```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
```

Distribución normal teórica

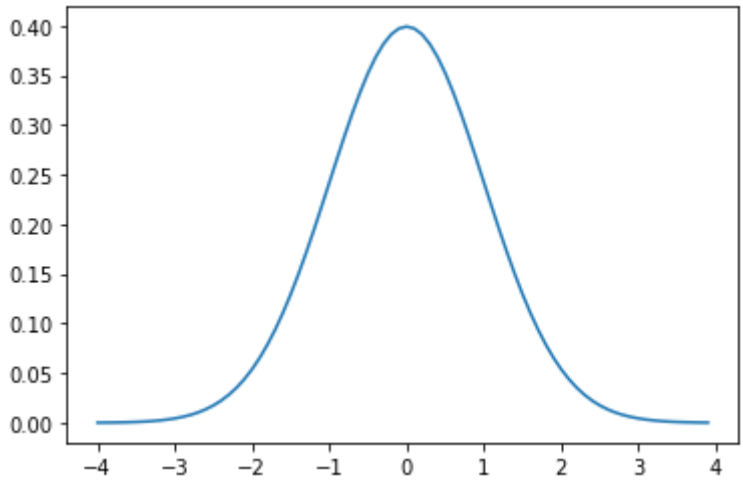
$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

- μ: media de la distribución
- σ: desviación estándar de la distribución

```
# definimos nuestra distribución gaussiana
def gaussian(x, mu, sigma):
  return 1/(sigma*np.sqrt(2*np.pi))*np.exp(-0.5*pow((x-mu)/sigma,2))
```

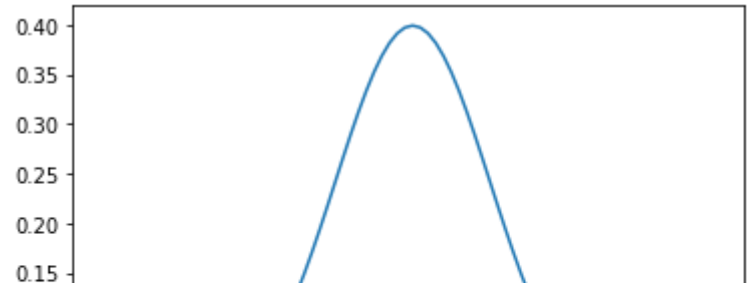
```
x = np.arange(-4,4,0.1)
y = gaussian(x, 0.0, 1.0)

plt.plot(x, y)
```

```
[<matplotlib.lines.Line2D at 0x7faae7c028d0>]
```



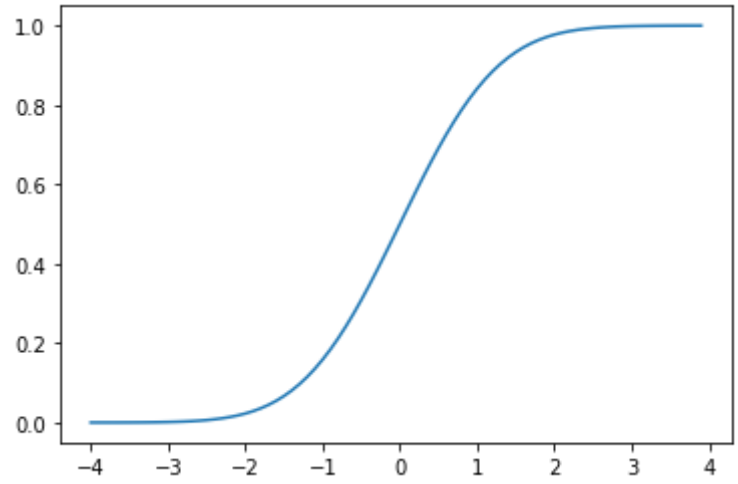

```
# usando scipy
dist = norm(0, 1)
x = np.arange(-4,4,0.1)
y = [dist.pdf(value) for value in x]
plt.plot(x, y)
```

```
[<matplotlib.lines.Line2D at 0x7faade103a90>]
```




```
# calculando la distribución acumulada
dist = norm(0, 1)
x = np.arange(-4,4,0.1)
y = [dist.cdf(value) for value in x]
plt.plot(x, y)
```

```
[<matplotlib.lines.Line2D at 0x7faade0ebfd0>]
```



Distribución normal (gausiana) a partir de los datos

- *El archivo excel* lo puedes descargar en esta página: https://seattlecentral.edu/qelp/sets/057/057.html

```
df = pd.read_excel('s057.xls')
arr = df['Normally Distributed Housefly Wing Lengths'].values[4:]
values, dist = np.unique(arr, return_counts=True)
print(values)
plt.bar(values, dist)
```

```
[37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55]
<BarContainer object of 19 artists>
```


```
# estimación de la distribución de probabilidad
mu = arr.mean()

#distribución teórica
sigma = arr.std()
dist = norm(mu, sigma)
x = np.arange(30,60,0.1)
y = [dist.pdf(value) for value in x]
plt.plot(x, y)

# datos
values, dist = np.unique(arr, return_counts=True)
plt.bar(values, dist/len(arr))
```

```
<BarContainer object of 19 artists>
```