

Creado por:

Isabel Maniega

```
In [9]: import pandas as pd
import math
```

Otras librerías de interes para calcular los estadísticos son:

- <https://docs.python.org/3/library/statistics.html>

## Ejemplo 1

Estamos interesados en la altura promedio de los árboles en un bosque en particular. Para obtener resultados rápidamente, hicimos que 5 estudiantes salieran y midieran una muestra de 20 árboles. Cada estudiante regresó con la altura promedio de los árboles de sus muestras. Resultados de la muestras:

- 35.23
- 36.71
- 33.21
- 38.2
- 35.54

Si se sabe que el promedio de la población de la altura de los árboles en el bosque es de 36 pies con la desviación estandar de 2 pies. ¿A cuántos errores estándar se aleja el promedio de los estudiantes de la media de la población?

Para resolver este problema, primero necesitamos encontrar el promedio de los promedios de estos estudiantes para que:

```
In [2]: df = pd.DataFrame({"Media_arboles": [35.23, 36.71, 33.21, 38.2, 35.54]})
df
```

```
Out[2]:  Media_arboles
0      35.23
1      36.71
2      33.21
3      38.20
4      35.54
```

```
In [3]: x = df.mean()
x
```

```
Out[3]: Media_arboles    35.778
dtype: float64
```

Ahora encontramos nuestro error estándar de la muestra:

SE = desviacion estandar / raiz cuadrada(numero de muestra)

```
In [5]: SE = 2 / math.sqrt(5)
SE
```

```
Out[5]: 0.8944271909999159
```

Ahora, para obtener el número de errores estándar de la media de nuestra observación, podemos utilizar la formula de la puntuación Z:

```
In [6]: mu = 36
```

```
In [7]: Zscore = (round((x - mu) / SE, 2))
Zscore
```

```
Out[7]: Media_arboles    -0.25
dtype: float64
```

Así que nuestra distribución de muestra es relativamente cercana a la distribución de la población

## Problema 1

Una población distribuida normalmente tiene una media de 100 y una desviacion estándar de 20. ¿Cuál es la puntuación Z de una media muestral de 110, tomada de una muestra de tamaño 4?

```
In [11]: mu = 100
s = 20
x = 110
n = 4
```

```
In [ ]: SE = s/math.sqrt(n)
```

## Problema 2

El tiempo medio conocido que se tarda en entregar una pizza es de 22.5 minutos con una desviación estándar de 2 minutos. Pedí pizza todas las semanas durante las últimas 10 semanas y obtuve un tiempo de 21.5 minutos. ¿Cuál es la probabilidad de obtener este promedio?

## Problema 3

Si sigo pidiendo pizzas por la toda la eternidad. ¿A qué nivel puedo esperar que se acerque este promedio?

## Ejemplo más claro

Supongamos que un fabricante de arandelas toma una arandela al azar y mide su diámetro. El fabricante sabe que el diámetro de las arandelas que fabrica sigue una distribución normal con promedio de 2,54 cm y desviación estándar de 0,01 cm. Al medir una arandela al azar, esta mide 2,52 cm. El fabricante desea saber si esta medida está significativamente alejada del promedio.

```
In [1]: mu = 2.54
s = 0.01
x = 2.52
```

```
In [4]: Zscore = (round((x - mu) / s, 2))
Zscore
```

```
Out[4]: -2.0
```

Este resultado significa que la medida de la arandela se encuentra 2 desviaciones estándar por debajo de la media

Supongamos que el mismo fabricante de arandelas anterior se preocupó porque el valor del diámetro de la arandela que midió resultó demasiado pequeño para ser atribuido únicamente a variaciones aleatorias del diámetro. Entonces decide investigar más a fondo ,por lo que toma una muestra aleatoria de arandelas y mide el diámetro de cada una de ellas para luego calcular la media muestral y ver si las evidencias experimentales sugieren que hay un problema con la línea de producción. En este caso, se desea llevar a cabo el mismo cálculo del puntaje Z anterior, pero utilizando el valor medio de la muestra de 100 arandelas en lugar del valor del diámetro de una sola arandela, la cual es poco probable que sea representativa de todas las arandelas que produce la fábrica. La media de la muestra resultó ser 2,53 cm.

En esta situación, se sabe que la media muestral ( $\bar{X}$ ) también sigue una distribución normal cuya media es la misma media poblacional ( $\mu$ ) pero que tiene una desviación estándar igual a la desviación estándar poblacional dividida entre la raíz cuadrada del tamaño muestral.

```
In [7]: mu = 2.54
s = 0.01
x = 2.53
n = 100
```

```
In [10]: SE = s/math.sqrt(n)
```

```
In [11]: Zscore = (round((x - mu) / SE, 2))
Zscore
```

```
Out[11]: -10.0
```

Este resultado significa que el valor observado de la media muestral está diez desviaciones estándar muestrales por debajo de la media poblacional.

Creado por:

Isabel Maniega