

# Repaso Probabilidad y Estadística Aplicada

2023

Sebastián Decuadro - Maglis Mujica - Pablo Dubourdieu

Hay 2 conceptos centrales de la teoría de probabilidad y estadística.

Sobre los que se basan los métodos de estimación y prueba de hipótesis.

Estos son:

- Teorema Central del Límite
- Ley de los Grandes Números

# Teorema Central del Límite

Bajo ciertas condiciones, la distribución de las medias de muestras aleatorias tiende a aproximarse a una distribución normal a medida que el tamaño de la muestra aumenta, independientemente de la forma de la distribución de la población original.

Condiciones del teorema:

- Independencia: Las observaciones dentro de cada muestra deben ser independientes entre sí.
- Tamaño de muestra suficientemente grande: A medida que el tamaño de la muestra aumenta, la aproximación a una distribución normal mejora.

Implicaciones del teorema:

- La media de las muestras se aproxima a la media de la población.
- La distribución de las medias muestrales tiende a ser normal, independientemente de la forma de la distribución de la población original.

# Ley de los Grandes Números

Establece que, a medida que el tamaño de la muestra aumenta, la media de una muestra tiende a acercarse a la media de la población.

Condiciones:

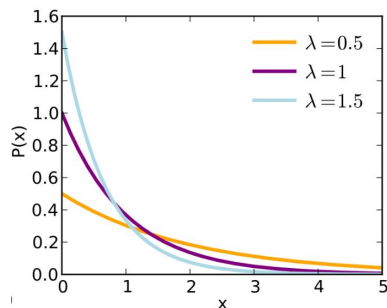
- Independencia: Las observaciones dentro de cada muestra deben ser independientes entre sí.
- Identidad de distribución: Las observaciones deben ser extraídas de la misma distribución de probabilidad.

Implicaciones del teorema:

- La aproximación a una distribución normal es más precisa a medida que el tamaño de la muestra aumenta.

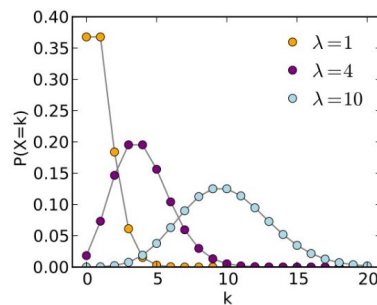
# Gráficos de Distribuciones

## Distribución Geométrica

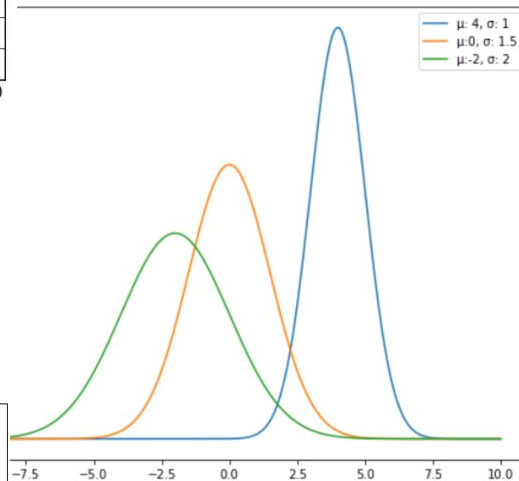


## Distribuciones asintóticamente normales

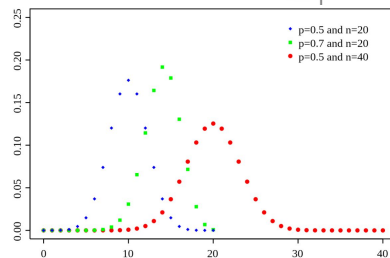
### Distribución Poisson



### Distribución Normal



### Distribución Binomial



# Estimación y Prueba de Hipótesis

Aplicaciones del teorema:

- Estimación de parámetros:

El teorema central del límite nos permite hacer suposiciones sobre la distribución de las medias muestrales y, por lo tanto, realizar estimaciones de parámetros poblacionales, como la media o la proporción.

- Prueba de hipótesis:

La aproximación a una distribución normal nos permite realizar pruebas de hipótesis sobre los parámetros poblacionales utilizando estadísticos de prueba basados en las medias muestrales.

# “Todos los modelos están mal pero algunos son útiles.”\*

- Confiar en una estimación implica aceptar cierta imprecisión.  
Es el precio a pagar por disponer de un conocimiento que de otra forma resultaría inalcanzable.
- El objetivo de las estimaciones no es conocer con exactitud sino proveer aproximaciones razonables y honestas que si bien difieren de la realidad, pueden dar información valiosa para la toma de decisiones.

# Algunos paquetes de Python

<code>import pandas as pd</code>	Análisis y manipulación de datos ( <a href="https://pandas.pydata.org/">https://pandas.pydata.org/</a> )
<code>import numpy as np</code>	Operaciones matemáticas y manipulación de arrays. ( <a href="https://numpy.org/">https://numpy.org/</a> )
<code>import matplotlib.pyplot as plt</code>	Visualizaciones y gráficos. ( <a href="https://matplotlib.org/">https://matplotlib.org/</a> )
<code>import scipy.stats as st</code>	Optimización, ecuaciones algebraicas y estadísticas. ( <a href="https://scipy.org/">https://scipy.org/</a> )
<code>import seaborn as sns</code>	Visualizaciones y gráficos. ( <a href="https://seaborn.pydata.org/">https://seaborn.pydata.org/</a> )
<code>import math</code>	Módulo incorporado en Python que proporciona una amplia gama de funciones matemáticas y constantes predefinidas. ( <a href="https://docs.python.org/3/library/math.html">https://docs.python.org/3/library/math.html</a> )

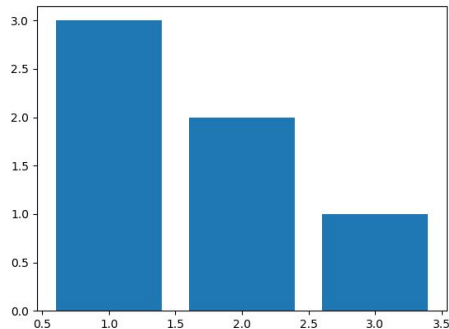
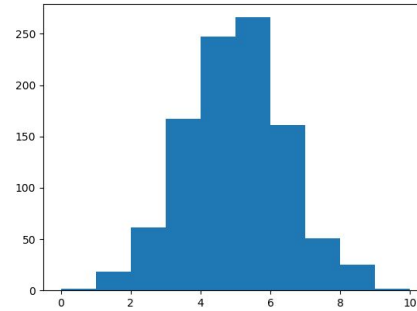
Nota: Este resumen no contempla todas las funcionalidades de los paquetes.



# Estadística Descriptiva

- Histograma

```
import numpy as np
import matplotlib.pyplot as plt
fig, ax = plt.subplots()
x = np.random.normal(5, 1.5, size=1000)
ax.hist(x, np.arange(0, 11))
plt.show()
```

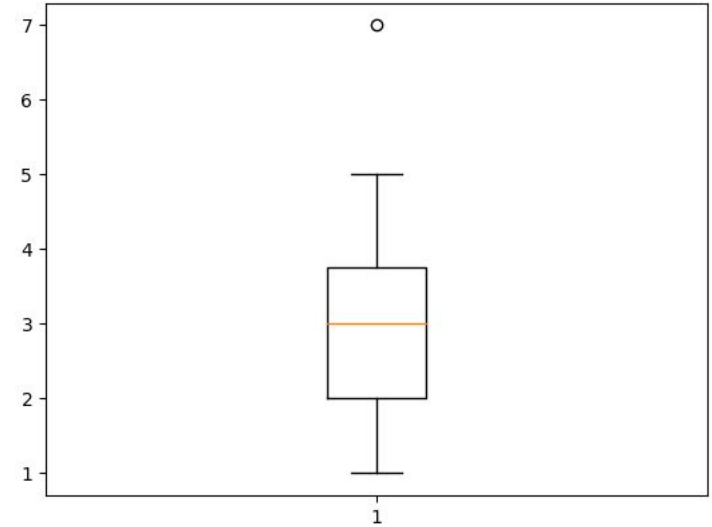


- Bar Plot

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots()
ax.bar([1, 2, 3], [3, 2, 1])
plt.show()
```

- Diagramas de caja y bigotes (Boxplot):

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots()
ax.boxplot([1, 2, 1, 2, 3, 4, 3, 3, 5, 7])
plt.show()
```



Es una representación gráfica que muestra la distribución de un conjunto de datos a través de los cuartiles.

Consiste en:

- una caja que representa el rango intercuartil (IQR)
- una línea que marca la mediana
- los "bigotes" que se extienden desde la caja para mostrar los valores extremos o atípicos.

Nota: Investigar el parámetro "whis" para excluir outliers.

# Prueba de Hipótesis

```
#H_0) mu = 5
#H_0) mu <> 5
popmean = 5
a = sample_mean
# Prueba de hipótesis
t_statistic, p_value = st.ttest_1samp (a, popmean)
#Calculate the T-test for the mean of ONE group of scores.
#Parameters:
#           a           array_like. Sample observation.
#           popmean     float or array_like. Expected value in null hypothesis.

# Comparar el valor p con el nivel de significancia
nivel_significancia = 0.05
if p_value < nivel_significancia :
    print("Se rechaza la hipótesis nula." )
else:
    print("No se rechaza la hipótesis nula." )
```

Algunas funciones útiles

`scipy.stats.ttest_ind`

[https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest\\_ind.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html)

`scipy.stats.ttest_1samp`

[https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest\\_ind.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html)

# Anexo

- Espacio Muestral
- Definición de Probabilidad Condicional
- Independencia de Eventos
- Ley de Probabilidad Total
- Fórmula de Bayes
- Variables Aleatorias
- Distribuciones

# Espacio muestral

## Definición (Espacio muestral)

*Al conjunto de resultados posibles de un experimento aleatorio se le llama espacio muestral, se le denota con la letra  $\Omega$ .*

## Ejemplo

En el experimento de tirar un dado se tiene que

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

# Definición de Probabilidad Condicional

## Definición

Sean  $A$  y  $B$  dos eventos tales que  $P(B) > 0$ . La probabilidad de  $A$  condicionada a la ocurrencia del evento  $B$  se define de la siguiente manera

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Independencia de eventos

## Definición

Decimos que dos eventos  $A$  y  $B$  son independientes si

$$P(A \cap B) = P(A)P(B)$$



# Ley de Probabilidad Total

## Teorema

*Sean  $A$  y  $B$  dos eventos con  $P(B) \neq 0$ . Entonces*

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

# Fórmula de Bayes

Es una fórmula muy útil para “invertir” probabilidades condicionales.

## Teorema

*Sean  $A, B$  eventos con probabilidad no nula. Entonces*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

# Variables Aleatorias

Es una función que asigna un número real a cada resultado posible de un experimento aleatorio. Los resultados posibles, también llamados eventos, son los posibles resultados de un experimento o situación aleatoria. Por ejemplo, al lanzar un dado, los resultados posibles son los números del 1 al 6.

Una variable aleatoria puede ser:

- discreta o continua.

Una variable aleatoria discreta toma valores aislados y contables, mientras que una variable aleatoria continua puede tomar cualquier valor dentro de un rango específico.

- determinística o probabilística.

En el caso de una variable aleatoria determinística, los valores numéricos están completamente determinados por los resultados posibles.

En el caso de una variable aleatoria probabilística, los valores numéricos están asociados con probabilidades.

# Parámetros de Distribuciones

Distribución	Media teórica	Varianza teórica
Geométrica	$1/p$	$(1-p)/(p^2)$
Binomial	$n \cdot p$	$n \cdot p \cdot (1-p)$
Poisson	$\lambda$	$\lambda$
Normal	$\mu$	$\sigma^2$