

Distribución de Probabilidad Normal

Escuela de Matemática

Semana 12

Giovanni
Sanabria Brenes

Luis Ernesto
Carrera Retana

Erick Chacón
Vargas

Mario Marín
Sánchez

Rebeca Solís
Ortega

Distribución de Probabilidad Normal

En general las distribuciones de probabilidad son herramientas muy necesarias en el estudio de problemas probabilísticos y estadísticos.

Entre las distribuciones de probabilidad continuas, la distribución *normal*, es la más utilizada y la más importante. Una variable que siga una distribución Normal suele tomar valores aleatorios que se concentra alrededor de su media, unos pocos están muy por arriba de la media y otros pocos muy por debajo de la media. Esto hace que su gráfico sea en forma de una campana centrada en la media.

Muchas mediciones dentro de poblaciones siguen distribuciones normales y en casos donde las poblaciones no se distribuyen normalmente, es común que ciertos promedios y ciertos valores acumulados si se distribuyan en forma normal, esta última observación se conoce como el *teorema del límite central* y lo estudiaremos más adelante.

En términos muy simples, una población sigue una distribución normal respecto a alguna medición cuando el grueso de los valores de la población se distribuyen cerca de la media y existe cierta simetría en la forma en que se distribuyen los datos alrededor de la media. Por ejemplo, el coeficiente intelectual de los seres humanos tiene una distribución normal con media de 100 puntos. En términos matemáticos la definición es la siguiente:

Definición 1 (Función de distribución de probabilidad normal).

Una variable aleatoria X sigue una distribución normal con parámetros μ y σ , lo que denotamos por $X \sim N(\mu, \sigma^2)$, si la función de densidad de probabilidad tiene la forma.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{Para } x \in \mathbb{R}$$

Se puede demostrar que la media de esta distribución es μ y la desviación es σ .

Teorema 1 (Función de distribución normal acumulada).

Si $X \sim N(\mu, \sigma^2)$ entonces la función de distribución de probabilidad acumulada es

$$F_X(x) = P[X \leq x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Note que los cálculos de probabilidades en las distribuciones normales son difíciles pues implican integrales complejas. Sin embargo, vamos que todas las distribuciones normales si pueden transformar en una distribución particular (distribución normal estándar) y para ella contaremos con tablas que nos permitirán realizar los cálculos.

Entre las distribuciones normales, la distribución más importante es la que se llama distribución normal estándar. Si X sigue este tipo de distribución normal se cumple que la media de X es 0 y la desviación estándar es 1. O sea:

$$\mu = 0 \quad \text{y} \quad \sigma = 1$$

Es decir: $X \sim N(0, 1)$

De hecho en estas mismas notas veremos que toda probabilidad que implique la distribución normal puede reducirse a una en que se utilice la normal estándar. De manera más formal tenemos:

Definición 2 (Función de distribución de probabilidad normal estándar).

Una variable aleatoria Z sigue una distribución normal estándar si la función de densidad de probabilidad tiene la forma.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{Para } -\infty \leq x \leq \infty.$$

Es decir

$$Z \sim N(0, 1)$$

En este caso la distribución de probabilidad acumulada es,

$$\Phi(x) = P[Z \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

La última expresión es una variante de una función que se conoce como la función error $erf(x)$, y solo hay formas numéricas de aproximar sus valores.

Los valores de la función $\Phi(x)$ se pueden obtener utilizando las tablas que se anexan al final. Eventualmente se necesita utilizar la función inversa de la distribución normal, es decir dada una probabilidad calcular un valor de la variable aleatoria que produciría tal probabilidad, también se puede aproximar con tablas de valores (ver Anexo #1).

Algunas Propiedades Importantes

Como la función de distribución de probabilidad es simétrica, y además el área total acumulada, sobre toda la recta real es 1, entonces para cualquier x real se obtiene la siguiente propiedad:

$$\Phi(x) + \Phi(-x) = 1, \tag{1}$$

Si X sigue una distribución normal con parámetros μ y σ entonces si aplicamos el cambio de variable $\omega = \frac{t - \mu}{\sigma}$ a la integral en

$$P[X \leq x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dt,$$

obtenemos

$$P[X \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Teorema 2.

Si X sigue una distribución de probabilidad normal con media μ y desviación estándar σ entonces

$$P[X \leq x] = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad (2)$$

Ejemplo 1.

Las notas finales de un curso se distribuyen en forma normal con una media de 75 y una desviación estándar de 10. Si la nota de aprobación es de 70 qué porcentaje de los estudiantes aprobarán el curso.

Solución:

Primero debe notar que la afirmación de que las notas siguen una distribución normal debe entenderse en el sentido aproximado. Así el porcentaje solicitado puede obtenerse al encontrar el valor $P[X \geq 70]$.

Dadas las propiedades de las distribuciones de probabilidad se tiene que

$$P[X \geq 70] = 1 - P[X \leq 70] = 1 - \Phi\left(\frac{70-75}{10}\right) = 1 - \Phi(-0.5)$$

Para hallar $\Phi(-0.5)$ como $\Phi(x) = 1 - \Phi(-x)$ entonces $\Phi(-0.5) = 1 - \Phi(0.5)$

Así

$$P[X \geq 70] = 1 - \Phi(-0.5) = \Phi(0.5)$$

Finalmente, para obtener $\Phi(0.5)$, el número 0.5 es igual a 0.50 donde sus unidades y décimas son 0.5 y sus centésimas son 0.00. Así cruzamos en la tabla la fila de 0.5 con la columna de 0.00 obteniendo 0.6915. Por lo tanto

$$P[X \geq 70] = \Phi(0.5) = 0.6915$$

Ejemplo 2.

La distribución de peso de ciertos bultos de papel para reciclaje es normal con media de 50 kilos y desviación estándar de 10 kilos. La persona que transporta los paquetes cobra 100 colones por bulto pero desea imponer un peso máximo después del cual cobrar un recargo. Cuál debería ser ese peso máximo para que los bultos tengan una probabilidad mayor al 10 % de pagar tal recargo.

Solución

Hay dos aspectos importantes que se deben notar, el primero de ellos es que si X es la variable aleatoria para el peso de cada paquete lo que se debe encontrar es un valor r tal que:

$$P[X \geq r] > 0.1,$$

lo que se reduce a encontrar un r que cumpla con:

$$P[X \leq r] \leq 0.9,$$

El problema es inverso en el sentido de que no se busca una probabilidad, sino un valor que permita obtener cierta probabilidad. El segundo aspecto que debe tenerse en cuenta es que para poder utilizar las tablas, la distribución debe normalizarse. Uniendo ese par de observaciones se debe resolver:

$$P[X \leq r] = \Phi\left(\frac{r-50}{10}\right) \leq 0.9.$$

Primero hallemos el valor Z (valor de la distribución estándar) que acumula una área de 0.9. Para ellos vemos todos los números dentro de la tabla buscando lo cercano a 0.9, este valor es 0.8997 y viendo su fila (1.2) y su columna (0.08) tiene asociado el valor Z de 1.28, por lo tanto $0.9 \approx \Phi(1.28)$

Así se tiene aproximadamente que

$$\Phi\left(\frac{r-50}{10}\right) \leq 0.9 \approx \Phi(1.28)$$

Dado que Φ es una función estrictamente creciente entonces

$$\Phi\left(\frac{r-50}{10}\right) \leq \Phi(1.28) \Rightarrow \frac{r-50}{10} \leq 1.28$$

de donde $r \leq 62.8$. Por lo que el peso máximo es de 62.8 kilogramos aproximadamente.

Ejercicios propuestos

1. Se sabe que la distribución de notas en un curso sigue una distribución normal. El 10 % de los exámenes tienen una nota por encima de los 80 puntos, y el 5 % tiene una nota por debajo de los 40 puntos. ¿Cuáles son el valor de la media y de la desviación estándar para esta distribución?

Solución:

No se conocen μ_X ni σ_X pero se sabe que:

$$P[X \leq 40] = 0.05 \Rightarrow \Phi\left(\frac{40-\mu}{\sigma}\right) = 0.05 \Rightarrow \frac{40-\mu}{\sigma} = -1.64$$

Por otra parte

$$P[X \leq 80] = 0.9 \Rightarrow \Phi\left(\frac{80-\mu}{\sigma}\right) = 0.9 \Rightarrow \frac{80-\mu}{\sigma} = 1.28$$

Resolviendo ese par de ecuaciones se obtiene la respuesta.

2. Suponga que el tiempo T que tarda una persona en resolver un examen sigue una distribución normal con media 50 minutos y desviación estándar de 12 minutos. Se quiere establecer un rango de tiempo centrado en la media en el cual se contesten el 90 % de los tests. ¿Cuál es ese rango?

Solución:

Lo que se quiere determinar es un valor ζ tal que $P[\mu - \zeta \leq X \leq \mu + \zeta] = 0.90$ esto es equivalente a que $P[X \leq 50 + \zeta] = 0.95$ normalizando esto es equivalente a

$$\Phi\left(\frac{\zeta - 50}{12}\right) = 0.95 \Rightarrow \frac{\zeta - 50}{12} = 1.65$$

de allí se despeja ζ .

3. La distribución de peso de cierto tipo de paquete sigue una distribución normal con media 10 libras y desviación estándar de 2 libras. Se desea establecer un valor de peso c , más allá del cual habrá cargo extra. ¿Cuál es el valor de c para que el 99 % de los paquetes no deban pagar cargo extra?

Solución:

Lo que se desea es determinar c tal que $P[X \leq c] = 0.99$ de donde el problema se reduce a:

$$\Phi\left(\frac{c - 10}{2}\right) = 0.99 \Rightarrow \frac{c - 10}{2} = 2.33$$

de allí se despeja c .

4. Se sabe que las notas en un curso siguen una distribución normal. Con media 75 puntos y desviación estándar 15 puntos, determine.

- El porcentaje de notas que queda sobre 80.

Solución:

$$1 - \Phi\left(\frac{80 - 75}{15}\right)$$

- El rango centrado en la media en el que queda el 45 % de las notas.

Solución:

Si μ es la media entonces lo que se desea resolver es:

$$P[\mu - c < X < \mu + c] = 0.45$$

lo cual es equivalente a resolver la ecuación

$$\Phi\left(\frac{c}{15}\right) = 0.725$$

de donde se tiene que

$$\frac{c}{15} = 0.6$$

- La nota sobre la que queda el 40 % de las mejores notas del curso.

Solución:

Lo que se busca es hallar ρ tal que $P[X > \rho] = 0.4$ y se resuelve encontrando la solución de

$$\Phi\left(\frac{70 + \rho}{15}\right) = 0.6$$

5. Suponga que el tiempo T que tarda una persona en realizar una reparación sigue una distribución normal con media 50 minutos y desviación estándar de 12 minutos. Un día esa persona debe realizar 8 reparaciones:

- Si las hace consecutivamente determine la probabilidad de que si empieza a las 8 de la mañana finalice antes de las 3:30 pm.

Solución:

Si $T_i, i = 1, 2, \dots, 8$ es el tiempo que tarde en la reparación i , entonces $T_A = T_1 + T_2 + \dots + T_8$, como suma de normales, es normal y la solución del problema se reduce a obtener:

$$\Phi\left(\frac{480 - 400}{12\sqrt{8}}\right)$$

- Determine la probabilidad de que en al menos 5 de las reparaciones haya durado menos de una hora.

Solución:

Se tiene de que la probabilidad de que una reparación dure menos de una hora es:

$$P[T \leq 60] = \Phi\left(\frac{60 - 50}{12}\right) = \Phi(0.83) \approx 0.7967$$

Luego sea Y el total de reparaciones en que se dura al menos de una hora, se busca:

$$P[Y \geq 5] = 1 - P[Y < 5] = 1 - \sum_{i=0}^4 b(i; 8, 0.7967)$$

- Determine la probabilidad de que la última reparación haya sido la primera en que tardó menos de una hora.

Solución:

Se sabe (por el ejercicio anterior) que la probabilidad de que una reparación dure menos de una hora es 0.7967.

Luego para determinar la probabilidad solicitada se recurre a una distribución geométrica:

$$(0.2033)^7(0.7967)$$

6. Demuestre la igualdad $\Phi(w) + \Phi(-w) = 1$

Solución:

$$1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_w^{\infty} e^{-\frac{x^2}{2}} dx$$

La primera de esas integrales es $\Phi(w)$ y en la segunda hacemos el cambio de variable $x = -t$ para obtener:

$$\int_w^{\infty} e^{-\frac{x^2}{2}} dx = - \int_{-w}^{-\infty} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{-w} e^{-\frac{t^2}{2}} dt = \Phi(-w)$$

Anexo: Tabla de distribución acumulada de la normal estándar $\Phi(x)$

[illegible]