



Clase 7: Distribución Normal y propiedades. Aproximación Normal de la Binomial

Profesora: Olga Alexandra Bustos Giraldo

Escuela de Estadística

Universidad Nacional de Colombia, Sede Medellín

oabustos@unal.edu.co



Distribución continua: Normal

La distribución Normal es la más importante en toda la Ciencia de Probabilidad y Estadística. Muchas poblaciones numéricas tienen distribuciones que pueden ser representadas muy fácilmente por una curva normal apropiada. Los ejemplos incluyen:

- a) Estaturas, pesos y otras características físicas.
- b) Errores de medición en experimentos científicos.
- c) Mediciones antropométricas en fósiles
- d) Tiempos de reacción en experimentos psicológicos.
- e) Mediciones de inteligencia y aptitud.



Distribución continua: Normal

La distribución Normal juega un papel clave en la aplicación de la Inferencia Estadística, puesto que las distribuciones de muchas estadísticas muestrales tienden hacia la distribución Normal conforme crece el tamaño de la muestra. Las pruebas de hipótesis tienen su fundamento en esta distribución.



Distribución Normal

Definición

Sea X una variable aleatoria continua; se dice que X tiene una distribución normal, si su f.d.p. es de la forma:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} ; \quad \begin{matrix} -\infty < x < \infty \\ \mu \in \mathbf{R} , \sigma > 0 \end{matrix} .$$

Por notación se escribe $X \sim N(\mu, \sigma^2)$. A las cantidades μ y σ^2 se les conoce como parámetros de localización y de escala, respectivamente.

Cuando $\mu = 0$ y $\sigma^2 = 1$, se obtiene una distribución normal especial, conocida como *Normal Estándar* y usualmente denotada con la letra Z .

Se escribe $Z \sim N(0, 1)$.



Distribución Normal

Sea X una variable aleatoria continua tal que $X \sim N(\mu, \sigma^2)$. Las siguientes afirmaciones se cumplen para la variable aleatoria X .

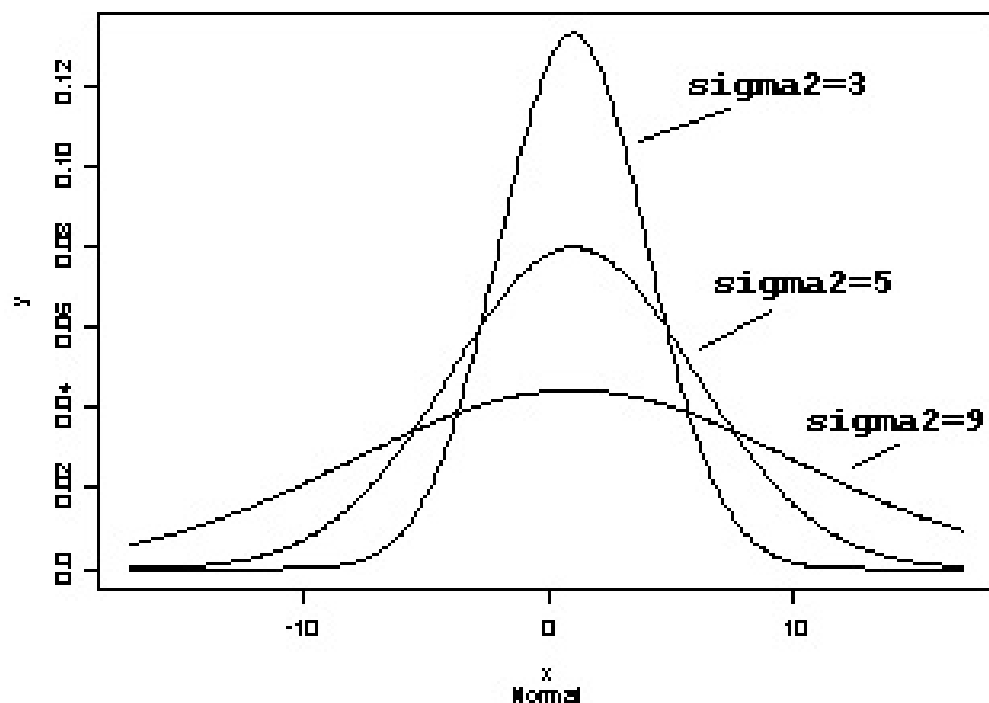
- El área bajo la curva normal comprendida entre $(-\infty, \infty)$ es igual a 1.
- La normal es simétrica respecto a μ . Como consecuencia de esto, el área bajo la curva en el intervalo $[\mu, \infty)$ es igual a 0.5.
- La distribución normal tiene forma de campana.
- $E[X] = \mu$ y $Var[X] = \sigma^2$.
- Cuanto más grande es el parámetro σ^2 el gráfico de la función es mas 'achatada' y de colas más largas.
- El centro de la campana (punto de simetría) es tanto la media de la distribución como la mediana (aquel valor que divide los datos en dos partes porcentualmente iguales).



Distribución Normal

A continuación se ilustran curvas normales con el mismo valor de $\mu = 0$ y diferentes valores de σ^2

Distribuciones Normales con distintas Varianzas





Distribución Normal

Suponga que $X \sim N(\mu, \sigma^2)$ y que se quiere hallar $P(x_1 < X < x_2)$. Observe que para hallar esta probabilidad debe resolverse la siguiente integral:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx .$$

Esta integral no puede ser resuelta de manera explícita; se requiere de métodos numéricos para aproximarla. Además, cada vez que se modifique alguno de los parámetros μ o σ^2 , se debe calcular de nuevo dicha integral.

Para evitar este problema, se utiliza el cambio de variable $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$. El proceso de transformar una v.a normal en una normal estándar se conoce como *Estandarización*.



Distribución Normal

Para el cálculo de probabilidades con una variable $N(0, 1)$. La f.d.p y la f.d.a están dadas por:

La f.d.p:

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

La f.d.a:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz .$$

De esta manera se tiene que:

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = \Phi(z_2) - \Phi(z_1) .$$



Propiedades de la Normal estándar

Sea Z una variable aleatoria tal que $Z \sim N(0, 1)$:

- $P(Z < -z) = P(Z > z)$.
- $P(Z \geq -z) = P(Z \leq z)$.
- $P(-z < Z < 0) = P(0 < Z < z)$.
- $P(Z < 0) = P(Z > 0) = \frac{1}{2}$.
- $P(-z_1 < Z < -z_2) = P(z_2 < Z < z_1)$.



Ejemplo de la Normal Estándar

Suponga que Z es una variable aleatoria tal que $Z \sim N(0, 1)$. Calcule las siguientes probabilidades:

- a) $P(Z < 1.32)$
- b) $P(Z < 3)$
- c) $P(Z > 1.45)$
- d) $P(Z > 2.15)$
- e) $P(1.76 < Z < 2.34)$
- f) $P(-1 < Z < 1)$



Ejemplo de la Normal estándar

Solución

a) $P(Z < 1.32) = 0.9066$, el código en R es: `pnorm(1.32)`

b) $P(Z < 3) = 0.9987$, el código en R es: `pnorm(3)`

c) $P(Z > 1.45) = 1 - P(Z \leq 1.45) = 1 - 0.9265 = 0.0735$, el código en R es: `1-pnorm(1.45)`

d) $P(Z > 2.15) = 1 - P(Z \leq 2.15) = 1 - 0.9842 = 0.0158$, el código en R es: `1-pnorm(2.15)`

e) $P(1.76 < Z < 2.34) = P(Z < 2.34) - P(Z < 1.76) = 0.9904 - 0.9608 = 0.0296$, el código en R es: `pnorm(2.34) - pnorm(1.76)`

$$\begin{aligned} \text{f) } P(-1 < Z < 1) &= P(Z < 1) - P(Z < -1) = P(Z < 1) - P(Z > 1) \\ &= P(Z < 1) - [1 - P(Z < 1)] = 2P(Z < 1) - 1 \end{aligned}$$

$$P(-1 < Z < 1) = 2P(Z < 1) - 1 = 2(0.8413) - 1 = 0.6826$$

En R: 2*pnorm(1) - 1

= $P(Z < 1) - P(Z < -1)$. **En R:** pnorm(1)-pnorm(-1)

pnorm: La función pnorm() calcula la probabilidad acumulada de una normal estándar.



Ejemplo de la Normal estándar

Suponga que Z es una variable aleatoria tal que $Z \sim N(0, 1)$. Determine el valor de a para el cual se cumple que:

a) $P(Z < a) = 0.9$

b) $P(Z > a) = 0.05$

c) $P(-1.24 < Z < a) = 0.8$



Ejemplo de la Normal

Solución

a) $P(Z < a) = 0.9 \longrightarrow a = 1.28$

El código en R es: `qnorm(0.9)`

b) $P(Z > a) = 0.05 \iff P(Z < a) = 0.95 \longrightarrow a = 1.645$

El código en R es: `qnorm(0.95)`

`qnorm(0.05, lower.tail = F)`

c) $P(-1.24 < Z < a) = 0.8$

$$P(-1.24 < Z < a) = 0.8 \iff P(Z < a) - P(Z < -1.24) = 0.8 .$$

$$P(Z < a) = 0.8 + P(Z < -1.24) = 0.8 + 0.107 = 0.907 .$$

$$P(Z < a) = 0.9075 \longrightarrow a = 1.32$$

El código en R es: `qnorm(0.907)`

`qnorm`: La función `qnorm()` determina el valor de z de una probabilidad dada.



Ejemplo de la Normal

La nota promedio obtenida por cada estudiante de cierto curso tiene una distribución aproximadamente normal con una nota promedio de 3.3 y una desviación estándar de 0.2. Si se desea que solo el 5 % de todos los estudiantes de dicho curso aprueben, ¿Cuál debe ser la nota mínima para que esto sea posible?

Sea X : nota obtenida por un estudiante. Del enunciado se tiene que $X \sim N(3.3, 0.04)$.



Ejemplo de la Normal

Sea a la nota mínima aprobatoria que satisface $P(X > a) = 0.05$

$$\begin{aligned} P(X > a) = 0.05 &\Rightarrow P\left(\frac{X - 3.3}{0.2} > \frac{a - 3.3}{0.2}\right) = 0.05 \\ \Rightarrow P\left(Z > \frac{a - 3.3}{0.2}\right) &= 0.05 \quad (\text{En } R: \text{qnorm}(0.05, \text{lower.tail} = \text{FALSE})) \\ \Rightarrow 1 - P\left(Z \leq \frac{a - 3.3}{0.2}\right) &= 0.05 \\ \Rightarrow P\left(Z \leq \frac{a - 3.3}{0.2}\right) &= 0.95 \end{aligned}$$

El código en R es: `qnorm(0.95)=1.64`

$\frac{a - 3.3}{0.2} = 1.64 \Rightarrow a = 3.628$. Nota mínima aprobatoria para que sólo el 5% apruebe.

También, puede calcularse directamente, sin estandarizar.

`qnorm(0.05, 3.3, 0.2, lower.tail = F)`



Ejemplo de la Normal

El diámetro de un cable eléctrico está distribuido normalmente con un diámetro medio de 0.8 plg y una desviación estándar de 0.02.

- a) ¿Qué proporción de cables tiene un diámetro superior a 0.81 plg?
- b) Un cable es defectuoso si su diámetro difiere del promedio en más de 0.025 plg. ¿Qué proporción de cables son defectuosos?



Ejemplo de la Normal

Solución

Sea X la variable aleatoria que representa el diámetro del cable. Se tiene que $X \sim N(0.8, 0.0004)$.

a)

$$\begin{aligned} P(X > 0.81) &= 1 - P(X \leq 0.81) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{0.81 - 0.8}{0.02}\right) \\ &= 1 - P(Z \leq 0.5) = 1 - 0.6915 = 0.3085 \end{aligned}$$

El código en R es: `1-pnorm(0.5)`

Sin estandarizar, `1-pnorm(0.81,0.8,0.02)`

El 30.85 % de los cables tienen un diámetro superior a 0.81 plg.



Ejemplo de la Normal

b)

$$P(|X - 0.8| > 0.025) = 1 - P(|X - 0.8| \leq 0.025)$$

$$= 1 - P(-0.025 \leq X - 0.8 \leq 0.025)$$

$$= 1 - P\left(\frac{-0.025}{0.02} \leq \frac{X - 0.8}{0.02} \leq \frac{0.025}{0.02}\right)$$

$$= 1 - P(-1.25 \leq Z \leq 1.25) = 0.21$$

$$= 1 - (P(Z \leq 1.25) - P(Z \leq -1.25)).$$

En R: 1-(pnorm(1.25) - pnorm(-1.25))

Sin estandarizar: 1-(pnorm(0.825,0.8,0.02)-pnorm(0.775,0.8,0.02))



Teorema de aproximación

Sea $X \sim \text{bin}(n, p)$, si n es grande y p no es muy cercano a cero o a 1, si $\mu = np$ y $\sigma^2 = np(1 - p)$, entonces la v.a. X la podemos aproximar con una v.a. $Y \sim N(\mu, \sigma^2)$, una regla conservativa es que la aproximación es adecuada si $np \geq 5$ y $n(1 - p) \geq 5$. Además se debe usar un factor de corrección por continuidad (la Normal es una distribución continua y la Binomial es discreta) de 0.5, así, sea $a \in \mathbb{Z}^+$, entonces:

$$P(X \leq a) \approx P(Y \leq a + 0.5)$$

$$P(X < a) \approx P(Y \leq a - 0.5)$$

$$P(X \geq a) = 1 - P(X < a) \approx 1 - P(Y < a - 0.5)$$

$$P(X > a) = 1 - P(X \leq a) \approx 1 - P(Y \leq a + 0.5)$$

$$P(X = a) = P(X \leq a) - P(X < a) \approx P(Y \leq a + 0.5) - P(Y < a - 0.5)$$



Ejemplo aproximación Normal

Un proceso de fabricación produce un 2% de chips defectuosos. Suponga que la determinación de esta característica es independiente para cada chip y 1000 de ellos son seleccionados. Calcule la probabilidad de que este lote contenga más de 25 chips defectuosos, usando la aproximación normal.

Solución

Sea X : # de chips defectuosos en los 1000 seleccionados. Entonces

$X \sim \text{bin}(1000, 0.02)$. Como $np = 1000(0.02) = 20 > 5$ y

$n(1 - p) = 1000(0.98) = 980 > 5$, entonces:

$$\begin{aligned} P(X > 25) &= 1 - P(X \leq 25) = 1 - P(Y \leq 25.5), \quad Y \sim N(20, 19.6) \\ &\approx 1 - P\left(Z \leq \frac{25.5 - 20}{\sqrt{19.6}}\right) = 1 - P(Z \leq 1.24) = 1 - 0.8925 = 0.1075 \end{aligned}$$

El código en R es: `1-pnorm(1.24)`

Sin estandarizar: `1-pnorm(25.5,20,4.42)`



Ejemplo aproximación Normal

Los alambres que se utilizan en una cierta computadora deben tener una resistencia entre 0.12 y 0.14 ohms. Por experiencia se sabe que dichas resistencias se distribuyen normalmente, con media de 0.13 ohms y desviación estándar de 0.005 ohms.

- a) ¿Cuál es la probabilidad de que un alambre seleccionado al azar de la producción de la compañía satisfaga las especificaciones?

- b) Si se utilizan cuatro de estos alambres de la compañía en el sistema, ¿Cuál es la probabilidad de que los cuatro satisfagan las especificaciones? Asuma independencia.



Ejemplo aproximación Normal

Solución

- a) Sea X : resistencia de un alambre de computadora de la compañía. Por el enunciado se tiene que $X \sim N(0.13, (0.005)^2)$. La probabilidad pedida es:

$$\begin{aligned} P(0.12 \leq X \leq 0.14) &= P\left(\frac{0.12 - 0.13}{0.005} \leq Z \leq \frac{0.14 - 0.13}{0.005}\right) \\ &= P(-2 \leq Z \leq 2) = 2P(Z \leq 2) - 1 \\ &= 2(0.9772) - 1 = 0.9544. \end{aligned}$$

El código en R es: `2 pnorm(2) - 1`

Sin estandarizar: `pnorm(0.14,0.13,0.005)-pnorm(0.12,0.13,0.005)`

La calidad de los alambres se puede considerar óptima.



Ejemplo aproximación Normal

b) Sea Y : número de alambres de la compañía que cumplen las especificaciones entre los 4 seleccionados.

Esta variable toma los valores $y = 0, 1, 2, 3, 4$.

éxito : cuando un alambre cumple las especificaciones.

Por el numeral anterior se sabe que $P(\text{éxito}) = 0.9544$.

Como se asume independencia la probabilidad de éxito es constante.

$Y \sim \text{bin}(4, 0.9544)$.

La probabilidad pedida es:

$$P(Y = 4) = \binom{4}{4} (0.9544)^4 (1 - 0.9544)^0 = 0.8297 .$$

O sea que la probabilidad de que todos los alambres funcionen en el sistema se puede considerar alta.



Ejercicios propuestos

1. Suponga que X tiene una distribución normal con media 10 y desviación estándar 2. Calcule lo siguiente:

a) $P(X < 13)$

b) $P(X > 9)$

c) $P(6 < X < 14)$

d) $P(2 < X < 4)$

e) $P(-2 < X < 8)$



Ejercicios propuestos

2. Suponga que X tiene una distribución normal con media 5 y desviación estándar 4. Obtenga el valor de x que resuelve cada una de las siguientes probabilidades:

a) $P(X > x) = 0.5$

b) $P(X > x) = 0.95$

c) $P(x < X < 9) = 0.2$

d) $P(3 < X < x) = 0.95$

e) $P(-x < X < x) = 0.99$



Ejercicios propuestos

3. La resistencia a la compresión de una serie de muestras de cemento puede modelarse con una distribución normal con media 6000 kilogramos por centímetro cuadrado, y una desviación estándar de 100 kilogramos por centímetro cuadrado.
- a) ¿Cuál es la probabilidad de que la resistencia de una muestra sea menor que 6250 kg/cm^2 ?
- b) ¿Cuál es la probabilidad de que la resistencia de una muestra se encuentre entre 5800 y 5900 kg/cm^2 ?



Ejercicios propuestos

4. El volumen de una máquina de llenado automático, que se deposita en latas de una bebida gaseosa, tiene una distribución normal con media 12.4 onzas de líquido y desviación estándar de 0.1 onzas de líquido.
- a) ¿Cuál es la probabilidad de que el volumen depositado sea menor que 12 onzas de líquido?
 - b) Si se desechan todas las latas que tienen menos de 12.1 o más de 12.6 onzas de líquido, ¿cuál es la proporción de latas desechadas?