Capítulo

$\boxed{5}$

Distribuciones Continuas

Las distribuciones continuas mas comunes son:

1. Distribución Uniforme

2. Distribución Normal

3. Distribución Exponencial

4. Distribución Gamma

5. Distribución Beta

6. Distribución Cauchy

7. Distribución Chi-Cuadrado

8. Distribución t-Student

9. Distribución F de Snedecor

5.1. Distribución Normal

5.1.1. Introducción

Es la distribución continua con mayor aplicación en estadística. Su propio nombre indica su extendida utilización, justificada por la frecuencia o normalidad con la que ciertos fenómenos tienden a parecerse en su comportamiento a esta distribución. Muchas variables aleatorias continuas presentan una función de densidad cuya gráfica tiene forma de campana. En otras ocasiones, al considerar por ejemplo distribuciones binomiales, $Bin(n, p)$, para un mismo valor de p y valores de n cada vez mayores, se ve que sus polígonos de frecuencias se aproximan a una curva en "forma de campana".
En resumen, la importancia de la distribución normal se debe principalmente a que hay muchas variables asociadas a fenómenos naturales que siguen el modelo de la normal. Como por ejemplo:

1. Caracteres morfológicos de individuos (personas, animales, plantas,etc) entre ellos: tallas, pesos, diámetros, perímetros.

2. Caracteres fisiológicos, por ejemplo: efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.

3. Caracteres sociológicos, por ejemplo: consumo de cierto producto por un mismo grupo de individuos, puntuaciones de examen.

4. Caracteres psicológicos, por ejemplo: cociente intelectual, grado de adaptación a un medio,...

5. Errores cometidos al medir ciertas magnitudes.

6. Valores estadísticos muestrales, por ejemplo : la media.

7. Otras distribuciones como la binomial o la de Poisson son aproximaciones normales, ...

El uso extendido de la distribución normal en las aplicaciones estadísticas puede explicarse, además, por otras razones. Muchos de los procedimientos estadísticos habitualmente utilizados asumen la normalidad de los datos observados. Aunque muchas de estas técnicas no son demasiado sensibles a desviaciones de la normal y, en general, esta hipótesis puede obviarse cuando se dispone de un número suficiente de datos, resulta recomendable contrastar siempre si se puede asumir o no una distribución normal. La simple exploración visual de los datos puede sugerir la forma de su distribución. No obstante, existen otras medidas, gráficos de normalidad y contrastes de hipótesis que pueden ayudarnos a decidir, de un modo más riguroso, si la muestra de la que se dispone procede o no de una distribución normal. Cuando los datos no sean normales, podremos o bien transformarlos o emplear otros métodos estadísticos que no exijan este tipo de restricciones (los llamados métodos no paramétricos).

5.1.2. La Distribución Normal

La distribución normal fue reconocida por primera vez por el francés Abraham de Moivre (1667-1754). Posteriormente, Carl Friedrich Gauss (1777-1855) elaboró desarrollos más profundos y formuló la ecuación de la curva; de ahí que también se la conozca, más comúnmente, como la çampana de Gauss".

La distribución de una variable normal está completamente determinada por dos parámetros, su media y su desviación estándar, denotadas generalmente por μ y σ. La función de densidad de la distribución normal se define de la siguiente manera:

Definición 5.1 *Se dice que una variable aleatoria X tiene una distribución de prob-*

abilidad normal con media μ y varianza σ^2, se representa por $X \sim N(\mu, \sigma^2)$ si y sólo si, la función de densidad de X es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} \qquad \begin{array}{l} -\infty < x < +\infty \\ -\infty < \mu < +\infty \\ \sigma > 0 \end{array} \tag{5.1}$$

5.1.3. Propiedades de la Distribución Normal

La distribución Normal tiene las siguientes propiedades

1. La distribución normal está completamente especificada por sus dos parámetros, la media μ y la varianza σ^2. La media es el parámetro de centralidad y la varianza mide su dispersion. Por lo tanto el gráfico de la función de densidad depende de dichos valores.

   En la figura 1, se muestran los gráficos de la distribución normal para varios valores de μ manteniendo fija σ^2, donde se aprecia que el gráfico de desplaza a la derecha a medida de que el valor de μ aumenta.

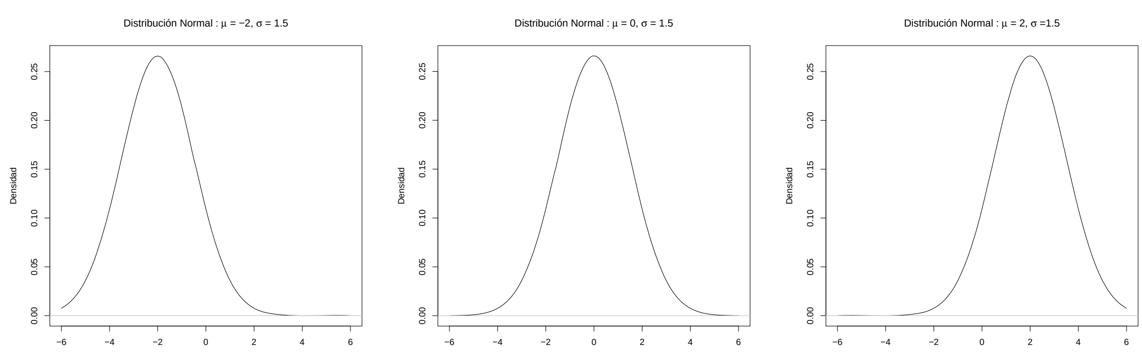

Figura 5.1: Distribución Normal: (a) $\mu = -2$ y $\sigma = 1{,}5$, (b) $\mu = 0$ y $\sigma = 1{,}5$, (c) $\mu = 2$ y $\sigma = 1{,}5$

Por otro lado, en la figura 2, se varía el valor de σ^2 manteniendo fijo el valor de

μ, provocando esta situación que a medida que aumenta el valor de σ^2 la gráfica se va extendiendo sobre el eje X, es decir las colas se hacen mas grandes.

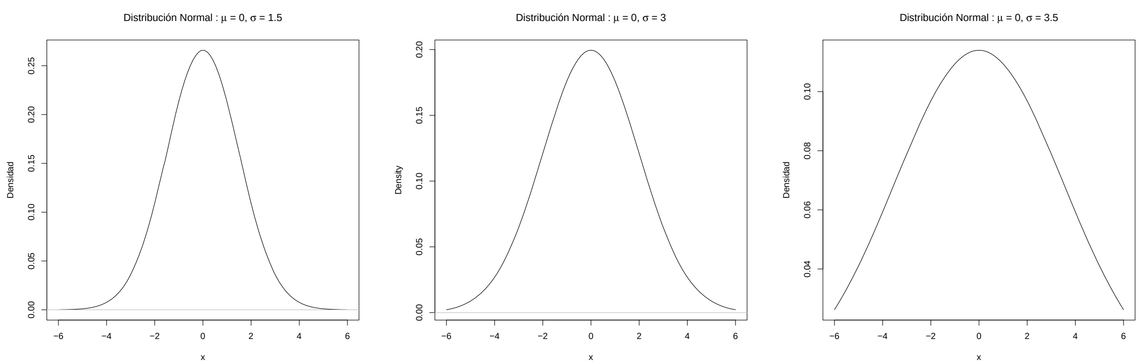

Figura 5.2: Distribución Normal: (a) $\mu = 0$ y $\sigma = 1{,}5$, (b) $\mu = 0$ y $\sigma = 2$, (c) $\mu = 0$ y $\sigma = 3{,}5$

2. La distribución es simétrica con respecto a la media, μ.

3. Tiene una única moda, que coincide con su media y su mediana.

4. La gráfica de la función de densidad tiene forma de campana.

5. La esperanza de una variable aleatoria normal es su media es decir, $E(X) = \mu$

6. La varianza de una variable aleatoria normal es σ^2

7. La forma de la campana de Gauss depende de los parámetros

8. La curva normal es asintótica al eje de abscisas. Por ello, cualquier valor entre $-\infty$ y ∞ es teóricamente posible. El área total bajo la curva es, por tanto, igual a 1.

9. El área bajo la curva da la distribución normal son probabilidades.

5.1.4. Cálculo de Probabilidades

Como se comentó antes las probabilidades en el caso de la distribución normal (en general si la variable aleatoria es continua) está representada por el área bajo la curva de la función de densidad. Entonces

1. $P(X = x) = 0$, ya que nu punto no tiene área.
2. $P(X < x)$ esta representada por al área bajo la curva a la izquierda de x como se ve en la figura ***
3. $P(X < x)$ esta representada por al área bajo la curva a la derecha de x como se ve en la figura ***
4. $P(a < X < b)$ esta representada por al área bajo la curva que esta entre los valores de a y b como se ve en la figura ***

Para hallar esa área se debe utilizar una herramienta del cálculo diferencial conocida como integración, de la siguiente manera,

1. $P(X < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} dx$
2. $P(X > x) \int_{x}^{-\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} dx$
3. $P(a < X < b) = \int_{a}^{b} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} dx$

Aún así, la integrales que habría que resolver tiene solución haciendo uso de técnicas numéricas de integración, la cuales están fuera del alcance de este texto. Para poder hallar las probabilidades debemos en primer lugar estandarizar la variable X, que es lo que viene en la siguiente sección.

5.1.5. Distribución Normal Estándar

Cuando $\mu = 0$ y $\sigma = 1$ la distribución se conoce con el nombre de normal estándar.

Definición 5.2 *Sea X una variable aleatoria normal con media μ y varianza σ^2 ($X \sim N(\mu, \sigma^2)$). Entonces la variable aleatoria definida como*

$$Z = \frac{X - \mu}{\sigma} \tag{5.2}$$

se dice tener una distribución normal estándar ($Z \sim N(\emptyset, 1)$). Otra manera de decirlo es, se ha estandarizado la variable X.

Definición 5.3 $Z \sim N(\emptyset, 1)$ *si y sólo si su función de densidad de probabilidad está dada por*

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad -\infty < z < +\infty \tag{5.3}$$

Una de las ventajas de la estandarización es que la distribución no depende de los parámetros, pues en este caso la media siempre será cero y la varianza uno. Por tanto, la distribución es única y el gráfico de la función de densidad también. La otras propiedades enunciadas sobre la distribución normal son heredadas por la estándar.

Esta propiedad resulta especialmente interesante en la práctica, ya que para la distribución N(0,1) existen tablas publicadas a partir de las que se puede obtener de modo sencillo la probabilidad de observar un dato menor o igual a un cierto valor z, y que permitirán resolver preguntas de probabilidad acerca del comportamiento de variables de las que se sabe o se asume que siguen una distribución aproximadamente normal.

La tabla que nosotros vamos a usar es la tabla *****, lo cuál se emplea de la siguiente manera

****

Ejemplo 5.1.1 *Sea $Z \sim N(0,1)$, calcular*

1. $P(Z < 2)$
2. $P(Z > 2)$
3. $P(-2 < Z < 2)$
4. $P(0 < Z < 1{,}73)$

Solución

1. *Buscamos este valor en la tabla de la siguiente manera, Z=2 es equivalente a Z=2.00, por lo tanto en la primera columna buscamos la fila con el valor 2.0 la cual la interceptamos con la columna de la primera fila con el valor 0. El valor obtenido es 0.9772, es decir,* $P(Z < 2) = 0{,}9772$
2. *Para calcular esta probabilidad no podemos usar directamente la tabla, primero debemos trabajar con el complemento de la siguiente manera,* $P(Z > 2) = 1 - P(Z \leq 2)$. *Ahora buscamos en la tabla como se hizo en el caso anterior y obtenemos que* $P(Z > 2) = 1 - P(Z \leq 2) = 1 - 0{,}9772 = 0{,}0228$
3. $P(-2 < Z < 2)$
4. $P(0 < Z < 1{,}73)$