La función de probabilidad normal: Características y aplicaciones

Guillermina Martín Reyes

Resumen: El objetivo de esta nota es mostar las características básicas de la función de probabilidad normal, así como su utilización.

Palabras clave: función de probabilidad normal; variable aleatoria.

Código JEL: C10.

1. Introducción

La función de probabilidad normal, también denominada función de Gauss, función de Gauss-Laplace o función de probabilidad de los errores, es una función de probabilidad para variables aleatorias continuas de gran aplicación en ingeniería, en física y en las ciencias sociales. De hecho, es la función de probabilidad con más aplicaciones al campo de la Economía y de la Empresa. Al ser la distribución normal una aproximación excelente a algunos fenómenos aleatorios continuos, como la estatura, el peso, el tiempo de servicio al cliente; y a fenómenos aleatorios discretos, a los cuales se les puede dar un tratamiento continuo, como la antigüedad de la vivienda, los impuestos anuales, la producción, etc, dicha distribución tiene mucha aplicación en ámbitos tan variados como el control de calidad, el marketing, y la gestión de la producción, entre otros. En general, se puede decir que la distribución normal es de vital importancia en estadística por tres razones esenciales:

— Existen muchos fenómenos continuos que parecen seguirla o se pueden aproximar mediante ella.

— Se puede utilizar para aproximar distribuciones discretas de probabilidad y de esta forma evitar cálculos engorrosos.

— Por su relación con el teorema central del límite es la base de la inferencia estadística clásica.

2. La Distribución Normal

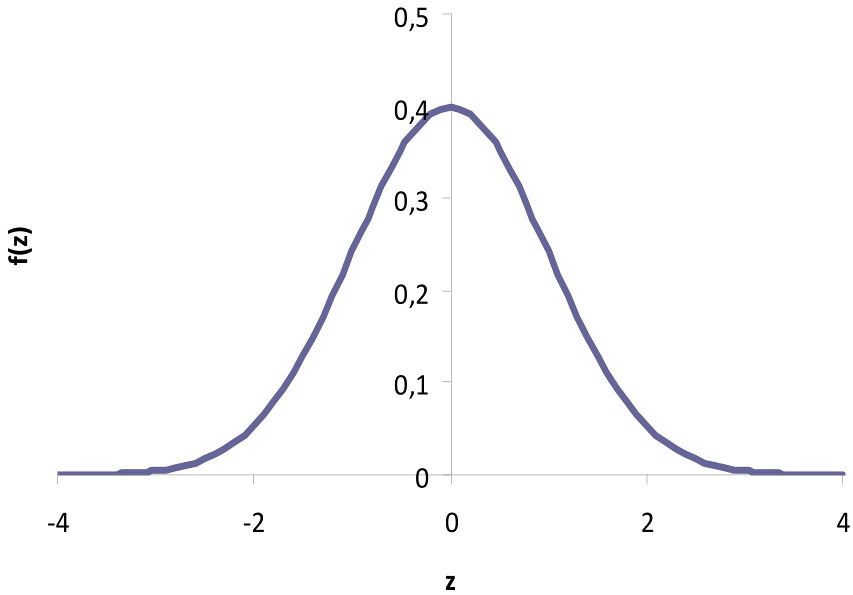
La distribución normal fue introducida por Carl Friedrich Gauss a principios del siglo XIX, al estudiar los errores de medida en los movimientos de los cuerpos celestes. Aunque algunos autores consideran que en 1733 ya había sido descrita por Moivre, y que Laplace también realizó sus aportaciones casi simultáneamente a la descripción de Gauss.

La función de densidad de una distribución normal tiene varios rasgos importantes: Es una distribución que tiene forma de campana, es simétrica y puede tomar valores entre menos infinito y más infinito (figura 1).

Es decir, los valores centrales de la variable se presentan con más frecuencia que los valores extremos. Así por ejemplo, en el caso de los errores de medida, los errores por defecto o por exceso pequeños en valor absoluto se presentan frecuentemente, mientras que los errores grandes, tanto los positivos como los negativos, se observan menos veces.

En segundo lugar, decimos que es simétrica. En efecto, es simétrica respecto a un valor central alrededor del cual toma valores con gran proba-

Gráfico 1: Distribución normal estándar



Fuente: Elaboración propia.

bilidad mientras que apenas aparecen los valores extremos. De manera que sus medidas de posición central (media, mediana y moda) coinciden.

Por último, presenta como asíntota el eje de abscisas al que se va aproximando sin llegar a cortarse.

La expresión matemática que representa una función de densidad con estas características es la siguiente:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2} \text{ para } -\infty\leq x\leq\infty \qquad (1)$$

Función de densidad que sólo depende de dos parámetros, la media μ, y la varianza de la distribución σ^2. Si conocemos la media y la varianza de una variable *X*, podemos definir la distribución normal utilizando la notación: $X\approx N(\mu,\sigma^2)$.

En la figura 2 se representan varias funciones de densidad normal con la misma media y diferentes varianzas.

Asimismo, la función de distribución acumulada de cualquier variable aleatoria *X*, se define como $F(x_0)= P(X<x_0)$, es decir la probabilidad de que dicha variable tome un valor menor que x_0. Matemáticamente esta función de distribución coincide con la integral de la función de densidad desde $-\infty$ hasta el valor x_0.

3. Distribución normal estandarizada

La distribución normal que tiene de media μ=0, y σ^2=1 se denomina distribución normal estándar, N(0,1), o tipificada. Su función de distribución se encuentra tabulada, siendo de gran utilidad para el cálculo de probabilidades de cualquier distribución N(μ,σ^2).

Sea una variable *X* que se distribuye como una normal con media μ y varianza σ^2.

Dicha variable se puede transformar en una variable normal tipificada, N(0,1), mediante la siguiente expresión:

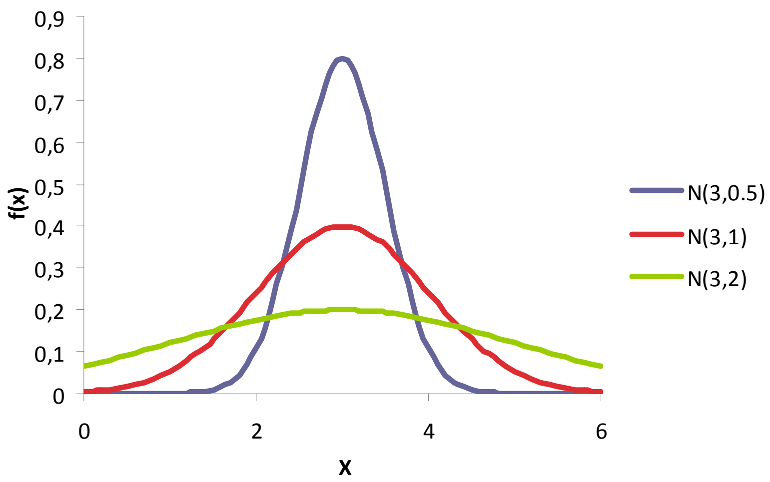
$$Z=\frac{X-\mu}{\sigma} \qquad (2)$$

De manera que la función de densidad de esta nueva variable es:

$$f(Z)=\frac{1}{\sqrt{2\pi}}e^{-(\frac{z^2}{2})} \qquad (3)$$

A partir de dicha expresión, las tablas de la distribución normal presentan los valores de o probabilidad de que la variable *Z* tome un valor inferior a z_0.

Gráfico 2. Distribuciones normales con distintas varianzas y la misma media



Fuente: Elaboración propia.

$$F(Z)=\int_{-\infty}^{z_0} f(Z)dz \qquad (4)$$

Las fórmulas más importantes para el cálculo de probabilidades de una N(0,1) son las siguientes:

$$\begin{aligned} &P(Z \geq z) = 1 - F(z) \\ &P(Z \leq -z) = P(Z \geq z) \\ &P(z_a \leq Z \geq z_b) = F(z_b) - F(z_a) \\ &P(|Z| \geq z_c) = P(Z \geq z_c) + P(Z \leq -z_c) = 2P(Z \geq z_c) = 2[1 - F(z_c)] \end{aligned} \qquad (5)$$

De forma que si se quiere obtener la probabilidad de que X se encuentre entre dos valores, x_a y x_b, primero se tipifican dichos valores con la expresión (2), y después se buscaría en las tablas de la distribución normal la probabilidad acumulada para esos dos nuevos valores, z_a y z_b. De manera que:

$$P(x_a \leq X \leq x_b) = P(z_a \leq Z \leq z_b) = F(z_b) - F(z_a) \qquad (6)$$

4. Algunas propiedades de la distribución normal.

— En toda distribución normal en el intervalo $\mu \pm 2\sigma$ se encuentra el 95,5 por ciento de la distribución; y en el intervalo $\mu \pm 3\sigma$ se tiene el 99,7 por ciento de la distribución.

— Una transformación lineal de una variable normal también se distribuye normalmente. Así si X sigue una distribución N(μ,σ^2), una nueva variable $Y = aX + b$ tiene una distribución normal con media $a\mu + b$ y desviación estándar $|a|\sigma$.

— Cualquier combinación lineal de variables aleatorias independientes e idénticamente distribuidas en forma normal, tiene una distribución normal. Esta propiedad implica que el promedio de variables independientes distribuidas normales tiene una distribución normal. Resultado que es crucial para la inferencia estadística sobre la media de una población normal.

— Si X e Y están distribuidas normalmente, estas dos variables son independientes si y sólo si la covarianza entre X e Y es cero.

— Las distribuciones que siguen una binomial o una distribución de Poisson se pueden aproximar a una distribución normal.

5. Aplicaciones

Suponga que en una determinada empresa se analiza el tiempo que lleva a los trabajadores la instalación de una determinada pieza del producto que fabrica, concluyendo que se distribuye como una normal con una media de 30 minutos y una desviación estándar de 5 minutos. Con estos datos se podrían contestar preguntas como:

1. ¿Cuál es la probabilidad de que un trabajador aleatoriamente seleccionado pueda montar la pieza en menos de 30 minutos?

La respuesta es la siguiente:

$$P(X \leq 30) = P(Z \leq \frac{30-30}{5}) = 0,5 \qquad (6)$$

Es decir, el 50 por ciento.

2. ¿Cuántos minutos tienen que pasar antes de que el 10 por ciento de los trabajadores monten la pieza?

Lo que debemos obtener aquí es el valor de la variable Z que deja en la cola izquierda de la distribución el 10 por ciento. En las tablas obtendríamos un valor z_0=-1,28. Este será el valor tipificado, de manera que el valor de X será: $x_0=\mu+z_0\,\sigma$=30-1,28x5=23,6 minutos.

Otro ejemplo en el que se puede aplicar el uso de la distribución normal es el siguiente: Suponga que el encargado de la tesorería de una empresa está interesado en el número de días que transcurre entre la facturación y el pago de las cuentas a crédito. Una vez recogidos los datos, se observa que el tiempo se distribuye aproximadamente como una normal con una media y una variancia determinada. En este caso se podría responder a preguntas como la siguiente: ¿Qué porcentaje de cuentas se pagará antes de x días? ¿Dentro de cuántos días se pagará el 80 por ciento de las cuentas?