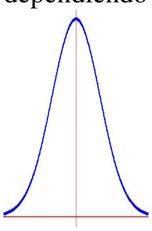
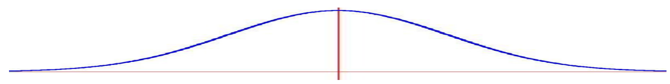
Cuestiones básicas sobre la distribución normal de probabilidad

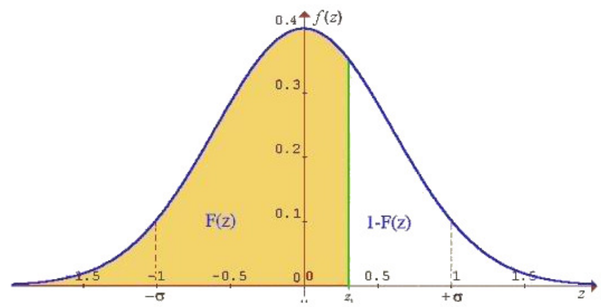
1. **¿En qué consiste la distribución normal y cómo se emplea?**
Se trata del ejemplo más conocido de variable aleatoria continua. Se ha comprobado que en infinidad de casos, determinadas distribuciones de probabilidad aplicadas a un gran número de casos tienden gráficamente a este perfil. Su función densidad tiene forma acampanada simétrica respecto de la vertical que pasa por su punto más alto y, por supuesto, el área definida bajo la curva con el eje X es 1. En estas condiciones hay una infinidad de posibilidades dependiendo del valor de x de su punto máximo y de su 'estilización'. Modificar la abscisa de su punto máximo es muy simple y consiste en una traslación a izquierda o derecha. Esto está directamente relacionado con la media μ de la variable. Estirar verticalmente la campana, manteniendo su área 1, está directamente relacionado con la desviación típica σ de la variable. Cuanto más estilizada, menor será la desviación típica, como se ve en la campana de la izquierda, y cuanto más aplanada, mayor será la desviación típica, como se ve en la campana de la derecha.





Todo ello se simboliza así: $N(\mu, \sigma)$

Los libros suelen disponer una tabla de la $N(0, 1)$, esto es, cero de media y 1 de desviación típica. A la variable aleatoria con esta distribución se le reserva la letra Z , mientras que a las restantes distribuciones normales $N(\mu, \sigma)$ se las nombra con la X.



La tabla de $N(0, 1)$ y algunas calculadoras nos dan el valor del área bajo la campana desde el extremo izquierdo hasta un concreto valor z de la variable aleatoria. Esto es: $F(z) = P[-\infty < Z \le z]$ que aparece sombreado en el gráfico de la derecha. A esta función densidad se le denomina $F(z)$. Para valores de z superiores a 4, se aproxima el área con 1.

Cualquier otro cálculo se realiza utilizando estos recintos, la simetría de la gráfica y el 1 del área total. Por ejemplo:

$P[Z > z] = 1 - F(z) \qquad P[z_1 < Z \le z_2] = F(z_2) - F(z_1)$

Si, como es habitual, la tabla no dispone de valores negativos de z, aplicando la simetría de la campana se pueden utilizar:

$P[Z \le -z] = 1 - F(+z) \quad P[Z > -z] = F(+z) \quad P[-z_1 < Z \le -z_2] = F(+z_1) - F(+z_2)$

$P[-z_1 < Z \le z_2] = F(z_2) - (1 - F(+z_1))$

También con la tabla, o con una calculadora que disponga de ello, se puede hacer una 'búsqueda inversa'. Esto es, conocida la probabilidad, hallar la abscisa correspondiente.

2. **¿Cómo se calculan las probabilidades en una distribución normal que no sea la $N(0, 1)$?**
Para trabajar con una $N(\mu, \sigma)$, salvo que dispongamos de una calculadora en la que se pueda definir cualquier media y desviación típica, tendremos que convertir la variable para poder usar los valores de la $N(0, 1)$. Esta conversión se denomina *Tipificación* y consiste en la fórmula: $Z = \dfrac{X - \mu}{\sigma}$. En ella, X representa la $N(\mu, \sigma)$ y Z la $N(0, 1)$.

De esta forma: $P[X \le x] = P\left[\dfrac{X-\mu}{\sigma} \le \dfrac{x-\mu}{\sigma}\right] = P\left[Z \le \dfrac{x-\mu}{\sigma}\right] = F\left(\dfrac{x-\mu}{\sigma}\right)$

3. **¿Influye en los cálculos de probabilidad que las desigualdades sean estrictas o no lo sean?**
Es obvio que sí influye en las distribuciones discretas ya que, por ejemplo, al cambiar un '*menor que*' por un '*menor o igual que*' estaríamos añadiendo un valor concreto de la variable con su valor de probabilidad. En cambio, en las distribuciones continuas no importaría, ya que la probabilidad de todos los valores concretos de la variable es nula y únicamente son evaluables las probabilidades de intervalos.

4. Ejemplos

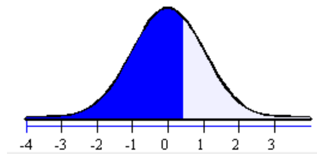
4.1 Siendo Z es una variable aleatoria normal N(0, 1) halla:

a) $P[Z \le 0{,}43]$ **b)** $P[Z \le -1{,}46]$ **c)** $P[Z > 1{,}61]$ **d)** $P[Z > -2{,}06]$

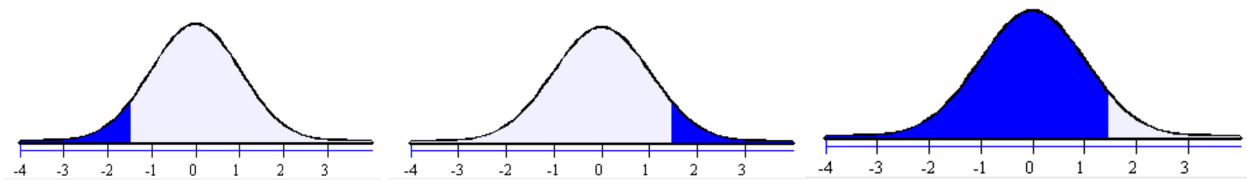
e) $P[0{,}91 < Z \le 2{,}3]$ **f)** $P[-1{,}72 < Z \le -0{,}23]$ **g)** $P[-0{,}74 < Z \le 1{,}5]$

Se trata de una búsqueda directa sobre la tabla de la distribución normal N(0,1). Cada probabilidad indicada en los cálculos aparece representada, en el mismo orden, mediante una región sombreada bajo la campana.

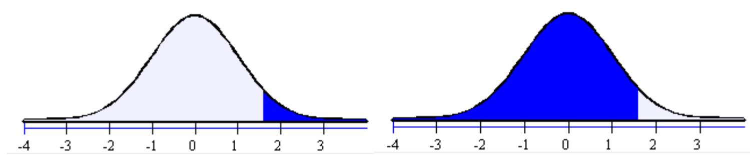
a) $P[Z \le 0{,}43] = 0{,}6664$



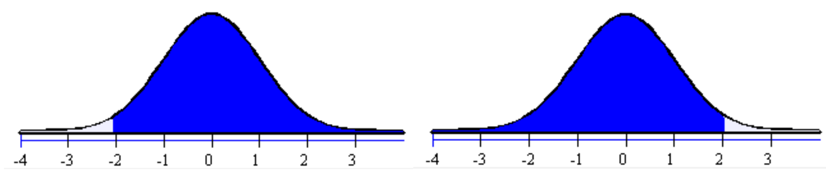
b) $P[Z \le -1{,}46] = P[Z > +1{,}46] = 1 - P[Z \le +1{,}46] = 1 - 0{,}9279 = 0{,}0721$



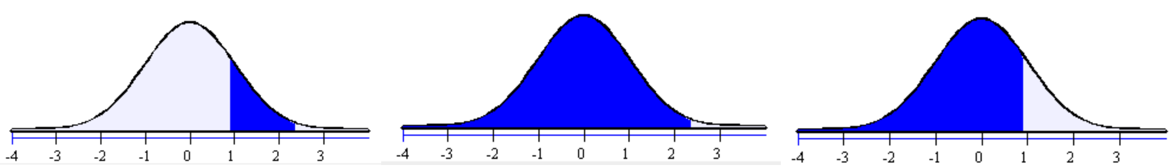
c)
$P[Z > 1{,}61] = 1 - P[Z \le 1{,}61] =$
$= 1 - 0{,}9463 = 0{,}0537$



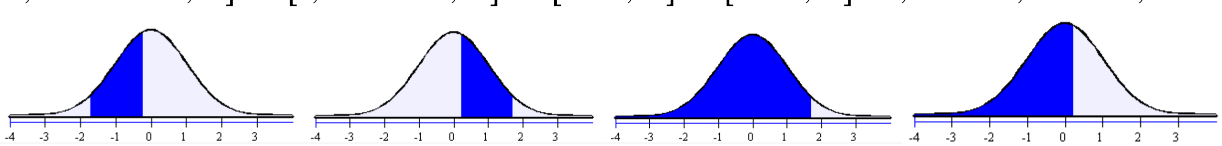
d) $P[Z > -2{,}06] = P[Z \le +2{,}06] = 0{,}9803$



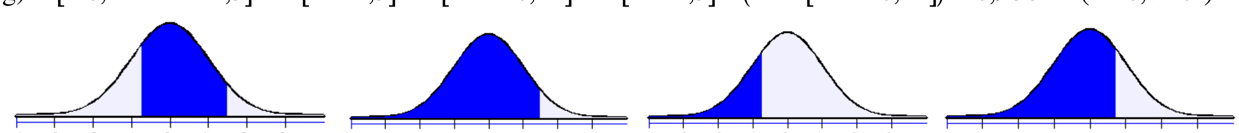
e) $P[0{,}91 < Z \le 2{,}3] = P[Z \le 2{,}3] - P[Z \le 0{,}91] = 0{,}9788 - 0{,}8186 = 0{,}1602$



f) $P[-1{,}72 < Z \le -0{,}23] = P[0{,}23 < Z \le 1{,}72] = P[Z \le 1{,}72] - P[Z \le 0{,}23] = 0{,}9474 - 0{,}5910 = 0{,}3564$



g) $P[-0{,}74 < Z \le 1{,}5] = P[Z \le 1{,}5] - P[Z \le -0{,}74] = P[Z \le 1{,}5] - (1 - P[Z \le +0{,}74]) = 0{,}9332 - (1 - 0{,}7704) = 0{,}7036$



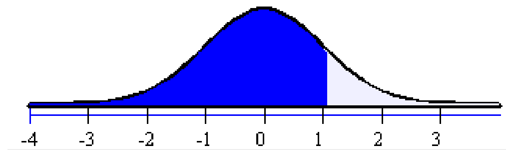
4.2 Siendo **Z es una variable aleatoria normal N(0,1). Halla el valor de *a* tal que:**

a) $P[Z \leq a] = 0'8599$ **b)** $P[Z \leq a] = 0'0392$ **c)** $P[Z > a] = 0'0951$

Se trata de una búsqueda inversa sobre la tabla de la distribución normal N(0,1). Cada probabilidad indicada en los cálculos aparece representada, en el mismo orden, mediante una región sombreada bajo la campana.

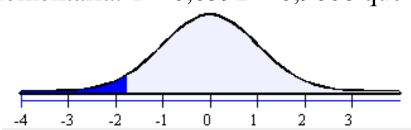
a) $P[Z \leq a] = 0{,}8599 \Rightarrow a = 1{,}08$

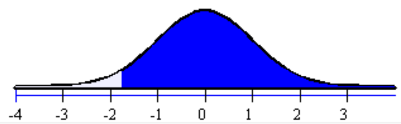
Buscamos en la tabla la probabilidad 0,8599, que nos muestra la abscisa 1,08. Por lo tanto *a* = 1,08

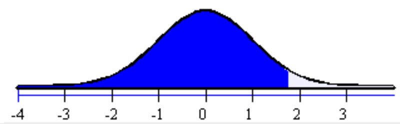


b) $P[Z \leq a] = 0{,}0392 \Rightarrow P[Z > a] = 1 - 0{,}0392 = 0{,}9608 \Rightarrow P[Z \leq -a] = 0{,}9608 \Rightarrow -a = 1{,}76 \Rightarrow a = -1{,}76$

La probabilidad 0,0392 es menor que 0,5 luego *a* es una cantidad negativa. Pasamos entonces a la probabilidad complementaria: 1 – 0,0392 = 0,9608 que buscada en la tabla nos muestra la abscisa 1,76. Por lo tanto *a* = – 1,76

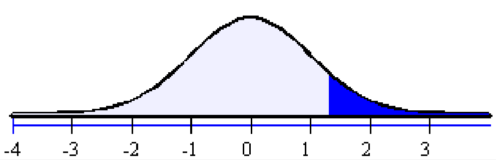


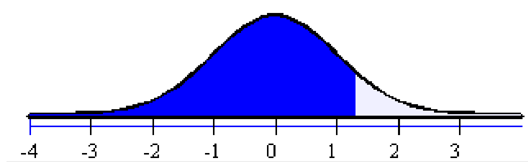




c) $P[Z > a] = 0{,}0951 \Rightarrow P[Z \leq a] = 1 - 0{,}0951 = 0{,}9049 \Rightarrow a = 1{,}31$

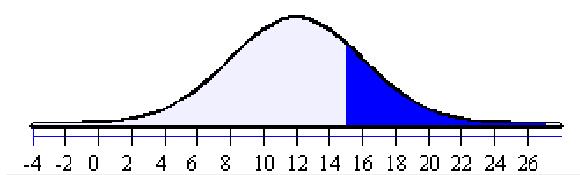
Como el recinto mira hacia la derecha, pasamos a la probabilidad complementaria: 1 – 0,0951 = 0,9049. Buscando en la tabla nos muestra la abscisa 1,31. Por lo tanto *a* = 1,31





4.3 La concentración de cierto agente contaminante del acuífero se encuentra normalmente distribuida con una media de 12 ppm (partes por millón) y desviación típica 4 ppm. Se considera tóxica una concentración superior a 15 ppm. ¿Con qué probabilidad encontraríamos concentraciones tóxicas?

Se trata de una distribución *N*(12, 4) y nos piden: $P[X > 15]$. La probabilidad pedida es la siguiente región sombreada.



Tipificamos haciendo $Z = \dfrac{X-12}{4}$, y como la región mira hacia la derecha, pasamos a la probabilidad complementaria.

Hacemos una búsqueda directa en la tabla, encontrando 0,7733, luego la respuesta es 1 – 0,7733 = 0,2267.

$$P[X > 15] = P\left[\frac{X-12}{4} > \frac{15-12}{4}\right] = P[Z > 0{,}75] = 1 - P[Z \leq 0{,}75] = 1 - 0{,}7733 = 0{,}2267$$

4.4 En una asignatura de Psicología evolutiva se ha podido determinar que las calificaciones de distribuyen según una *N*(5; 1'5).
a) ¿A partir de qué nota se encontrará el 75% de los alumnos mejor calificados?
b) ¿Hasta qué nota obtendrían el 5% de los alumnos peor calificados?

En todos los apartados tenemos que tipificar la normal N(5; 1'5) para poder usar la tabla de la normal ***N***(0;1). Los dos apartados consisten en búsquedas inversas. Si aparecen recintos que miran hacia la derecha o tienen abscisas negativas actuamos como en los ejercicios anteriores.

a)

$$P[X>a]=0{,}75 \Rightarrow P\left[\frac{X-5}{1{,}5}>\frac{a-5}{1{,}5}\right]=0{,}75 \Rightarrow P\left[Z>\frac{a-5}{1{,}5}\right]=0{,}75 \Rightarrow P\left[Z\le-\frac{a-5}{1{,}5}\right]=0{,}75 \Rightarrow$$

$$\Rightarrow -\frac{a-5}{1{,}5}=0{,}675 \Rightarrow \frac{a-5}{1{,}5}=-0{,}675 \Rightarrow a=-1{,}5\cdot 0{,}675+5=3{,}9875$$

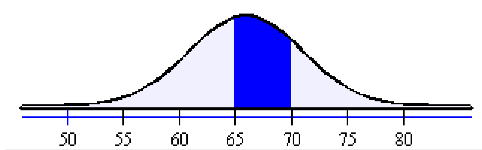
b)

$$P[X\le a]=0{,}05 \Rightarrow P\left[\frac{X-5}{1{,}5}\le\frac{a-5}{1{,}5}\right]=0{,}05 \Rightarrow P\left[Z\le\frac{a-5}{1{,}5}\right]=0{,}05 \Rightarrow P\left[Z>\frac{a-5}{1{,}5}\right]=0{,}95 \Rightarrow$$

$$\Rightarrow P\left[Z\le-\frac{a-5}{1{,}5}\right]=0{,}95 \Rightarrow -\frac{a-5}{1{,}5}=1{,}645 \Rightarrow \frac{a-5}{1{,}5}=-1{,}645 \Rightarrow a=-1{,}5\cdot 1{,}645+5=2{,}5325$$

4.5 Las tallas de 800 recién nacidos se distribuyen normalmente con una media de 66 cm. y una desviación típica de 5 cm. ¿Cuántos recién nacidos cabe esperar con tallas comprendidas entre 65 y 70 cm?

Se trata de una distribución ***N***(66; 5) que tipificaremos para poder usar la tabla de la normal ***N***(0;1) y actuaremos como en los ejercicios anteriores. La siguiente región sombreada muestra $P[65 < X \le 70]$ y que obviamente necesita dos búsquedas directas.



$$P[65<X\le 70]=P\left[\frac{65-66}{5}<\frac{X-66}{5}\le\frac{70-66}{5}\right]=P[-0{,}2<Z\le 0{,}8]=0{,}7881-(1-0{,}5793)=0{,}3674=36{,}74\%$$

Por último, sólo quedaría aplicar dicho porcentaje al total de los 800 recién nacidos: $800\cdot\frac{36{,}74}{100}=293{,}94$

Aproximadamente, 294 recién nacidos.

5. Consejos finales

Es muy útil, casi indispensable, visualizar cada cálculo que se haga con una región concreta de la campana de gauss y recordar las siguientes obviedades:

- No debemos confundir las abscisas, que son valores de la variable aleatoria, con las probabilidades. El error más frecuente sucede cuando aparece un valor 0'**** que puede ser ambas cosas.
- La desigualdad < alude a una región que se extiende desde el extremo izquierdo de la campana.
- La desigualdad > alude a una región que se extiende hasta el extremo derecho de la campana.
- Una región que ocupe menos de media campana tiene una probabilidad menor que 0,5 por lo que es posible que su probabilidad se calcule restando de 1 la lectura de una probabilidad en la tabla sin que importe que la desigualdad sea < o >.
- Una región que ocupe más de media campana tiene una probabilidad mayor que 0,5 por lo que es posible que su probabilidad se lea directamente de la tabla sin que importe que la desigualdad sea '*menor que*' o '*mayor que*'.
- Una probabilidad mayor que 0,5 tiene dos opciones: o le corresponde una región que comienza en una abscisa negativa (que probablemente hallaremos poniendo un signo menos a una abscisa de la tabla) hasta el extremo derecho de la campana o le corresponde una región que comienza en el extremo izquierdo de la campana hasta una abscisa positiva (que probablemente leeremos directamente en la tabla).
- Una probabilidad menor que 0,5 tiene dos opciones: o le corresponde una región que comienza en una abscisa positiva (que probablemente leeremos directamente en la tabla) hasta el extremo derecho de la campana o le corresponde una región que comienza en el extremo izquierdo de la campana hasta una abscisa negativa (que probablemente hallaremos poniendo un signo menos a una abscisa de la tabla).