

```
In [8]: import sqlite3
import pandas as pd
import sqlalchemy
import numpy as np
import scipy
from scipy import stats
from scipy.stats import norm

probability_pdf = norm(0,1).cdf(1.5)
print (probability_pdf)

0.9331927987311419
```

S07 T01: Test de hipótesis.

- Ejercicio 1

Coge un conjunto de datos de tema deportivo que te guste y selecciona un atributo del conjunto de datos.
Calcula el **p-valor**

y di si rechaza la **hipótesis nula** cogiendo un alfa de 5%.

```
In [10]: con= sqlite3.connect("database.sqlite") # importamos la base de datos.

df = pd.read_sql_query( "SELECT * from Player_Attributes", con)
df.head()
```

```
Out[10]:
```

	id	player_fifa_api_id	player_api_id	date	overall_rating	potential	preferred_foot	attacking_work_rate	defen:
0	1	218353	505942	2016-02-18 00:00:00	67.0	71.0	right	medium	
1	2	218353	505942	2015-11-19 00:00:00	67.0	71.0	right	medium	
2	3	218353	505942	2015-09-21 00:00:00	62.0	66.0	right	medium	
3	4	218353	505942	2015-03-20 00:00:00	61.0	65.0	right	medium	
4	5	218353	505942	2007-02-22 00:00:00	61.0	65.0	right	medium	

5 rows × 42 columns

Escogemos la variable " **pierna dominante** "

```
In [11]: dff= df [ ["preferred_foot"] ]
```

Según un estudio realizado en 2009 un 12% de la población mundial es zurda. Aunque si miramos los países,

este porcentaje varia

en función del país, siendo holanda el país con más personas zurdas(13,23%), mientras que en ciertos países asiáticos el número

de zurdos se reduce por debajo del 5%, seguramente a una estigmatización social, y al intento de intentar usa la diestra.

<https://es.statista.com/grafico/22552/porcentaje-de-personas-zurdas-en-paises-seleccionados/#:~:text=Se%20calcula%20que%20aproximadamente%20un,elaborado%20una%20comparaci%C3%B>

En el fútbol, debido al propio juego y a que es un juego en que las simetrías respecto a las longitudinal del campo son

importantes, es posible que el número de zurdos sea mucho mayor al de la población mundial (12%). En el fútbol se da mucha

importancia a tener un lateral zurdo, y en menor medida o un central zurdo y un extremo. El primero es fundamental en casi

cualquier equipo tanto para la fase defensiva como ofensiva. Mientras que para los otros dos puestos, es más fácil ver a

diestros. Teniendo en cuenta que si hay 11 jugadores por equipo, y al menos habrán dos diestros, Así que mi hipótesis nula es

- $H_0: p > p_0 \quad p_0 = 0.2$

Siendo la hipótesis alternativa .

- $H_1: p < p_1$

Realizaremos el test de hipótesis unilateral de Bernoulli, que es como el test de Student o el Z, pero aplicado a identificar

un valor p de probabilidad. Cómo es sabido, la media coincide con p, tal que el estadístico será la media y la región de

rechazo:

- $\bar{x} < p_0 - Z_\alpha \sqrt{\frac{p_0(1 - p_0)}{n}}$

Z_α es el valor de la variable aleatoria , normal estándar $N(0,1)$, cuya probabilidad de que $P(Z > Z_\alpha) = \alpha$, y n es el tamaño de la muestra

Para los test de hipótesis de Bernoulli el tamaño de la muestra ha de ser grande. $n > 30$ (que 30 suele ser aceptado en

literatura metmática como un muestreo mínimamente aceptable). α es el nivel de significación

In [21]:

```
# hacemos un muestreo aleatorio simple de n= 1000
n= 1000
muestra =dff.sample(n = 1000)

muestra.head()
```

Out[21]:

	preferred_foot
22606	left
46729	right
17424	right
76489	left
124239	right

```
#calculamos el valor de la media(p)
pq= muestra.value_counts()
pq
```

Out[22]:

preferred_foot	
right	749
left	244

dtype: int64

```
In [52]: pqarr= pq.to_numpy()
p= pqarr[1]/ (pqarr[1]+pqarr[0])
q= 1-p
# en una prueba de hipótesis de Bernoulli, la esperanza coincide con p y la variación con p

#calcularemos la posibilidad el p-valor para saber si aceptamos la hipótesis o no.

print ( "el valor p es :" , p )
```

el valor p es : 0.2457200402819738

```
In [30]: #definimos las siguientes variables.
p0= 0.2 # de la hipótesis nula

# definimos el a0 como el valor mínimo de alfa por el que se rechazará la hipótesis.
```

despejando α de la ecuación encontramos el valor mínimo de α , sea a_0

$$\bullet \bar{x} < p_0 - Z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$$

```
In [54]: z= -(p-p0)* (( p0*(1-p0)/n )**(-0.5)) # sólo es una variable auxiliar para evitar la fórmula
alpha= 0.05
```

```
In [55]: a0= norm(0,1).cdf(z) # es el valor mínimo de alpha por el que será aceptada la hipótesis nula
print (a0)
```

0.00015047171885920055

```
In [56]: scipy.stats.binom_test ( pqarr[1], n=(pqarr[1]+pqarr[0]), p= 0.2, alternative='greater' )
```

Out[56]: 0.00025195130700063595

```
In [57]: if (alpha >= a0) and ( p>= p0 ) :
        print ( " la hipótesis nula no es rechazada")
```

```
else:  
    print ( " la hipótesis nula es rechaza")
```

la hipótesis nula no es rechazada