MÉTODOS

ESTADÍSTICOS Y NUMÉRICOS

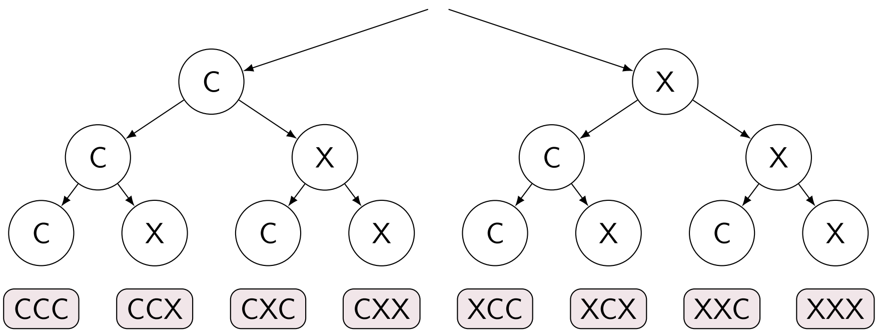
Índice

PROBABILIDAD

COMBINATORIA

La **combinatoria** es una rama de las Matemáticas que estudia las ordenaciones o agrupaciones de un determinado número de elementos.
Para ayudarnos a calcular la cantidad de ordenaciones o agrupaciones posibles, utilizaremos **diagramas de árbol** como apoyo.

Lanzamos una moneda 3 veces consecutivas y anotamos el resultado (cara o cruz) en el orden en el que aparecen. Escribiremos C para representar cara y X para representar cruz.



1. *a*) Lanzamos una moneda 4 veces consecutivas. ¿Cuántos son los resultados posibles?

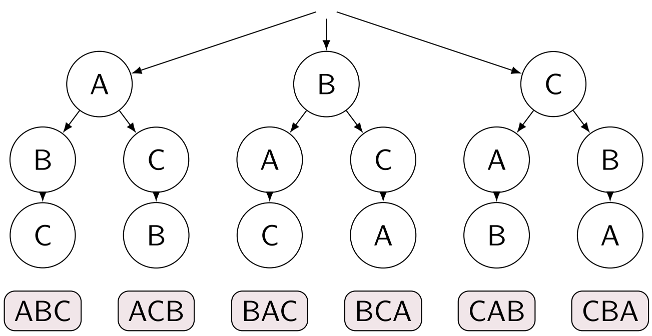
   b) Lanzamos la moneda 7 veces consecutivas. ¿Cuántos son los resultados posibles?

2. Tenemos tiras de cinco colores distintos (rojo, blanco, azul, verde y negro) para elaborar diseños de banderas.

   a) Con dos franjas, ¿cuántas banderas diferentes pueden crearse?

   b) ¿Y con tres franjas? ¿Y con diez franjas?

Ángela, Bárbara y Carolina se disputan los primeros puestos de una carrera. Veamos en qué orden podrían llegar.



El **factorial de un número natural** n es el producto de todos los números naturales desde 1 hasta n.

$$n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot (n-1) \cdot n$$

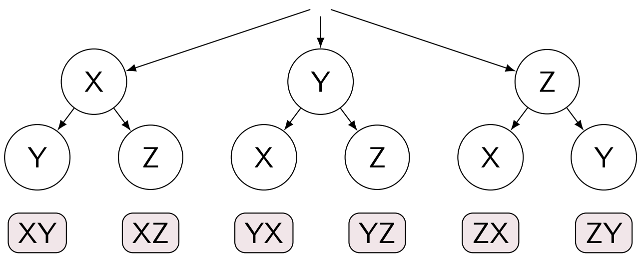
En la calculadora científica puede utilizarse la tecla $\boxed{x!}$.

El valor de $0!$ se define según el convenio de producto vacío.

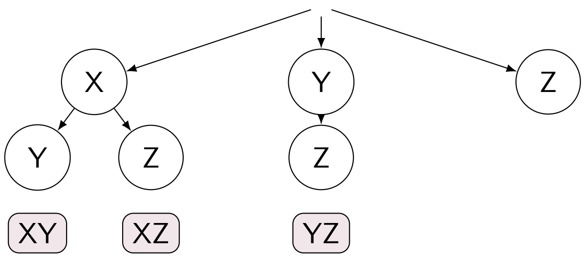
$$0! = 1$$

3. En la categoría junior de una competición participan 7 jóvenes deportistas. ¿De cuántas formas pueden ordenarse en la tabla de clasificación?
4. En una asignatura hay matriculadas trece personas. ¿De cuántas formas pueden ponerse en fila?
5. Tras mezclar las cartas de una baraja de 40 naipes, ¿de cuántas formas distintas podrían aparecer ordenadas?
6. ¿De cuántas formas diferentes se pueden sentar cinco personas en un banco de cuatro asientos?
7. ¿De cuántas formas diferentes se pueden sentar cinco personas en un banco de cinco asientos?

Veamos de cuántas formas diferentes se pueden sentar tres personas (X, Y, Z) en un banco de dos asientos.



Pero si solo nos interesa quienes se sientan pero no en que orden se sientan, observamos que en realidad varias de esas combinaciones aparecen repetidas.



Llamamos **número combinatorio** o **coeficiente binomial** de n sobre r a:

$$\binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$$

Para que tenga sentido definirlo de ese modo, es requisito imprescindible que ambos sean números naturales y que $n \geq r$.

En la calculadora científica puede utilizarse la tecla nCr .

8. ¿Cuántas combinaciones diferentes de 7 personas podríamos hacer en un banco de 3 asientos?
9. 5 personas juegan a las sillas musicales, así que hay colocadas 4 sillas. ¿Cuántas combinaciones posibles de personas sentadas hay?
10. Para aprobar un examen de cinco preguntas hay que contestar bien tres de ellas. ¿De cuántas formas diferentes se pueden elegir las tres preguntas?
11. Cinco personas desean cruzar un río en una barca en la que sólo pueden subir dos personas. ¿Cuántas formas diferentes tienen de cruzar el río?

En general, utilizando diagramas de árbol y ayudándonos con la calculadora podremos resolver multitud de problemas de combinatoria, pero debemos tener presente que no todas las situaciones se podrán plantear con las fórmulas vistas anteriormente.

Utiliza diagramas de árbol para hallar la solución:

12. ¿Cuántas elecciones distintas de delegado/a y subdelegado/a se pueden realizar en una clase de 25 personas?
13. *a*) ¿Cuántas maneras diferentes hay de formar un tren con cuatro coches de pasajeros y un coche cafetería?

    b) ¿Cuántas maneras diferentes hay de formar un tren con tres coches de pasajeros, un coche cafetería y un vagón de correos?
14. En una clase de 30 alumnos se elige a 5 al azar para realizar una tarea. ¿Cuántas elecciones distintas hay?
15. Con los dígitos $\{1, 3, 5, 7, 9\}$.

    a) ¿Cuántos números de tres cifras se pueden formar?

    b) ¿Cuántos números de tres cifras diferentes se pueden formar?

    c) ¿Cuántos números de cinco cifras se pueden formar?

    d) ¿Cuántos números de cinco cifras diferentes se pueden formar?

    e) ¿Cuántos de los números de cinco cifras diferentes son menores que $70\,000$?
16. ¿Cuántos números de tres cifras se pueden formar con los dígitos $\{0, 2, 4, 6, 8\}$? ¿Y de cinco?
17. En el juego de la primitiva hay 49 bolas numeradas, de las que los jugadores escogen seis. ¿Entre cuántas posibles combinaciones escogen al jugar?

Álgebra de conjuntos

Experimentos aleatorios y deterministas

Un **experimento** consiste en analizar un fenómeno en determinadas circunstancias.

- Un **experimento determinista** es aquel en los que se puede llegar a predecir el resultado, pues está determinado por las condiciones iniciales.

- Un **experimento aleatorio** es aquel en el que incluso con las mismas condiciones iniciales los resultados pueden ser diversos.

A cada uno de los posibles resultados de un experimento aleatorio se le llama **suceso elemental**.

El **espacio muestral** es el conjunto formado por todos los sucesos elementales.
Lo representaremos con la letra E o bien la letra griega Ω.

	E
Lanzamiento de un dado	$\{1, 2, 3, 4, 5, 6\}$
Lanzamiento de dos monedas	{XX, XC, CX, CC}

Un **suceso** es un subconjunto del espacio muestral, es decir, un conjunto formado por sucesos elementales. Un suceso puede estar formado por un solo suceso elemental, o por varios. Incluso todos, o ninguno.

Tipos de sucesos

- **Suceso elemental** es el formado por un único elemento del espacio muestral.

- **Suceso compuesto** es el formado por varios elementos del espacio muestral.

- **Suceso imposible** es el que nunca se verifica, pues no incluye ningún valor del espacio muestral. Se denota por el símbolo $\varnothing$.

18. Tenemos dado ordinario con sus caras numeradas. Nuestro experimento consiste en lanzarlo una vez, obteniendo el espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$.

    a) ¿Cuántos sucesos elementales tiene? ¿Cuáles son?

    b) ¿Cuántos sucesos compuestos de 2 elementos tiene?

    c) ¿Cuántos sucesos compuestos de 3 elementos tiene?

    d) ¿Cuántos sucesos compuestos de 4 elementos tiene?

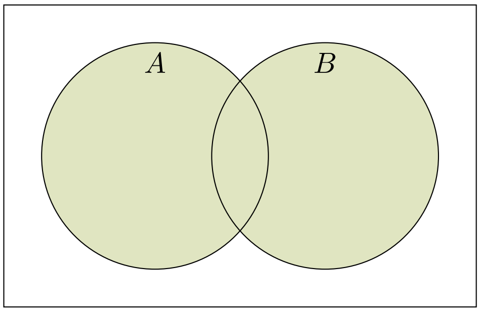
    e) ¿Cuántos sucesos compuestos de 5 elementos tiene?

    f) Entonces, ¿cuántos sucesos son posibles?

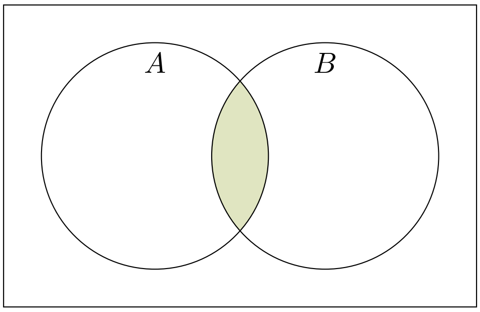
Operaciones con sucesos

Teniendo en cuenta que los sucesos son conjuntos, podemos realizar con ellos todas las operaciones que se definen para conjuntos:

El suceso **unión** ($A \cup B$) está formado por todos los elementos de A y todos los elementos de B. Es decir, son los elementos que pertenecen a A **o** pertenecen a B.

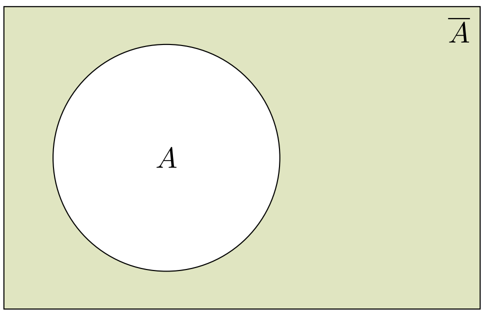


El suceso **intersección** ($A \cap B$) está formado por los elementos comunes de A y B. Es decir, son los elementos que pertenecen a A **y** pertenecen a B.

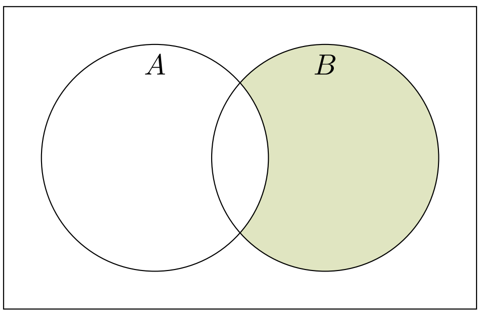


19. El espacio muestral del lanzamiento de dos monedas es $\Omega = \{$XX, XC, CX, CC$\}$.
    a) Indica cuáles son los elementos de los sucesos $A =$«obtener alguna cara», $B =$«obtener alguna cruz».
    b) Indica cuáles son los elementos de $A \cup B$ y los de $A \cap B$.
20. El espacio muestral del lanzamiento de un dado es $\Omega = \{1, 2, 3, 4, 5, 6\}$.
    a) Indica los elementos de los sucesos $A =$«obtener impar», $B =$«obtener menor que 5», $C =$«obtener mayor que 2».
    b) Indica los elementos de $A \cap B$, $A \cap B$ y $B \cap C$.
    c) Indica los elementos de $A \cup B$, $A \cup B$ y $B \cup C$.

El suceso **complementario** ($\overline{A}$) está formado por los elementos que no están en A.



El suceso **diferencia** ($B - A$) está formado por aquellos elementos de B que no están en A. Coincide con $B \cap \overline{A}$.



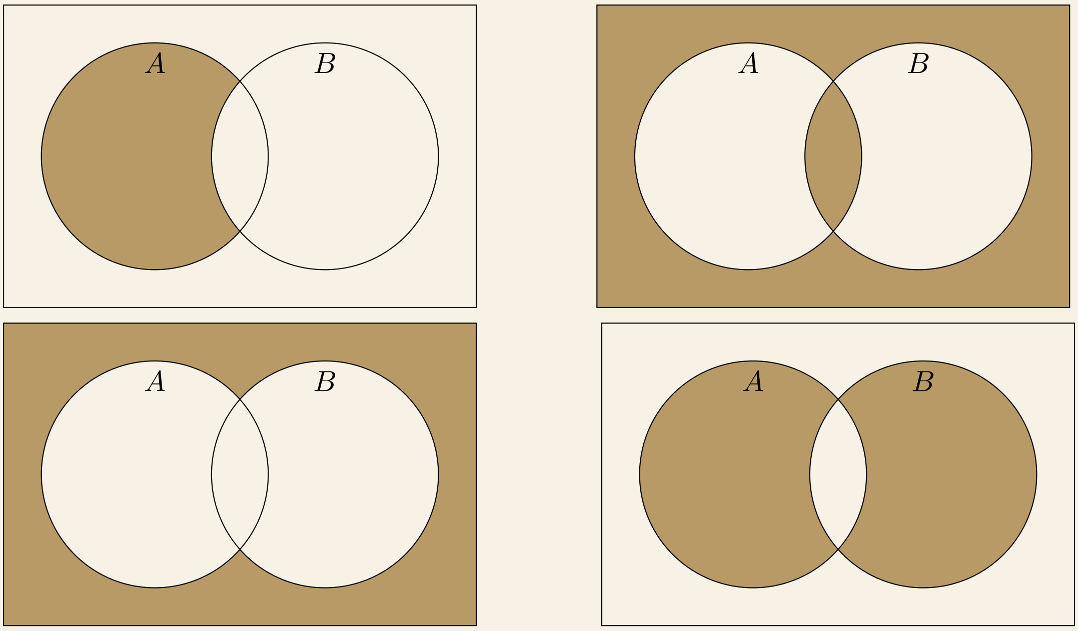
21. a) ¿Cuál es el espacio muestral del lanzamiento de tres monedas?
    b) Indica cuáles son los elementos de los sucesos $A =$«obtener alguna cara», $B =$«obtener alguna cruz», $C =$«obtener dos caras», $D =$«obtener una cruz».

c) Describe (con palabras o elementos) los sucesos:

a) $\overline{A}$ c) $\overline{C}$ e) $A - B$ g) $A - C$ i) $D - A$

b) $\overline{B}$ d) $\overline{D}$ f) $B - A$ h) $B - C$ j) $D - B$

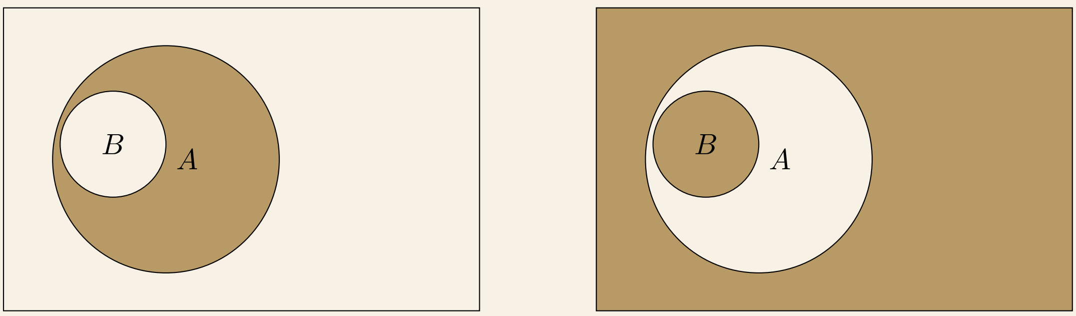
22. Identifica los siguientes sucesos:



23. El espacio muestral $\Omega = \{a, b, c, d, e\}$ contiene los sucesos $A = \{a, b, c, d\}$ y $B = \{a, e\}$. Halla:

a) $A \cup B$ b) $A \cap B$ c) $\overline{A \cup B}$ d) $\overline{A \cap B}$ e) $\overline{A} \cup \overline{B}$ f) $\overline{A} \cap \overline{B}$

24. Identifica los siguientes sucesos:



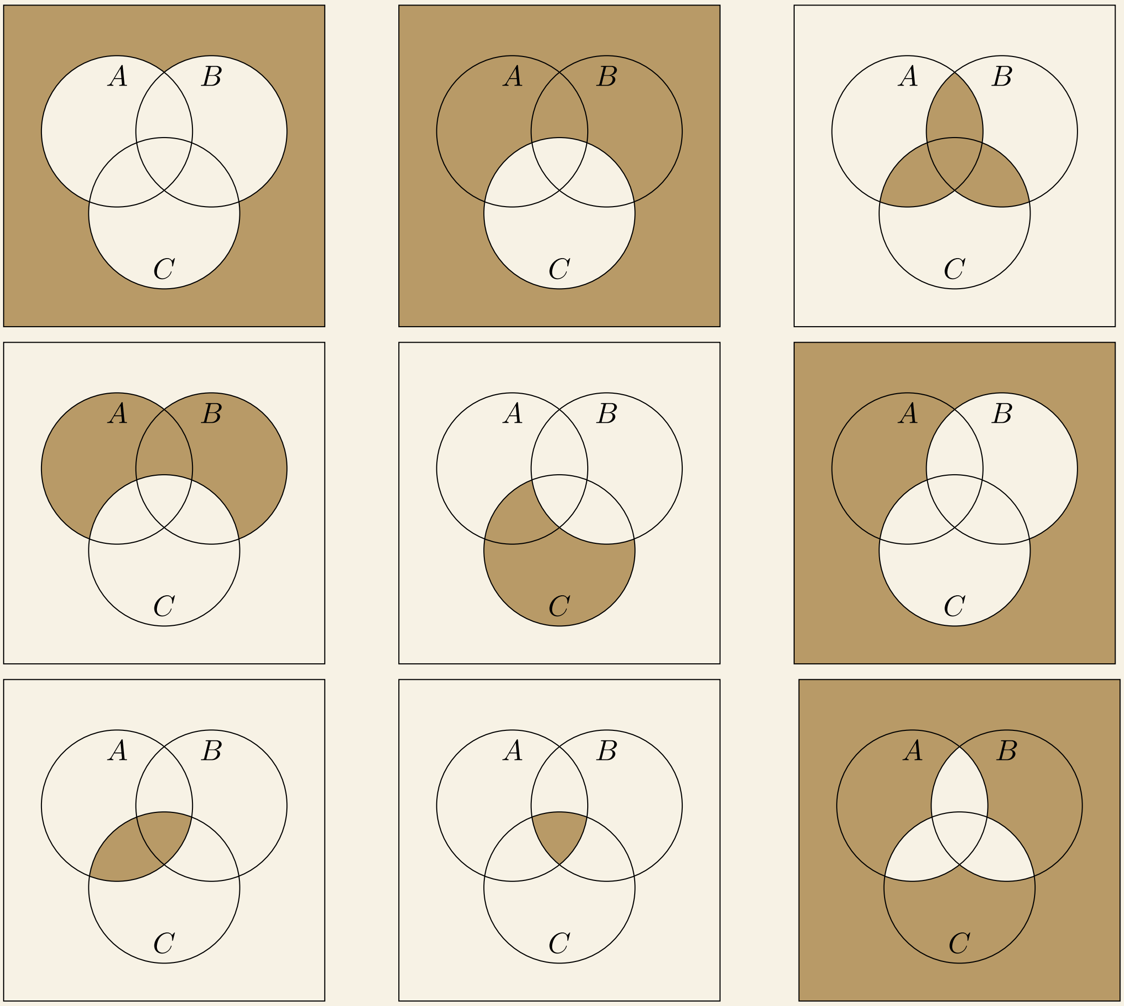
25. Al espacio muestral Ω = «letras del abecedario» pertenecen los sucesos A =«vocales», $B = \{m, a, t, e, s\}$ y $C = \{r, u, i, d, o\}$. Describe los siguientes sucesos:

a) $A \cup B$ d) $\overline{A \cap B}$ g) $A \cup C$ j) $\overline{B \cap C}$

b) $A \cap B$ e) $\overline{A} \cup \overline{B}$ h) $A \cap C$ k) $A \cap B \cap C$

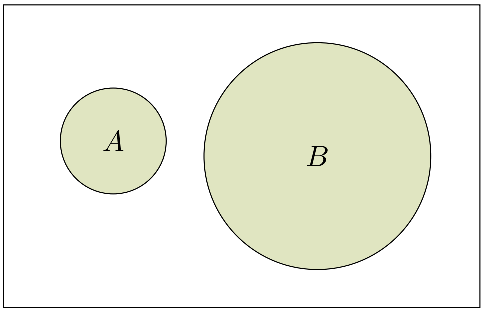
c) $\overline{A \cup B}$ f) $\overline{A} \cap \overline{B}$ i) $B \cup C$ l) $\overline{A} \cap (B \cup C)$

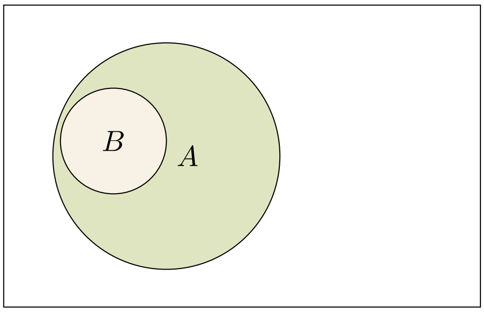
26. Identifica los siguientes conjuntos:



Decimos que dos sucesos son **incompatibles** cuando no tienen ningún elemento en común, es decir, $A \cap B = \varnothing$.

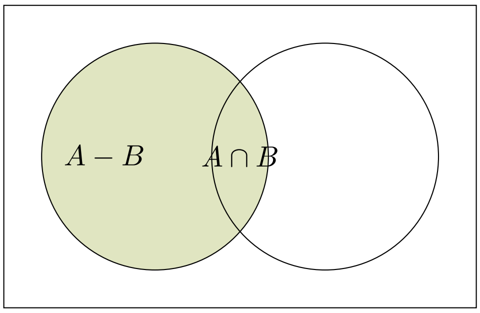
Por otro lado, decimos que un suceso está **contenido** ($A \subset B$) en otro si todos los elementos del primero pertenecen también al segundo.



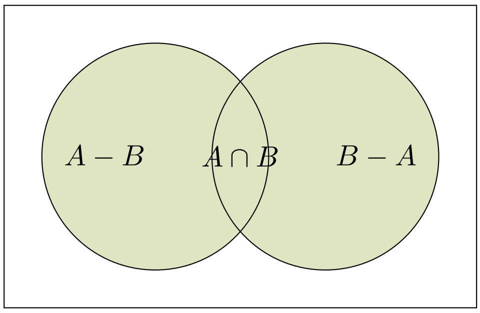


Observa que dos sucesos complementarios son siempre incompatibles ($A \cap \overline{A} = \varnothing$) pero el recíproco no es cierto en general.
Puede resultar interesante definir un conjunto en función de la unión de conjuntos incompatibles, pues el cálculo de probabilidades es más sencillo en ese caso.

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



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



27. En una urna hay 10 bolas numeradas del 1 al 10. Se saca una bola y se considera los sucesos:

$A =$ «Sacar múltiplo de 3»
$B =$ «Sacar múltiplo de 5»
$C =$ «Sacar número par»
$D =$ «Sacar número mayor que 10»
$F =$ «Sacar número menor o igual que 10»
$G =$ «Sacar número que no sea múltiplo de 3»

a) Indica cuál es el suceso seguro.
b) Indica cuál es el suceso imposible.
c) Halla dos sucesos complementarios.
d) Halla dos sucesos incompatibles.

28. Tenemos una bolsa con 9 bolas, numeradas del 1 al 9, y realizamos el experimento que consiste en sacar una bola de la bolsa, anotar el número y devolverlo a la bolsa.
Consideremos los sucesos $A =$«salir un número primo» y $B =$«salir un cuadrado perfecto».

a) Halla los sucesos $A \cup B$ y $A \cap B$.
b) ¿Son compatibles los sucesos A y B?
c) Halla $\overline{A}$ y $\overline{B}$.

29. En familias de tres hijos, se estudia la distribución de sus sexos.
Por ejemplo (H, M, M) significa que el mayor es varón y las otras dos son mujeres.

a) ¿Cuántos elementos tiene el espacio muestral?
b) Describe los sucesos $A =$ «la menor es mujer» y $B =$ «el mayor es hombre».
c) Halla $A \cap B$. ¿Son sucesos compatibles o incompatibles?

30. Realizamos un experimento que consiste en lanzar un dado y una moneda.

a) Describe el espacio muestral.
b) Describe los sucesos $A =$ «sacar menos de 3 en el dado» y $B =$ «sacar cruz en la moneda».
c) Halla $A \cap B$ y $A \cup B$.
d) Si además consideramos el suceso $C = \{1C, 2X, 3C, 4X, 5C, 6X\}$, describe $A \cap \overline{C}$ y $\overline{B} \cap C$.

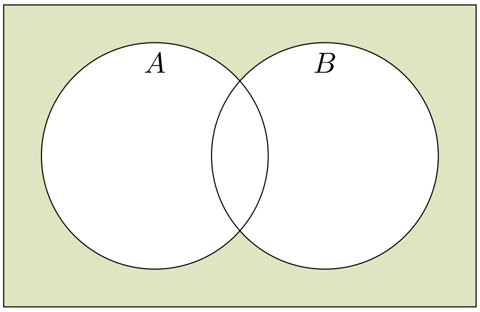
31. Sean A, B y C tres sucesos de un mismo espacio muestral. Expresa en función de ellos los siguientes sucesos:

a) Se realiza alguno de los tres.
b) No se realiza ninguno de los tres.
c) Se realizan los tres.
d) Se realizan, por lo menos, dos de los tres.
e) Se realiza B pero no se realiza C.
f) Se realiza solo A.

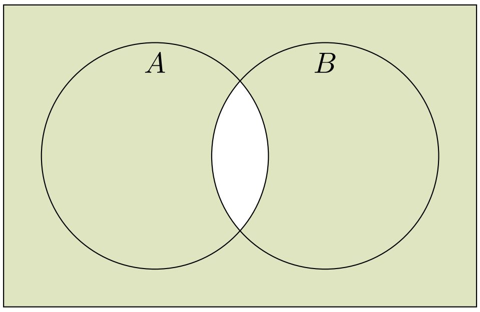
Leyes de Morgan

Las leyes de De Morgan son un par de reglas de transformación en lógica proposicional y álgebra de Boole que permiten expresar formalmente la relación entre la unión y la intersección de conjuntos bajo complementarios.

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$



$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$



32. De una baraja española de 40 cartas se extrae una carta. Si definimos los sucesos $A =$ «extraer un rey» y $B =$ «extraer una copa», describe los siguientes:

a) $A \cup B$
b) $A \cap B$
c) $A \cup \overline{B}$
d) $\overline{A} \cup \overline{B}$
e) $A - B$
f) $\overline{B} - \overline{A}$

Cálculo de probabilidades

En la probabilidad clásica se define la **frecuencia relativa** de un suceso en un experimento realizado n veces:

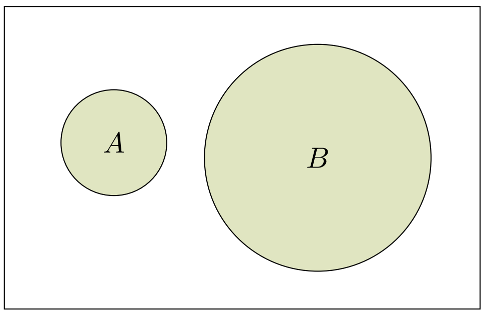
$$f(A) = \frac{\text{n}^{\text{o}} \text{ casos favorables}}{n}$$

Estos valores típicamente sufren oscilaciones, pero según crece n se van estabilizando en torno a un valor determinado. Así que buscamos definir la probabilidad de tal modo que refleje este hecho.

$$P(A) = \lim_{x \to \infty} f(A)$$

Para que además podamos construir un sistema probabilístico coherente, le exigimos que cumpla dos condiciones:

- La probabilidad total es un 100 % , es decir, $P(E) = 1$.
- Si dos sucesos son incompatibles, conocer la probabilidad de cada uno de ellos nos permite calcular la probabilidad de la unión.
  $A \cap B = \varnothing \Longrightarrow P(A \cup B) = P(A) + P(B)$



Por ello definimos la **probabilidad** de un experimento regular (es decir, en el que los sucesos elementales son equiprobables) del modo siguiente:

Regla de Laplace

$$P(A) = \frac{\text{n}^\text{o}\text{ sucesos } \textbf{elementales} \text{ contenidos en } A}{\text{n}^\text{o}\text{ sucesos elementales}}$$

33. ¿Cuál es la probabilidad al lanzar un dado de obtener un número impar?

34. Se lanza una moneda dos veces. ¿Cuál es la probabilidad de obtener cara una única vez?

35. En una baraja española de 40 cartas, estudia la probabilidad de que:

    a) salga un as   *b*) salga oros   *c*) salga el as de oros

36. En una urna hay 10 bolas numeradas del 1 al 10. Se saca una bola y se considera los sucesos:

    $A =$ «Sacar múltiplo de 3»
    $B =$ «Sacar múltiplo de 5»
    $C =$ «Sacar número par»
    $D =$ «Sacar número mayor que 9»
    $E =$ «Sacar número menor o igual que 10»
    $F =$ «Sacar número que no sea múltiplo de 3»

    a) Calcula la probabilidad de los sucesos descritos.

    b) Calcula la probabilidad de $A \cap B$, $A \cup B$, $A \cap F$, $A \cap F$.

37. Calcula la probabilidad de tener un boleto ganador al jugar a la lotería.

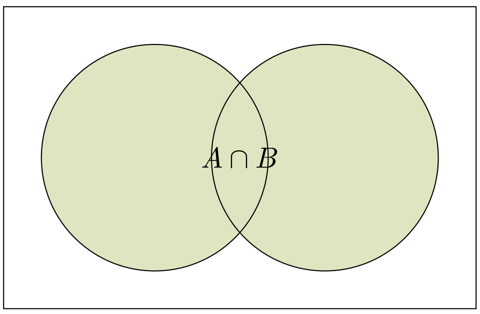
38. Calcula la probabilidad de tener una combinación ganadora al jugar a la primitiva.

Propiedades

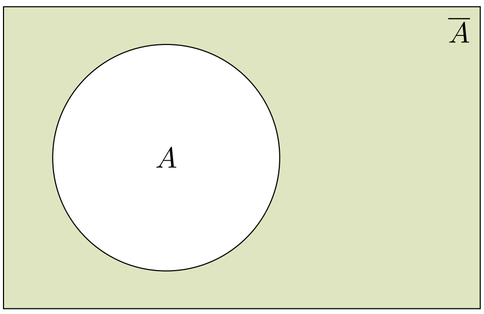
Es importante observar que la probabilidad de un suceso toma siempre valores entre 0 y 1. Será 0 cuando se trate de un sucesos imposible y será 1 cuando se trate de un suceso seguro; en general tomará cualquier otro valor intermedio.

Las propiedades fundamentales que nos permitirán resolver multitud de problemas son las siguientes:

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



- $P(\overline{A}) = 1 - P(A)$



Además podemos considerar interesantes las siguientes propiedades:

- $A \subset B \Longrightarrow P(B - A) = P(B) - P(A)$
- $P(A - B) = P(A \cap \overline{B}) = P(A) - P(A \cap B)$

39. Sean A y B dos sucesos aleatorios con probabilidades $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, $P(A \cap B) = \frac{1}{4}$.
Calcula:

    a) $P(\overline{A})$ b) $P(\overline{B})$ c) $P(A - B)$ d) $P(B - A)$ e) $P(A \cup B)$

40. Sean A y B dos sucesos aleatorios incompatibles con probabilidades $P(A) = \frac{3}{4}$, $P(B) = \frac{1}{5}$.
Calcula:

    a) $P(\overline{A})$ b) $P(\overline{B})$ c) $P(A - B)$ d) $P(B - A)$ e) $P(A \cup B)$

41. Sean A y B dos sucesos aleatorios con probabilidades $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{4}$, $P(A \cup B) = \frac{7}{10}$.
Calcula:

    a) $P(A \cap B)$ b) $P(A - B)$ c) $P(B - A)$ d) $P(\overline{A} \cap \overline{B})$ e) $P(\overline{A} \cup \overline{B})$

42. Sean A y B dos sucesos aleatorios con probabilidades $P(A) = \frac{3}{4}$, $P(B) = \frac{2}{5}$.
¿Son compatibles o incompatibles?

43. En un concesionario han observado las preferencias de los compradores respecto al color del coche. Un 25 % lo prefieren negro, un 10 % rojo, un 20 % azul y un 15 % gris.

    a) ¿Cuál es la probabilidad de que un comprador elija un coche negro o gris?
    b) ¿Cuál es la probabilidad de que elija un coche que no sea azul?
    c) ¿Cuál es la probabilidad de que elija un coche que no sea ni rojo ni azul?

44. Un 20 % del alumnado de una clase juega a balonmano, y el 40 % practica remo.
Si el 53 % de los alumnos practican alguno de los dos deportes, ¿cuál es el porcentaje que practica los dos deportes?

PROBABILIDAD CONDICIONADA

Si se dispone de información adicional sobre un experimento, esta puede modificar la probabilidad de un suceso.

Consideremos el experimento consistente en extraer una carta de una baraja española de 40 cartas y los sucesos $A =$ «obtener figura» y $B =$ «obtener rey».

$$P(A) = \frac{12}{40} = 0'3 \qquad P(B) = \frac{4}{40} = 0'1$$

Pero si sabemos que la carta extraída es una figura (es decir, se verifica A) la probabilidad de que sea un rey pasa a ser:

$$P(B \text{ sabiendo que se cumple } A) = \frac{4}{12} = 0'\widehat{3} \neq 0'1$$

Llamaremos **probabilidad de B condicionada por A** a la probabilidad de que ocurra B sabiendo que ha ocurrido A.

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

De ella se deduce inmediatamente la **regla del producto**:

$$P(A \cap B) = P(B/A) \cdot P(A)$$
$$P(A \cap B) = P(A/B) \cdot P(B)$$

45. El mismo profesor de matemáticas da clase, en tres grupos diferentes, a un total de 93 alumnos. En una evaluación los aprobados y suspensos de cada grupo fueron los siguientes:

	A	B	C
Aprobado	16	18	17
Suspenso	16	12	14

   a) Halla la probabilidad de que un alumno escogido al azar sea del grupo A.

   b) Halla la probabilidad de que un alumno escogido al azar sea del grupo A y haya aprobado.

   c) Halla la probabilidad de que un alumno del grupo A haya aprobado.

46. Sean A y B dos sucesos aleatorios con probabilidades $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, $P(A \cap B) = \frac{1}{4}$. Emplea las propiedades de las probabilidad y calcula:

   a) $P(B/A)$ b) $P(A/B)$ c) $P(\overline{B}/A)$ d) $P(B/\overline{A})$ e) $P(\overline{B}/\overline{A})$

47. En un instituto, el 80 % de alumnos de 2º de Bachillerato estudia Matemáticas, y el 60 % de los que estudian Matemáticas también estudia Física.
    Si se elije al azar un estudiante de ese curso, ¿cuál es la probabilidad de que estudie Matemáticas y Física?

48. Halla la probabilidad de que al extraer 3 cartas de una baraja española (de 40 cartas) todas sean de distinto palo.

49. Se extrae una carta de una baraja española de 40 cartas y sin reintegrarla a la baraja se extrae otra. Calcula las probabilidades de:

    a) Dos cartas de oros.
    b) Solo la segunda es de oros.

Decimos dos sucesos son **independientes** cuando la probabilidad de que ocurra uno no está influida por que haya ocurrido la otra, es decir, $P(A) = P(A/B)$ y $P(B) = P(A/B)$.

En ese caso, necesariamente se cumple que $P(A \cap B) = P(A) \cdot P(B)$.

50. Halla la probabilidad de que al extraer tres cartas, con reemplazamiento, de una baraja española, sean todas espadas.

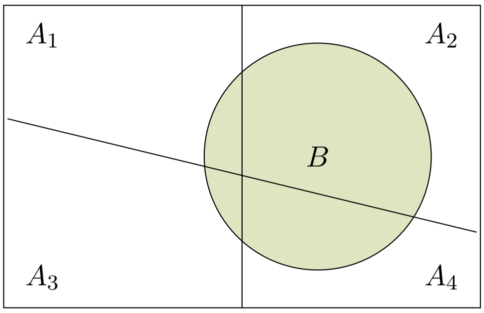
51. En una urna tenemos 4 bolas blancas, 3 bolas rojas y 3 bolas negras. Calcula la probabilidad de que en la primera extracción la bola sea blanca y en la segunda extracción la bola sea roja, en cada uno de estos casos:

    a) Devolvemos la bola a la urna tras la primera extracción.
    b) No devolvemos la bola a la urna tras la primera extracción.

Teorema de la probabilidad total

Si sabemos que los sucesos $A_1, A_2, \ldots, A_n$ son sucesos incompatibles y su unión es todo el espacio muestral, podemos utilizarlos para particionar el espacio.

Entonces se puede calcular la probabilidad de cualquier otros suceso B como:



$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_n)$$

Cuando conocemos la relación entre probabilidades de intersecciones y probabilidades condicionadas, podemos expresarlas del modo siguiente:

$$P(B) = P(B/A_1) \cdot P(A_1) + \cdots + P(B/A_n) \cdot P(A_n)$$

$$\boxed{P(B) = \sum_{i=1}^{n} P(B/A_i) \cdot P(A_i)}$$

52. Un gato persigue a un ratón. Este puede entrar en la puerta A, B o C, con probabilidades del 30 %, 50 % y 20 %. La probabilidad de que el gato cace al ratón si entra en la puerta A es del 40 %, en la B del 60 % y en la C del 10 %.
Calcula la probabilidad de que el ratón sea cazada.

53. En una clase hay 22 alumnos y 18 alumnas. El número de chicos que aprobaron un examen es 14, y número de chicas que aprobaron es también 14.
Si se escoge un alumno al azar, ¿cuál es la probabilidad de que haya suspendido?

54. En una fábrica de tornillos, una máquina fabrica el 40 % de los tornillos (de los que un 2 % son defectuosos), otra el 25 % (de los que el 5 % son defectuosos) y la última el resto (de los que el 3 % son defectuosos).

    a) Expresa esta información de forma esquemática o con una tabla de datos.

    b) Halla la probabilidad de que un tornillo sea defectuoso.

Teorema de Bayes

Resulta fundamental entender la diferencia entre las condiciones A/B y B/A, pues es frecuente cometer el error en el lenguaje no científico de confundir la una con la otra.

Supongamos de nuevo que los sucesos $A_1, A_2, \ldots, A_n$ son incompatibles y que su unión el todo el espacio muestral. Si deseamos conocer la probabilidad condicionada de uno de los sucesos A_k condicionada a otro suceso B, y conocemos la probabilidad de otro suceso B condicionada por cada uno de los sucesos de la partición, podemos relacionarlos.

$$P(A_k/B) = \frac{P(A_k \cap B)}{P(B)}$$

Si aplicamos la regla del producto en el numerador y la ley de probabilidad total en el denominador, sus probabilidades condicionadas se relacionan del modo siguiente:

$$P(A_k/B) = \frac{P(B/A_k) \cdot P(A_k)}{\displaystyle\sum_{i=1}^{n} P(B/A_i) \cdot P(A_i)}$$

55. En una zapatería hay tres estanterías. La primera tiene 50 pares de zapatos negros y 25 marrones, la segunda tiene 40 de cada color y la última 20 negros y 30 marrones.
Si cogemos un par de zapatos marrones, calcula la probabilidad de que proceda de la segunda estantería.

56. Un fabricante de coches desea lanzar un modelo de coches el próximo año. Al estudiar la posible situación económica que habrá, contempla tres alternativas económicas que considera equiprobables: inflación, estabilidad o depresión. La probabilidad de que se lance al mercado el coche es del 70 % si hay inflación, 40 % si hay estabilidad y 10 % si hay depresión.

    a) Calcula la probabilidad de que el coche se lance al mercado.

b) Si se lanza al mercado, determina la probabilidad de que exista inflación.

57. Disponemos de una prueba diagnóstica con una sensibilidad (proporción de pacientes enfermos que obtuvieron un resultado positivo) es del 85 % y cuya especificidad (proporción de pacientes sanos que obtuvieron un resultado negativo) es del 90 %.

    a) En la que la enfermedad es endémica y tiene una prevalencia de 12 %:

        (I) ¿Qué probabilidad tiene un paciente de obtener un falso positivo?

        (II) ¿Qué probabilidad tiene un paciente de obtener un falso negativo?

    b) En una zona en la que la enfermedad es rara y tiene una prevalencia de 0'1 %:

        (I) ¿Qué probabilidad tiene un paciente de obtener un falso positivo?

        (II) ¿Qué probabilidad tiene un paciente de obtener un falso negativo?

58. Disponemos de una prueba diagnóstica excelente, con una sensibilidad y una especificidad ambas de 95 %, para una enfermedad muy rara con una prevalencia de 0'02 %.

    a) ¿Qué probabilidad tiene un paciente de obtener un falso positivo?

    b) ¿Qué probabilidad tiene un paciente de obtener un falso negativo?

59. En la sala de pediatría de un hospital, el 60 % de los pacientes son niñas. El 35 % de los niños y el 20 % de las niñas son menores de 24 meses. La pediatra llama al primer paciente de la lista.

    a) Determina la probabilidad de que el paciente sea menor de 24 meses.

    b) Si tiene menos de 24 meses, ¿cuál es la probabilidad de que sea una niña?

Estadística descriptiva unidimensional

Muestreo

La **estadística** es la ciencia que estudia la técnica o método que se sigue para recoger, organizar, resumir, representar, analizar, generalizar y predecir resultados de las observaciones de fenómenos aleatorios.

El grupo de individuos u objetos que deseamos estudiar se llama **población**.

Una población puede ser finita o infinita.

- Finita es aquella que tiene una cantidad limitada de elementos.
  P.ej. los países del mundo en la actualidad.
- Infinita es aquella que tiene una cantidad ilimitada (o imposible de manejar) de elementos.
  P. ej. el número de estrellas que ha habido en el universo.

Con frecuencia es imposible o nada práctico recoger datos relativos a las características de toda la población, por lo que se examina una pequeña parte del grupo llamada **muestra**.

Una **muestra representativa** debe tener las mismas propiedades y proporciones que la de la población a la que pertenece, y por ello es posible inferir conclusiones sobre la población a partir del análisis de la muestra.

Variables estadísticas

Se llama **variable estadística** a cada una de las cualidades o propiedades que poseen los individuos de una población, que pueden variar adoptando distintos valores medibles u observables.
El objetivo de un proceso estadístico es estudiar las distintas variables, relacionándolas con el problema o fenómeno que está siendo investigado.

Tipos de variables estadísticas

- Variable **cualitativa** es la que expresa una cualidad, que no se puede cuantificar.
  - dicotómica: Hay dos respuestas posibles (sí/no)
  - ordinal: Hay una secuenciación o jerarquía (medalla de oro, plata, bronce)
  - nominal: No se agrupa de ningún modo.

- Variable **cuantitativa** es la que se expresa numéricamente.
  - **discreta**: Solo puede tomar valores aislados.
  - **continua**: Puede tomar valores dentro de un intervalo.

1. Di, en cada caso, cuál es la población y cuál la variable que se quiere estudiar. Especifica si es una variable cualitativa o cuantitativa, determinando en este último caso si es discreta o continua:

   a) Estudios que quiere hacer el alumnado de un centro escolar al acabar la ESO.

   b) Horas que dedica a estudiar el alumnado de Bachillerato.

   c) Número de televisores que hay en los hogares españoles.

   d) Intención de voto en unas elecciones municipales.

   e) Peso al nacer de los bebés que nacieron en Galicia durante el año pasado.

   f) Profesiones que quieren tener los estudiantes de un centro escolar.

   g) Número de animales de compañía que hay en los hogares españoles.

   h) Tiempo semanal que dedican a la lectura los menores de edad en España.

   i) Número de tarjetas amarillas mostradas en los partidos de fútbol de la temporada pasada.

2. Clasifica las siguientes variables:

   a) Estado civil.

   b) Número de parados de un país.

   c) Color de ojos.

   d) Temperaturas medias en un observatorio cada media hora.

   e) Número de acciones vendidas en un día en la Bolsa de Valores.

   f) Longitud de los 1000 tornillos producidos en una fábrica en una hora.

   g) Ingresos anuales de los empleados de una empresa.

   h) Provincia de nacimiento.

   i) Número de coches fabricados en España en un día.

3. Responde, razonadamente, si cada uno de los siguientes estudios se puede hacer tomando las respectivas poblaciones completas o es necesario tomar muestras:

   a) La profesión que desea tener el alumnado de tu clase.

   b) Intención de voto de cada español con derecho a votar.

   c) El número de horas diarias que ven la TV las personas de entre 14 y 16 años de edad de una misma ciudad.

Estadística descriptiva y estadística inferencial

Estadística descriptiva

La **estadística descriptiva o deductiva** es la parte de la estadística que se ocupa de describir y analizar a un grupo dado sin sacar conclusiones sobre un grupo mayor.
Es el método que seguimos para recoger datos, clasificarlos, ordenarlos y compararlos. Es decir, es un método que sirve para acumular información, analizarla o sintetizarla para describir un fenómeno.
Su instrumento fundamental es la **encuesta**.

Estadística inferencial

La **estadística inferencial o inductiva** es la parte de la estadística que se basa en el estudio y análisis de muestras para obtener conclusiones sobre la población y analiza los resultados para aceptar o rechazar una hipótesis.
Este análisis tiene en cuenta el azar, apareciendo así el lenguaje de las probabilidades.

Técnicas de muestreo

En estadística se conoce como **muestreo** al procedimiento realizado para elegir la muestra.
Los aspectos fundamentales en el muestreo son:

- El tamaño de la muestras, puesto que si es demasiado pequeña no obtendremos de ellas información concluyente.
- El método empleado para la selección de los individuos que la forman.

Llamamos **sesgo** al error sistemático en un proceso de estimación, y frecuentemente es resultado de un muestreo inadecuado.
Estos errores pueden dar lugar a interpretaciones que no se corresponden en modo alguno con la realidad estudiada.

Idealmente se buscan estimaciones sin sesgo, y cuando esto no es posible se prefieren aquellos procedimientos que dan lugar a estimaciones **consistentes**, es decir, que cuyo sesgo tiende a cero conforme crece el tamaño muestral.

Muestreo aleatorio

Una muestra es aleatoria cuando todos los individuos de la muestra se eligen al azar, de modo que todos los individuos de la población tienen la misma probabilidad de ser elegidos.
Este método controla la representatividad de la muestra de una manera muy precisa pero tiene el inconveniente de ser muy costosa o impracticable, ya que a veces es imposible acceder a un individuo concreto que fue seleccionado para la muestra.

El **muestreo aleatorio simple (M.A.S.)** es el tipo de muestreo aleatorio más sencillo y en él se basan todos los demás.
Para obtener una muestra se enumeran los elementos de una población y se seleccionan al azar los n elementos que debe contener la muestra.

Pueden utilizarse tablas aleatorias que indican que individuos tomar (de una lista previamente enumeradas) o bien pueden sortearse mediante un bombo.

Para un **muestreo aleatorio sistemático** se enumeran los individuos y a partir de uno de ellos al azar se toman los siguientes mediante saltos numéricos iguales.

Procedimiento

1. Elaboramos una lista ordenada de los N individuos de la población.
2. Para una muestra de tamaño n, dividimos esa lista ordenada en n fragmentos iguales.
   Para ello cada uno de esos fragmentos medirá d:
$$d = \frac{N}{n}$$
3. La posición del primer elemento (a_1) se elige al azar en el primer fragmento.
4. La posición de los elementos restantes se obtiene mediante una progresión aritmética cuya diferencia es d.
$$\begin{cases} a_2 = a_1 + d \\ a_3 = a_2 + d \\ \quad\quad \dots \\ a_n = a_{n+1} + d \end{cases}$$

Ejemplo de m.a. sistemático

Deseamos obtener una muestra de 75 individuos de una población de 1500.
$$d = \frac{N}{n} = \frac{1500}{75} = 20$$
El primer individuo se seleccionará aleatoriamente entre el 1 y el 20.
Para este ejemplo, consideremos el 8.
$$a_1 = 8 \quad \Rightarrow \quad a_2 = 28 \quad \Rightarrow \quad a_3 = 48 \quad \Rightarrow \quad \dots \quad \Rightarrow \quad a_{75} = 1488$$

4. En un centro escolar hay 1300 alumnos. Explicar como se eligen una muestra de tamaño 100:
   a) Mediante muestreo aleatorio simple.
   b) Mediante muestreo aleatorio sistemático.
5. Una ganadería tiene 3000 vacas. Explica como se obtiene una muestra de tamaño 120:
   a) Mediante muestreo aleatorio simple.
   b) Mediante muestreo aleatorio sistemático.

Para un **muestreo aleatorio estratificado con reparto proporcional** la población es dividida en distintos subgrupos (también llamados **estratos**) que han de ser incompatibles entre sí, es decir, no pueden tener individuos en común.
Estratos comúnmente utilizados son: la edad, el género, el nivel socio-económico, la nacionalidad y el nivel de estudios.

Se establece la proporción de cada estrato dentro de la población total, y se busca una muestra que mantenga las mismas proporciones que el conjunto de la población. En cada estrato, los individuos de la muestra se escogen aleatoriamente.

Ejemplo de m.a. estratificado

En un pueblo habitan 700 hombres adultos, 800 mujeres adultas y 500 menores. De él se quiere seleccionar una muestra de 80 personas, utilizando, para ello, muestreo estratificado con reparto proporcional.

$$700 + 800 + 500 = 2000 \text{ individuos (población)}$$
$$\frac{80}{2000} = 0{,}04 = 4\,\% \text{ (proporción de la muestra)}$$
$$4\,\% \text{ de } 700 = 28 \quad \text{hombres}$$
$$4\,\% \text{ de } 800 = 32 \quad \text{mujeres}$$
$$4\,\% \text{ de } 500 = 20 \quad \text{menores}$$

6. ¿Cómo se elegiría una muestra de 100 alumnos mediante muestreo aleatorio estratificado en un centro con el siguiente censo?

   426 de 1º   359 de 2º   267 de 3º   133 de 4º   115 de Bach.

   Solución: 33, 28, 21, 10, 9.

7. Una fábrica realiza 2000 tornillos diarios de distintos tipos y queremos extraer una muestra de 120.

   850 de tipo A   512 de tipo B   324 de tipo C   204 de tipo D   110 de tipo E

   a) ¿Cuántos tornillos hay que elegir de cada tipo para que el muestreo sea estratificado con reparto proporcional?

   b) ¿Cómo tiene que ser la elección dentro de cada estrato?

   Solución: 51, 31, 19, 12, 7.

En un muestreo **por conglomerados** se forman lotes o conglomerados colectivos equivalentes dentro de la población total. Dentro de estos colectivos se elige uno al azar, que será el objeto de investigación.

Ejemplo de m. por conglomerados

Se desea hacer un estudio estatal académico de los estudiantes de bachillerato. Entonces se divide el territorio nacional en regiones. De entre esas regiones, se seleccionan unas cuantas al azar y luego dentro de ellas se selecciona también al azar un par de institutos que son los que se estudian con detalle.

En un **muestreo discrecional** se seleccionan aquellos individuos que por sus características son más adecuados. Esto sucede con frecuencia en estudios de medicina y ciencias sociales.

Ejemplo de m. discrecional

Para estudiar el efecto de un tratamiento para curar una cierta enfermedad, se elige un conjunto de pacientes que solo tengan esa enfermedad, sin otras enfermedades que puedan enmascarar los resultados.

Tablas de frecuencias y gráficos estadísticos

Intervalos de clase

Cuando una variable estadística puede tomar un número elevado de valores (por ejemplo, porque es continua) es útil agrupar los datos en intervalos. A cada uno de esos intervalos le llamamos **clase**.

Por convenio, en los intervalos de clase se incluyen los extremos inferiores pero no los superiores.

Por ejemplo, al estudiar las alturas de una serie de personas, es conveniente dividir en intervalos las posibles estaturas de los individuos investigados.

$$[150,160) \quad [160,170) \quad [170,180) \quad [180,190) \quad [190,200)$$

El número de clases es variable, pero debemos tener en cuenta que si el número de intervalos es muy reducido puede conducir a una pérdida de información.
Los conceptos fundamentales son:

- El **rango** o **recorrido** de una variable es la diferencia entre el valor más grande y el valor más pequeño que toma la variable.
- La **amplitud del intervalo** es la diferencia entre el extremo superior y el extremo inferior de la clase. La amplitud de los distintos intervalos puede ser variable, pero en la práctica es frecuente elegir clases de amplitud constante.
- Se llama **marca de clase** al punto medio del intervalo.

En trabajos formales de estadística se utilizan algunas reglas para calcular el número o la amplitud de los intervalos.
Antiguamente era frecuente el uso de la regla de Sturges (1926, ampliamente superada pero aún utilizada en computación); hoy se prefieren las reglas de Scott (1979) y de Diaconis (1981).

Frecuencia absoluta y frecuencia relativa

Consideremos una población de tamaño N sobre la que estudiamos una variable aleatoria X que puede tomar varios valores $\{X_1, X_2, X_3, \dots\}$.

- Llamamos **frecuencia absoluta** de X_i al número de veces que aparece dicho valor en el total de casos posibles que se presentan en la muestra. La denotamos por n_i.

- Llamamos **frecuencia relativa** de X_i al cociente entre la frecuencia absoluta de X_i y el número de individuos de la población. La denotamos por f_i.

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

Es muy habitual expresar las frecuencias relativas como porcentajes y, en ese caso, podemos referirnos a ellas como **frecuencias porcentuales**.

A partir de esas frecuencias pueden calcularse las respectivas **frecuencias acumuladas**, que son la suma de las frecuencias de todos los valores inferiores o iguales al valor considerado. Se denotan con las letras mayúsculas correspondientes, es decir, N_i y F_i.

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

Llamamos **distribución de frecuencias** a la aplicación que asocia a cada valor de la variable estadística sus frecuencias.
La forma más habitual de disponer los datos de una distribución de frecuencias es empleando una tabla.

Intervalo	X_i	n_i	N_i	f_i	F_i
$[0, 5)$	2'5	1	1	0'025	0'025
$[5, 10)$	7'5	1	2	0'025	0'050
$[10, 15)$	12'5	3	5	0'075	0'125
$[15, 20)$	17'5	3	8	0'075	0'200
$[20, 25)$	22'5	3	11	0'075	0'275
$[25, 30)$	27'5	6	17	0'150	0'425
$[30, 35)$	32'5	7	24	0'175	0'600
$[35, 40)$	37'5	10	34	0'250	0'850
$[40, 45)$	42'5	4	38	0'100	0'950
$[45, 50)$	47'5	2	40	0'050	1
Total:		40		1	

Puedes ver este mismo ejemplo hecho en una hoja de cálculo, en Google Drive.

8. Hemos recopilado las calificaciones finales en Matemáticas de 40 estudiantes. ¿De qué tipo de variable estadística se trata?

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

Haz la tabla de frecuencias y halla:

a) La calificación más alta.

b) La calificación más baja.

c) El rango.

d) El número de estudiantes con calificación 5.

e) El número de estudiantes con calificaciones menores que 7.

f) El número de estudiantes con calificaciones mayores que 6, pero no superiores a 8.

g) El porcentaje de alumnos con calificación 3.

9. Muchas personas experimentan reacciones alérgicas a las picaduras de insectos. Estas reacciones varían paciente a paciente, no sólo en la gravedad sino también en el tiempo de aparición de la reacción, del que han tomado nota en un hospital cercano durante el verano.
¿De qué tipo de variable estadística se trata?

10'5	11'2	9'9	15	11'4	12'7	16'5	10'1	12'7	11'4
11'6	6'2	7'9	8'3	10'9	8'1	3'8	10'5	11'7	8'4
12'5	11'2	9'1	10'4	9'1	13'4	12'3	5'9	11'4	8'8
7'4	8'6	13'6	14'7	11'5	11'5	10'9	9'8	12'9	9'9

Construye una tabla de frecuencias con los datos obtenidos.

10. Los pesos de 30 alumnos de una clase, en kilogramos, son los siguientes:

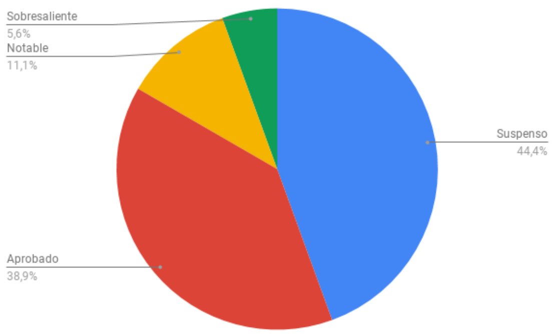
57	62	70	59	57	52	48	48	50	55
59	54	60	60	64	65	69	60	53	57
53	45	54	47	59	60	57	66	62	55

Clasifica la variable estadística y crea una tabla de frecuencias.

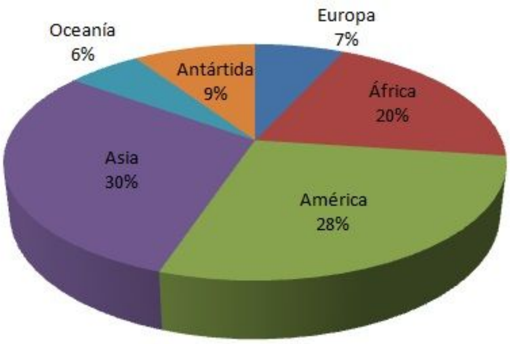
Gráficos estadísticos

Para presentar de modo visual los datos estadístico podemos apoyarnos en distintos tipos de gráficos:

- Para representar variables cualitativas frecuentemente se utiliza un **gráficos de sectores circulares**, en el que a cada valor se le asigna un sector de tamaño proporcional a su frecuencia.



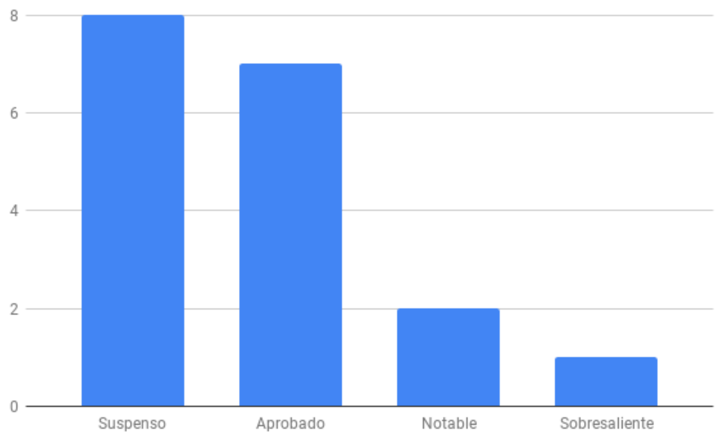
Es fundamental que los sectores circulares se realicen con una vista cenital, ya que las representaciones 3D manipulan la sensación de tamaño.



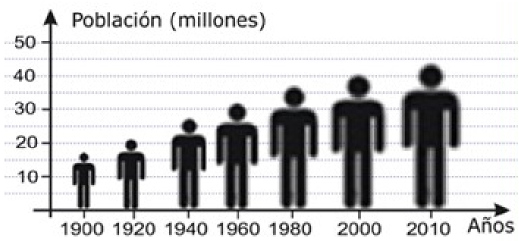
En una hoja de cálculo de Google Drive debemos:

1. Seleccionar el valor X_i y la frecuencia absoluta (n_i).
2. `Insertar` $\longrightarrow$ `Gráfico`
3. En las opciones, seleccionar gráfico circular y usar la columna X_i como etiquetas.

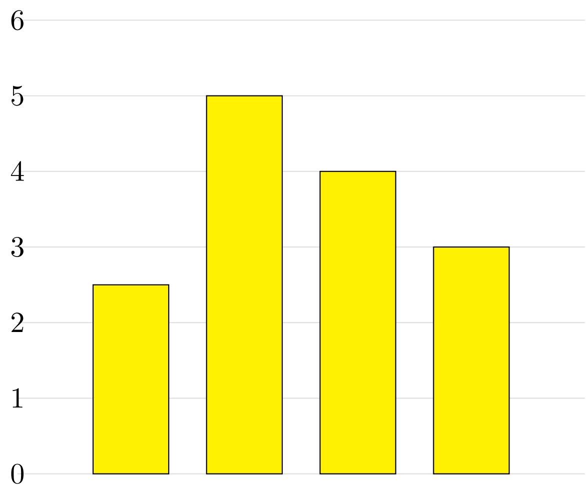
- Para representar variables cualitativas y cuantitativas discretas se puede utilizar un **diagrama de barras** en el que a cada valor se le asigna una barra de tamaño proporcional a su frecuencia.

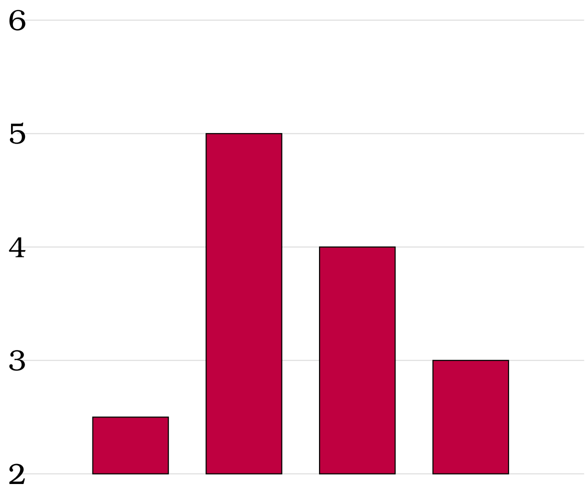


Las barras clásicas pueden ser sustituidas por algún gráfico concreto, en cuyo caso podríamos referirnos a ellas como **pictograma**.



La manipulación más frecuente de diagramas de barras consiste en no utilizar como referencia el valor 0, lo que altera la sensación de diferencia entre frecuencias.

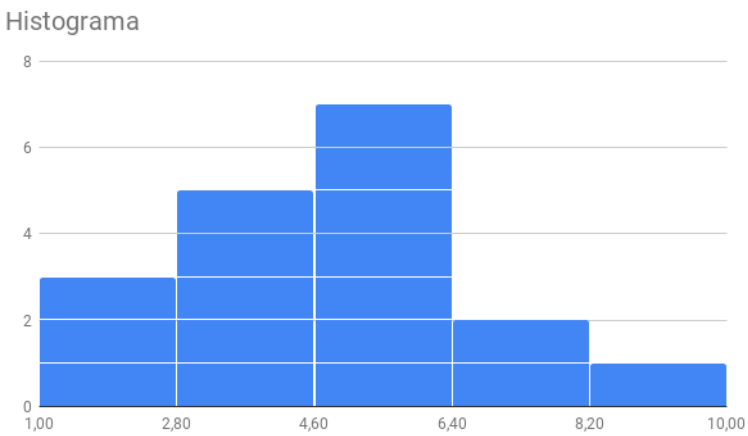




En una hoja de cálculo de Google Drive debemos:

1. Seleccionar la frecuencia absoluta (n_i).
2. `Insertar` $\longrightarrow$ `Gráfico`
3. En las opciones, seleccionar gráfico de columnas y añadir como eje X la columna de valores X_i.

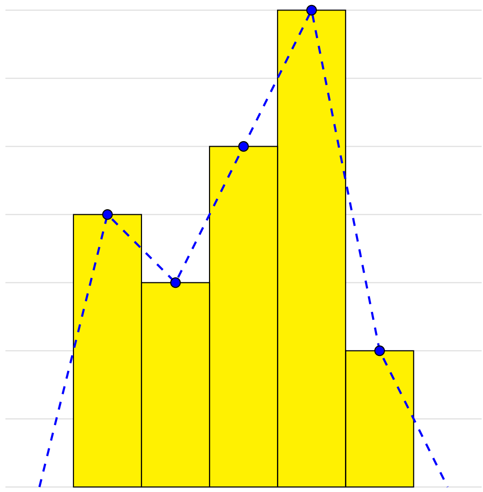
- Para representar variables cuantitativas con datos agrupados se puede utilizar un **histograma**, similar a un diagrama de barras en el que la base de cada barra corresponde a la amplitud de la clase.



Si los intervalos de clase tienen diferentes amplitudes, ajustamos la altura de cada barra del histograma de modo inversamente proporcional a la amplitud:

$$\text{altura} = \frac{n_i}{\text{amplitud}}$$

- Un **polígono de frecuencias** se forma uniendo los extremos medios de las barras de un diagrama de barras o de un histograma mediante segmentos, y puede representarse en combinación con ellos o bien solo.



En una hoja de cálculo de Google Drive debemos:

1. Seleccionar la marca de clase (X_i) y la frecuencia absoluta (n_i).
2. `Insertar` $\longrightarrow$ `Gráfico`
3. En las opciones, seleccionar gráfico de áreas escalonadas y usar la columna de marcas de clase como Eje X.

Existe también un gráfico llamado histograma, que se aplica a listas de datos sin tabular.

11. Las calificaciones obtenidas por los 32 alumnos de una materia fueron las siguientes. Construye la tabla de frecuencias y representa de dos formas distintas la distribución.

	Nº alumnos
Muy deficiente	0
Insuficiente	5
Suficiente	6
Bien	4
Notable	12
Sobresaliente	5

12. Las edades de las personas que acuden a un médico a lo largo de un mes son:

3	2	11	13	4	3	2	4	5	6	7	3
4	5	3	2	5	6	27	15	4	21	14	4
3	6	29	13	6	17	6	13	6	5	12	26

a) Construye la tabla de frecuencias, agrupando los datos en clases de amplitud 5.

b) Representa esta distribución mediante un histograma y un polígono de frecuencias.

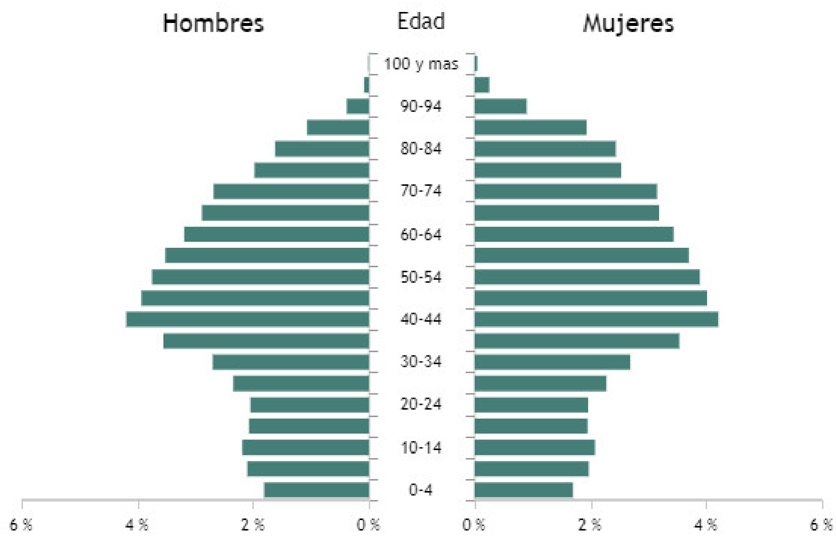
13. Considera la siguiente distribución de frecuencias, en la que no todos los intervalos tienen la misma amplitud.
Representa esta distribución mediante un histograma y un polígono de frecuencias.

Miles de km por año	Nº de coches
$[0, 15)$	120
$[15, 30)$	159
$[30, 35)$	89
$[35, 40)$	78
$[40, 50)$	66
$[50, 60)$	52

Otros gráficos estadísticos

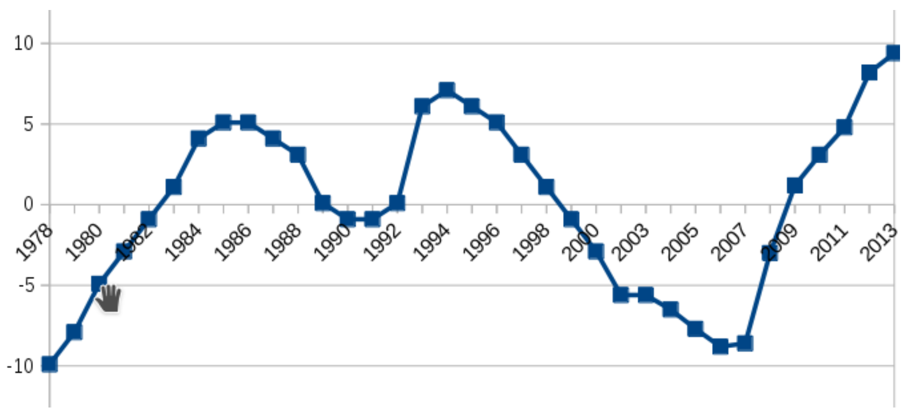
Existen otros gráficos cuyo uso está más restringido a situaciones concretas.

Las **pirámides de población** son histogramas dobles dispuestos de forma horizontal, en los que a cada lado se muestran los datos para un caso de una variable dicotómica (frecuentemente el sexo) y en el otro lado la otra.



Fuente: Estadística del Padrón Continuo

Una **serie temporal o cronológica** es una sucesión de observaciones de una variable tomadas en el transcurso del tiempo. Habitualmente se representan a través de una gráfica de función temporal.



Parámetros estadísticos

Las tablas de frecuencias y los gráficos estadísticos recogen la información relativa a una distribución estadística, pero parece razonable sintetizar toda esa información en unos pocos valores que nos indiquen el comportamiento de todos los individuos de una población respecto a una estadística variable estudiada.

Por ello en los estudios estadísticos se emplean algunas medidas que describen de modo conciso el comportamiento y las características generales de los datos estudiados.
Cada uno de esos parámetros ofrece una información diferente, y dependiendo del caso puede resultar de interés uno u otro.

Parámetros de centralización

Los parámetros de centralización aportan datos acerca de la cohesión de la muestra y su comportamiento más frecuente.

Promedio

El **promedio** o **media aritmética simple** ($\overline{x}$) es el resultado de sumar todos los valores posibles y dividirlo entre el número de datos.

4 4 3'5 5 4'5 3 5 6 3'5 5

En total tenemos 10 datos, así que los sumamos todos y dividimos entre 10.

$$\overline{x} = \frac{4+4+3'5+5+4'5+3+5+6+3'5+5}{10} = 4'35$$

En una tabla de frecuencias, se obtiene multiplicando cada valor posible (o marca de clase) por su frecuencia absoluta, y dividiendo entre el número total de valores.
O bien, multiplicando cada valor (o marca de clase) por la frecuencia relativa.

$$\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i \cdot n_i) = \sum_{i=1}^{N} (x_i \cdot f_i)$$

Otras medias

La **media geométrica** es recomendada para datos de progresión geométrica, para promediar razones, interés compuesto y números índice.
Es menos sensible que la media aritmética a los valores extremos, pero además de la complejidad de cálculo tiene la desventaja de anularse cuando hay un valor nulo y no ser siempre factible si hay valores negativos.

$$g = \sqrt[N]{\prod_{i=1}^{N} x_i} = \sqrt[N]{x_1 \cdot \cdots \cdot x_N}$$

La **media armónica** es la inversa del promedio de los inversos.
Es más representativa que el promedio, pero demasiado sensible si hay valores muy pequeños y no factible si hay algún valor nulo.

$$H = \frac{1}{\frac{1}{N} \cdot \sum_{i=1}^{N} \frac{1}{x_i}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

Además las medias pueden ser **ponderadas**.

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i \cdot p_i}{\sum_{i=1}^{N} p_i} \qquad \text{y} \qquad g = \left(\prod_{i=1}^{N} (x_i)^{p_i} \right)^{\frac{1}{\sum p_i}}$$

Moda

La **moda** (M_o) es el valor más frecuente de la variable. Una variable con una única moda se llama unimodal, con dos bimodal; a partir de ahí se le suele llamar con el genérico multimodal.

La moda no se ve afectada por valores extremos, pero es muy poco susceptible de realizar operaciones algebraicas con ella.

4 4 3'5 5 4'5 3 5 6 3'5 5

La moda de estos datos es $M = 5$, porque es el valor que más veces se repite.

En una tabla de frecuencias se identifica la moda por ser el valor que tiene una mayor frecuencia absoluta (n_i).

En una tabla de frecuencias con intervalos agrupados, el intervalo de clase que tiene mayor frecuencia absoluta se llama **intervalo modal**.
La moda propiamente dicha se calcula considerando que se encontrará en una posición dentro del intervalo modal que dependerá de la frecuencia de las clases más próximas.

extremo inferior intervalo

amplitud intervalo

$$Mo = L_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \cdot a_i$$

Mediana

La **mediana** es el valor central de la variable, obtenida tras ordenar todos los valores de menor a mayor.

- Si los datos son impares es sencillo, pues al ordenarlos de menor a mayor es evidente cuál es el valor central.

- Si los datos son pares habrá dos valores centrales, así que hallamos la media aritmética simple de ambos.

Cuando hay grandes variaciones la mediana puede ser mucho más significativa que la media, pero tiene el inconveniente de no prestarse a planteamientos algebraicos.

4 4 3'5 5 4'5 3 5 6 3'5 5

Antes de nada, tenemos que ordenar los datos de menor a mayor.

3 3'5 3'5 4 4 4'5 5 5 5 6

Como en el centro hay dos números, les hacemos la media aritmética:

$$Me = \frac{4 + 4'5}{2} = 4'25$$

Para datos presentados en tablas de frecuencias, localizaremos la mitad del número de valores de la muestra $\left(\frac{N}{2}\right)$ y lo compararemos con las frecuencias acumuladas de nuestra tabla para hallar el primer valor $N_i \geq \frac{N}{2}$.

Para datos agrupados en intervalos, el **intervalo mediano** será aquel al que pertenezca la mediana. La mediana propiamente dicha se calcula considerando que se encontrará en una posición dentro del intervalo mediano directamente proporcional a la cantidad de elementos que se han superado.

f. acumulada anterior

extremo inferior intervalo

amplitud intervalo

$$Me = L_i + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i$$

1. Las dianas logradas en un campeonato por 25 tiradores fueron:

8	10	12	12	10	10	11	11	10	13	10	8	9
11	10	9	9	11	12	9	10	9	10	9	10	

Calcula la media, la mediana y la moda.

2. Los resultados obtenidos al lanzar un dado 200 veces vienen reflejados en la siguiente tabla:

Nº puntos	1	2	3	4	5	6
Repeticiones	x	32	35	33	y	35

   a) Determina las frecuencias que faltan sabiendo que el puntuación media es 3'6.

   b) Calcula la mediana y la moda.

3. Un médico atendió en 20 días las siguientes urgencias:

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

   a) Resume los datos en una tabla que muestre frecuencias absolutas y porcentajes.

   b) Dibuja el correspondiente diagrama de barras.

   c) Calcula la media, moda y la mediana.

Parámetros de posición

Si un conjunto de datos está ordenado, el valor central que divide al conjunto en dos mitades iguales es la mediana. Extendiendo esa idea, podemos pensar en aquellos valores que dividen el conjunto otro número de subconjuntos iguales.
En general se llaman **cuantiles**, y veremos tres tipos en particular.

Cuartiles

Si la mediana se obtiene dividiendo la lista de datos ordenados a la mitad, los **cuartiles** se obtienen dividiéndola en cuartos de forma análoga.

- Q_1 es el primer cuartil, es decir, hay un cuarto de valores menores que él. Corresponde al primer valor cuya frecuencia acumulada cumple $N_i \geq \frac{N}{4}$.
- Q_2 es el segundo cuartil. Coincide con la mediana.
- Q_3 es el tercer cuartil, es decir, hay un cuarto de valores mayores que él. Corresponde al primer valor cuya frecuencia acumulada cumple $N_i \geq \frac{3N}{4}$.

3 3'5 3'5 4 4 4'5 5 5 5 6

Como en total hay 10 datos al partir a la mitad nos quedan 5 datos a la izquierda y 5 datos a la derecha, por eso $Q_2 = Me = 4'25$.
Al repartir esos 5 nuevamente en dos mitades, el valor central será Q_1 y Q_3 respectivamente.

3 3'5 3'5 4 4 | 4'5 5 5 5 6

$Q_1 = 3'5 \qquad Q_3 = 5$

Para datos agrupados en intervalos, utilizaremos el procedimiento análogo al de la mediana.

$$Q_1 = L_i + \frac{\frac{N}{4} - N_{i-1}}{n_i} \cdot a_i \qquad Q_3 = L_i + \frac{\frac{3N}{4} - N_{i-1}}{n_i} \cdot a_i$$

Deciles y percentiles

Los **deciles** son aquellos valores que dividen la serie de datos en diez partes iguales. El decil D_k es aquel que deja por debajo $\frac{k}{10}$ de los datos.
Se pueden calcular los deciles desde D_1 hasta D_9, siendo justamente estos dos los más utilizados. Además, es trivial observar que $D_5 = Q_2 = Me$.

Para datos agrupados en intervalos, también utilizaremos el procedimiento análogo al de la mediana y los cuartiles.

$$D_k = L_i + \frac{\frac{kN}{10} - N_{i-1}}{n_i} \cdot a_i$$

Por su parte, los **percentiles** son aquellos valores que dividen la serie de datos en cien partes iguales. El percentil P_k es aquel que deja por debajo $\frac{k}{100} = k\,\%$ de los datos.
Se pueden calcular los percentiles desde P_1 hasta P_{99}, habiendo múltiples coincidencias con los deciles así como las hay con los cuartiles ($P_{25} = Q_1$, $P_{50} = Q_2 = Me$, $P_{75} = Q_3$).

Para datos agrupados en intervalos, también utilizaremos el procedimiento análogo al de la mediana y los cuartiles.

$$P_k = L_i + \frac{\frac{kN}{100} - N_{i-1}}{n_i} \cdot a_i$$

Parámetros de dispersión

Las medidas de centralización aportan individualmente una gran cantidad de información. Sin embargo, en muchas ocasiones esa información no solo es incompleta sino que puede inducir errores en su interpretación.
Los parámetros de dispersión informan acerca de la diversidad de valores de la variable estudiada y los extremos de la muestra.

Rango

El **rango o recorrido** es la diferencia entre el mayor valor y el menor valor.
Es la medida de dispersión más sencilla y la que proporciona menos información.
A menudo se prefiere el **rango intercuartículo**, que es la diferencia entre el tercer cuartil y el primero.

3	3'5	3'5	4	4	4'5	5	5	5	6

$R = 6 - 3 = 3 \qquad R_I = 5 - 3'5 = 1'5$

Desviación absoluta promedio

Para distintos parámetros de dispersión tomamos como referencia la **desviación** que cada valor (o marca de clase) presenta con respecto a la media, es decir, a la diferencia $x_i - \overline{x}$.

El problema que plantea considerar las desviaciones de todos los valores e intentar agruparlas (mediante una suma) es la posibilidad de que desviaciones con signo positivo y otras con signo negativo se anulen mutuamente.
Una de las opciones para evitar este problema es considerar dichas desviaciones en valor absoluto, lo que da lugar a la **desviación absoluta promedio** o **desviación media**.

$$D_m = \frac{\displaystyle\sum_{i=1}^{N} |x_i - \overline{x}| \cdot n_i}{N}$$

Si la desviación absoluta promedio es muy alta indica gran dispersión, y si es muy baja refleja mucho agrupamiento (es decir, valores son parecidos entre sí).

Varianza y desviación típica

Otro modo de evitar el problema provocado por signos negativos es elevar cada una de las desviaciones al cuadrado, dando lugar a la **varianza**.

$$s^2 = \frac{\displaystyle\sum_{i=1}^{N} (x_i - \overline{x})^2 \cdot n_i}{N}$$

Aplicando identidades notables y reordenando adecuadamente los términos obtenemos una expresión más cómoda para el cálculo de la varianza:

$$s^2 = \frac{\displaystyle\sum_{i=1}^{N} x_i^2 \cdot n_i}{N} - \overline{x}^2$$

El valor numérico de la varianza no es un indicativo claro de la dispersión de los datos, debido a que al elevar todas las desviaciones al cuadrado los cálculos arrojan un resultado muy grande.
Por ello empleamos la **desviación típica** (s), que es la raíz cuadrada positiva de la varianza.

$$s = +\sqrt{s^2}$$

Coeficiente de variación

La desviación promedio y la desviación típica son términos absolutos, y por ello es difícil establecer comparaciones respecto a las dispersiones de distintas muestras.

En esos casos se suele utilizar una razón que permite expresar la dispersión en términos relativos (y por lo tanto carente de unidades) llamada **coeficiente de variación**.
Suele expresarse como un porcentaje.

$$C_V = \frac{s}{\bar{x}}$$

4. Dados los datos de la tabla adjunta que resume la observación hecha a 30 niños de la edad en meses en que empiezan a andar:

Meses	9	10	11	12	13	14	15
Frecuencia	1	2	4	13	6	3	1

   a) Calcula la promedio, la moda y la mediana.

   b) Calcula el tercer cuartil y el octavo decil.

   c) Calcula el coeficiente de variación.

5. Dada la distribución de frecuencias:

Intervalos	(0,3)	(3,6)	(6,9)	(9,12)	(12,15)	(15,18)
Frecuencias	2	7	12	13	4	3

   a) Dibuja el histograma y el polígono de frecuencias.

   b) Calcula promedio, moda y mediana.

   c) Calcula primer y tercer cuartil.

   d) Calcula el 7º decil y el percentil 15.

   e) Calcula el coeficiente de variación.

6. Los salarios anuales, en miles de euros, de 100 empleados de una empresa vienen dados por la tabla adjunta.

Miles €	Nº empleados
40 – 70	13
70 – 100	30
100 – 130	32
130 – 160	15
160 – 190	10

   a) Construye el histograma.

   b) Calcula el promedio, la moda y la mediana.

7. Juan tiene 19 años de edad y mide 1’90 m de estatura.
   Se dice que su talla está en el percentil 92 para jóvenes de 19 años. ¿Qué quiere decir esto?

8. En un test psicotécnico de 60 preguntas realizado por 40 trabajadores de una empresa se presentaron estos resultados

Puntuación	Nº empleados
$[0, 10)$	1
$[10, 20)$	6
$[20, 30)$	13
$[30, 40)$	12
$[40, 50)$	7
$[50, 60)$	1

   a) Calcula las medidas de centralización, cuartiles, decil 2 y percentil 87.

   b) ¿A partir de qué puntuación se encuentran el 30 % de los trabajadores que realizaron mejor el test?

9. El peso medio de los alumnos de una clase es de 58'2 kg y su desviación típica 3'1 kg. El de las alumnas de esa misma clase es de 52'4 kg y su desviación típica es 5'2 kg. Compara la dispersión de ambos grupos.

10. Completa los datos que faltan en la siguiente tabla estadística.
    Calcula la moda, promedio, mediana, cuartiles, varianza y desviación típica.

x_i	n_i	N_i	f_i	F_i
1	4		0'09	
2	4			
3		16	0'18	
4	7		0'16	
5	5	28		
6		38		
7	7	45	0'15	
TOTAL				

Distribuciones de probabilidad

Una **variable aleatoria** sobre un espacio muestral E es una función que asigna un valor numérico a cada uno de los sucesos que componen el espacio muestral.

Denotaremos las variables aleatorias con letras mayúsculas y a sus posibles valores letras minúsculas.

$$\begin{array}{rccc} X: & E & \longrightarrow & \mathbb{R} \\ & e_i & \hookrightarrow & X(e_i) = x_i \end{array}$$

Sobre un mismo conjunto se pueden definir distintas funciones, y del mismo modo sobre un mismo espacio muestral se pueden definir distintas variables aleatorias.

En el experimento «lanzar dos monedas» el espacio muestral es $E = \{CC, CX, XC, XX\}$.
Si definimos la variable $A =$«número de caras»:

$$A(CC) = 2 \quad A(CX) = 1 \quad A(XC) = 1 \quad A(XX) = 0$$

Si definimos la variable $B =$«número de cruces»:

$$B(CC) = 0 \quad B(CX) = 1 \quad B(XC) = 1 \quad B(XX) = 2$$

Podemos clasificar una variable aleatoria en:

- **discreta** si toma solo valores aislados
- **continua** si toma todos los valores de un intervalo

Consideramos como espacio muestral los ordenadores del aula de informática y sobre él la variable $X =$«capacidad nominal de la memoria RAM» y la variable $Y =$«tiempo de arranque».
Aún estando definidas sobre el mismo espacio, X es discreta e Y es continua.

Funciones de variables discretas

La **función de masa de probabilidad** de una variable discreta es una función que asocia a cada elemento del espacio muestral la probabilidad de que este ocurra.

$$f(x_i) = P(X = x_i)$$

Por definición de probabilidad, se ha de cumplir que:

- $0 \leq f(x_i) \leq 1$
- $\displaystyle\sum_{i=1}^{k} f(x_i) = 1$

1. Determina el valor del parámetro k en la siguiente función de masa de probabilidad:

x	1	2	3	4
$f(x)$	$0'3$	$3k$	k	$0'3$

2. La función de masa de probabilidad de una variable aleatoria discreta es:

x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{1}{4}$	k	$\frac{1}{4}$	$\frac{1}{16}$

   a) Halla el valor del parámetro k.

   b) Calcula $P(X \leq 2'7)$.

   c) Calcula $P(X \geq 2)$.

   d) Calcula $P(1 \leq X < 4)$.

La **función de distribución acumulada** es la función que asocia a cualquier valor real x la probabilidad de que la variable aleatoria asuma un valor inferior o igual a x.

$$F(x) = P(X \leq x)$$

Por lo tanto ha de cumplir que:

- $0 \leq F(x) \leq 1$
- F es creciente

Además podemos observar que la probabilidad en un intervalo podría escribirse como:

$$P(a < x \leq b) = F(b) - F(a)$$

Medidas características

Esperanza (promedio)

La **esperanza matemática** de una variable aleatoria discreta es el valor promedio de sus probabilidades.

$$\mu = \mathbb{E}[X] = \sum_{i=1}^{n} x_i \cdot f(x_i)$$

En el experimento «lanzar dos monedas» el espacio muestral es $E = \{CC, CX, XC, XX\}$.
Consideramos la variable «número de cruces» (que puede tomar valores 0, 1 o 2) y hallamos su esperanza:

$$\mathbb{E}[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = \boxed{1}$$

Varianza

La **varianza** de una variable aleatoria discreta es una medida del grado de dispersión de la distribución en torno a la media.

$$\sigma^2 = \mathbb{E}[(X-\mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i) = \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - \mu^2$$

Desviación típica

La **desviación típica** es la raíz cuadrada de la varianza.

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

En el experimento «lanzar dos monedas», la esperanza de la variable «número de cruces» es 1.

$$\sigma^2 = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{2}{4} + 2^2 \cdot \frac{1}{4} - 1^2 = \boxed{\frac{1}{2}}$$

$$\sigma = \sqrt{0'5} \approx \boxed{0'707}$$

3. La función de masa de probabilidad de una variable aleatoria discreta es:

x	0	1	2	3	4
$f(x)$	$0'1$	a	b	c	$0'2$

Sabiendo que $P(X \leq 2) = 0'7$ y $P(X \geq 2) = 0'75$,

   a) Halla el valor de los parámetros a, b y c.
   b) Calcula la esperanza y la desviación típica.

4. En una urna hay 3 bolas blancas y 7 negras. Se extraen con devolución 3 bolas y se observan cuántas son de color blanco.

   a) Halla la función de masa de probabilidad.
   b) Calcula $P(X \leq 2)$ y $P(X \geq 1)$.
   c) Calcula la esperanza, la varianza y la desviación típica.

5. Sacamos al azar tres bolas de una bolsa que contiene cinco bolas rojas y tres azules. Consideramos la variable aleatoria «número de bolas azules en cada extracción».

   a) Calcula $P(X \leq 1)$ y $P(1 \leq X \leq 2)$.
   b) Escribe la función de masa de probabilidad.
   c) Calcula la esperanza, la varianza y la desviación típica.

6. Tres jugadores de fútbol aciertan en el lanzamiento de penaltis con probabilidades 0'5, 0'8 y 0'3 respectivamente. Si cada uno de ellos hace un lanzamiento y tomamos como variable aleatoria X el número total de goles:

   a) Escribe la función de masa de probabilidad.
   b) Calcula $P(X \leq 2)$ y $P(2 \leq X \leq 3)$
   c) Calcula la media, la varianza y la desviación típica.

Distribución binomial

Ensayos de Bernouilli

La **distribución dicotómica o de Bernouilli** se plantea en un experimento (llamado ensayo) en el que solo hay dos posibilidades: éxito o fracaso.
La probabilidad de éxito se denomina típicamente p y por lo tanto la probabilidad de fracaso es igual a $1-p$.

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1-p & \text{si } x = 0 \\ 0 & \text{si } x \text{ toma otro valor} \end{cases}$$

Por lo tanto sus medidas características son:

- $\mu = \mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = \boxed{p}$
- $\sigma^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p - p^2 = \boxed{p \cdot (1-p)}$

7. Durante un control de calidad, se detecta que el 2 % de las pastillas de freno que produce una máquina tienen algún defecto y deben ser retiradas.
Calcula las probabilidades de sus sucesos, su esperanza y su desviación típica.

Función de masa de la distribución binomial

La **distribución binomial** cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí con una probabilidad de éxito p en cada uno de ellos.

Depende por lo tanto de los parámetros n (número de ensayos) y p (probabilidad de éxito), por lo que la denominaremos $B(n,p)$.

Para construir la función de masa debemos conocer probabilidad de que en n experimentos haya k éxitos (y por lo tanto $n-k$ fracasos), para cualquier valor natural de k menor o igual que n. Debemos tener en cuenta que:

- Hay $\binom{n}{k}$ combinaciones de n elementos de tal forma que k sean éxitos.
- Cada una de ellas tiene probabilidad $p^k \cdot (1-p)^{n-k}$.

Por lo tanto la función de masa de probabilidad será:

$$f(k) = P(X=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad \forall k \in \{0, \dots, n\}$$

Revisamos 500 tornillos fabricados por una máquina que elabora un 10 % de piezas defectuosas y observamos si representan o no alguna anomalía.
Queremos calcular la probabilidad de que menos de tres tornillos sean defectuosos.

$$X \sim B(500, 0'1)$$

$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$

$P(X = 0) = \binom{500}{0} \cdot 0'1^0 \cdot 0'9^{500} = 1'3221 \cdot 10^{-23} \approx 0$

$P(X = 1) = \binom{500}{1} \cdot 0'1^1 \cdot 0'9^{499} = 7'3448 \cdot 10^{-22} \approx 0$

$P(X = 2) = \binom{500}{2} \cdot 0'1^2 \cdot 0'9^{498} = 2'0362 \cdot 10^{-20} \approx 0$

La probabilidad de que sean menos de tres es prácticamente nula.

8. Lanzamos una moneda al aire 5 veces. Calcula la probabilidad de obtener:

   a) Exactamente 2 veces cara.

   b) Como máximo 2 veces cara.

   c) Más de 2 veces cara.

   d) Como mínimo 2 veces cara.

Medidas características de la distribución binomial

Teniendo en cuenta que la distribución binomial se corresponde con n ensayos de Bernouilli y que la esperanza de cumple que $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$, es trivial observar que para la suma los n ensayos:

$$\mu = n \cdot p$$

Además para variables independientes (y los ensayos de Bernouilli considerados lo son) también se cumple $Var[X + Y] = Var[X] + Var[Y]$, así que podemos concluir que:

$$\sigma^2 = n \cdot p \cdot (1 - p)$$

Consideremos 500 ensayos con un 50 % de probabilidad de éxito, es decir, $X \sim B(500, 0'5)$

- $\mu = 500 \cdot 0'5 = 250$
- $\sigma^2 = 500 \cdot 0'5 \cdot (1 - 0'5) = 125 \Longrightarrow \sigma = \sqrt{125} \approx 11'18$

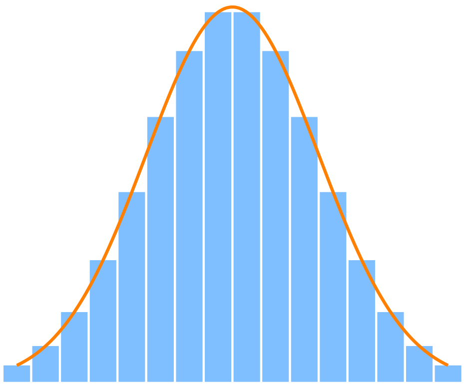
Funciones de variables continuas

Una variable continua puede tomar todos los valores de un intervalo, es decir, puede tomar infinitos valores distintos.
Por ello no podemos definir la probabilidad en cada uno de esos puntos del mismo modo que haríamos en una variable discreta.

$$P(X = x_i) = 0$$

Para definir una función análoga a la función de masa de una variable discreta, consideraremos el polígono de frecuencias de una variable continua.

Si aumentamos indefinidamente el número de observaciones aumenta el número de intervalos y disminuye su amplitud.
La poligonal que une los puntos medios del cada barra del histograma se va suavizando y coincide en el límite con una curva que llamaremos **función de densidad**.

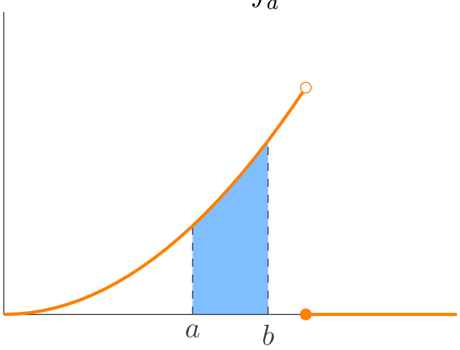


Esta función cumple que:

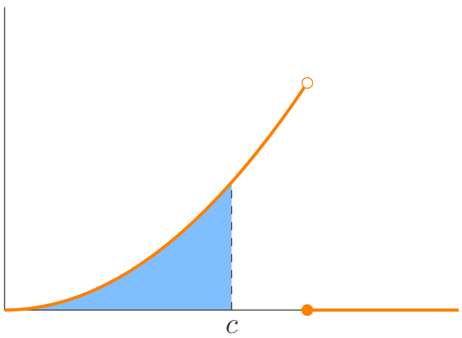
- $f(x) \geq 0 \; \forall x \in \mathbb{R}$
- El área encerrada bajo la curva es 1.

Tiene entonces sentido interpretar el área total como la probabilidad total, y por lo tanto el área de un fragmento como la probabilidad de un intervalo.

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$



$$P(X \leq c) = \int_{-\infty}^{c} f(x)\,dx$$



La **función de distribución acumulada** es la función que asocia a cualquier valor real x la probabilidad de que la variable aleatoria asuma un valor inferior o igual a x, lo que en el caso de una variable continua quiere decir que

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

Por lo tanto las función de distribución de una variable continua cumple las mismas propiedades que la de una variable discreta:

- $0 \leq F(x) \leq 1$
- F es creciente

Y además la probabilidad en un intervalo también podría escribirse de modo análogo, aunque debemos tener en cuenta que la probabilidad en cualquier punto es 0 por lo tanto incluir o no los extremos del intervalo es irrelevante.

$$P(a \leq x \leq b) = F(b) - F(a)$$

Medidas características

Esperanza (promedio)

La **esperanza matemática** de una variable aleatoria continua nos da una idea de donde se encuentra el «centro» de la variable.

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$$

Varianza

La **varianza** de una variable aleatoria continua es una medida del grado de dispersión de la distribución en torno a la media.

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu^2$$

Desviación típica

La **desviación típica** es la raíz cuadrada de la varianza.

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

Sea X una variable con función de densidad $f(x) = \dfrac{x+1}{6}$ en el intervalo $(1,3)$.

- $\mu = \displaystyle\int_1^3 x \cdot \frac{x+1}{6}\,dx = \frac{1}{6}\int_1^3 x^2 + x\,dx = \frac{1}{6}\left[\frac{x^3}{3} + \frac{x^2}{2}\right]_1^3 = \frac{1}{6}\left(\frac{27}{3} + \frac{9}{2} - \frac{1}{3} - \frac{1}{2}\right) = \boxed{\frac{19}{9}}$
- $\sigma^2 = \displaystyle\int_1^3 x^2 \cdot \frac{x+1}{6}\,dx - \left(\frac{19}{9}\right)^2 = \frac{1}{6}\int_1^3 x^3 + x^2\,dx - \left(\frac{19}{9}\right)^2 = \frac{1}{6}\left[\frac{x^4}{4} + \frac{x^3}{3}\right]_1^3 - \left(\frac{19}{9}\right)^2 = \boxed{\frac{26}{81}}$
- $\sigma = \sqrt{\dfrac{26}{81}} \approx \boxed{0'567}$

Distribución normal

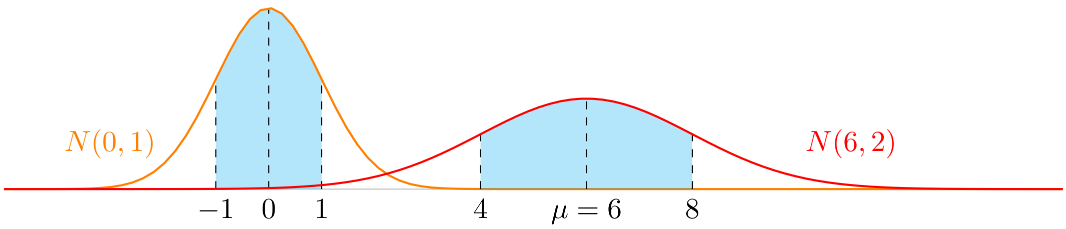
La distribución normal es la más común entre todas las distribuciones de probabilidad utilizadas en Estadística y tiene importantes aplicaciones en la modelización de variables estadísticas asociadas a los elementos de una población.
Por ejemplo siguen distribuciones normales: las medidas físicas del cuerpo humano en una población, las características psíquicas medidas por test de inteligencia o personalidad, las medidas de calidad en muchos procesos industriales, los errores de las observaciones astronómicas, etc.

Función de densidad de la distribución normal

La función de densidad de una distribución normal de media μ y varianza σ^2 es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in \mathbb{R}$$

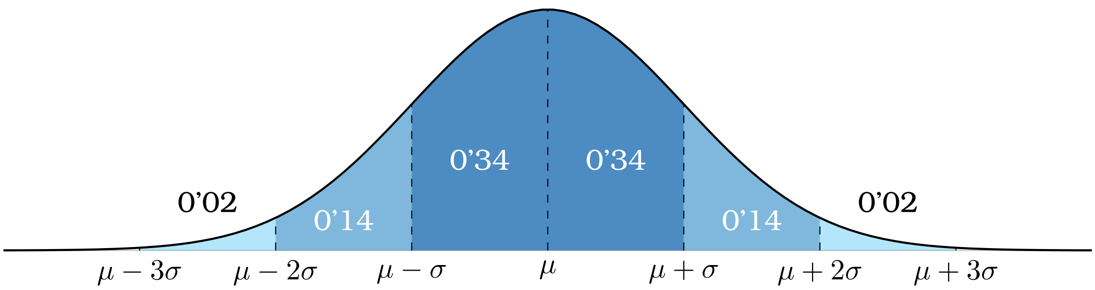
y se representa gráficamente como una **campana de Gauss**, más o menos aplanada dependiendo de la desviación típica.



Puede observarse en la gráfica que en el punto $x = \mu$ se halla su único máximo relativo y absoluto, y que además es simétrica con respecto al eje $x = \mu$.

Tipificación de la variable normal

Independientemente del valor de μ y σ las áreas bajo la curva en los intervalos de amplitud σ siempre miden lo mismo.



Esta propiedad nos permite utilizar la función de densidad **normal tipificada** $N(0,1)$ como referencia para cualquier otra función normal.
Ese proceso recibe el nombre de **tipificación**.

Para evitar confusión, generalmente llamamos X a una variable normal con media μ y desviación típica σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in \mathbb{R}$$

y llamamos Z a una variable normal tipificada $N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \qquad \forall z \in \mathbb{R}$$

así que necesitamos hacer el cambio de variable

$$Z = \frac{X-\mu}{\sigma}$$

para poder utilizar una tabla de la normal tipificada.

En una normal $N(2,4)$ podemos plantear $P(X<3)$ del siguiente modo:

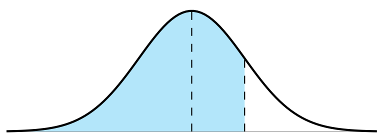
$$P(X<3) = P\left(\frac{X-2}{4} < \frac{3-2}{4}\right) = P\left(Z < \frac{1}{4}\right)$$

Uso de las tablas de la normal tipificada

Existen distintos tipos de tablas para el cálculo de probabilidades de sucesos, así que es importante tener claro cuál de ellas es la que estamos utilizando.

Nosotros emplearemos la misma que se utiliza en la ABAU, que define la probabilidad de que $P(Z \leq z)$ para valores de z positivos.

$P(Z \leq z)$



Para calcular la probabilidad $P(Z \leq z)$ o $P(Z < z)$ para un valor z positivo debemos buscarlo en la tabla.

Para ello localizaremos las unidades y décimas en la columna de la izquierda y las centésimas en la fila superior.

Queremos calcular $P(X < 3)$ en una normal $N(2,4)$.

$$P(X<3) = P\left(\frac{X-2}{4} < \frac{3-2}{4}\right) = P\left(Z < \frac{1}{4}\right) = \boxed{P(Z < 0'25)}$$

Para calcular la probabilidad de los valores de Z menores que $0'25$, en la columna izquierda de mi tabla busco el valor hasta las décimas ($0'2$) y en la fila superior las centésimas ($0'05$).

Trazo las líneas correspondientes a cada uno de ellos en mi tabla, y selecciono el valor que se haya en la intersección de ambas.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879

Por lo tanto $P(X < 3) = P(Z < 0'25) = 0'5987$

9. Calcula las probabilidades en las distribuciones normales indicadas:

a) $P(X \leq 3)$ en $N(2,5)$

b) $P(X \leq 7)$ en $N(4,1)$

c) $P(X < 12)$ en $N(9,9)$

d) $P(X \leq 10)$ en $N(1,3)$

e) $P(X \leq 0)$ en $N(-3,4)$

f) $P(X < 7)$ en $N(4,6)$

g) $P(X < 5)$ en $N(4,6)$

h) $P(X < 97)$ en $N(70,15)$

i) $P(X < 87)$ en $N(80,7)$

j) $P(X \leq -2)$ en $N(-4,7)$

k) $P(X \leq 3)$ en $N(0,4)$

l) $P(X < 31'2)$ en $N(15,5)$

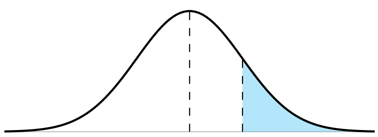
m) $P(X \leq 25)$ en $N(9,9)$

n) $P(X < 12'5)$ en $N(10'5,8)$

ñ) $P(X \leq 0'3)$ en $N(0'1,0'7)$

o) $P(X \leq 18'3)$ en $N(7'4,9'2)$

$P(Z \geq z)$



Para calcular la probabilidad $P(Z \geq z)$ o $P(Z > z)$ para un valor z positivo tendremos en cuenta que es el complementario de $P(Z < z)$ o $P(Z \leq z)$.

$$P(Z \geq z) = 1 - P(Z < z)$$

Queremos calcular $P(X > 80)$ en una normal $N(66, 8)$.

$P(X > 80) = P\left(\frac{X-66}{8} > \frac{80-66}{8}\right) = P\left(Z > \frac{7}{4}\right) = P(Z > 1'75) = \boxed{1 - P(Z < 1'75)} = 1 - 0'9599 = 0'0401$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

10. Calcula las probabilidades en las distribuciones normales indicadas:

a) $P(X > 3)$ en $N(2, 5)$

b) $P(X \geq 7)$ en $N(6, 6)$

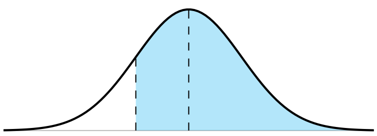
c) $P(X > 12)$ en $N(10, 7)$

d) $P(X \geq 10'3)$ en $N(9'2, 1'3)$

e) $P(X > -8)$ en $N(-10, 6)$

f) $P(X \geq 0)$ en $N(-1'2, 2'5)$

$P(Z \geq -z)$



Para calcular la probabilidad referida a un valor negativo tendremos en cuenta que la distribución normal es simétrica respecto a su media.
Por lo tanto $P(Z \geq -z) = P(Z \leq z)$.

Queremos calcular $P(X \geq 3)$ en una normal $N(5, 2)$.

$P(X \geq 3) = P\left(\frac{X-5}{2} \geq \frac{3-5}{2}\right) = P(Z \geq -1) = \boxed{P(Z \leq 1)} = 0'8413$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

11. Calcula las probabilidades en las distribuciones normales indicadas:

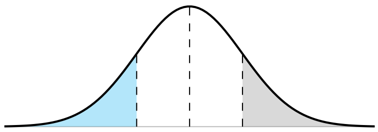
a) $P(X > -1)$ en $N(2, 5)$

b) $P(X \geq 4)$ en $N(6, 6)$

c) $P(X > 5)$ en $N(10, 7)$

d) $P(X \geq 7'3)$ en $N(9'2, 1'3)$

$P(Z \leq -z)$



Razonando también por simetría
$P(Z \leq -z) = P(Z \geq z) = 1 - P(Z < z)$.

Queremos calcular $P(X < 6)$ en una normal $N(9, 2)$.

$P(X < 6) = P\left(\frac{X-9}{2} < \frac{6-9}{2}\right) = P\left(Z < \frac{-3}{2}\right) = P(Z < -1'5) = P(Z > 1'5) = \boxed{1 - P(Z < 1'5)} = 1 - 0'9332 = 0'0668$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

12. Calcula las probabilidades en las distribuciones normales indicadas:

   a) $P(X < -4)$ en $N(2, 5)$
   b) $P(X \leq 2)$ en $N(6, 6)$
   c) $P(X < 2)$ en $N(10, 7)$
   d) $P(X \leq -1'3)$ en $N(2'2, 4'3)$

Teorema central del límite

Una justificación de la frecuente aparición de la distribución normal es el **teorema central del límite**.
Este establece que cuando los resultados de un experimento sean debidos a un conjunto muy grande de causas independientes, que actúan sumando sus efectos, siendo cada efecto individual de poca importancia respecto al conjunto, es esperable que los resultados sigan una distribución normal.

$$\sum_{i=1}^{n} X_i \xrightarrow{n \to \infty} N(\mu, \sigma)$$

La suma de distribuciones independientes tiende a un distribución normal que tiene:

- como media la suma de las medias
- como varianza la suma de las varianzas

En el caso particular en que n variables independientes tengan la misma media μ y la misma varianza σ^2, la suma de todas ellas será una distribución normal:

- media $n \cdot \mu$
- varianza $n \cdot \sigma^2$, y por lo tanto desviación típica $\sqrt{n} \cdot \sigma$

13. Los parámetros de una distribución son $\mu = 20$ y desviación típica $\sigma = 10$, de la cual tomamos una muestra de 100 individuos.
Calcula la probabilidad de que la suma de todos los valores muestrales sea inferior a 1 500.

14. El salario anual de los trabajadores de una gran multinacional se distribuye mediante una distribución de media $\mu = 25\,000$ € y desviación típica $\sigma = 3\,000$ €.
Calcula la probabilidad de que al seleccionar al azar a 100 trabajadores la suma de sus rentas supere los 2 millones de euros.

Distribución de la media muestral

Podemos aplicar el mismo teorema cuando necesitamos trabajar con la **media muestral** ($\overline{X}$) de valores tomados de forma independiente. En este caso:

- La esperanza es $E\left[\frac{X_1+\cdots+X_n}{n}\right] = \dfrac{1}{n} \cdot E[X_1 + \cdots + X_n] = \frac{1}{n} \cdot n \cdot \mu = \boxed{\mu}$
- La varianza es $V\left[\frac{X_1+\cdots+X_n}{n}\right] = \dfrac{1}{n^2} \cdot V[X_1 + \cdots + X_n] = \dfrac{1}{n^2} \cdot n \cdot \sigma^2 = \dfrac{\sigma^2}{n}$
  y por lo tanto la desviación típica es $\sqrt{\dfrac{\sigma^2}{n}} = \boxed{\dfrac{\sigma}{\sqrt{n}}}$

Aplicando el teorema central del límite podemos afirmar que:

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

15. Los parámetros de una distribución normal son $\mu = 20$ y desviación típica $\sigma = 10$, de la cual tomamos una muestra de 100 individuos.
Calcula la probabilidad de que la media muestral se halle entre 15 y 25.

16. El salario anual de los trabajadores de una gran multinacional se distribuye mediante una distribución de media $\mu = 25\,000$ € y desviación típica $\sigma = 3\,000$ €.
¿Cuál es la esperanza y la varianza del salario anual medio de la empresa?

17. En el control de calidad de una fábrica de latas de atún, se envasan latas de 100 gramos con una desviación típica de 2 gramos.

    a) Determina la probabilidad de que un lote de 400 latas pese más de 40 100 gramos.

    b) Se empaquetan en cajas de 50 latas. Calcula la probabilidad de que la media de las latas de una caja sea menor que 99 gramos.

18. Los pesos de las ovejas de una cierta ganadería tienen una media de 50 kg con una desviación típica de 4 kg. Elegimos al azar una muestra aleatoria simple de 100 ovejas.
Determina la probabilidad de que su media sea:

    a) superior a 51 kg *b*) inferior a 56 kg *c*) entre 48 kg y 52 kg

Aproximación de la binomial por la normal

Hemos definido la distribución binomial como la suma de n ensayos de Bernouilli independientes con la misma media y varianza.

Así vista es inmediato que podemos aplicar el teorema central del límite para un número suficientemente de ensayos ($n > 30$), y se considera una buena aproximación cuando se cumple que $np \geq 5$ y $n(1-p) \geq 5$.

Recordemos que la esperanza de la binomial es np y su varianza $np(1-p)$.

$$B(n,p) \sim N\left(np, \sqrt{np(1-p)}\right)$$

En la fabricación de piezas de alta presión, el 17 % de las piezas son defectuosas.

		Probabilidad
Éxito	Pieza correcta	$p = 0'83$
Fracaso	Pieza defectuosa	$1 - p = 0'17$

Se trata de una binomial $B(50, 0'83)$, por lo tanto:

- $\mu = np = 41'5$
- $\sigma^2 = np(1-p) = 7'055 \quad \Longrightarrow \quad \sigma \approx 2'66$

Así que puede aproximarse por una normal $N(41'5, 2'66)$.

19. Comprueba si las siguientes distribuciones binomiales pueden ser aproximadas razonablemente por una distribución normal y en caso afirmativo hállala.

    a) $B(100, 0'3)$ b) $B(23, 0'75)$ c) $B(52, 0'4)$ d) $B(85, 0'03)$

Si aproximamos una distribución binomial mediante una normal, estamos convirtiendo una variable discreta (toma un número determinado de valores) en una continua (que en valores aislados tiene probabilidad cero) por lo que debemos emplear algún método de corrección para evitar ese problema.

En 2º de Bachillerato empleamos la **corrección de continuidad de Yates**, sustituyendo cada valor fijo por un intervalo de amplitud 1 centrado en el valor.

Caso	$\longrightarrow$	Corrección	
$P(X = a)$	$\longrightarrow$	$P(a - 0'5 \leq Y \leq a + 0'5)$	
$P(X \leq a)$	$\longrightarrow$	$P(Y \leq a + 0'5)$	para que contenga a a
$P(X < a)$	$\longrightarrow$	$P(Y \leq a - 0'5)$	para que no contenga a a
$P(X \geq a)$	$\longrightarrow$	$P(Y \geq a - 0'5)$	para que contenga a a
$P(X > a)$	$\longrightarrow$	$P(Y \geq a + 0'5)$	para que no contenga a a

Una vez que hemos aplicado la corrección de continuidad procederemos a tipificar para poder utilizar nuestra tabla de la distribución normal $N(0, 1)$.

En la fabricación de piezas de alta presión el 17 % de las piezas son defectuosas, lo cual puede aproximarse por una $N(41'5, 2'66)$. Calcula las siguientes probabilidades:
a) Hay al menos 40 correctas. b) Hay exactamente 10 correctas.

a) $P(X \geq 40) \xrightarrow{\text{corrección Yates}} P(Y \geq 39'5) = P\left(\frac{Y - 41'5}{2'66} \geq \frac{39'5 - 41'5}{2'66}\right) \approx P(Z \geq -0'75) = P(Z \leq 0'75) = 0'7734$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

b) $P(X = 10) \xrightarrow{\text{corrección Yates}} P(9'5 \leq Y \leq 10'5) = P\left(\frac{9'5 - 41'5}{2'66} \leq Z \leq \frac{10'5 - 41'5}{2'66}\right) = P(-12'03 \leq Z \leq -11'65) =$
$P(Z \leq -11'65) - P(Z \leq -12'03) = P(Z \leq 12'03) - P(Z \leq 11'65) = 1 - 1 = 0$

20. El 2 % de los tornillos fabricados por una máquina presentan defectos. Si tenemos un lote de 2 000 tornillos, ¿cuál es la probabilidad de que haya menos de 50 defectuosos?

21. En un test de resistencia se comprueba que un objeto en caída libre desde cierta altura tiene una probabilidad del 20 % de romperse. Si dejamos caer 800 de esos objetos desde esa altura, ¿cuál es la probabilidad de que exploten exactamente 160 de ellos?

22. Se lanza una moneda al aire 400 veces. Calcula la probabilidad de que salga cara entre 180 y 210 veces.

23. Un tirador acierta en el blanco el 70 % de los tiros. Si en una competición tira 25 veces, ¿cuál es la probabilidad de que acierte más de 10 tiros?

24. Un cierto equipo electrónico está formado por 100 componentes conectados. Si cada componente tiene una probabilidad de 0.02 de romperse cunado el equipo es lanzado en un cohete, hallar la probabilidad de que al hacerlo se rompan 10 o más componentes.

25. Un examen tipo test consta de 38 preguntas a contestar verdadero o falso. El examen se aprueba si se contesta correctamente al menos 20 preguntas. Un alumno responde al examen lanzando al aire una moneda y contestando verdadero si sale cara y falso si sale cruz. Halla:

    a) La probabilidad de aprobar el examen.

    b) Probabilidad de acertar más de 24 y menos de 31.

Distribución de la proporción muestral

En ocasiones se plantea la necesidad de estimar una proporción o bien un porcentaje. Para ello tenemos en cuenta que se trata de la observación de una serie de datos muestrales

que toman solamente dos valores diferentes (éxito o fracaso), es decir, que siguen una distribución binomial.
En consecuencia la **proporción muestral** $\widehat{X} = \frac{X}{n}$ hallada a partir de n experimentos de Bernouilli con probabilidad de éxito p tiene:

- La esperanza es $E\left[\frac{X}{n}\right] = \dfrac{1}{n} \cdot V[X] = \dfrac{np}{n} = \boxed{p}$
- La varianza es $V\left[\frac{X}{n}\right] = \dfrac{1}{n^2} \cdot V[X] = \dfrac{n \cdot p \cdot (1-p)}{n^2} = \dfrac{p \cdot (1-p)}{n}$

  y por lo tanto la desviación típica es $\boxed{\sqrt{\dfrac{p \cdot (1-p)}{n}}}$

Aplicando el teorema central del límite podemos afirmar que:

$$\widehat{X} \sim N\left(p, \sqrt{\frac{p \cdot (1-p)}{n}}\right)$$

Aunque técnicamente estamos aproximando una variable discreta por una continua, en este tipo de cálculos no se emplea la corrección de Yates pues los valores que toma la proporción muestral son tan pequeños que una modificación de medio punto nos llevaría a cometer un error mayor que el de continuidad.

De una población de 120 alumnos, hay 48 que tienen 2 o más hermanos. Si de dicha población se toman muestras de tamaño 54, ¿cuál es la probabilidad de que se encuentre en dicha muestra una proporción de más del 60 % de alumnos con 2 o más hermanos?

Tenemos $n = 54$ muestras, con una probabilidad $p = \dfrac{48}{120} = 0'4$ cada una.
Por lo tanto aproximaremos la proporción muestral por una normal de media $p = 0'4$ y desviación típica $\sqrt{\dfrac{p \cdot (1-p)}{n}} = \sqrt{\dfrac{0'4 \cdot 0'6}{54}} = \dfrac{1}{15}$

$$P(X \leq 0'6) = P\left(\frac{X - 0'4}{\frac{1}{15}} \leq \frac{0'6 - 0'4}{\frac{1}{15}}\right) = P(Z \leq 3) = 0'9987$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995

26. Una fábrica de pasteles fabrica, en su producción habitual, un 3 % de pasteles defectuosos. Un cliente recibe un pedido de 500 pasteles de la fábrica. Calcula la probabilidad de que encuentre más del 5 % de pasteles defectuosos.

27. En una determinada población se sabe que el 20 % de las personas usan gafas graduadas y el resto no. Tomamos una muestra de 256 personas. ¿Cuál es la probabilidad de que el porcentaje

de personas encuestadas que usan gafas esté entre el 15 % y el 25 %?

28. Se sabe que el 10 % de los habitantes de una determinada ciudad va regularmente al teatro. Se toma una muestra al azar de 100 habitantes de esta ciudad, ¿cuál es la probabilidad aproximada de que al menos el 13 % de ellos vaya regularmente al teatro?

29. Un estudio realizado por una compañia de seguros de automóviles establece que una de cada cinco personas accidentadas es mujer. Si se contabilizan, por término medio, 169 accidentes cada fin de semana:

    a) ¿Cuál es la probabilidad de que, en un fin de semana, la proporción de mujeres accidentadas supere el 24 %?

    b) ¿Cuál es la probabilidad de que, en un fin de semana, la proporción de hombres accidentados supere el 85 %?

Programación lineal

Inecuaciones

Si una ecuación es la igualdad entre expresiones algebraicas, una **inecuación** es la desigualdad entre expresiones algebraicas.

La solución de una inecuación es frecuentemente un intervalo o varios intervalos.

$$\begin{gathered} 2x+7<3 \\ 2x<3-7 \\ 2x<-4 \\ x<-2 \end{gathered}$$

Al resolver inecuaciones sencillas hay una diferencia fundamental a tener en cuenta: multiplicar (o dividir) por un número negativo modifica el tipo de desigualdad.

$$\begin{gathered} 2<5 \\ -2>-5 \end{gathered} \qquad \begin{gathered} -40>-70 \\ 4<7 \end{gathered} \qquad \begin{gathered} -2x\geq 5 \\ x\leq -2'5 \end{gathered}$$

1. Resuelve las siguientes inecuaciones y expresa la solución como intervalos:

   a) $2x-4<5$
   b) $3x+8>20$
   c) $2x-6\leq 5x$
   d) $-4x+10\geq -6x$
   e) $3x+7>5x-11$
   f) $3x+6<30$
   g) $8x-5\leq 13+10x$
   h) $\dfrac{x}{2}-4\geq \dfrac{x}{4}+1$

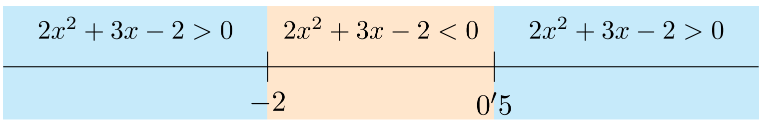
2. Resuelve las siguientes inecuaciones y expresa la solución como intervalos:

   a) $x^2\leq 4$
   b) $3x^2\geq 75$
   c) $45-5x^2>0$
   d) $12-3x^2<0$

Las inecuaciones polinómicas no siempre pueden resolverse de forma directa.
En ese caso procederemos a resolver la ecuación asociada para averiguar en que valores se cumple exactamente la igualdad, y razonar el tipo de desigualdad que se cumple en los intervalos entre ellos.

$2x^2+3x-2>0$

$$2x^2+3x-2=0 \Longrightarrow x=\frac{-3\pm\sqrt{3^2-4\cdot 2\cdot(-2)}}{2\cdot 2}=\frac{-3\pm 5}{4}\Longrightarrow \begin{cases} x=\dfrac{1}{2} \\ x=-2 \end{cases}$$



Por lo tanto la solución es $(-\infty,-2)\cup(0'5,+\infty)$.

3. Resuelve las siguientes inecuaciones de segundo grado:

a) $x^2 - x - 20 \geq 0$ b) $-x^2 - 2x + 8 > 0$ c) $x^2 - 3x - 10 \leq 0$ d) $x^2 + x - 2 < 0$

4. Factoriza y resuelve las siguientes inecuaciones:

a) $x^3 - 3x > -2$ b) $x^3 + x^2 - 9x - 9 \leq 0$ c) $6x^3 + x^2 \geq x$

5. Opera y resuelve las siguientes inecuaciones:

a) $1 - 2(x+2) < 4(x-9)$

b) $5(4x-1) \geq -4(x+3)$

c) $x + 3(7-x) < 4x - 5(3+x)$

d) $\dfrac{x+5}{4} > 2x + \dfrac{x}{3}$

Sistemas de inecuaciones

Los sistemas de inecuaciones más sencillos son aquellos que tienen solo una incógnita. En ese caso la solución será el intervalo o intervalos que cumplan todas desigualdades.

6. Resuelve los siguientes sistemas:

a) $\begin{cases} 3x+1 > -2 \\ 4-2x \leq 0 \end{cases}$

b) $\begin{cases} x > 0 \\ 2x \leq 4 \end{cases}$

c) $\begin{cases} x+3 < 2 \\ 2x-5 < 3 \end{cases}$

d) $\begin{cases} 2x-5 \leq 1 \\ 4-2x > -4 \end{cases}$

e) $\begin{cases} \dfrac{x}{2} \geq -2 \\ 5x-4 \leq 2 \end{cases}$

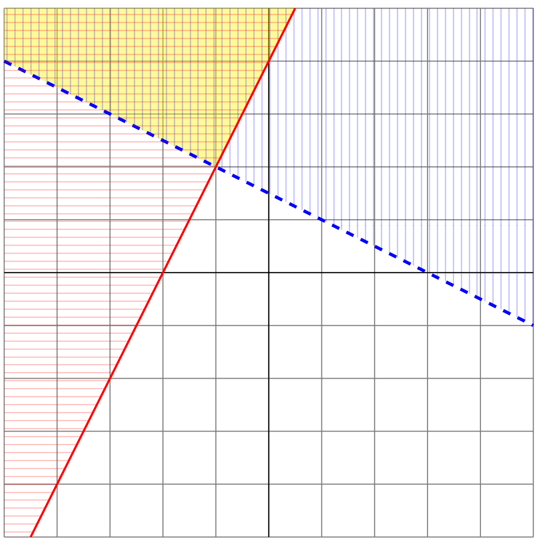
f) $\begin{cases} 6x-3 \geq x+7 \\ 7x+3 \leq 15+3x \end{cases}$

Para resolver sistemas de inecuaciones con dos incógnitas emplearemos el método gráfico. Las soluciones que obtendremos darán lugar a una figura plana (acotada o no acotada) a la que llamamos **región factible**.

$\begin{cases} x+2y > 3 \\ 2x-y \leq -4 \end{cases}$

- Dibujamos la recta $x + 2y = 3$ y marcamos el semiplano que corresponde a la inecuación $x + 2y > 3$.
- Dibujamos la recta $2x - y = -4$ y marcamos el semiplano que corresponde a la inecuación $2x - y \leq -4$.

La región que pertenece a ambos semiplanos es la región factible de nuestro problema.



7. Halla la región factible de:

a) $\begin{cases} x+4y \leq 1 \\ x-2y \geq 2 \end{cases}$ b) $\begin{cases} 2x+y < -1 \\ 4x-y > -3 \end{cases}$ c) $\begin{cases} 2x+3x \leq 4 \\ x+y \leq 3 \end{cases}$

8. Halla la región factible de:

a) $\begin{cases} x > 0 \\ y > 0 \\ x+2y \geq 3 \end{cases}$ b) $\begin{cases} x < 2 \\ y < 3 \\ x+2y \geq 3 \end{cases}$ c) $\begin{cases} y > 0 \\ x+y < 2 \\ 5x+3y > 0 \end{cases}$

Optimización

Llamamos **programación lineal** a un conjunto de técnicas que pretenden **optimizar** (es decir, minimizar o maximizar) una función lineal de dos variables llamada **función objetivo** sujeta a un conjunto de **restricciones** definido como un sistema de inecuaciones lineales de dos incógnitas.

Teorema fundamental de la Programación Lineal

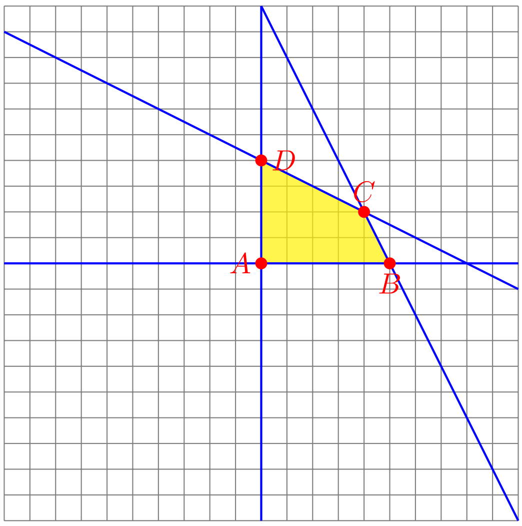
Si un problema de programación lineal tiene una región factible acotada y existe una solución única que optimice la función objetivo, ésta se encuentra en un **vértice** de la región factible.

Si una función alcanza el valor óptimo en dos vértices consecutivos de la región factible, entonces alcanza también dicho valor óptimo en todos los puntos del **segmento** que determinan ambos vértices.

Queremos maximizar la función $f(x,y) = 3x+4y$ teniendo en cuenta las restricciones:

- $x \geq 0$
- $2x+y \leq 10$
- $y \geq 0$
- $x+2y \leq 8$

Para ello dibujamos la región factible:



Calculamos las coordenadas de los vértices de la región factible, que corresponden a la solución del sistema de ecuaciones de las rectas dibujadas:

$$A=(0,0) \quad B=(5,0)$$
$$C=(4,2) \quad D=(0,4)$$

Y evaluamos la función objetivo en dichos puntos:

- $f(0,0) = 3\cdot 0 + 4\cdot 0 = 0$
- $f(5,0) = 3\cdot 5 + 4\cdot 0 = 15$
- $f(4,2) = 3\cdot 4 + 4\cdot 2 = 20$
- $f(0,4) = 3\cdot 0 + 4\cdot 4 = 16$

Por lo tanto el máximo se alcanza en $x=4$, $y=2$.

9. Maximiza la función $f(x, y) = 5x - y$, sujeta las restricciones:

   $x \geq -2; \qquad y \leq 3; \qquad 2x - y \leq 3$

10. Maximiza la función $f(x, y) = 5x - y$, sujeta las restricciones:

    $2x - y \leq 3; \qquad x + y \geq -23; \qquad y - x \leq 3.$

11. Minimiza la función $f(x, y) = 7x - 9y$, sujeta las restricciones:

    $x \leq 5; \qquad x + y \geq 0; \qquad x - y \geq 3.$

12. Maximiza la función $f(x, y) = 3x + 2y$, sujeta las restricciones:

    $x \geq 1; \qquad 3y \leq 24 - 2x;$

    $y \geq 2; \qquad y + 2x \leq 12.$

13. Maximiza la función $f(x, y) = x + y$ sabiendo que deben cumplirse las restricciones:

$$\begin{cases} x + 3y \leq 26 \\ 4x + 3y \leq 44 \\ 2x + 3y \leq 28 \\ x \geq 0 \\ y \geq 0 \end{cases}$$

14. Minimiza la función $f(x, y) = 3x + 2y$ sabiendo que deben cumplirse las restricciones:

$$\begin{cases} 3x + 4y \leq 12 \\ 3x + 2y \geq 2 \\ x \geq 0 \\ y \geq 0 \end{cases}$$

En el caso de que la región factible no esté acotada es posible que no haya solución. or ello, debemos comprobar siempre el valor de la función objetivo en los puntos interiores de la región factible.

15. Maximiza la función $f(x, y) = 4y - x$ sabiendo que deben cumplirse restricciones:

$$\begin{cases} -2x + y \leq 3 \\ 2x - y \leq 2 \\ x + 2y \leq 4 \end{cases}$$

16. Minimiza la función $f(x, y) = 5x + 2y$ sabiendo que deben cumplirse restricciones:

$$\begin{cases} x \leq 1 \\ 4x - y \leq 0 \\ 3x + 2y \geq -6 \end{cases}$$

Problemas de producción

Los **problemas de producción** consisten en maximizar los beneficios o minimizar los costes de producción de artículos que están sometidos a restricciones.

17. Para fabricar dos tipos de cables, A y B, que se venderán a 150 € y 100 € el hectómetro, respectivamente, se emplean 16 kg de plástico y 4 kg de cobre para cada hectómetro del tipo A y 6 kg de plástico y 12 kg de cobre para cada hectómetro del tipo B. El cable fabricado del tipo B no puede ser mayor que el doble del tipo A y, además, solo tenemos 252 kg de plástico y 168 kg de cobre. Determina la longitud, en hectómetros, de cada tipo de cable para que la cantidad de dinero obtenida en su venta sea máxima.

18. Un sastre tiene $80\ \mathrm{m}^2$ de tela de algodón y $140\ \mathrm{m}^2$ de tela de lana para confeccionar trajes y vestidos. Un traje requiere $1\ \mathrm{m}^2$ de algodón y $3\ \mathrm{m}^2$ de lana y un vestido $2\ \mathrm{m}^2$ de cada tela. Halla el número de trajes y vestidos que debe confeccionar para obtener máximos beneficios. ¿Cuál será el beneficio obtenido si el precio de venta es de 50 € por cada prenda?

19. Una fábrica tiene 800 kg de guisantes para conservar en dos tipos de latas. La lata pequeña contiene 200 g y aporta un beneficio de 10 céntimos por lata; la grande, 500 g y aporta 30 céntimos. Si en el almacén hay 2000 latas de tamaño pequeño y 1000 grandes, determina la cantidad de latas de cada tamaño que hay que producir para maximizar el beneficio.

20. Se desea fabricar dos tipos de bombones que llamaremos A y B. Las cajas del tipo A contienen 1 kg de chocolate y 2 kg de cacao; las del tipo B contienen 2 kg de chocolate, 1 kg de cacao y 1 kg de almendras. Disponemos de 500 kg de chocolate, 400 de cacao y 225 de almendras. Por cada caja del tipo A se ganan 2 euros y por cada caja del tipo B 3 euros. ¿Cuántas cajas de cada tipo hay que fabricar para que la ganancia se máxima?

21. Una casa empacadora de alimentos recibe diariamente 700 kg de café de tipo C y 800 kg de café de tipo K. Hace con ellos dos mezclas. Una que consta de 2 partes de café de tipo C y 1 de tipo K, en la que gana 22 céntimos por kilo, y otra que consta de una parte de tipo C y 2 del tipo K, en la que gana 26 céntimos por kilo. Halla la cantidad de mezcla que la casa debe preparar de cada clase para que la ganancia sea máxima.

22. Con 80 kg de acero y 120 kg de aluminio se quieren fabricar bicicletas de montaña y de paseo que se venderán a 200 euros y 150 euros, respectivamente. Para la de montaña son necesarios 1 kg de acero y 3 kg de aluminio y para la de paseo, 2 kg de cada uno de los materiales. ¿Cuántas bicicletas de cada tipo se deben fabricar para maximizar el beneficio?

23. Los abonos A y B se obtienen mezclando cierto sustrato con dos fertilizantes F1 y F2 en las siguientes proporciones:

	F1	F2
A	100 g/kg	50 g/kg
B	70 g/kg	80 g/kg

La cantidad disponible de los fertilizantes F1 y F2 son 39 kg y 24 kg respectivamente. El beneficio que producen los abonos A y B son 75 céntimos/kg y 60 céntimos/kg. ¿Cuántos kilos se deben fabricar del abono A y del abono B para maximizar el beneficio?

Problemas de dietas

Los **problemas de dietas** consisten en minimizar los costes de los alimentos que constituyen la dieta diaria de un colectivo (animales o personas) y que aportan los niveles de nutrientes necesarios.

24. Un veterinario desea dar a sus animales una dieta que contenga un mínimo de 30 unidades de pienso tipo A y 20 unidades de pienso tipo B. En el mercado se encuentran dos productos, P_1 y P_2, que se elaboran con dichos piensos. Cada bolsa de P_1, que cuesta $2,50$ €, contiene 4 unidades del tipo A y 2 unidades de B, mientras que cada bolsa de P_2, cuyo coste es de $3,25$ €, contiene 5 unidades de A y 5 unidades de B. ¿Qué cantidad de P_1 y P_2 deberá comprar para que la dieta sea de coste mínimo?

25. Un deportista necesita diariamente consumir $36\ \mathrm{g}$ de una sustancia $M, 24g$ de N y $8\ \mathrm{g}$ de P. En la farmacia ha encontrado dos tipos de cápsulas que contienen estas sustancias. Las cápsulas A tienen $6\ \mathrm{g}$ de $M, 2\ \mathrm{g}$ de N y $18\ \mathrm{g}$ de P, y cuestan 3 céntimos por cápsula. Las cápsulas B tienen $3\ \mathrm{g}$ de $M, 4\ \mathrm{g}$ de N y $18\ \mathrm{g}$ de P, y cuestan 4,5 céntimos por cápsula. ¿Cuántas cápsulas de cada tipo necesita para que el coste sea mínimo?

26. Los animales de una granja deben tomar, al menos, $60\mathrm{mg}$ de vitamina Ay, al menos, $90\mathrm{mg}$ de vitamina B. Existen dos compuestos con estas vitaminas. El compuesto X contiene $10\mathrm{mg}$ de vitamina A y $15\mathrm{mg}$ de $\mathrm{B, y}$ cada dosis cuesta $0,50$ €. El compuesto Y contiene $10\mathrm{mg}$ de cada vitamina, y cada dosis cuesta $0,30$ €. Además, se recomienda no tomar más de 8 dosis diarias. Calcula qué dosis tiene que tomar para que el coste sea mínimo.

27. Se quiere elaborar una dieta diaria para ganado que satisfaga unas condiciones minimas de contenido vitaminico al dia: $2\mathrm{mg}$ de vitamina A, $3\mathrm{mg}$ de vitamina $\mathrm{B}, 30$ de la C y 2 de la D. Para ello se mezclan piensos de los tipos P y Q cuyo precio por kilogramo es para ambos de 30 céntimos, y cuyo contenido vitaminico por kilo se recoge en la tabla adjunta.

Problemas de transporte

Los **problemas de transporte** consisten en minimizar los costes del traslado de mercancías o personas a un destino para satisfacer una demanda sin excederla.

28. Se quiere transportar, entre dos ciudades, a 1600 personas y 96 toneladas de equipaje. Los aviones disponibles son de dos tipos: 11 del tipo A y 8 del tipo B. La contratación de un avión del tipo A cuesta 4000 € y puede transportar a 200 personas y 6 toneladas de equipaje; los aviones del tipo B cuestan 1000 € y pueden transportar a 100 personas y 15 toneladas. ¿Cuántos aviones de cada tipo deben utilizarse para que el coste sea mínimo?

29. Una empresa se dedica a elaborar lotes de productos. El lote de tipo A tiene 1 queso y 2 botellas de vino, y su transporte cuesta $0,90$ €. El lote de tipo B tiene 3 quesos y 1 botella de vino, y cuesta $1,50$ € transportarlo. La empresa dispone de 200 quesos y 100 botellas de vino, y necesitan elaborar, al menos, 10 lotes del tipo A y 25 del tipo B. ¿Cuántos lotes de cada clase han de elaborar para que los gastos en transporte sean mínimos?

30. Una compañía aérea tiene dos aviones A y B para cubrir un determinado trayecto. El avión A debe hacer más veces el trayecto que el avión B pero no puede sobrepasar 120 viajes. Entre los dos aviones deben hacer 60 o más vuelos, pero no más de 200. En cada vuelo A consume 900 litros de combustible y B 700 litros. En cada viaje del avión A la empresa gana 2000 euros y 1500 por cada viaje del B. ¿Cuántos viajes debe hacer cada avión para obtener el máximo de ganancias? ¿Cuántos vuelos debe hacer cada avión para que el consumo de combustible sea mínimo?

Otros problemas

Los anteriores son los tipos de problemas más populares de programación lineal, pero hay muchísimas cuestiones de logística en las que se puede aplicar.

31. Un astillero recibe un encargo para reparar barcos de la flota de un armador, compuesta por pesqueros de 500 toneladas y yates de 100 toneladas. Cada pesquero se tarda en reparar 100 horas y cada yate 50 horas. El astillero dispone de 1600 horas para hacer las reparaciones. Por política de empresa, el astillero no acepta encargos de más de 12 pesqueros ni más de 16 yates. Las reparaciones se pagan a 100 euros la tonelada, independientemente del tipo de barco. ¿Cuántos barcos de cada clase debe reparar el astillero para maximizar el ingreso con este encargo? ¿Cuál es dicho ingreso máximo?

32. Para abonar una parcela de huerta se necesitan por los menos 8 kg de nitrógeno y 12 kg de fósforo. Se dispone de un producto A cuyo precio es de 30 céntimos/kg y que contiene un 10 % de nitrógeno y un 30 % de fósforo. Existe en el mercado otro producto B que contiene un 20 % de nitrógeno y un 20 % de fósforo, y cuyo precio es de 40 céntimos/kg. ¿Qué cantidad se deben tomar de A y B para abonar la parcela con el menor gasto posible?

33. En un almacén se guarda aceite de girasol y de oliva. Para atender a los clientes se ha de tener almacenado un mínimo de 20 bidones de aceite de girasol y 40 de aceite de oliva y, además, el número de bidones de aceite de oliva no debe ser inferior a la mitad del número de bidones de aceite de girasol. La capacidad total del almacén es de 150 bidones. Sabiendo que el gasto de almacenaje de un bidón de aceite de oliva es de 100 euros y de uno de girasol de 50 euros, ¿cuántos bidones de cada tipo habrá que almacenar para que el gasto sea mínimo?