Probabilidad:Variables aleatorias continuas. Distribución normal

Matemáticas Aplicadas a las Ciencias Sociales I

1.º Bachillerato

Contenidos

Probabilidad

Variables aleatorias continuas. Distribución normal

1. Introducción

Aunque a lo largo de todos los temas intentamos aplicar al máximo las matemáticas a las Ciencias Sociales, es este tema quizás, el que más se preste a ello, pues es enorme la cantidad de situaciones financieras, sociológicas, psicológica, biológicas,..., que se ajustan al modelo de probabilidad que vamos a estudiar.

En los dos primeros temas de la unidad has visto cómo calcular probabilidades, y en el anterior, como algunas situaciones se pueden ajustar a un modelo concreto, el binomial, que facilita en gran medida el cálculo de estas probabilidades. Pues bien, en este tema, vamos a introducir un nuevo modelo y verás que hay muchas situaciones que se pueden resolver haciendo referencia a él.

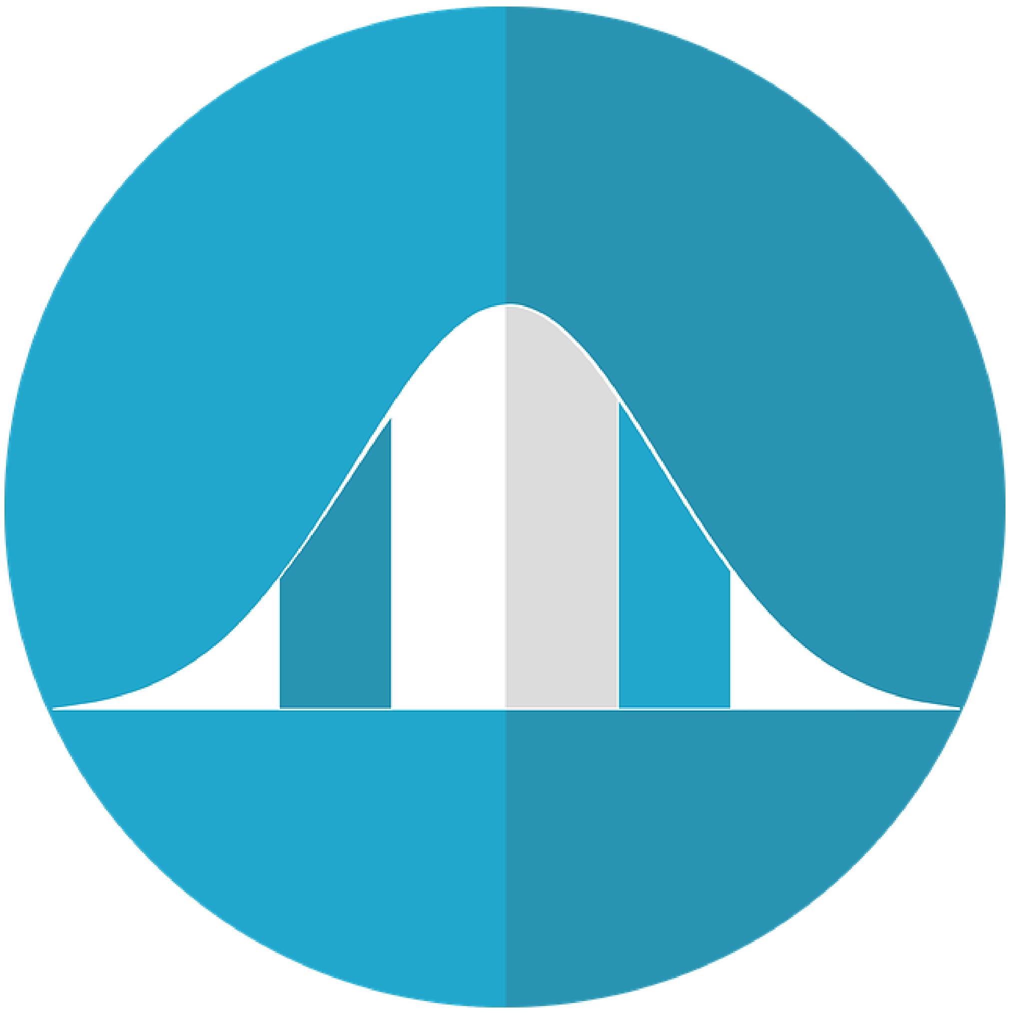


Imagen de mcmurryjulie en Pixabay. Dominio Público

2. Variable continua

¿Recuerdas que al empezar los temas de estadística dividíamos las variables estadísticas en dos tipos, discretas y continuas?

Así por encima, las variables discretas eran las que tomaban valores aislados mientras que las continuas eran las que podían tomar cualquier valor entre dos concretos.

Pues bien, en el cálculo de probabilidades ocurre lo mismo. Ya en el tema anterior has visto que muchos experimentos aleatorios se pueden transformar en variables aleatorias, siempre que el resultado del experimento se pueda expresar mediante un número, pero en todos los casos que vimos, las variables que nos salían eran discretas. Sólo podían tomar algunos valores. Pero, esto no siempre es así, ¿no?

Imagínate que al azar elijo la primera persona que se me cruza por la calle, y me pregunto: *¿cuánto medirá?*

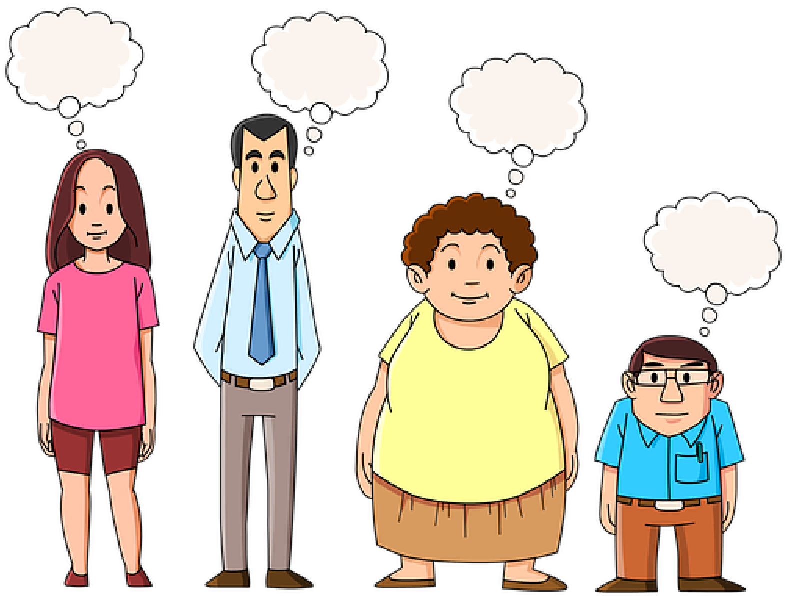


Imagen de MoteOo en Pixabay. Pixabay License

Las respuestas podrían ser muchas. Infinitas incluso. Bueno, pues ejemplos como éste, nos van a llevar al concepto de variable aleatoria continua.

Importante

Una **variable aleatoria** es **continua** si al realizar el experimento aleatorio, entre cada dos valores, el número de valores que puede tomar es infinito.

Por ejemplo, la altura de una persona, la longitud del dedo índice, el peso de un perro, el caudal de un río...

La **media,** μ, y la **desviación típica,** σ , indican los mismos parámetros que en las distribuciones estadísticas. La media indica el centro de gravedad de la distribución y la desviación típica la medida de dispersión.

Caso práctico

Imagínate que hemos recogido los datos de la altura de 50 personas y los hemos anotado en esta tabla:

1,65	1,84	1,57	1,77	1,70	1,54	1,94	1,65
1,63	1,79	2,00	1,83	1,53	1,85	1,53	2,06
1,97	1,67	2,01	2,02	1,62	1,91	1,90	1,52
2,03	1,98	2,05	1,71	1,72	1,89	2,05	1,99
2,06	1,87	2,01	1,96	1,98	1,87	1,78	1,60

1,63	**1,82**	**1,85**	**2,09**	**1,92**	**1,78**	**1,78**	**1,53**
1,98	**1,91**						

¿Cuál sería la probabilidad de que al elegir una persona al azar, ésta mida 1,81 m? ¿Y 1,73 m?

Pues si te fijas, estos valores no parecen en esa tabla, así que, si aplicamos la regla de Laplace, casos favorables entre casos posibles, tenemos que ambas probabilidades son cero, pues el número de casos favorables es 0.

Si definimos la variable aleatoria X = altura de la persona, tendríamos que:

P(X= 1,81) = 0 = P(X = 1,73)

Claro dirás, coge algún valor de la tabla. Pues de acuerdo, vamos a coger uno. ¿Cuál es la probabilidad de que al elegir una persona al azar ésta mida 1,92 m? ¿Y 1,77 m? ¿Y 1,63 ?

En los dos primeros casos, estos valores aparecen una sola vez, luego el número de casos favorables sería 1 y aplicando otra vez la Regla de Laplace, obtenemos que ambas probabilidades son 1 / 50, o lo que es lo mismo, 0,02.

P(X = 1,92) = 0,02 y P(X = 1,77) = 0,02

En el tercer caso, el valor 1,63 aparece dos veces, luego la probabilidad de que tome ese valor sería 0,04.

P(X = 1,63) = 0,04

Fíjate que en ambos casos la probabilidad es prácticamente cero, y eso, que sólo nos hemos quedado en los centímetros. Si hubiéramos afinado hasta los milímetros sería casi imposible que el valor por el que nos preguntáramos estuviera en la tabla.

Así, si una variable es continua, la probabilidad de que la variable tome un valor concreto es cero.

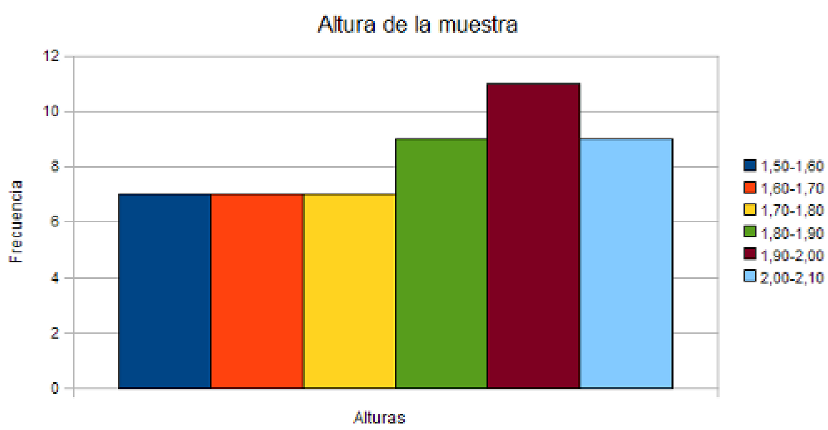
Importante

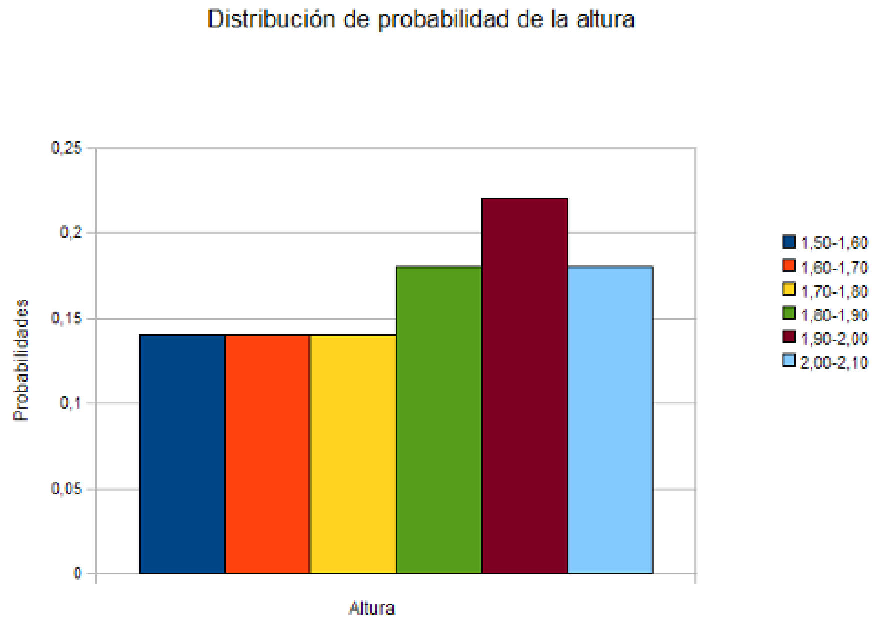
Si X es una variable aleatoria continua, la probabilidad de que tome un valor concreto es cero.

P[X = a] = 0, para cualquier valor de a.

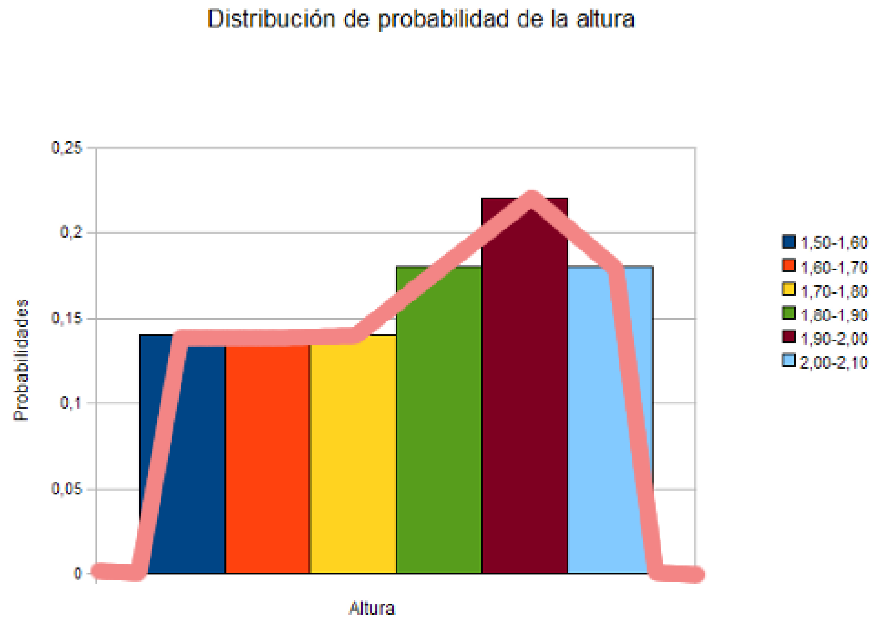
Si en lugar de preguntarnos por valores sueltos nos preguntamos por la probabilidad de un grupo ya la cosa cambia.

Fíjate, en la imagen de la derecha hemos agrupado los mismos datos de la tabla de arriba en intervalos de 10 centímetros de amplitud.
Ahora sin mucha dificultad, podemos calcular la probabilidad de que al elegir una persona al azar ésta esté entre 1,70 y 1,80; sería 7/50.
O la probabilidad de que una persona mida menos de 1,80; sería 21/50.
Si te fijas, estas probabilidades son en realidad las frecuencias relativas, así que en lugar de este gráfico vamos a usar el de las frecuencias relativas.





Pues bien, en una variable aleatoria continua, las probabilidades se calculan a través de la llamada **función de densidad**, función que juega un papel similar al polígono de frecuencias, y éstas se van a calcular como el área del recinto que está bajo esta curva.
En este ejemplo que estamos viendo, la función de densidad sería la función dibujada en rosa:



Aunque a decir verdad, la función de densidad es el polígono de frecuencias que surge cuando tomamos intervalos muy muy pequeñitos, es decir, con una amplitud muy pequeña.

Importante

En una variable aleatoria continua, las probabilidades que calculamos están siempre asociadas a intervalos.
Si X es una variable aleatoria continua, la probabilidad de que X esté en un intervalo es el área del recinto limitado por el intervalo y la **función de densidad** f **(o función de probabilidad).**
f debe ser siempre mayor o igual que cero y el área debajo de su gráfica tiene que ser 1.
En el caso de tener una variable aleatoria continua X, ya hemos visto que las probabilidades de valores concretos son 0. Así, al describir un intervalo, no influye que la desigualdad sea estricta o no para calcular la probabilidad.

P (X ≤ a) = P (X < a)

P (a< X < b) = P (a ≤ X < b) = P (a < X ≤ b) = P (a ≤ X ≤ b)

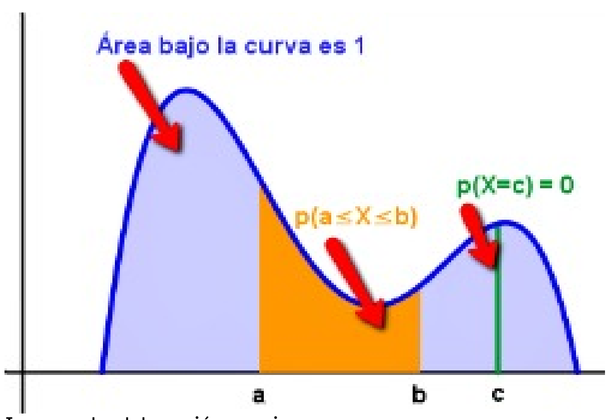


Imagen de elaboración propia

Se define la **función de distribución, F(x)**, como: **F(x)= P [X≤ x]**

Comprueba lo aprendido

Contesta verdadero o falso a las siguientes cuestiones. Repasa lo comentado anteriormente y propiedades generales de la probabilidad.

1) Si X es una variable aleatoria continua, $P(X=2) = 0$

○ Verdadero ○ Falso

Verdadero

Estás calculando la probabilidad de un valor concreto

2) Si X es una variable aleatoria continua, $P(X > 2) = 0$

○ Verdadero ○ Falso

Falso

No tiene porqué. Depende de los valores de X.

3) El área máxima bajo la función de densidad es 1.

○ Verdadero ○ Falso

Verdadero

Claro, ten en cuenta que la probabilidad máxima o total es 1, que es la probabilidad de todo el espacio muestral.

4) Si X es una variable aleatoria continua, $P(X > 2) = P(X \geq 2)$

○ Verdadero ○ Falso

Verdadero

Sólo tienes que revisar el último apartado "Importante"

5) La función de densidad de una variable aleatoria continua puede tomar valores por debajo del eje OX, es decir, valores negativos.

○ Verdadero ○ Falso

Falso

Cómo mínimo tiene que estar sobre el eje OX, para que el área sea cero y también la probabilidad, ¿o hay probabilidades negativas?

3. Distribución normal

Del mismo modo que entre las variables discretas hemos introducido el modelo binomial, entre las variables continuas vamos a introducir un modelo al que se ajusta un gran número de variables de nuestro entorno. Este modelo se va a llamar **distribución normal**, y su nombre se debe ni más ni menos a que la gran mayoría de variables que se refieren a aspectos físicos, psicológicos, sociológicos, biológicos, etc. se ajustan a este modelo, o sea, que lo normal, es que sea una variable normal.

Imagen de geralt en Pixabay. Pixabay License

Cuando comenzaron los estudios estadísticos modernos, se encontró que la proporción de variables continuas que seguían esta distribución era tan elevada, que se llegó a pensar que todas estas variables la seguían.
Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- Caracteres morfológicos de individuos (no sólo personas, sino también animales o plantas) como la estatura, el peso, perímetros, etc.
- Caracteres fisiológicos como el efecto de una dosis de un fármaco o el producido por un abono en un cultivo.
- Caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos.
- Caracteres psicológicos como el cociente intelectual o nivel de adaptación a un medio.
- Nivel de ruido en telecomunicaciones.
- Errores cometidos al medir ciertas magnitudes.

Y en general, cualquier característica que se obtenga como suma de muchos factores independientes se ajustará a un modelo normal.
La distribución Normal se expresa por $N(\mu, \sigma)$, donde el primer parámetro μ, representa la media de la variable aleatoria y σ la desviación típica.

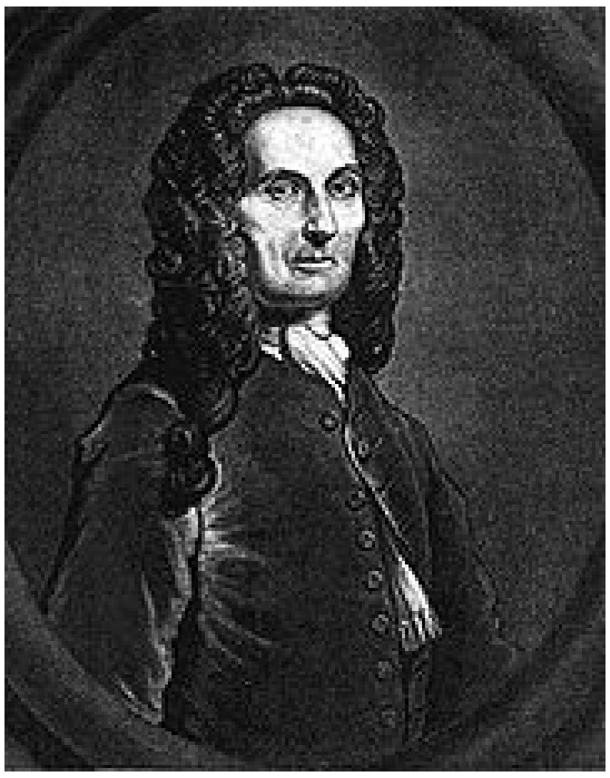
Curiosidad

Antes de continuar, vamos a dar una pequeña pincelada histórica sobre el descubrimiento y evolución de esta importantísima distribución de probabilidad.

La distribución normal fue presentada por primera vez por Abraham De Moivre en un artículo del año 1733, que fue reimpreso en la segunda edición de su The Doctrine of Chances, de 1738, en el contexto de cierta aproximación de la distribución binomial para grandes valores de n. Su resultado fue ampliado por Laplace en 1812, y en la actualidad el resultado obtenido se llama Teorema de De Moivre-Laplace, teorema que veremos en el último punto del tema.

Laplace usó la distribución normal en el análisis de errores de experimentos. Gauss, que afirmaba haber usado el método desde 1794, lo justificó rigurosamente en 1809 asumiendo una distribución normal de los errores. El nombre de Gauss se ha asociado a esta distribución porque la usó con profusión cuando analizaba datos astronómicos, llegando a dar la expresión de la función de densidad de estas variables.

Esta distribución también se conoce como *"distribución de Gauss" o "campana de Gauss"*, ¿pillas ahora el jeroglífico del principio del tema?, y en el siguiente apartado verás el porqué. El nombre de "campana" viene de Esprit Jouffret que usó el término "bell



Abraham De Moivre.
Imagen en Wikimedia Commons. Dominio Público

surface" (superficie campana) por primera vez en 1872 para una distribución normal.

Algunos autores le atribuyen a Gauss un descubrimiento de esta distribución independiente del de De Moivre. Esta atribución del nombre de la distribución a una persona distinta de su primer descubridor es un claro ejemplo de la Ley de Stigler.

El nombre de "distribución normal" fue otorgado independientemente por Charles S. Peirce, Francis Galton y Wilhelm Lexis hacia 1875.
Galton llegó a decir de esta distribución, *"Si los griegos la hubieran conocido la habrían adorado como a un Dios"*

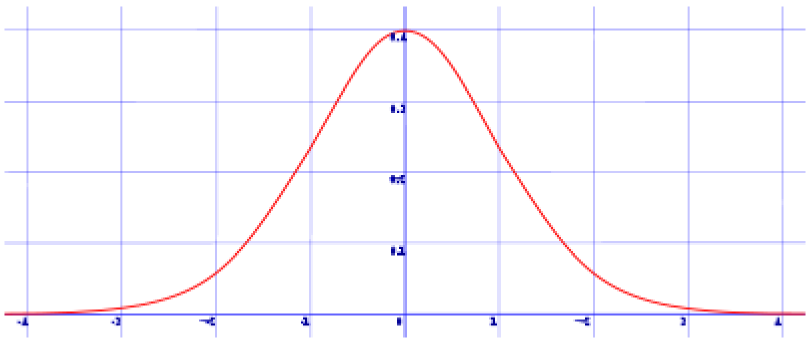
3.1. Definición y características

Y ahora ya sí, ha llegado el momento de conocer con todo lujo de detalles a esta distribución continua:

Una variable aleatoria continua **X** sigue una distribución **normal** de **media μ** y **desviación típica σ**, y se representa **X ~ N(μ,σ)** si se cumple que su función de densidad viene dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

La gráfica de esta función es similar a esta:



Debido a la forma que adopta se le llama campana de Gauss.

Importante

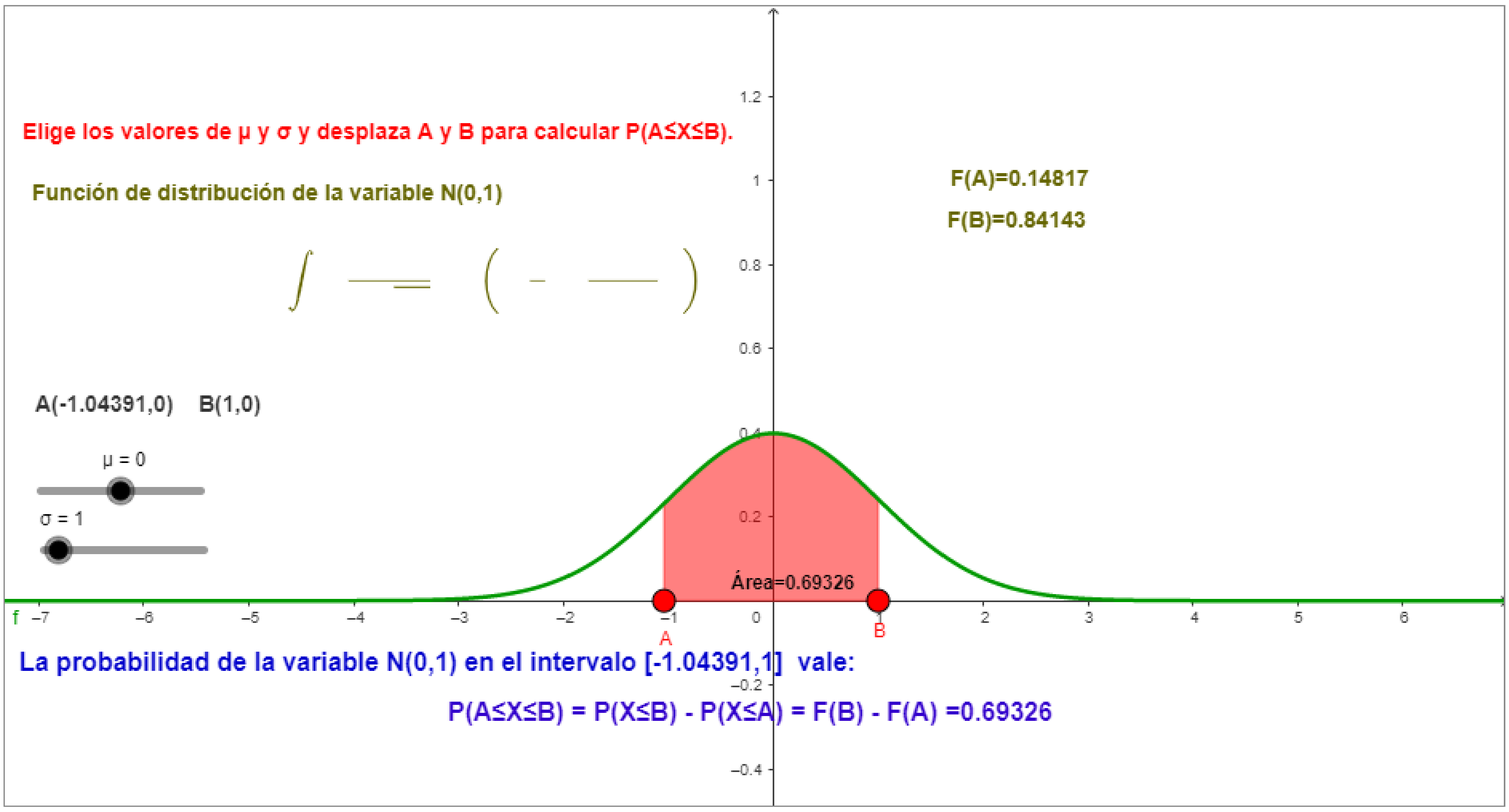
La función de densidad de una variable **aleatoria Normal** es la curva Normal o campana de Gauss. Su gráfica es una función continua, simétrica, cuyo máximo coincide con la **media de la distribución**.

Para cada valor μ de la media y cada valor σ de la desviación típica, existe una curva Normal y su distribución asociada que llamaremos $N(\mu,\sigma)$.

Para saber más

¿Te has recuperado ya del susto? Bueno, no te preocupes que tampoco es que tengamos que utilizar en la práctica, esa función para lo que nos interesa.
Ya hemos comentado que en estas distribuciones, la probabilidad de un intervalo se calcula cómo el área que hay bajo la curva dentro de ese intervalo.
Matemáticamente, esto se hace usando el concepto de integral, algo similar a la operación contraria de la derivada, pero para esta función es tan sumamente complicado que se utilizan otras herramientas para calcular estas probabilidades, aunque en el fondo de esos cálculos están las integrales.

En la siguiente escena, aparece representada la función de densidad de una distribución normal. En los controles "media (μ)" y "desv. tip (σ)" puedes variar el valor de los dos parámetros de la distribución normal y ver sombreada una parte de esa distribución. Además, dos puntos A y B que puedes mover libremente.



Escena de iedamatematicas en Geogebra. Licencia CC

Reflexiona

Contesta a las siguientes cuestiones manipulando la escena anterior:

1. ¿Qué efecto produce sobre la función de densidad aumentar la media? ¿Y disminuirla?
2. ¿Qué efecto produce sobre la función de densidad aumentar la desviación típica de la Normal? ¿Y disminuirla?
3. Si X es N(0,1), ¿Cuál es la probabilidad de que X esté entre 2,1 y 3,5?
4. Si X ~ N(0,1), calcula P(-3<X<0)
5. Si X ~ N(0,1), calcula P(X ≥ -1) (Aleja el punto B)
6. Si X ~ N(1,5 ; 0,6), calcula P(-2<X<2)
7. Si X ~ N(-0,5 ; 1,5), calcula P(-2<X<2)

Las soluciones deben haberte coincidido más o menos, decimal arriba decimal abajo con estas:

1. Al aumentar la media la función se desplaza a la derecha y al disminuirla a la izquierda.
2. Al aumentar la desviación típica, la función de densidad se achata, mientras que al aumentarla se hace más puntiaguda.
3. 0,0172
4. 0,4988
5. 0,8429
6. 0,7988
7. 0,7940

Curiosidad

Si te fijas bien en la gráfica de esta función, puedes ver que:
El dominio es el conjunto de todos los números reales.
Respecto a la media, la función es simétrica.
Tiene asíntota horizontal y = 0, tanto cuando x se va a -∞ como a +∞.
La función es creciente hasta que x llega al valor de la media y decreciente a partir de él. Por tanto, en x = μ la función alcanza su máximo.
Otra propiedad curiosa que cumple esta distribución es que las tres medidas de centralización coinciden, es decir, moda, mediana y media coinciden en el valor de μ.

Y ahora dirás, vale muy bien, pero en la escena anterior la media solo llega hasta 5, ¿y si tengo que resolver una cuestión en la que la media sea mayor? o, y si no tengo esa escena, ¿cómo calculo una probabilidad de una distribución normal?
Pues tienes toda la razón, nos faltan recursos.
Existen tablas donde están calculadas las áreas para distintos valores de x en una distribución normal, pero ojo, en una distribución N(0,1), pero como vamos a ver en el siguiente apartado, cualquier distribución normal se puede transformar en una del tipo N(0,1).

Caso práctico

Vamos a familiarizarnos antes de seguir con esta tabla y cómo hay que usarla. En el enlace, te puedes descargar la tabla de valores de la distribución Normal N(0,1). >> Documento de descarga

Lo primero que tienes que observar es que con esta tabla se calculan probabilidades del tipo P(Z≤a), donde Z representa la variable aleatoria que sigue la N(0,1) y "a" el valor extremo. Es decir, se calcula el área de la zona dibujada en azul.

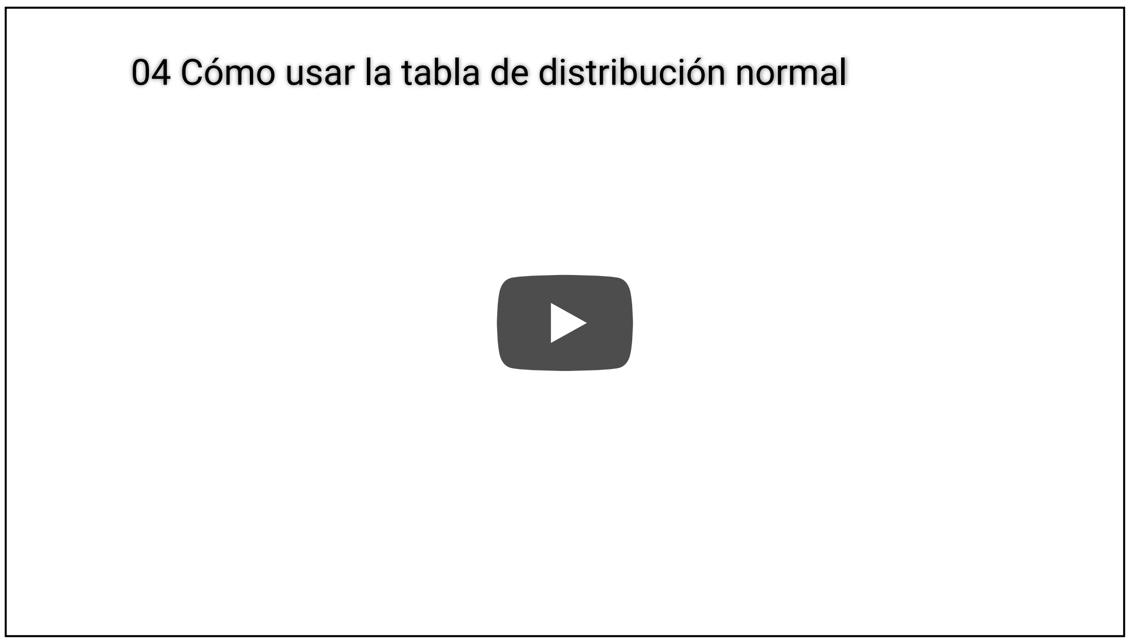
Por ejemplo, vamos a calcular la probabilidad de que si Z es una variable N(0,1), un valor tomado al azar sea menor que 1,37.

Para calcularla, en la columna donde pone z_0 hemos de buscar la fila en la que aparece el valor entero y la primera cifra decimal, es decir, 1,3. Una vez encontrado, nos vamos a las columnas y buscamos en el encabezamiento la segunda cifra decimal, o sea, donde aparezca 0,07.

El lugar donde se cruzan esas dos búsquedas me indica la probabilidad. En nuestro caso, aparece, 0,9147.

Luego P(Z ≤ 1,37) = 0,9147.

Si te ha quedado alguna duda, viendo el siguiente vídeo seguro que se te aclara.



Vídeo de Píldoras Matemáticas alojado en Youtube.

Comprueba lo aprendido

Calcula las siguientes probabilidades a partir de la tabla.

Usa la coma (,) para escribir los decimales.

P(Z ≤ 0,95) =

P(Z ≤ 3,14) =

P(Z ≤ 1) =

P(Z < 2,92) =

P(Z ≤ 3,07) =

3.2. Tipificación


Imagen de mohamed_hassanen Pixabay. Pixabay License

Ya que conocemos la tabla, imagino que algunas cuestiones se te habrán pasado por la cabeza:

¿Y si quiero calcular la probabilidad de que la variable sea mayor que un valor concreto?

¿Puede haber números negativos?

¿Y si queremos calcular la probabilidad de que esté entre dos valores?

¿Y si la media y la desviación típica no son 0 y 1 respectivamente?

Pues a todo esto vamos a responder en este apartado del tema, y vamos a empezar por la última pregunta, viendo cómo cualquier distribución normal se puede transformar en una Normal de media 0 y desviación típica 1.

Jorge Está equipando la cocina del piso que se ha comprado y entre otras cosas está ahora comparando lavavajillas. Entre los diverso folletos, está ojeando el de la marca electro-elec, y ahí ve la siguiente información sobre la vida útil del lavavajillas: "vida media: 14,6 años. Desviación típica: 0,5"

Jorge se dice; si suponemos que la vida útil del lavavajillas se distribuye según un modelo normal, ¿qué probabilidad tengo de que como mucho me dure 16 años?

Si definimos X = tiempo de vida del lavavajillas, tenemos que X ~ N(14,6 ; 0,5). Pues bien, si restamos el valor de la media, tenemos una nueva variable con media 0 y si dividimos entre la desviación típica, obtenemos otra nueva variable con desviación típica 1.

Así, la variable $Z=\frac{X-14,6}{0,5}$ sigue una distribución N(0,1).

Por tanto, para dar respuesta a la cuestión que se nos plantea, hacemos esa transformación en la variable y en el valor del que queremos determinar la probabilidad.

$$P(X\leq 16) \ = \ P(\frac{X-14,6}{0,5}\leq\frac{16-14,6}{0,5}) \ = \ P(Z\leq 2,8)$$

Para finalizar nos queda buscar 2,8 en la tabla >> Documento de descarga y obtenemos que dicha probabilidad es: 0,9974

Importante

Para transformar una variable aleatoria normal en Normal N(0,1), primero se le resta la media de la variable, y el resultado se divide entre la desviación típica:

$$X\sim N(\mu,\sigma)\rightarrow \ Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

A este proceso se le llama **tipificar** la variable.

Siguiendo con el lavavajillas que Jorge se va a comprar, ha visto que la probabilidad de que le dure menos de 16 años es muy grande, pero ¿y de que le dure más de 15?
Ahora la cuestión es que tiene que calcular la P(X > 15), o haciendo ya el proceso de tipificar (recuerda restar 14,6 y dividir entre 0,5), P(Z > 0,8) ¿Y cómo se hace esto, si en la tabla siempre viene para menor?
Pues muy fácil, el suceso contrario a Z > 0,8 es Z ≤ 0,8, y este suceso sí que se puede calcular a través de la tabla. Además, si recuerdas del primer tema del bloque, la probabilidad del suceso contrario es 1- probabilidad del suceso: $P(\overline{A}) \ = \ 1-P(A)$
Por tanto, P(Z > 0,8) = 1 - P(Z ≤ 0,8) = 1 - 0,7881 = 0,2119. Luego no debe confiar demasiado en que le dure más de 15 años.

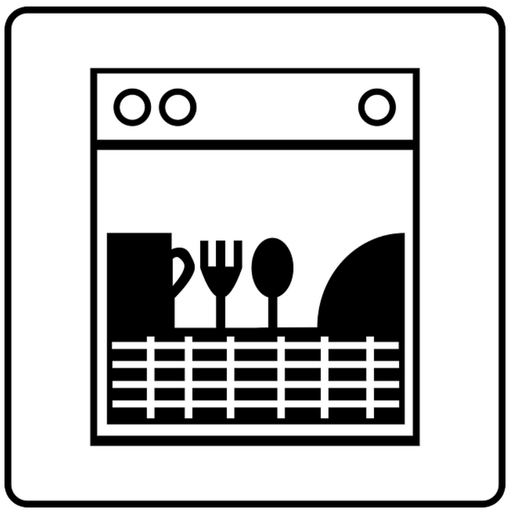


Imagen de OpenClipart-Vectors en Pixabay. Pixabay License

Siguiente cuestión. Ahora Jorge en vista del resultado que ha tenido, se pregunta un poco desolado, ¿será pequeña la probabilidad de que dure menos de 13 años?
Ahora habría que calcular la probabilidad P(X < 13), y haciendo el proceso de tipificar, ésta se transforma en P(Z < -3,2). Un número negativo. ¿Y si en la tabla sólo hay positivos?
Utilizaremos la probabilidad del suceso contrario: P(Z < -a) = 1 - P(Z < a).
Así que ya tenemos la respuesta a nuestra cuestión. P(X < 13) = P(Z< -3,2) = 1- P(Z < 3,2) = 1 - 0,99931 = 0,00069. Por tanto, la probabilidad de que el lavavajillas nos dure menos de 13 años es prácticamente nula, 0,069%.

Si queremos calcular la probabilidad de Z mayor que un número negativo, viendo nuevamente que es el suceso contrario al menor tendríamos que:
P(Z > -a) = 1 - P(Z < -a) = 1 - [1 - P(Z< a)] = P (Z < a).

La última cuestión. Calcular una probabilidad entre dos valores. ¿Cuál sería la probabilidad de que el lavavajillas durara más de 14 años pero menos de 16?
Como podemos observar en la escena de Geogebra del apartado anterior: P(a < Z < b) = P(Z < b) - P(Z < a).
En el lavavajillas de Jorge, tendríamos que P(14 < X < 16) = P(X < 16) - P (x < 14). Tipificando cada valor, esta probabilidad es igual a:
$P(Z<\frac{16-14,6}{0,5})-P(Z<\frac{14-14,6}{0,5}) \; = \; P(Z<2,8)-P(Z<-1,2)$
El primer valor se obtiene de la tabla y el segundo como es negativo, hacemos 1 - P(Z < 1,2).
Así, el resultado es: 0,9974 - (1 - 0,8849) = 0,9974 - 0,1151 = 0,8464. Luego hay una probabilidad bastante alta de que el lavavajillas dure entre 14 y 16 años.

Importante

Cálculo de probabilidades en una distribución Normal.

Si "**a**" es un número positivo y **Z** sigue una distribución **N(0,1)**:

- P(Z < a) → A partir de la tabla de probabilidades de una distribución N(0,1) >> Documento de descarga
- P(Z > a) = 1 - P(Z < a)
- P(Z < -a) = 1 - P(Z < a)
- P(Z > -a) = P (Z < a)
- P(a < Z < b) = P(Z < b) - P(Z < a)

Comprueba lo aprendido

Ahora te toca a ti. Calcula las siguientes probabilidades usando las propiedades vistas arriba y la tabla de probabilidades de la distribución Normal N(0,1): >> Documento de descarga

1. Z sigue una distribución N(0,1). Calcula:

    1. P(Z ≥ 0,32)
    2. P(Z ≤ 0)
    3. P (Z > 0,7)
    4. P (-0,51 ≤ Z ≤ 0,51)
    5. P (Z > - 2,63)

2. Si X es N(5,2), ¿cuál es la probabilidad de que X sea mayor o igual que 13,5?

Las soluciones deben haberte salido estas:

1.

    1. P(Z ≥ 0,32) = 1 - P(Z < 0,32) = 1 - 0,6255 = 0,3745
    2. P(Z ≤ 0) = 0,5
    3. P(Z > 0,7) = 1 - P (Z < 0,7) = 1 - 0,7580 = 0,242
    4. P(-0,51 ≤ Z ≤ 0,51) = P(Z ≤ 0,51) - P(Z < -0,51) = P(Z ≤ 0,51) - [1 - P(Z < 0,51)] = 0,695 - (1 - 0,695) = 0,39
    5. P(Z > - 2,63) = P(Z < 2,63) = 0,9957

2. P(X ≥ 13,5) = $P(Z \geq \frac{13,5-5}{2})$ = P (Z ≥ 4,25) = 1 - P (Z < 4,25) = 1 - 1 = 0.

Fíjate que en la tabla el último valor que aparece es 3,99 con una probabilidad 0,99997. Cualquier valor de Z mayor que 3,99 podemos suponer que tiene ya probabilidad 1.

Caso práctico

Imagen en Flickr de Diego López bajo CC

El precio del kilogramo de gambas frescas en el mercado de abastos de Huelva sigue una distribución Normal de media 16.5 euros y desviación típica 2 euros.

Responde a las siguientes cuestiones:

a) ¿Cuál es la probabilidad de que un kilogramo de gambas nos cueste menos de 15 euros?

Llamemos X = "precio del kg de gambas frescas (en euros)"

Sabemos que $X \sim N(16.5\,,\,2)$

Pero esta variable X hay que tipificarla y convertirla en una nueva variable $Z \sim N(0\,,\,1)$, es decir, en una Normal estándar.

Para ello, restamos a X el valor de la media (en este caso $\mu = 16.5$), así tendremos una nueva variable con media $\mu = 0$ (16.5-16.5=0) y luego dividimos entre la desviación típica (en este caso $\sigma = 2$) y obtendremos otra nueva variable con desviación típica $\sigma = 1$ (2/2=1).

Así, a partir de X, obtendremos la variable $Z = \frac{X-16.5}{2}$ que sigue una distribución $N(0\,,\,1)$.

Por tanto, para dar respuesta a la cuestión hacemos esta transformación tanto a la variable como al valor del que queremos determinar su probabilidad. Vamos estableciendo una cadena de igualdades (haciendo uso de las propiedades vistas en el apartado anterior) hasta llegar a la probabilidad buscada.

$$p(X<15) \;=\; p\left(\frac{X-16.5}{2}<\frac{15-16.5}{2}\right) = p(Z<-0.75) = p(Z>0.75) =$$
$$= 1-p(Z \leq 0.75) = 1-0.7734 = 0.2266$$

La $p(Z \leq 0.75)$ la hemos obtenido de la tabla de la $N(0\,,\,1)$.

Por tanto, la probabilidad de que un kilogramo de gambas frescas nos cueste menos de 15 euros es 0.2266.

(Interpretación: Hay un 22.66 % de probabilidad que 1 kg de gambas nos cueste menos de 15 €)

b) ¿Y la probabilidad de que nos cueste entre 15 y 17 euros?

Seguiremos el mismo esquema empleado en la anterior cuestión.

Llamemos X= "Precio del kg de gambas frescas (en euros)"

Sabemos que $X \sim N(16.5\,,\,2)$ y, por tanto, $Z = \frac{X-16.5}{2} \sim N(0\,,\,1)$

Por tanto, para dar respuesta a la cuestión hacemos esta transformación tanto a la variable como a los valores entre los que queremos determinar su probabilidad. Vamos estableciendo una cadena de igualdades (haciendo uso de las propiedades vistas en el apartado anterior) hasta llegar a la probabilidad buscada.

$$p(15<X<17) = p(X<17) - p(X<15)$$

$$p(X<17) = p\left(\frac{X-16.5}{2}<\frac{17-16.5}{2}\right) = p(Z<0.25) = 0.5987$$

$$p(X<15) \;=\; p\left(\frac{X-16.5}{2}<\frac{15-16.5}{2}\right) = p(Z<-0.75) = p(Z>0.75) =$$
$$= 1-p(Z \leq 0.75) = 1-0.7734 = 0.2266$$

Las probabilidades $p(Z<0.25)$ y $p(Z<0.75)$ las hemos obtenido de la tabla de la $N(0\,,\,1)$.

Por tanto:

$p(15 < X < 17) = p(X < 17) - p(X < 15) = 0.5987 - 0.226 = 0.3721$

De lo anterior deducimos que la probabilidad de que un kilogramo de gambas frescas nos cueste entre 15 y 17 € vale 0.3721.

(Interpretación: Hay un 37.21 % de probabilidad de que 1 kg de gambas nos cueste entre 15 y 17 €)

c) ¿Y la probabilidad de que un kilogramo nos cueste más de 17 euros?

Llamemos X = "Precio del kg de gambas frescas (en euros)"

Sabemos que $X \sim N(16.5\,,\,2)$, y por tanto, $Z = \frac{X - 16.5}{2} \sim N(0\,,\,1)$

Por tanto, para dar respuesta a la cuestión hacemos esta transformación tanto a la variable como a los valores entre los que queremos determinar su probabilidad. Vamos estableciendo una cadena de igualdades (haciendo uso de las propiedades vistas en el apartado anterior) hasta llegar a la probabilidad buscada.

$$p(X > 17) = p\left(\frac{X - 16.5}{2} > \frac{17 - 16.5}{2}\right) = p(Z > 0.25) = 1 - p(Z \leq 0.25) =$$
$$= 1 - 0.5987 = 0.4013$$

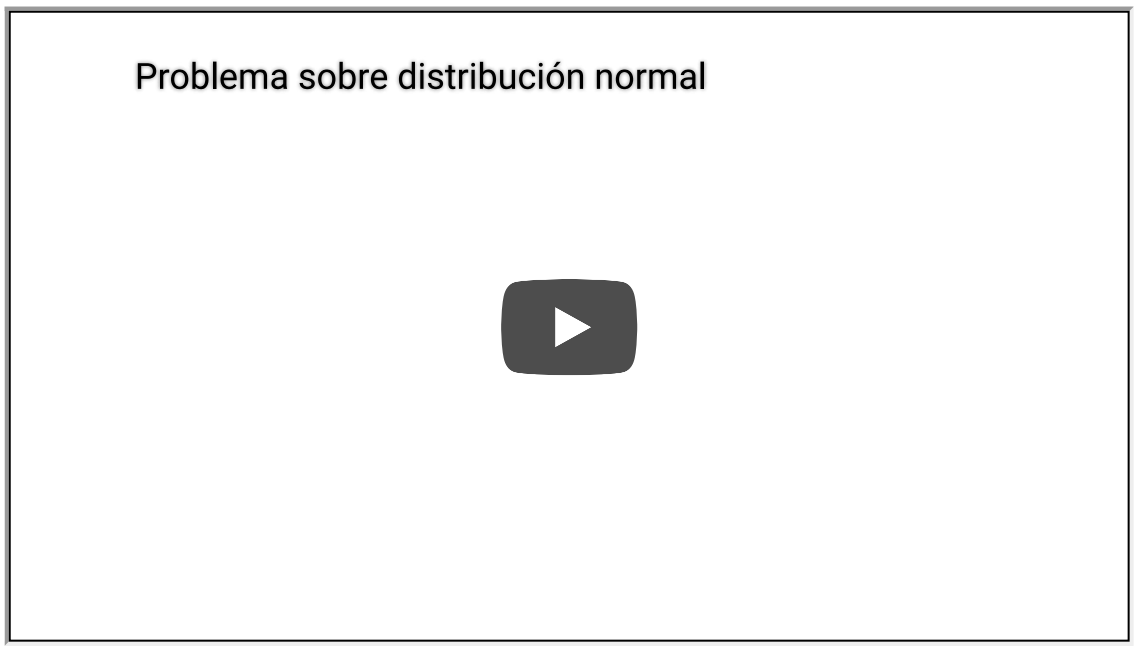
La $p(Z < 0{,}25)$ la hemos obtenido de la tabla de la $N(0\,,\,1)$.

La probabilidad de que un kilogramo de gambas frescas nos cueste más de 17 € vale $0{,}4013$.

(Interpretación: Hay un 40,13 % de probabilidad que 1 kg de gambas nos cueste más de 17 €)

4. Aplicaciones

La distribución normal como ya hemos comentado se ajusta a un montón de situaciones y casos concretos de la naturaleza, la economía, la sociología,...
En este apartado te presentamos unas cuantas situaciones que se ajustan y se resuelven aplicando el modelo normal.
En primer lugar, en este vídeo tienes un primer ejemplo en el que la distribución normal se aplica al estudio del peso de una población:



Problema de probabilidad sobre la distribución normal.
Vídeo de lasmatematicas.es alojado en Youtube

Caso práctico

Imagen de OpenClipart-Vectors en Pixabay. Pixabay License

Ahora aplicamos la distribución normal al estudio de las capacidades intelectuales, pues uno de los indicadores más fiables, el cociente intelectual se distribuye según una ley normal:
Las puntuaciones de cociente intelectual (IQ) están distribuidas normalmente con una media de 100 y una desviación típica de 15. Mensa es una organización para personas con cociente intelectual elevado y sólo acepta a personas con un IQ mayor que 131,5. Si se elige una persona al azar, ¿qué probabilidad hay de que satisfaga el requisito que exige Mensa? Si en una ciudad viven 75000 personas, ¿cuántos podrían entrar en Mensa?

Si definimos X = puntuación obtenida en el test IQ, parece claro que la probabilidad que tenemos que calcular es P(X > 131,5)

En primer lugar, tipificamos, pues X ~ N(100 , 15):

$$P(X>131{,}5)=P(Z>\frac{131{,}5-100}{15})=P(Z>2{,}1)$$

Como tenemos que calcular la probabilidad de Z mayor que un número positivo y aplicamos la regla vista en el punto anterior: 1 - probabilidad de Z menor que ese número:

$$P(Z>2{,}1)=1-P(Z<2{,}1)$$

Buscamos en la tabla el valor 2,1 o lo que es lo mismo, 2,10.

$$P(Z>2{,}1)=1-P(Z<2{,}1)=1-0{,}9821=0{,}0179$$

Luego la probabilidad de que una persona cumpla el requisito para entrar en Mensa es 0,0179, o expresando el resultado en tanto por ciento, 1,79%

Para la segunda pregunta, lo único que hay que hacer es multiplicar la probabilidad de que una persona cumpla el requisito por el número de personas que viven en esa ciudad:

Número de personas que cumplen el requisito = 0,0179 · 75000 = 1342,5. O sea, se espera que 1342 personas de esa ciudad puedan entrar en Mensa.

Caso práctico

En este ejemplo hablamos ahora de colesterol. La distribución normal tiene muchísimas aplicaciones en la medicina pues muchos indicadores se distribuyen según un modelo normal. En este ejemplo, vemos que el nivel de colesterol se distribuye según una normal.
En una ciudad de 300.000 habitantes el nivel de colesterol en sangre sigue una distribución normal de media 192mg/dl y varianza 144$(mg/dl)^2$.

a) ¿Qué probabilidad hay de que una persona elegida al azar en esa ciudad, tenga un nivel de colesterol comprendido entre 186 y 200 mg/dl ?

b) ¿Número de personas con nivel de colesterol elevado, considerándose colesterol elevado si el nivel que presenta es superior a 235?

c) ¿Qué nivel de colesterol podemos decir que supera 84,2% de la población adulta sana?

- Primero a partir de la varianza calculamos la desviación típica

$\delta = \sqrt{144} = 12$

- Luego tipificamos y buscamos la probabilidad en la tabla normal

$$X \sim N(192, 12) \rightarrow Z = \frac{X-\mu}{\delta} \sim N(0,1)$$

a) Calcular la probabilidad de que una persona elegida al azar tenga un nivel de colesterol comprendido entre 186 y 200.

$P(186 < X < 200) = P[(186-192)/12 < Z < (200-192)/12)] = P(-0{,}5 < Z < 0.67) =$

$P(Z < 0.67) - P(Z < -0.5) = P(Z < 0.67) - P(Z > 0.5) =$

$P(Z < 0.67) - [1- P(Z < 0.5)] = 0{,}7486-(1-0{,}6915) =$ **0,4401**

Se corresponde con un porcentaje de 44,01%

b) El número de personas con nivel de colesterol elevado, considerándose colesterol elevado si es superior a 235 y por último un nivel de colesterol de forma que el 84,2% de la población tenga un nivel superior a ese.

$P(X \geq 235) = P[Z \geq (235-192)/12] = P(Z \geq 3{,}58) = 1-P(Z \leq 3{,}58) = 1-0{,}99983 = 1{,}7 \cdot 10^{-4}$

$=0{,}00017$

Número de personas: 0.00017·300000= **51**

c) Y por último un nivel de colesterol de forma que el 84,2% de la población tenga un nivel superior a ese.

P(X ≥ **k**)=0.842

Busco en la tabla P(Z ≤ **a**)=0.842 y por la simetría de la gráfica P(Z ≥ -a)=0.842.

para averiguar la **a** usamos la tabla de manera inversa a como la hemos usado en los otros casos

z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z_0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	$\boxed{0,8438}$	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4

El valor mas próximo a 0,842 es 0,8438 y la **a** será igual a: 1,01

Ahora para calcular el valor de la variable X, "deshacemos la tipificación" : P(Z ≥ -a)=0.842

-1,01=(k-192)/12 y resolvemos la ecuación: K= -1,01·12+192

k=179,88.

Luego el nivel de colesterol que supera el 84,2% de la población es **179,88 mg/dl** .

Comprueba lo aprendido

Resuelve tú ahora las siguientes cuestiones. Procede de forma similar a como lo hemos hecho en los ejemplos resuelto: en primer lugar piensa qué probabilidad hay que calcular, después tipifica y transforma la probabilidad para que puedas buscarla en la tabla.
El tiempo de vida de una bombilla sigue una distribución normal N(180, 15), donde el tiempo se mide en horas. ¿Cuál es la probabilidad de que al comprar una bombilla, luzca más de 207 horas?

- 0,8810
- 0,119
- 0,9641
- 0,0359

Repasa los cálculos al tipificar y el resultado en la tabla.

Repasa los cálculos al tipificar y el resultado en la tabla.

Tienes que calcular la probabilidad de Z mayor que un positivo. Repasa la regla.

Correcto

Solución

1. Incorrecto
2. Incorrecto

3. Incorrecto
4. Opción correcta

Un laboratorio farmacéutico prepara pastillas circulares con un diámetro medio de 12 mm y una desviación típica de 0,8 mm, pero si la pastilla fabricada tiene un diámetro inferior a 9,5 mm o superior a 14,7 mm, ésta se rechaza por no tener la cantidad adecuada de medicamento. Sabemos además que el diámetro sigue una distribución normal. ¿Cuál es la probabilidad de que al fabricar una pastilla, ésta esté en condiciones de ser utilizada?

Sugerencia

- 0,99872
- 0,99877
- 0,00051
- 1,00051

No has redondeado bien los resultados al tipificar.

Correcto

Aplica bien las reglas de cálculo de probabilidades de la normal.

¿Una probabilidad mayor que 1?

Solución

1. Incorrecto
2. Opción correcta
3. Incorrecto
4. Incorrecto

Para saber más

Como ya hemos visto, la distribución normal es una distribución de probabilidad simétrica respecto a la media, concentrando la máxima probabilidad en torno también a la media μ.

Pues bien sean cuales sean son parámetros de la distribución normal, siempre se cumple que los siguientes intervalos centrados en la media tienen las probabilidades que aparecen:

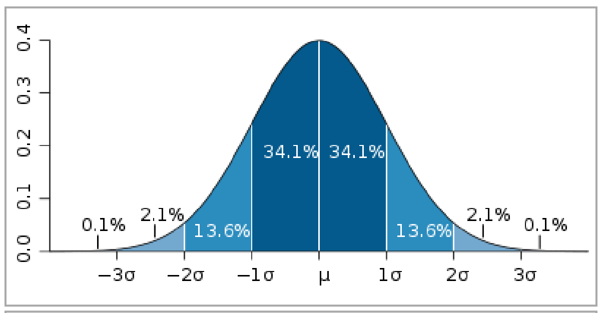


Imagen en Wikimedia Commons bajo CC

- En el intervalo [μ - σ, μ + σ] se encuentra comprendida, aproximadamente, el 68,26% de la distribución. Es decir, la probabilidad de que X pertenezca a ese intervalo es del 68,26%.
- En el intervalo [μ - 2σ, μ + 2σ] se encuentra, aproximadamente, el 95,44% de la distribución.
- Por su parte, en el intervalo [μ -3σ, μ + 3σ] se encuentra comprendida, aproximadamente, el 99,74% de la distribución.

Estas propiedades son de gran utilidad para el establecimiento de intervalos de confianza como veremos en el próximo curso. Por otra parte, el hecho de que prácticamente la totalidad de la distribución se encuentre a tres desviaciones típicas de la media justifica los límites de las tablas empleadas habitualmente en la normal estándar.

5. Aproximación de la distribución binomial por la normal

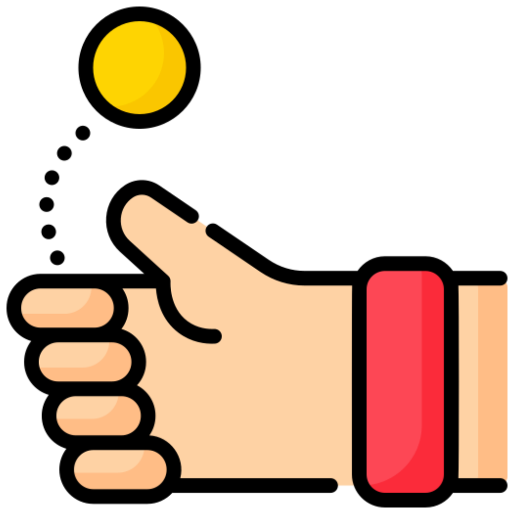


Imagen de Freepik en flaticon. Términos de uso.

¿Te acuerdas de la binomial? Sí, el tema anterior. El modelo en el que se repetía una serie de veces un experimento, había solamente dos posibles resultados, éxito y fracaso, y contábamos el número de éxitos.
Por ejemplo, si lanzabas 8 veces una moneda y contabas el número de caras, eso era una binomial con n = 8 y p, probabilidad de éxito en un lanzamiento igual a 0,5;
X= n.° de caras → B(8; 0,5)
Si sigue recordando, la fórmula para calcular la probabilidad de un valor en este modelo era:

$$P[X=k]=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$$

Esta fórmula era un poco pesada de aplicar y si teníamos que calcular la probabilidad de que X fuese mayor que un cierto valor o menor, había que aplicarla varias veces.
¿Te imaginas que repetimos la prueba 50 veces y queremos calcular la probabilidad de que al menos obtengamos 20 caras? Al menos 20 significa 20 o más, o sea, que tendríamos que calcular la probabilidad de X=20, de X=21, de X=22,..., y así hasta X=50. ¡Casi nada!
Pues bien, cuando hacemos que el tamaño de la muestra o de las repeticiones sea suficientemente grande, el modelo binomial se puede aproximar al normal, pero claro, ¿a qué normal? Pues al que tiene como media la media de esa distribución binomial y como desviación típica la desviación típica del modelo binomial. Y entonces todos esos cálculos se van a facilitar.
Pero otra cuestión que hay que tener en cuenta es el valor del parámetro "p". Repite lo anterior con "p" muy pequeño (menor que 0,1) y después con "p" muy próximo a 1 (entre 0,9 y 1). Ya no se parece tanto, ¿verdad?
Luego tenemos que fijar un criterio para considerar que la aproximación es buena, y ese criterio es que "n" tiene que ser suficientemente grande, y por ello entendemos al menos 30 y que "p" no sea ni muy grande ni muy chico, y para eso, hemos de comprobar que n·p sea por lo menos 5 y n·(1-p) también.

Importante

Si X es una variable aleatoria discreta que sigue un modelo binomial de parámetros n y p (X→ B(n , p)), X se puede aproximar a un modelo normal de parámetro n·p y $\sqrt{n\cdot p\cdot(1-p)}$ si se cumplen las condiciones para "n" y "p":

- n ≥ 30
- n·p ≥ 5 y n·(1-p) ≥ 5

Si se cumple esto, la variable quedaría aproximada por este modelo normal:

$$X \to N(np, \sqrt{npq})$$

A este resultado se le conoce como Teorema de De-Moivre

Caso práctico

¿Recuerdas a Jorge?, pues el próximo puente lo va a pasar de turismo en Sevilla y de camino probará suerte en el Gran Casino Aljarafe, aunque lo de Jorge más que probar es ganar casi seguro.
Con suficiente antelación hizo su reserva de hotel para los tres días que iba a pasar en la capital de andaluza.
Los hoteleros son conscientes de que mucha gente reserva plaza por si acaso, pero que después cuando llegue el momento no van a hacer uso de ella y le van a dejar colgada la habitación. Manejan la cifra de que aproximadamente el 15% de las personas que reservan una plaza, luego no aparecen.
Pues bien, en el hotel donde Jorge reservó la habitación se han aceptado 104 reservas para ese fin de semana, pese a que el hotel sólo dispone de 97 habitaciones. ¿Qué probabilidad hay de que Jorge se quede sin habitación y tenga que buscar otro hotel?

Para que Jorge se quede sin habitación, tiene que suceder que al menos vayan 97 personas de las que tenían hecha la reserva, así que, comenzamos definiendo la variable aleatoria X = n.° de personas que cumplen la reserva y van al hotel. Así, lo que tenemos que calcular es la probabilidad de que X sea igual o mayor que 97.

Para cada persona, hay dos posibles resultado, o va o no va al hotel, puesto que estamos interesados en contar las personas que van, consideramos que el éxito es que la persona va al hotel y el fracaso, que no va.

Así, nuestra variable X sigue un modelo binomial en el que n = 104, que son las personas que hay con reserva y p = 0,85; pues si el 15% no va, el 85% por tanto, sí lo hará.

X ~ B(104; 0,5) y hay que calcular P(X≥92)

Puesto que "n" es bastante grande, vamos a ver si podemos hacer la aproximación al modelo normal:

- n = 104, y por tanto, mayor que 30
- n·p = 104 · 0,85 = 88,4. Mayor que 5
- n·(1-p) = 104 · 0,15 =15,6.Mayor que 5

Por tanto, como se cumplen las condiciones, podemos suponer que X sigue una normal de media 104 · 0,85 y de desviación típica la raíz cuadrada de 104 · 0,85 · 0,15;

$$X \sim N(104{\cdot}0{,}85\ ;\ \sqrt{104{\cdot}0{,}85{\cdot}0{,}15})$$

$$X \sim N(88{,}4\ ;\ 3{,}64)$$

Y ya podemos calcular la probabilidad partiendo de esa distribución normal como en los apartados anteriores:

$$P(X \geq 97)$$

$$P(Z \geq \frac{97-88{,}4}{3{,}64})$$

$$P(Z \geq 2{,}36)$$

$1-P(Z<2{,}36)$

$1-0{,}9901\ =\ 0{,}0091$

Luego la probabilidad de que Jorge se quede sin habitación es mínima, ni un 1%; 0,91%

Comprueba lo aprendido

¿Hace mucho que aprobaste el carnet de conducir? ¿Te acuerdas del teórico? Había que contestar 40 preguntas tipo test y en cada pregunta había 4 posibles respuestas de las que sólo una era verdadera y para aprobar podías tener como máximo 4 fallos. Supongo que cuando lo hiciste estudiaste bastante el código, pero, ¿te preguntaste alguna vez que posibilidades hay de aprobar sin estudiar absolutamente nada, o lo que es lo mismo, contestando al azar?

Pues ahora lo vas a calcular. Ve eligiendo la respuesta correcta en las siguientes cuestiones:

Imagen de AnnaliseArt en Pixabay. Pixabay License

1) La variable a definir será:

- X = Número de preguntas acertadas
- X = Número de veces que me presento al examen.

Claro

No creo que influya mucho

Solución

1. Opción correcta
2. Incorrecto

2) Está claro que esa variable sigue un modelo binomial, pues en cada pregunta o aciertas o fallas. n será 40, pues hay 40 preguntas, pero ¿y p?

- 0,5
- 0,25
- 0,75

No parece que sea igual de probable acertar que fallar.

Exacto, la probabilidad de acertar una pregunta es una de cuatro.

Piénsalo mejor.

Solución

1. Incorrecto
2. Opción correcta
3. Incorrecto

3) Este modelo binomial puede aproximarse a uno normal

Sugerencia

- Sí, y la aproximación es N(10 ; 2,74)
- Sí, y la aproximación es N(10; 7,5)
- No

Correcto

Repasa los cálculos

Repasa las condiciones que se tienen que cumplir

Solución

1. Opción correcta
2. Incorrecto
3. Incorrecto

4) La probabilidad que hay que calcular es:

- $P(X > 4)$
- $P(X = 36)$
- $P(X \leq 36)$
- $P(X \geq 36)$

Vuelve a ver cómo hemos definido X y vuelve a leer el enunciado.

Valdría pero hay más posibilidades.

No es correcto. Así puedes suspender.

Correcto

Solución

1. Incorrecto
2. Incorrecto
3. Incorrecto
4. Opción correcta

5) La probabilidad de aprobar entonces es:

- 0,25
- 0,00738
- 0,0378
- 0

Incorrecto

Incorrecto

Incorrecto

Exacto, es imposible que apruebes contestando al azar

Solución

1. Incorrecto
2. Incorrecto
3. Incorrecto
4. Opción correcta

Para saber más

Corrección por continuidad

Como ya sabes, la distribución binomial es discreta, y por tanto, tiene sentido calcular probabilidades puntuales (P[X=a]), mientras que en la normal, al ser continua esto carece de sentido.
La aproximación de una variable discreta X por una continua a la que llamaremos X', genera un cierto error que se corrige modificando el intervalo cuya probabilidad se quiere calcular. Estas son situaciones y correcciones posibles:

- $P(X=a) = P(a - 0{,}5 \le X' \le a + 0{,}5)$
- $P(X \le a) = P(X' \le a + 0{,}5)$
- $P(X < a) = P(X' \le a - 0{,}5)$
- $P(X > a) = P(X' \ge a + 0{,}5)$
- $P(X \ge a) = P(X' \ge a - 0{,}5)$

A esta corrección se le conoce como corrección de Yates, debido a que su autor fue el matemático inglés Frank Yates (1902-1994)

Resumen

Importante

Una **variable aleatoria** es **continua** si al realizar el experimento aleatorio, entre cada dos valores, el número de valores que puede tomar es infinito.

Por ejemplo, la altura de una persona, la longitud del dedo índice, el peso de un perro, el caudal de un río...

Si X es una variable aleatoria continua, la probabilidad de que tome un valor concreto es cero.

P[X = a] = 0, para cualquier valor de a.

Importante

En una variable aleatoria continua, las probabilidades que calculamos están siempre asociadas a intervalos.
Si X es una variable aleatoria continua, la probabilidad de que X esté en un intervalo es el área del recinto limitado por el intervalo y la **función de densidad** f **(o función de probabilidad)**.
f debe ser siempre mayor o igual que cero y el área debajo de su gráfica tiene que ser 1.
En el caso de tener una variable aleatoria continua X, ya hemos visto que las probabilidades de valores concretos son 0. Así, al describir un intervalo, no influye que la desigualdad sea estricta o no para calcular la probabilidad.

P (X ≤ a) = P (X < a)

P (a< X < b) = P (a ≤ X < b) = P (a < X ≤ b) = P (a ≤ X ≤ b)

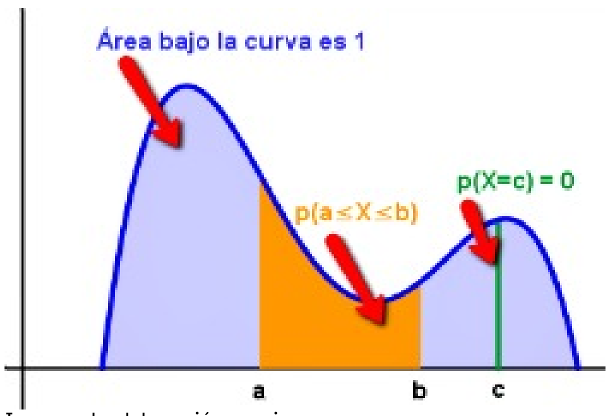


Imagen de elaboración propia

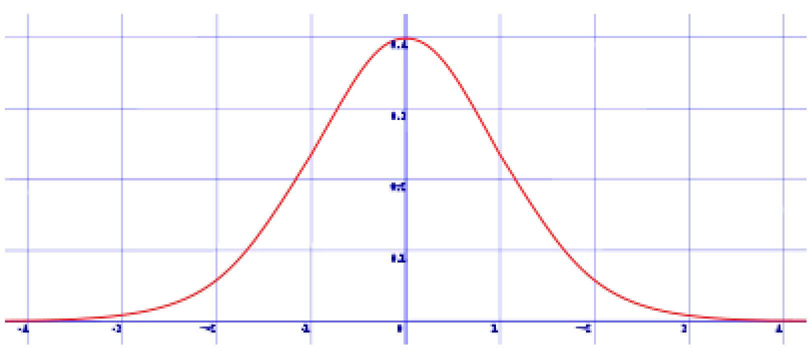
Se define la **función de distribución, F(x)**, como: **F(x)= P [X≤ x]**

Importante

Una variable aleatoria continua **X** sigue una distribución **normal** de **media μ** y **desviación típica σ**, y se representa **X ~ N(μ,σ)** si se cumple que su función de densidad viene dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

La gráfica de esta función es similar a esta:



Debido a la forma que adopta se le llama campana de Gauss

La función de densidad de una variable **aleatoria Normal** es la curva Normal o campana de Gauss. Su gráfica es una función continua, simétrica, cuyo máximo coincide con la **media de la distribución**.

Para cada valor μ de la media y cada valor σ de la desviación típica, existe una curva Normal y su distribución asociada que llamaremos $N(\mu,\sigma)$.

Importante

Para transformar una variable aleatoria normal en Normal N(0,1), primero se le resta la media de la variable, y el resultado se divide entre la desviación típica:

$$X \sim N(\mu,\sigma) \rightarrow \ Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

A este proceso se le llama **tipificar** la variable.

Cálculo de probabilidades en una distribución Normal.

Si "**a**" es un número positivo y **Z** sigue una distribución **N(0,1)**:

- P(Z < a) ⟶ A partir de la tabla de probabilidades de una distribución N(0,1) >> Documento de descarga
- P(Z > a) = 1 - P(Z < a)
- P(Z < -a) = 1 - P(Z < a)
- P(Z > -a) = P (Z < a)
- P(a < Z < b) = P(Z < b) - P(Z < a)

Importante

Si X es una variable aleatoria discreta que sigue un modelo binomial de parámetros n y p (X⟶ B(n , p)), X se puede aproximar a un modelo normal de parámetro n·p y $\sqrt{n\cdot p\cdot(1-p)}$ si se cumplen las condiciones para "n" y "p":

- n ≥ 30
- n·p ≥ 5 y n·(1-p) ≥ 5

Si se cumple esto, la variable quedaría aproximada por este modelo normal:

$$X \rightarrow N(np, \sqrt{npq})$$

A este resultado se le conoce como Teorema de De-Moivre
