

Introducción a la Probabilidad y la Estadística



Martes y Jueves Aula B17
Dra Ana Georgina Flesia

Variable aleatoria $N(\mu, \sigma^2)$

La distribución normal es especialmente relevante en el análisis de datos debido a varias razones:

1. Modelo ampliamente utilizado: Muchos métodos y técnicas estadísticas se basan en la suposición de que los datos siguen una distribución normal.
2. Facilita la inferencia estadística: Al asumir una distribución normal, podemos aplicar herramientas estadísticas estándar para realizar inferencias sobre los datos. Esto incluye realizar pruebas de hipótesis, construir intervalos de confianza y realizar análisis de regresión, entre otros.

Variable aleatoria $N(\mu, \sigma^2)$

La distribución normal es especialmente relevante en el análisis de datos debido a varias razones:

1. Simplifica los cálculos: La distribución normal tiene propiedades matemáticas bien conocidas y establecidas. Esto facilita los cálculos y permite realizar estimaciones y predicciones más precisas.
2. El Teorema Central del Límite: Establece que la suma o promedio de un gran número de variables aleatorias independientes tiende a seguir una distribución normal, incluso si las variables individuales no son normalmente distribuidas. Esto es fundamental en el análisis de muestras grandes y permite utilizar la distribución normal como una aproximación útil en muchos escenarios.

Variable aleatoria $N(0, 1)$

Se dice que una variable aleatoria continua Z tiene distribución Normal Patrón si su densidad de probabilidad φ está definida como

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

en donde aparecen tres de los números más famosos de la historia de la matemática, $\sqrt{2}$, π y e .

Distribución Normal Patrón

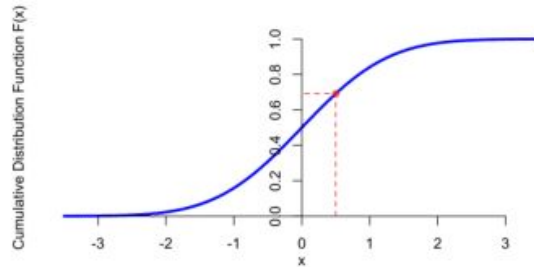
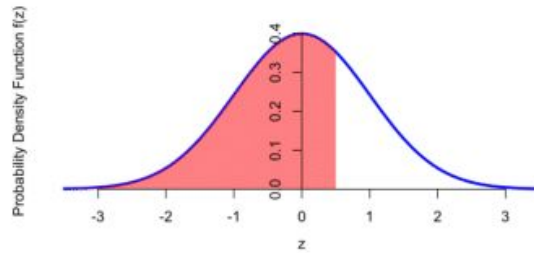
La distribución acumulada correspondiente a φ , denotada por Φ no tiene fórmula cerrada,

$$\Phi(z) = \int_{-\infty}^z \varphi(y) dy = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

por lo cual no es posible calcular probabilidades integrando la curva. Sin embargo, existen aproximaciones tabuladas de diferentes maneras.

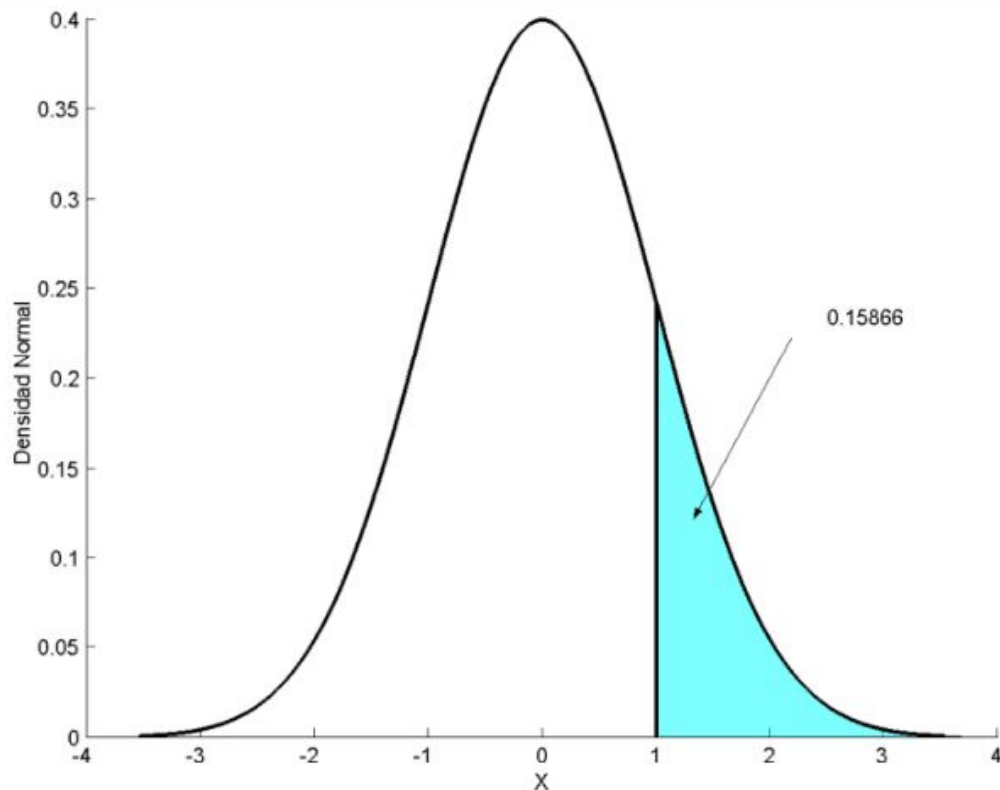
Distribución Normal

Densidad y Distribución



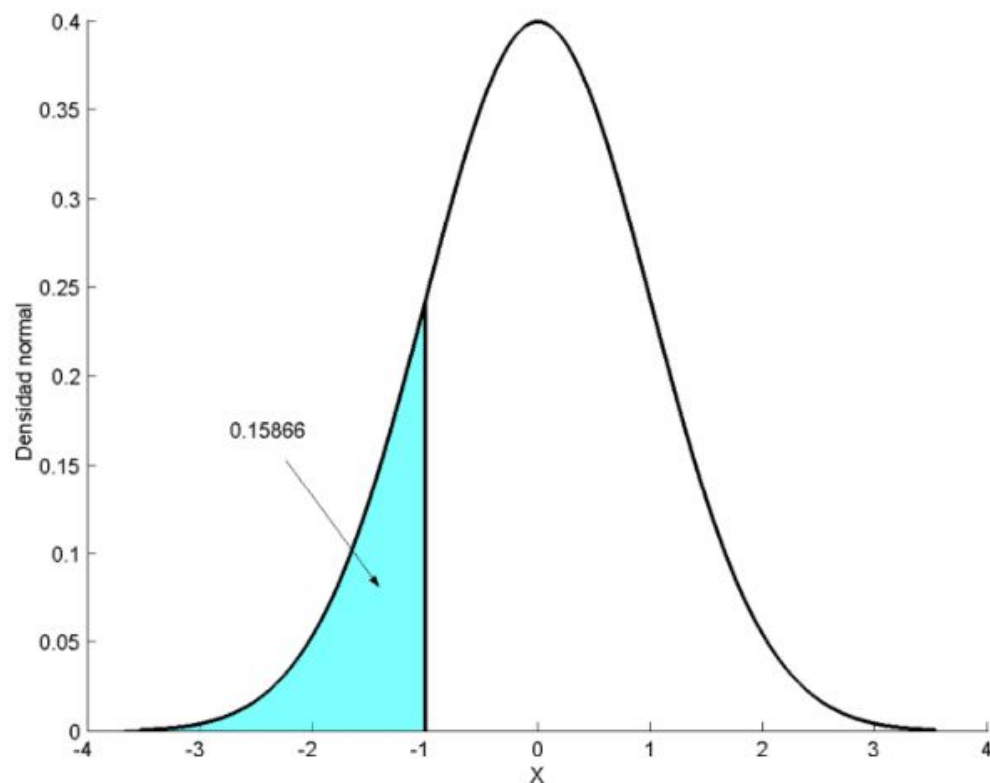
Si Z tiene distribución normal estándar

$$z > 0 \quad P(Z \geq z) = 1 - P(Z \leq z) = 1 - \Phi(z)$$



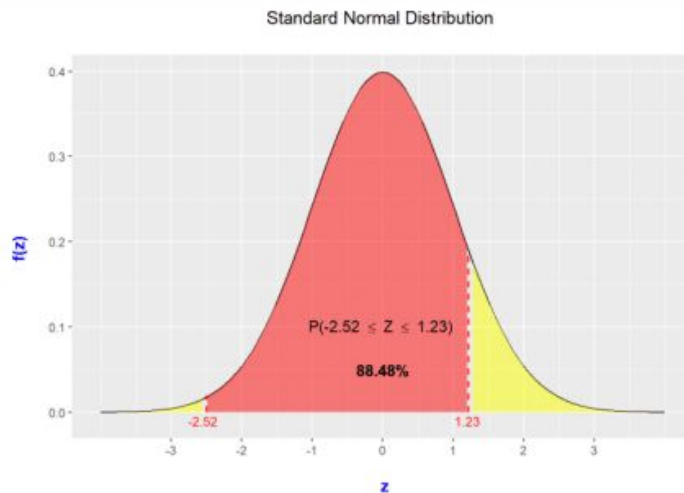
Si Z tiene distribución normal estándar

$$z > 0 \quad P(Z \leq -z) = P(Z \geq z) = 1 - P(Z \leq z) = 1 - \Phi(z)$$



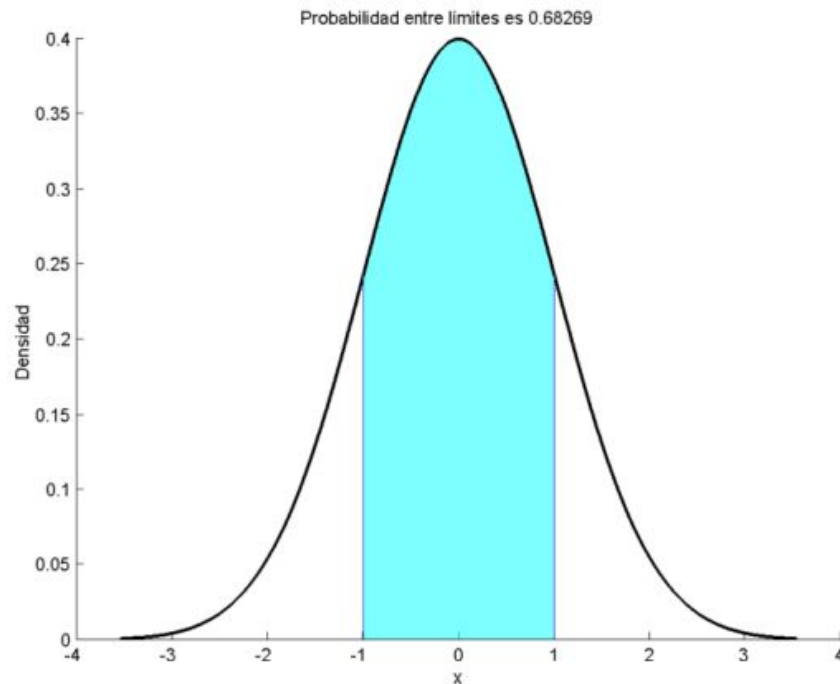
Si Z tiene distribución normal estándar

$$\begin{aligned} z_1, z_2 > 0 \quad P(-z_1 \leq Z \leq z_2) &= P(Z \leq z_2) - P(Z \leq -z_1) \\ &= P(Z \leq z_2) - P(Z \geq z_1) \\ &= P(Z \leq z_2) - [1 - P(Z \leq z_1)] \\ &= \Phi(z_2) + \Phi(z_1) - 1 \end{aligned}$$



Si Z tiene distribución normal estándar

$$z > 0 \quad P(-z \leq Z \leq z) = 2\Phi(z) - 1$$



Distribución Normal

Manejo de la tabla

1. Para el calculo de probabilidades de una normal estándar hay tablas generadas por métodos de integración numérica.
2. La tabla de acumulación a cola a izquierda tabula los valores de

$$\alpha = \Phi(z_\alpha) = P(Z \leq z_\alpha)$$

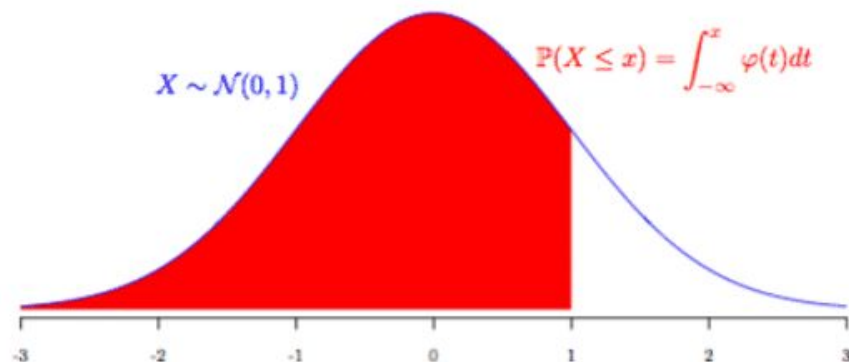
$$z_\alpha > 0 \text{ y } Z \sim N(0, 1).$$

Distribución Normal

Manejo de la tabla

- ▶ Supongamos tener una tabla acumulada de cola a izquierda.
- ▶ En el centro de la tabla están las columnas con probabilidades α y en los márgenes de la tabla los valores de los cuantiles $z_\alpha > 0$.
- ▶ El margen izquierdo tiene los valores entero y primer decimal de z_α y el margen superior el segundo decimal de z_α .

área normal tabla



	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
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	0.7704	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
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621

Distribución Normal

Manejo de la tabla

- ▶ Para encontrar

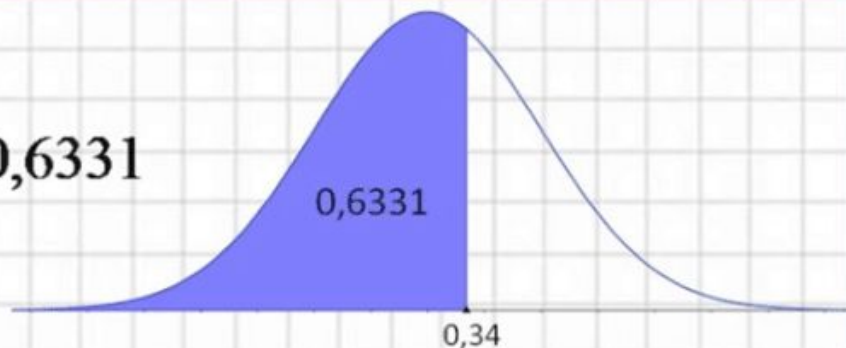
$$\alpha = P(Z < z_{\alpha}) = P(Z < 0.34)$$

se busca la fila del $z_{\alpha} = 0.3$ y luego se sigue esa fila hasta la columna del segundo decimal 0.04.

- ▶ La intersección de fila y columna da el valor $\alpha = 0.6331$.

área normal tabla

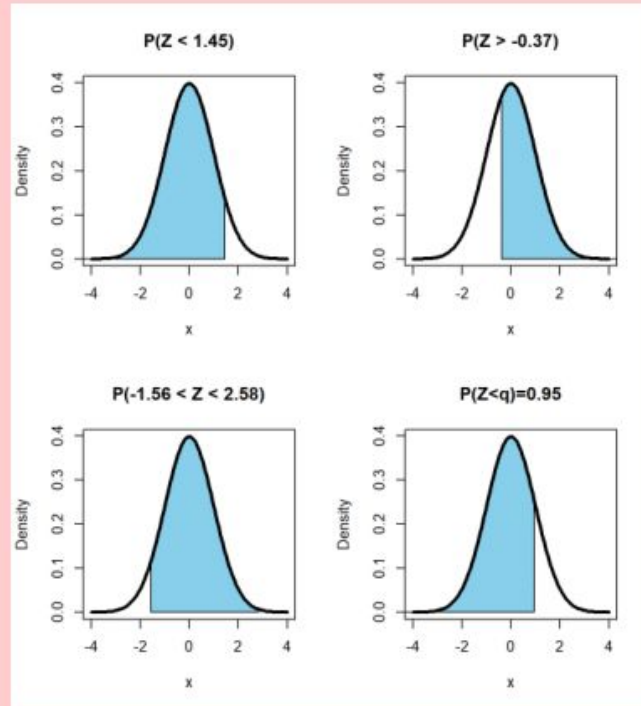
$$P(Z \leq 0,34) = 0,6331$$



z	,00	,01	,02	,03	<u>,04</u>	,05	,06
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026
<u>0,3</u>	<u>0,6179</u>	<u>0,6217</u>	<u>0,6255</u>	<u>0,6293</u>	<u>0,6331</u>	0,6368	0,6406
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772

Distribución Normal

Densidades



Distribución Normal

Ejercicio: Manejo de la tabla

1. Calcular las siguientes probabilidades:

1.1 $P(Z \leq 1.45)$

1.2 $P(Z \geq -0.37)$

1.3 $P(-1.56 \leq Z \leq 2.58)$

2. Determinar el valor de la constante q tal que:

2.1 $\Phi(q) = 0.95$

3. Determinar los percentiles 25, 50 y 75

Ejemplo: Resolución

- ▶ La probabilidad de que Z sea menor o igual que 1.45.

$$P(Z \leq 1.45) = 0.9265$$

- ▶ La probabilidad de que Z sea mayor o igual que -0.37.

$$\begin{aligned} P(Z \geq -0.37) &= 1 - P(Z \leq -0.37) \\ &= 1 - [1 - P(Z \leq 0.37)] \\ &= P(Z \leq 0.37) = 0.6443 \end{aligned}$$

- ▶ La probabilidad de que X sea mayor que -1.56 y menor que 2.58 es

$$\begin{aligned} P(-1.56 \leq Z \leq 2.58) &= P(Z \leq 2.58) + P(Z \leq 1.56) - 1 \\ &= 0.9951 + 0.9406 - 1 = 0.9357 \end{aligned}$$

Ejemplo: Resolución

- ▶ El valor de q tal que $\Phi(q) = 0.95$ es 1.645
- ▶ Los percentiles 25, 50 y 75 son

$$0.675 = \eta(0.75) \quad 0 = \eta(0.5) \quad -0.675 = \eta(0.25)$$

Variable aleatoria $N(\mu, \sigma^2)$

Sea Y una variable aleatoria con densidad dada por

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} = \frac{1}{\sigma} \varphi\left(\frac{y-\mu}{\sigma}\right) \quad -\infty < y < \infty$$

La distribución de Y se llama normal con parámetros μ y σ^2 y se denota $Y \sim N(\mu, \sigma^2)$.

Distribución Normal

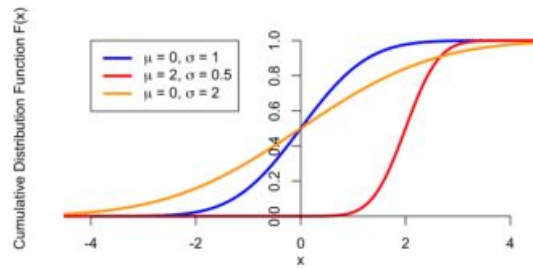
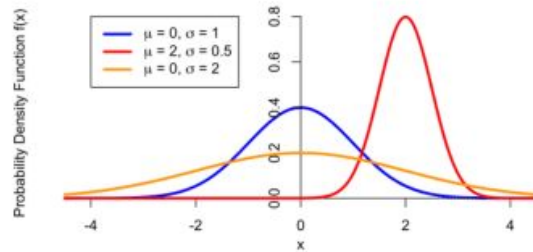
La distribución acumulada correspondiente a f_Y , no tiene formula cerrada,

$$F_Y(y) = \int_{-\infty}^y f_Y(t)dt = \int_{-\infty}^y \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2} dt$$

por lo cual no es posible calcular probabilidades integrando la curva.

Distribución Normal

Densidades



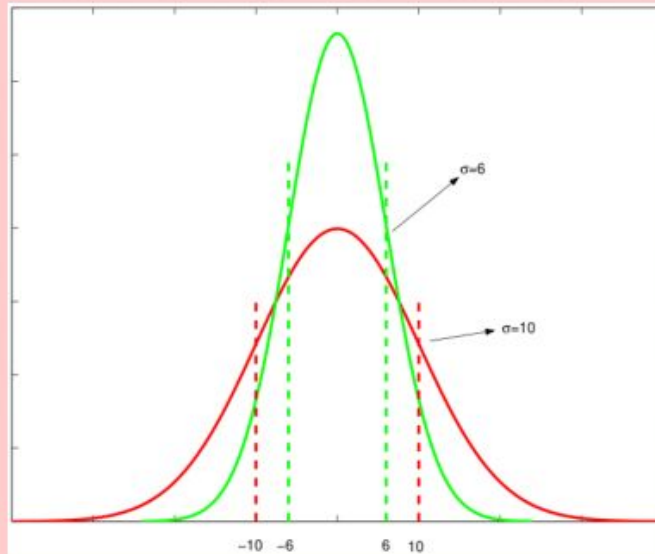
Propiedades $N(\mu, \sigma^2)$

Sea Y variable aleatoria con densidad $N(\mu, \sigma^2)$

1. $f_Y(x) \geq 0 \ \forall x \in \mathbb{R}$ y $\int_{-\infty}^{\infty} f_Y(x)dx = 1$
2. f_Y es simétrica con respecto a μ , o sea $f_Y(\mu - x) = f_Y(\mu + x)$
3. $\lim_{x \rightarrow -\infty} f_Y(t) = \lim_{x \rightarrow \infty} f_Y(t) = 0$
4. La gráfica tiene forma de campana.

Distribución Normal

Tendencia central



Distribución Normal

Propiedades

1. Si Y tiene distribución normal con parámetro μ y σ^2 entonces $Z = \frac{Y-\mu}{\sigma}$ tiene distribución normal estándar.
2. si Z es una variable normal estándar $Y = \mu + \sigma Z$ tiene distribución normal con parámetros μ y σ^2 .

Observación

Si X es una variable con distribución normal $N(\mu, \sigma^2)$ entonces para calcular $P(a \leq X \leq b)$ primero se debe estandarizar la variable aleatoria X creando $Z = \frac{X - \mu}{\sigma}$ y después usar la tabla de la $N(0, 1)$. Por ejemplo

$$P(a \leq X < b) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

si $a < b$ son números reales, y Φ es la función de distribución acumulada de una variable aleatoria $N(0, 1)$.

Ejemplo

Si X es una variable normal con media $\mu = 50$ y varianza $\sigma^2 = 100$, calcule las siguientes probabilidades utilizando la distribución normal estándar.

- ▶ La probabilidad de que X sea menor o igual que 40.
- ▶ La probabilidad de que $|X|$ (o sea, el valor absoluto de X) sea menor o igual que 60.
- ▶ La probabilidad de que $|X|$ (o sea, el valor absoluto de X) sea mayor o igual que 60.

Ejemplo: Resolución

- La probabilidad de que X sea menor o igual que 40.

$$P(X \leq 40) = P\left(Z \leq \frac{40 - 50}{10}\right) = P(Z \leq -1) = 0.1587$$

- La probabilidad de que $|X|$ (o sea, el valor absoluto de X) sea menor o igual que 60.

$$\begin{aligned}P(|X| \leq 60) &= P(-60 \leq X \leq 60) = P(X \leq 60) - P(X \leq -60) \\&= P\left(Z \leq \frac{60 - 50}{10}\right) - P\left(Z \leq \frac{-60 - 50}{10}\right) \\&= P(Z \leq 1) - P(Z \leq -11) = 0.8413 - 0 = 0.8413\end{aligned}$$

- La probabilidad de que $|X|$ (o sea, el valor absoluto de X) sea mayor o igual que 60.

$$P(|X| \geq 60) = 1 - P(|X| \leq 60) = 1 - 0.8413 = 0.1587$$

Ejemplo

- ▶ En una empresa hay dos maquinas disponibles A y B que cortan corchos, para ser usados en botellas de vino, tal que el 40 % son cortados con la maquina A y el resto con la B.
- ▶ Se sabe que la maquina A corta corchos con diámetro que están normalmente distribuido con una media de 3 cm y una desviación estándar de 0,1 cm y la maquina B corta corchos con diámetro que están normalmente distribuido con una media de 3,04 cm y una desviación estándar de 0,02 cm.
- ▶ Para que un corcho sea aceptable para su uso debe cumplir que su diámetro este comprendido entre 2,9 y 3,1 cm.
 1. Cual de las dos maquinas producen mayor porcentaje de corchos aceptables?
 2. Se selecciona al azar un corcho de la producción, cual es la probabilidad de que el corcho sea aceptable?

Ejemplo: resolución

Datos del Problema

► Máquina A:

1. Media (μ_A) = 3 cm
2. Desviación estándar (σ_A) = 0.1 cm

► Máquina B:

1. Media (μ_B) = 3.04 cm
2. Desviación estándar (σ_B) = 0.02 cm

► Rango aceptable de diámetro: entre 2.9 cm y 3.1 cm

► Proporción de corchos cortados por cada máquina:

1. Máquina A: 40%
2. Máquina B: 60%

Ejemplo: Resolución

a) Porcentaje de corchos aceptables por cada máquina

- Para la Máquina A, la probabilidad de que un corcho sea aceptable es

$$\begin{aligned}P(2.9 \leq X \leq 3.1) &= P\left(Z \leq \frac{3.1 - 3}{0.1}\right) - P\left(Z \leq \frac{2.9 - 3}{0.1}\right) \\&= P(Z < 1) - P(Z < -1) \\&= 0.8413 - 0.1587 = 0.6826\end{aligned}$$

Por lo tanto, el porcentaje de corchos aceptables producidos por la máquina A es 68.26%.

Ejemplo

- Para la Máquina B, la probabilidad de que un corcho sea aceptable con la máquina B es:

$$\begin{aligned}P(2.9 \leq X \leq 3.1) &= P\left(\frac{2.9 - 3.04}{0.02} \leq Z \leq \frac{3.1 - 3.04}{0.02}\right) \\&= P(Z < 3) - P(Z < -7) \approx 0.9987 - 0 = 0.9987\end{aligned}$$

Por lo tanto, el porcentaje de corchos aceptables producidos por la máquina B es 99.87%.

- Conclusión: La máquina B produce un mayor porcentaje de corchos aceptables.

Ejemplo

b) Probabilidad de que un corcho seleccionado al azar sea aceptable

- Usamos el teorema de la probabilidad total considerando las proporciones de corchos producidos por cada máquina.

$$P(\text{aceptable}) = P(\text{aceptable} \mid A) \cdot P(A) + P(\text{aceptable} \mid B) \cdot P(B)$$

Donde:

1. $P(\text{aceptable} \mid A) = 0.6826$
2. $P(\text{aceptable} \mid B) = 0.9987$
3. $P(A) = 0.40$
4. $P(B) = 0.60$

- Entonces:

$$\begin{aligned} P(\text{aceptable}) &= (0.6826 \times 0.40) + (0.9987 \times 0.60) \\ &= 0.27304 + 0.59922 = 0.87226 \end{aligned}$$

- La probabilidad de que un corcho seleccionado al azar sea aceptable es aproximadamente 0.872 o 87.2%.

Ejemplo: Continuación

Para la maquina que produce mayor porcentaje de corchos aceptables:

1. cual es la probabilidad de que el diámetro este comprendido entre 2,99 y 3,05 cm?
2. Si se seleccionan al azar 10 corchos de la producción, ¿cuál es la probabilidad de que el diámetro este comprendido entre 2,99 y 3,05 cm en por lo menos dos de los 10?
3. Hallar los percentiles 30 y 70 para la variable diámetro del corcho cortado con la máquina. elegida.

Ejemplo: Resolución

Para la máquina que produce mayor porcentaje de corchos aceptables (Máquina B):

1. La probabilidad de que el diámetro esté entre 2.99 y 3.05 cm es:

$$\begin{aligned}P(2.99 \leq X \leq 3.05) &= P\left(\frac{2.99 - 3.04}{0.02} \leq X \leq \frac{3.05 - 3.04}{0.02}\right) \\&= P(Z < 0.5) - P(Z < -2.5) \\&= 0.6915 - 0.0062 = 0.6853\end{aligned}$$

esto es, 68.53%

2. Observemos que la probabilidad de que un corcho tenga un diámetro en ese intervalo es $p = 0.685$. Si tenemos una muestra de tamaño 10, X el número de corchos con diámetro en el intervalo tiene distribución binomial $B(n = 10, p = 0.685)$.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$$

Ejemplo: Resolución

$$P(X = 0) = \binom{10}{0} \cdot 0.685^0 \cdot (1 - 0.685)^{10} = (0.315)^{10} \approx 0.0007$$

$$P(X = 1) = \binom{10}{1} \cdot 0.685^1 \cdot (1 - 0.685)^9 \approx 10 \cdot 0.685 \cdot 0.0007 = 0.0048$$

Entonces:

$$P(X \geq 2) = 1 - (0.0007 + 0.0048) = 1 - 0.0055 = 0.9945$$

La probabilidad de que el diámetro esté entre 2.99 y 3.05 cm en por lo menos dos de los 10 corchos es aproximadamente 0.9945 o 99.45%.

Ejemplo: Resolución

3. Hallar los percentiles 30 y 70 para el diámetro de los corchos Para la máquina B, usamos la distribución normal con media 3.04 y desviación estándar 0.02.

- El valor Z correspondiente al percentil 30 es aproximadamente -0.524.

$$X_{30} = \mu + Z \cdot \sigma = 3.04 + (-0.524) \cdot 0.02 = 3.04 - 0.01048 = 3.02952$$

El percentil 30 es aproximadamente 3.0295 cm.

- El valor Z correspondiente al percentil 70 es aproximadamente 0.525.

$$X_{70} = \mu + Z \cdot \sigma = 3.04 + 0.525 \cdot 0.02 = 3.04 + 0.0105 = 3.0505$$

El percentil 70 es aproximadamente 3.0505 cm.