AZAR Y PROBABILIDAD

[Es posible medir la probabilidad de un suceso.]

Experimentos aleatorios y deterministas

- **Experimentos aleatorios:** aquellos cuyo resultado no puede predecirse.

Ejemplo: lanzamiento de un dado. Salvo que el dado esté trucado es imposible saber si el resultado va a ser {1}, {2}, {3}, {4}, {5} o {6}.

- **Experimentos deterministas:** aquellos cuyo resultado puede predecirse con una alta probabilidad.

Ejemplo: disparo de una bala de fusil contra un huevo de gallina. Es casi seguro que la bala romperá el huevo en casi todos los experimentos.

Probabilidad de un suceso

Ley de los grandes números

- **Suceso:** cada uno de los posibles resultados de un experimento aleatorio.

Ejemplo: al lanzar un dado los sucesos posibles son {1}, {2}, {3}, {4}, {5}, {6}.

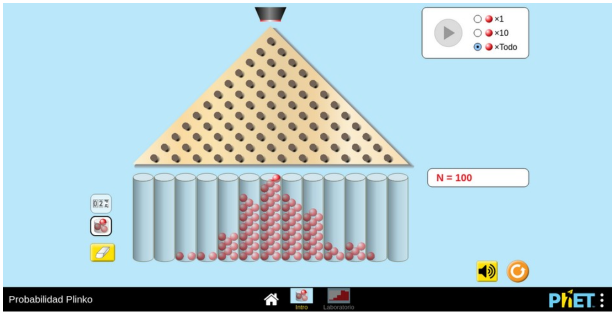
- **Probabilidad:** valor numérico que indica la frecuencia relativa esperable en un suceso.
- **Sucesos equiprobables:** sucesos cuya probabilidad es la misma.

Ejemplo: la probabilidad de obtener {3} al lanzar un dado es la misma que la de obtener {5}.

- **Ley de los grandes números:** a medida que el número de repeticiones de un experimento aumenta la frecuencia relativa de un suceso tiende a coincidir con su probabilidad.

Actividades

1. Intenta medir, mediante repetidos lanzamientos de chinchetas, la probabilidad de que una chincheta caiga de lado y la probabilidad de que caiga con la punta hacia arriba.
2. Intenta medir la probabilidad de que una bola caiga en cada una de las distintas casillas del dispositivo de la siguiente simulación:



(Imagen: Probabilidad Plinko, CC-By PhET Interactive Simulations University of Colorado Boulder https://phet.colorado.edu)

Ley de Laplace

La ley de Laplace dice que la probabilidad de un suceso puede calcularse dividiendo el número de casos favorables entre el número de casos posibles (siempre que la probabilidad de todos los casos posibles sea la misma).

Actividades

1. Tenemos dados de seis caras como los de la fotografía.

a) ¿Qué probabilidad hay de que, al lanzar uno de ellos, obtengamos un {3}.

b) ¿Qué probabilidad hay de que, al lanzar uno de ellos, obtengamos un número par?

c) ¿Qué probabilidad hay de que, al lanzar uno de ellos, obtengamos un número mayor que 2?

2. Si lanzamos un dado que tiene tres caras azules y tres caras rosas ¿Qué probabilidad hay de obtener una cara rosa? ¿Y qué probabilidad hay de obtener una cara azul?

3. Lanzamos un dado que tiene dos caras con el símbolo de "zona segura" (un triángulo rosa en un círculo blanco con borde verde) y las otras cuatro con los colores de las banderas arcoiris (LGTBIQ+), trans, bisexual y lesbiana ¿Qué probabilidad hay de que salga una de las banderas? ¿Y de que salga el símbolo de "zona segura"?

4. Calcula la probabilidad de que, al sacar una ficha de las de la foto al azar ocurra lo siguiente:

a) Qué la ficha sea morada.

b) Que la ficha sea azul.

c) Que la ficha sea blanca.

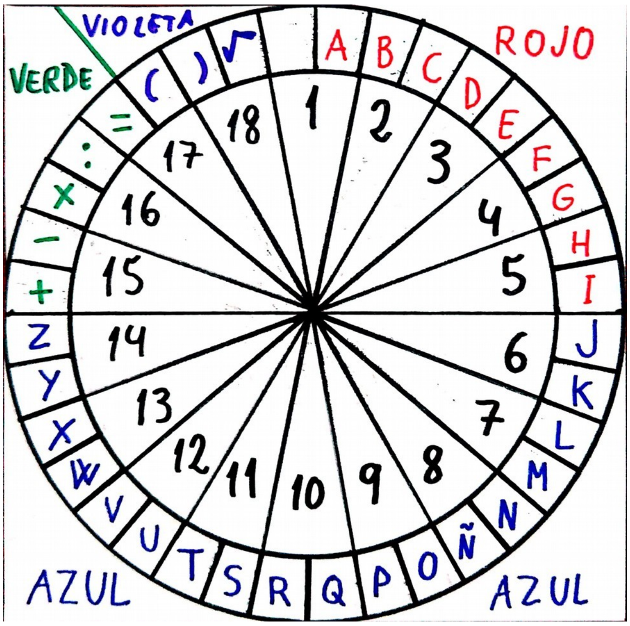
d) Que la ficha sea amarilla.

e) Que la ficha sea una bandera o contenga un triángulo.

5. Las fichas del problema anterior solo están decoradas por un lado. Si lanzamos una de ellas al aire ¿Qué probabilidad hay de que caiga por el lado decorado? ¿Y por el lado sin decorar?

6. Calcula la probabilidad de que, en una tirada de la ruleta de la foto, salga lo siguiente:

a) Rojo.
b) Azul.
c) Verde.
d) Violeta.
e) Par.
f) Impar.
g) Letra.
h) Vocal.
i) Letra.
j) Número.
k) Símbolo matemático.
l) Casilla en blanco.
m) Múltiplo de 3.



7. En el juego "Herstóricas Pioneras" hay seis parejas de naipes marcadas en color rojo, otras seis en color marrón claro, otras seis en marrón oscuro, otras seis en verde y otras seis en azul. Calcula la probabilidad de que, al sacar dos cartas al azar, **sin devolverlas al mazo**:

a) Las dos sean del mismo color.
b) Las dos sean de colores distintos.
c) Las dos sean de la misma pareja.
d) Las dos sean de parejas distintas.

El hecho de devolverlas o no al mazo influye sobre el número de cartas en juego y, por lo tanto, en la probabilidad de cada suceso.

Experimentos compuestos. Tablas de contingencia y diagramas en árbol

Sucesos dependientes e independientes

- **Sucesos independientes:** aquellos en los que el resultado de uno no influye en la probabilidad del otro.

Ejemplos: sacar un {3} en un dado de puntos rojos y sacar un {5} en un dado de puntos verdes, o sacar un {2} en un dado y, al volver a lanzarlo, sacar otro {2}.

- **Sucesos dependientes:** aquellos en los que el resultado de uno influye en la probabilidad del otro.

Ejemplo: en una baraja francesa de 52 cartas hay cuatro reinas. Si sacamos una reina y, sin devolverla, sacamos otra carta ¿Qué probabilidad hay de que se trate de otra reina?¿Y si la primera carta no fuera una reina?

Suma de probabilidades

A veces hay que sumar probabilidades de sucesos.

La probabilidad de un suceso compuesto de dos sucesos A y B es:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Ejemplo: "sacar par o más de 3 al lanzar un dado".

"Sacar par": $A=\{2,4,6\}$

"Sacar más de 3": $B=\{4,5,6\}$

$A \cap B = \{4,6\}$

$A \cup B = \{2,4,5,6\}$

La conjunción "o" suele indicar que el suceso ocurre cuando se cumple alguna de las condiciones o las dos. En esos casos se pueden sumar las probabilidades de los sucesos teniendo cuidado de no sumar varias veces los resultados comunes a distintos sucesos.

Actividades

1. En una baraja española hay 50 naipes: 12 de oros, 12 de copas, 12 de bastos, 12 de espadas y 2 comodines. Calcula la probabilidad de que, al sacar una carta, esta sea:

   a) De oros.

   b) De oros o de copas.

   c) De oros, copas o bastos.

   d) De oros, copas, bastos o espadas.

   e) De oros, o un caballo.

   f) De oros, o una sota o un caballo.

   g) Una sota o un caballo.

2. Calcula la probabilidad de sacar par o más de 3 al lanzar un dado.

3. En la baraja de cartas "Women in Science" hay 52 naipes repartidos en los siguientes equipos:

   - Biólogas y médicas: 3 de ellas son astronautas y 13 no lo son.
   - Astrónomas e ingenieras: 5 de ellas son astronautas y 16 no lo son.
   - Ecólogas y geólogas: 10 (ninguna es astronauta).
   - Matemáticas e informáticas: 11 (ninguna es astronauta).
   - Psicólogas y antropólogas: 8 (ninguna es astronauta).

   Calcula la probabilidad de que, al sacar una carta, sea:

   a) No astronauta.

   b) Bióloga o médica.

   c) A la vez astronauta y astrónoma o ingeniera.

   d) A la vez astronauta y ecóloga o geóloga.

e) A la vez psicóloga o antropóloga y no astronauta.

4. En una baraja francesa de 52 naipes, al sacar una carta ¿Qué probabilidad hay de que sea de corazones o una reina?

Producto de probabilidades

A veces hay que multiplicar probabilidades.

La probabilidad de un suceso formado por una serie de sucesos simples puede calcularse como el producto de las probabilidades de cada uno de esos sucesos.

Ejemplo: tenemos 1 carta amarilla, 2 naranjas, 3 verdes y 4 moradas. Calcula la probabilidad de sacar, sin devolverlas al mazo, 1 carta amarilla, luego 1 naranja, luego 1 verde y luego 2 moradas.

El adverbio "luego" y la conjunción "y" suelen indicar una serie de condiciones que deben cumplirse obligatoriamente. En esos casos se pueden multiplicar las probabilidades de los sucesos.

Actividades

1. Calcula la probabilidad del ejemplo anterior.
2. Calcula la probabilidad del ejemplo anterior, pero suponiendo que devolvemos cada carta al mazo y barajamos después de verla.
3. Tenemos 2 cartas amarillas, 3 naranjas, 4 verdes y 5 moradas. Calcula la probabilidad de sacar, sin devolverlas al mazo, 1 carta morada, luego 1 verde, luego 2 naranjas y luego 2 amarillas.
4. Resuelve el problema anterior, pero suponiendo que devolvemos cada carta al mazo justo después de ver su color.
5. Tenemos una baraja española con 12 cartas de oros, 12 de copas, 12 de bastos, 12 de espadas y 2 comodines (50 naipes en total). Calcula la probabilidad de sacar, sin devolverlas al mazo, 1 carta de oros, luego 2 cartas de copas, luego tres cartas de bastos y, al final, una carta de espadas.

6. Resuelve el ejercicio anterior suponiendo que devolvemos cada carta al mazo justo después de ver a que palo corresponde.

Uso de árboles y tablas

Actividades

1. En el juego "Tutty" hay 9 naipes marcados en color rojo teja, otras 9 en color marrón claro, 10 en verde y 10 en azul. Cuatro naipes representan a mascotas, siete a menores de edad, diez a personas adultas (jóvenes-maduros) y siete a personas mayores: en total 28 personajes. Calcula la probabilidad de:
    a) Sacar dos cartas seguidas del mismo color, suponiendo que no las devolvemos al mazo.
    b) Sacar dos mascotas seguidas, suponiendo que no las devolvemos al mazo.
    c) Sacar dos cartas de personas (personas humanas, no animales, plantas ni seres extraños no humanos) del mismo grupo de edad, suponiendo que no devolvemos las cartas al mazo.
2. Resuelve el problema anterior suponiendo que sí devolvemos las cartas al mazo.
3. En el juego "Eureka" el 48.08% de los naipes corresponden a mujeres, y el resto a hombres. Un 24% de las mujeres están marcadas en verde, otro 24% en azul, otro 24% en rojo y las demás en amarillo. De los hombres un 25.93% están marcados en azul, otro 25.93% en rojo, otro 25.93% en verde y los demás en amarillo. Si sacamos una carta al azar Calcula la probabilidad hay de que sea:
    a) Una mujer.
    b) Un hombre.
    c) Una mujer marcada en verde.
    d) Un hombre marcado en verde.
    e) Una mujer marcada en verde, azul o rojo.
    f) Un hombre marcado en verde, azul o amarillo.
    g) Una carta marcada en verde.
    h) Una carta marcada en amarillo.

Más problemas:

Más problemas: