Módulo 4 – Inferencia Estadística

Distribución de Probabilidad Continua

Ciencia de Datos

Objetivos de Aprendizaje

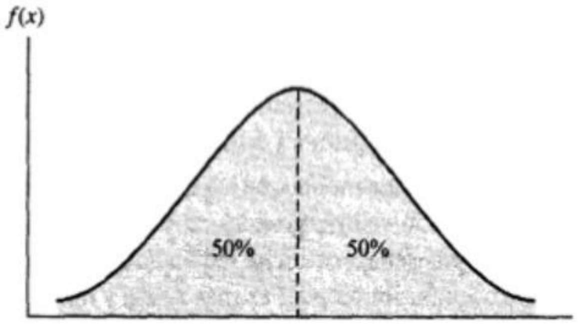
- Calcular probabilidades a partir de la distribución normal.
- Resolver problemas utilizando la distribución normal.

Contenido

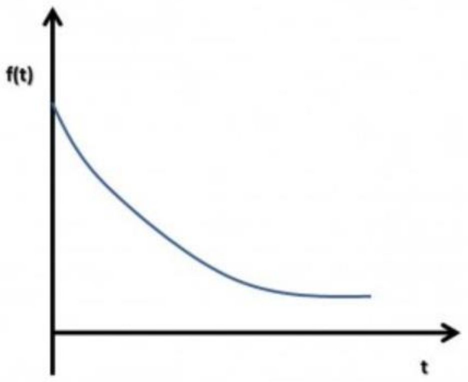
1. Distribuciones de probabilidad continuas.
2. Distribución normal.

Distribuciones de Probabilidad Continuas

Una función de densidad de probabilidad es una expresión matemática que define la distribución de los valores para una variable aleatoria continua.



Distribución Normal



Distribución Exponencial

Ejemplos de distribuciones de probabilidad continuas son:

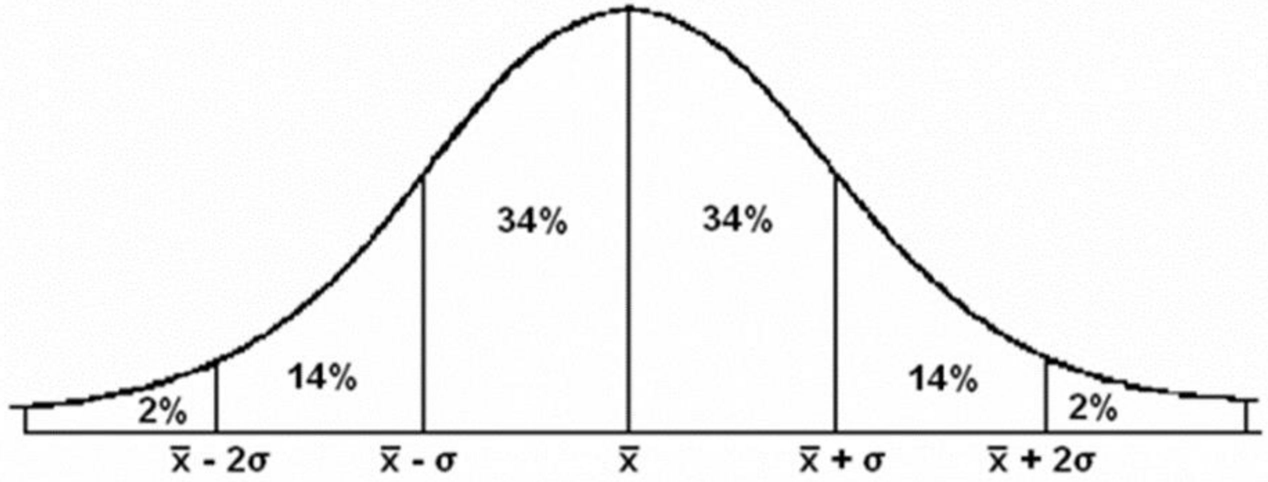
- Distribución uniforme continua.
- Distribución normal.
- Distribución chi-cuadrado.
- Distribución t de Student.
- Distribución exponencial.
- Distribución beta.
- Distribución gamma.
- Distribución de Pareto.

Distribución Normal

Distribución Normal

- La distribución normal, también conocida como distribución gaussiana, es una distribución de probabilidad continua que se utiliza comúnmente en estadística para modelar muchos fenómenos naturales y sociales.

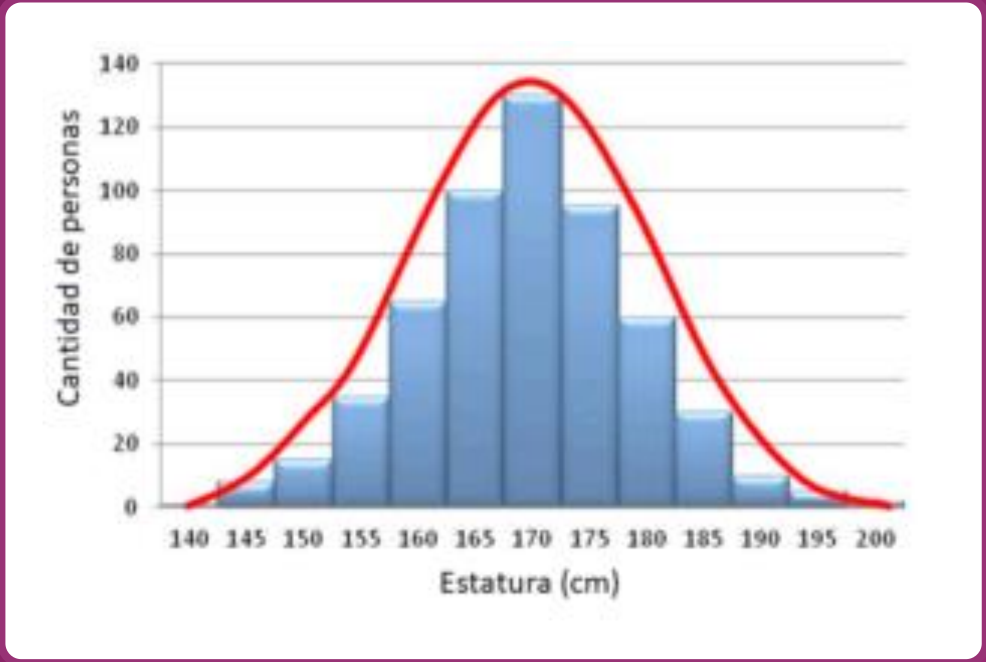
- La distribución normal tiene una forma de campana simétrica y está completamente determinada por dos parámetros: la **media (µ)** y la **desviación estándar (σ).** La media define el centro de la distribución y la desviación estándar indica qué tan extendida está la distribución alrededor de la media.



Fenómenos que pueden seguir una Distribución Normal

La altura de la población:
Si tomamos una muestra aleatoria de personas de una población, es probable que la altura de la mayoría de ellas siga una distribución normal.

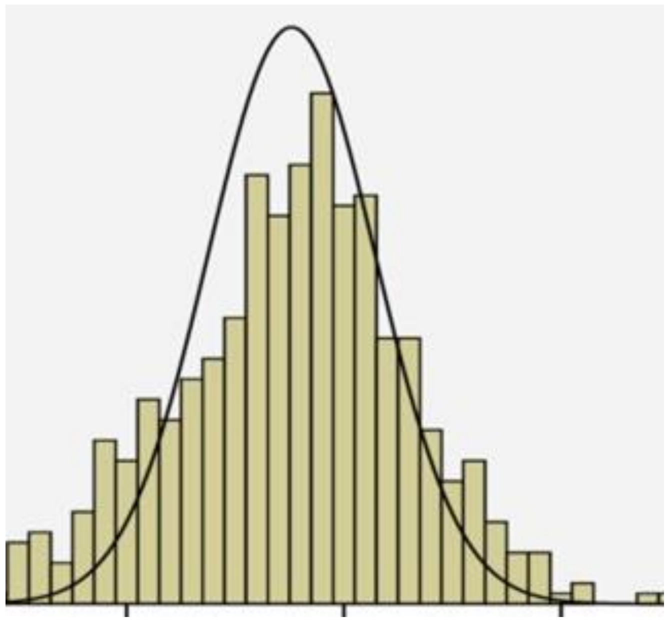
Los puntajes en un examen estandarizado:
Si los puntajes en un examen siguen una distribución normal, podemos utilizar la media y la desviación estándar para interpretar los resultados y comparar el rendimiento de los estudiantes.



Fenómenos que pueden seguir una Distribución Normal

El tiempo que tarda en atenderse a los clientes en una tienda.
Si medimos el tiempo que tarda en atenderse a los clientes en una tienda, es posible que los datos sigan una distribución normal, lo que podría ayudar a los gerentes a planificar y mejorar el servicio al cliente.

Los errores de medición en un instrumento de medición.
Si medimos una cantidad utilizando un instrumento de medición, es probable que los errores de medición sigan una distribución normal. Esto puede ser importante para evaluar la precisión del instrumento y la incertidumbre de las mediciones.



Importancia de la Distribución Normal

La distribución normal es muy importante para la estadística debido a las siguientes razones:

- Diversas variables continuas que son comunes en la administración tienen distribuciones que se parecen a la distribución normal.
- La distribución normal se puede utilizar para aproximar varias distribuciones de probabilidad discretas.
- La distribución normal conforma las bases para la inferencia estadística clásica debido a su relación con el teorema del límite central.

Características de la Distribución Normal

Principales características de la distribución normal:

- Es simétrica, por lo que su media y su mediana son iguales.
- A la distribución normal le corresponde un media cero y una desviación típica o estándar de 1. La desviación típica o estándar indica la separación que existe entre un valor cualquiera de la muestra y la media.
- Tiene la apariencia de una campana.
- En una distribución normal, se puede determinar con exactitud qué porcentaje de los valores estará dentro de cualquier rango específico.
- Tiene un rango infinito, es decir, $-\infty < X < \infty$.

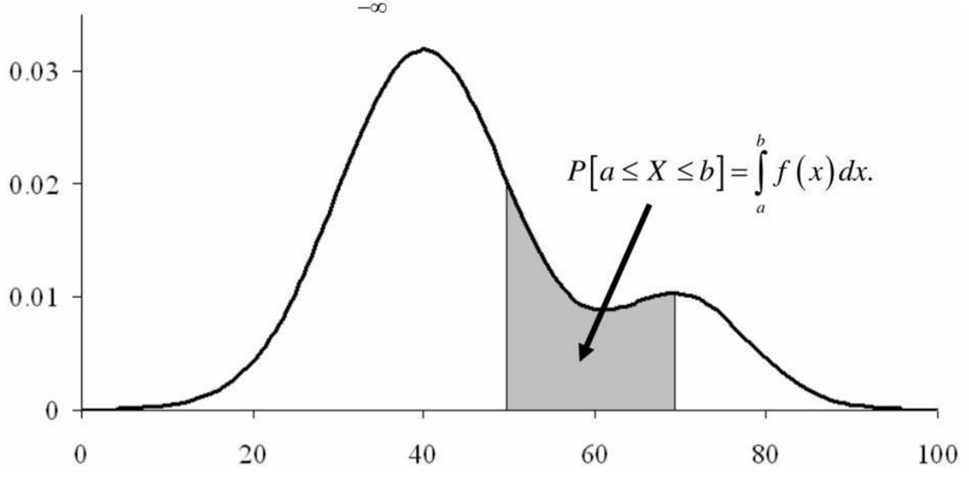
Función de Densidad de Probabilidad

Una función de densidad de probabilidad es una función matemática que se utiliza para describir la distribución de probabilidad de una variable aleatoria continua. En otras palabras, **indica la probabilidad relativa de que una variable aleatoria tome un valor en un intervalo determinado**.

La función de densidad de probabilidad está definida para cada valor posible de la variable aleatoria y se integra sobre todo el rango de **valores para dar una probabilidad total de 1**. Es importante tener en cuenta que la **Función de Densidad de Probabilidad** no da la probabilidad de un valor específico de la variable aleatoria, sino la probabilidad relativa de que la variable aleatoria tome un valor en un intervalo dado.

Función de Densidad de Probabilidad , $f(x)$

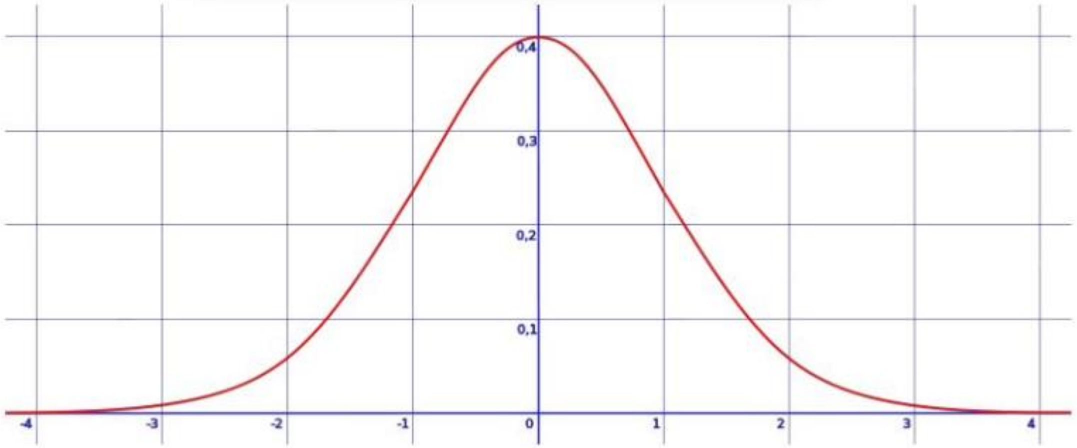
$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$



Función de Densidad de Probabilidad Normal

El símbolo $f(X)$ se utiliza para representar una función de densidad de probabilidad. **La función de densidad de probabilidad para la distribución normal está dada por**:

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{1}{2}\right)\left[\frac{X-\mu}{\sigma}\right]^2}$$



Cálculo de Probabilidades con la Distribución Normal

Para calcular probabilidades con la distribución normal, primero se convierte la variable aleatoria distribuida normalmente, X, en una **variable aleatoria normal estándar**, Z, utilizando la siguiente fórmula de transformación:

$$Z = \frac{X - \mu}{\sigma} \sim N(0; 1)$$

La aplicación de esta fórmula permite la búsqueda de valores en una tabla de probabilidades de la normal para evitar los cálculos tediosos y complejos que exige la definición de la normal.

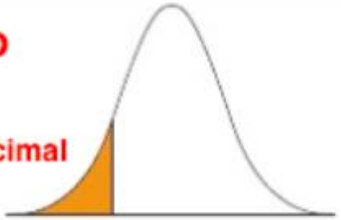
Cálculo de Probabilidades con la Distribución Normal

DISTRIBUCIÓN NORMAL TIPIFICADA

El número que busco

Parte entera
y primer decimal

Segundo decimal



$P(Z \leq z)$

Probabilidades

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
–3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
–3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
–3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
–3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
–3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
–3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
–2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
–2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
–2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
–2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
–2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048

KIBERNUM

Ejemplo 1

Supongamos que la puntuación en un examen de matemáticas sigue una distribución normal con una **media de 75** y una **desviación estándar de 10**. Queremos calcular la probabilidad de obtener una **puntuación mayor o igual a 85** en el examen (considere que las notas van de 0 a 100).

Ejemplo 1

Para hacer esto, primero estandarizamos la variable aleatoria utilizando la fórmula de estandarización para la distribución normal:

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

Donde:
X = la puntuación en el examen que queremos evaluar (en este caso, 85)
μ = la media de la distribución normal (75)
σ = la desviación estándar de la distribución normal (10)

Entonces, sustituimos los valores conocidos:

$$Z = \frac{85 - 75}{10} = 1$$

Ejemplo 1

Una vez que hemos estandarizado la variable aleatoria, podemos buscar la probabilidad correspondiente en la tabla de distribución normal estándar o utilizar una calculadora que nos proporcione esta información (o bien consultar una librería estadística en Python).

La probabilidad de que Z sea mayor o igual a 1 es de aproximadamente 0.1587.

Por lo tanto, la **probabilidad de obtener una puntuación mayor o igual a 85 en el examen es de aproximadamente 0.1587 o 15.87%**

Ejemplo 1

Implementación en Python

```python
import numpy as np
import matplotlib.pyplot as plt
# importamos librería scipy
from scipy.stats import norm
```

✓ 0.6s Python

```python
# Definir los parámetros de la distribución normal
mu = 75
sigma = 10

# Calcular la probabilidad de obtener una nota mayor o igual a 85 en el examen
probabilidad = 1 - norm.cdf(85, mu, sigma)

# Mostrar la probabilidad
print("La probabilidad de obtener una nota mayor o igual a 85 en el examen es:", round(probabilidad, 4))
```

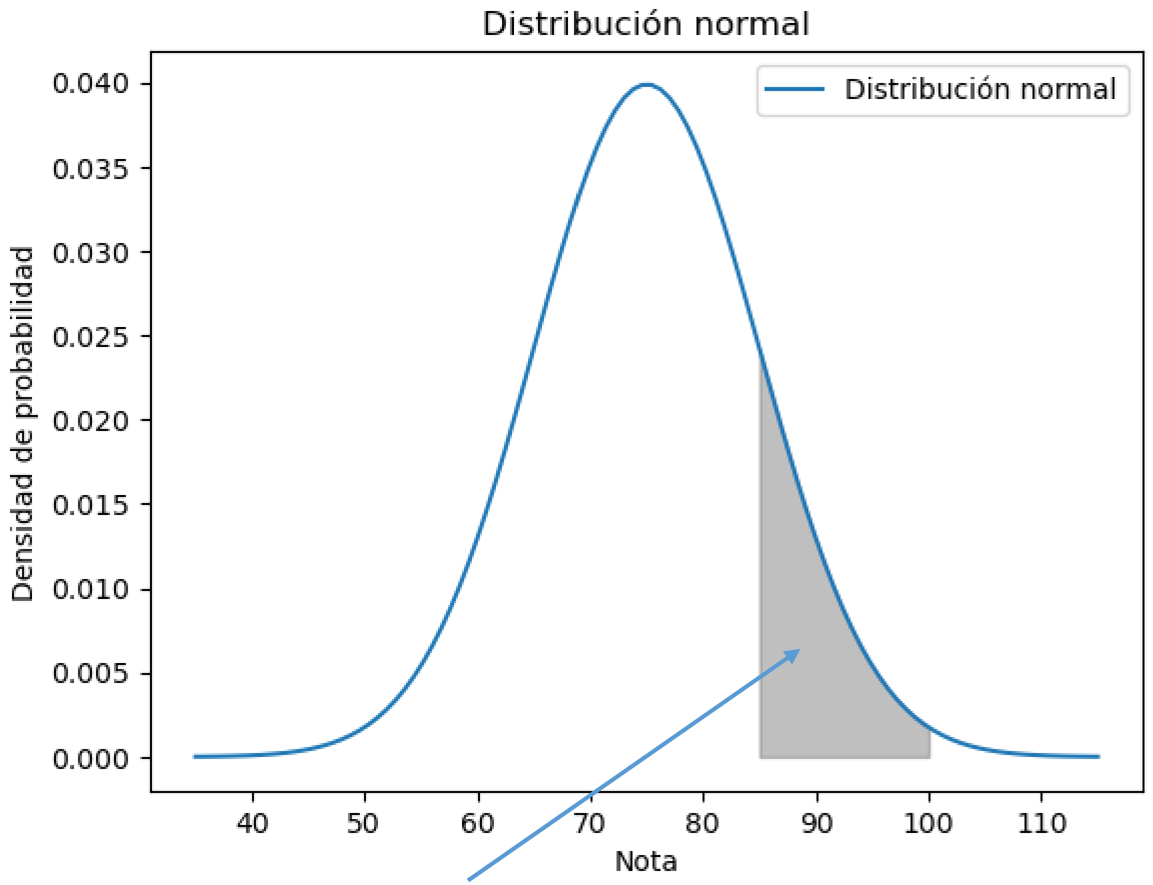
✓ 0.0s Python

```
La probabilidad de obtener una nota mayor o igual a 85 en el examen es: 0.1587
```

Ejemplo 1

Visualización

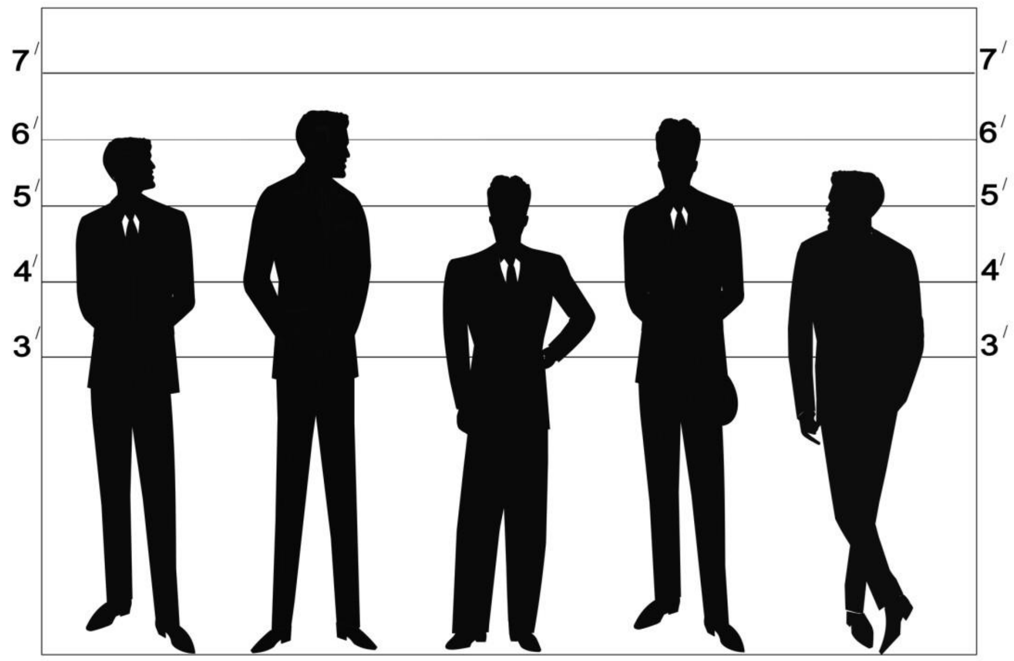
```python
# Crear un arreglo de valores para la variable aleatoria
x = np.linspace(mu - 4*sigma, mu + 4*sigma, 100)

# Calcular la función de densidad de probabilidad
pdf = norm.pdf(x, mu, sigma)

# Graficar la distribución normal
plt.plot(x, pdf, label='Distribución normal')

# Sombras para el intervalo de interés
x_fill = np.linspace(85, 100, 100)
y_fill = norm.pdf(x_fill, mu, sigma)
plt.fill_between(x_fill, y_fill, color='grey', alpha=0.5)

# Configurar la gráfica
plt.title('Distribución normal')
plt.xlabel('Altura (cm)')
plt.ylabel('Densidad de probabilidad')
plt.legend()

# Mostrar la gráfica
plt.show()
```
✓ 0.1s



El área representa la probabilidad de 15.87

Ejemplo 2

Supongamos que la altura de una población de hombres adultos sigue una distribución normal con una media de 175 cm y una desviación estándar de 7 cm. Queremos calcular la probabilidad de que un hombre adulto seleccionado al azar tenga una altura entre 170 cm y 180 cm.



KIBERNUM

Ejemplo 2

Para hacer esto, primero estandarizamos las dos puntuaciones de altura utilizando la fórmula de estandarización para la distribución normal:

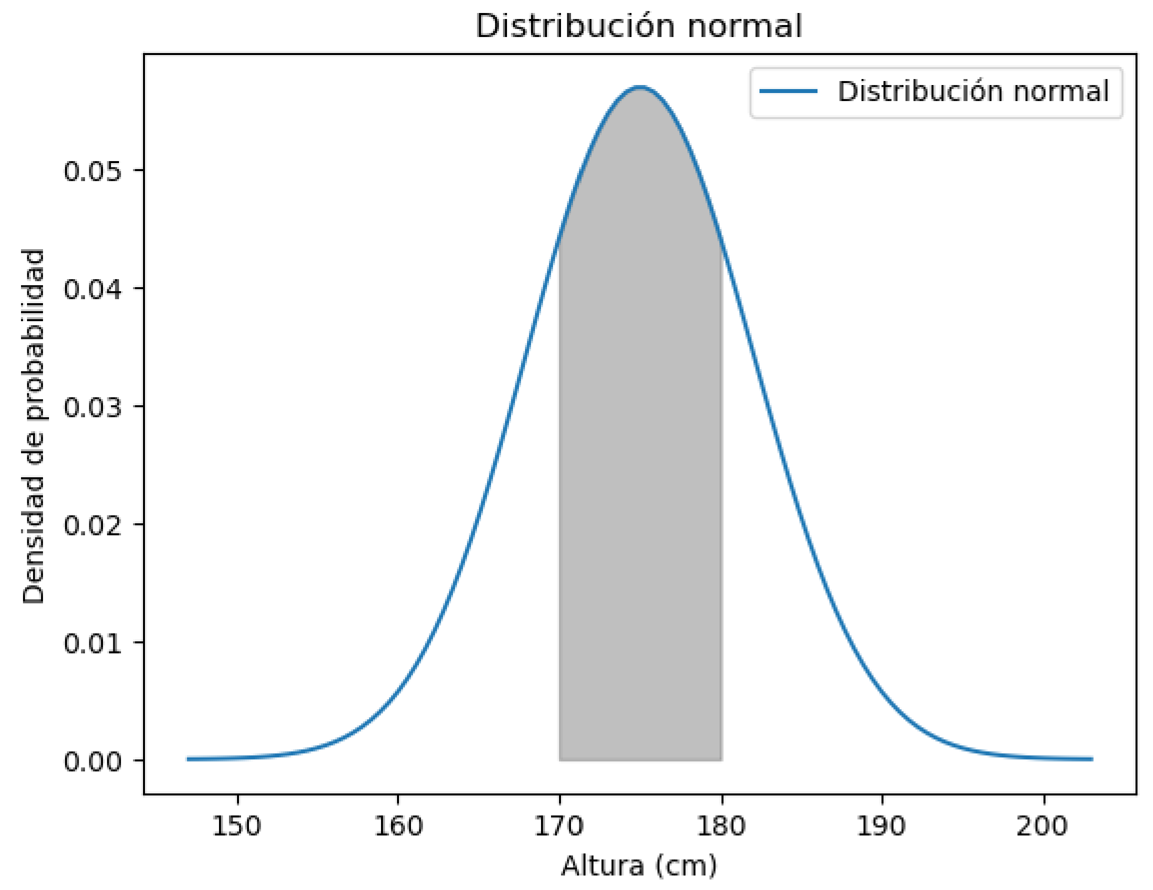
$$Z_1 = \frac{170 - 175}{7} = -0.71 \qquad Z_2 = \frac{180 - 175}{7} = 0.71$$

Una vez que hemos estandarizado ambas variables aleatorias, podemos buscar las probabilidades correspondientes en la tabla de distribución normal estándar o programar unas líneas de código en Python.

En este caso, queremos encontrar la probabilidad de que Z1 sea mayor o igual a -0.71 y Z2 sea menor o igual a 0.71, lo que representa el área debajo de la curva de la distribución normal entre estas dos puntuaciones.

$$Prob = P(Z_2) - P(Z_1) = 0.7625 - 0.2375 = 0.5249$$

Ejemplo 2



Podemos encontrar estas probabilidades en la tabla de distribución normal estándar o utilizar una calculadora de probabilidad. En este caso, la probabilidad es aproximadamente 0.5328 o 53.28%

Por lo tanto, la probabilidad de que un hombre adulto seleccionado al azar tenga una altura entre 170 cm y 180 cm es de aproximadamente 0.5328 o 53.28%

KIBERNUM

Dudas y consultas

Fin presentación