

▼ TRABAJO PRACTICO N 3

TEMAS A DESARROLLAR:

Nociones de Probabilidad, simulación de procesos aleatorios, cálculo de parámetros estadísticos: Media o Valor Esperado(PROMEDIO), Mediana, Desvio Estandar. DISTRIBUCIONES DE PROBABILIDAD más frecuentes, por ejemplo la Normal.

▼ Funciones estadísticas para análisis exploratorio de datos

A continuación se presentan ejemplos de funciones estadísticas para el análisis exploratorio de datos utilizando NumPy:

▼ Crear array aleatorio

En esta celda, se utiliza la función `np.random.rand()` de NumPy para crear un array de números aleatorios con una forma específica. Luego, se imprime el array.

```
random_array = np.random.rand(3, 3)
print("Array aleatorio:", random_array)

Array aleatorio: [[0.09784478 0.1368385  0.45138056]
 [0.69213679 0.88689074 0.46316395]
 [0.04395057 0.29633477 0.64388957]]
```

▼ Media

En esta celda, se crea un array 'array' y se utiliza la función `np.mean()` de NumPy para calcular la media del array. Luego, se imprime el resultado.

```
array = np.array([1, 2, 3, 4, 5])
media = np.mean(array)
print("Media:", media)
```

```
Media: 3.0
```

▼ Desviación estándar

En esta celda, se crea un array 'array' y se utiliza la función `np.std()` de NumPy para calcular la desviación estándar del array. Luego, se imprime el resultado.

```
array = np.array([1, 2, 3, 4, 5])
desviacion_estandar = np.std(array)
print("Desviación estándar:", desviacion_estandar)
```

Desviación estándar: 1.4142135623730951

▼ Varianza

La función `np.var()` se utiliza para calcular la varianza de un array. Ejemplo de uso:

```
array = np.array([1, 2, 3, 4, 5])
varianza = np.var(array)
print("Varianza:", varianza)
```

Varianza: 2.0

▼ Valor mínimo

En esta celda, se crea un array 'array' y se utiliza la función `np.min()` de NumPy para encontrar el valor mínimo del array. Luego, se imprime el resultado.

```
array = np.array([1, 2, 3, 4, 5])
minimo = np.min(array)
print("Valor mínimo:", minimo)
```

Valor mínimo: 1

▼ Valor máximo

En esta celda, se crea un array 'array' y se utiliza la función `np.max()` de NumPy para encontrar el valor máximo del array. Luego, se imprime el resultado.

```
maximo = np.max(array)
print("Valor máximo:", maximo)
```

Valor máximo: 5

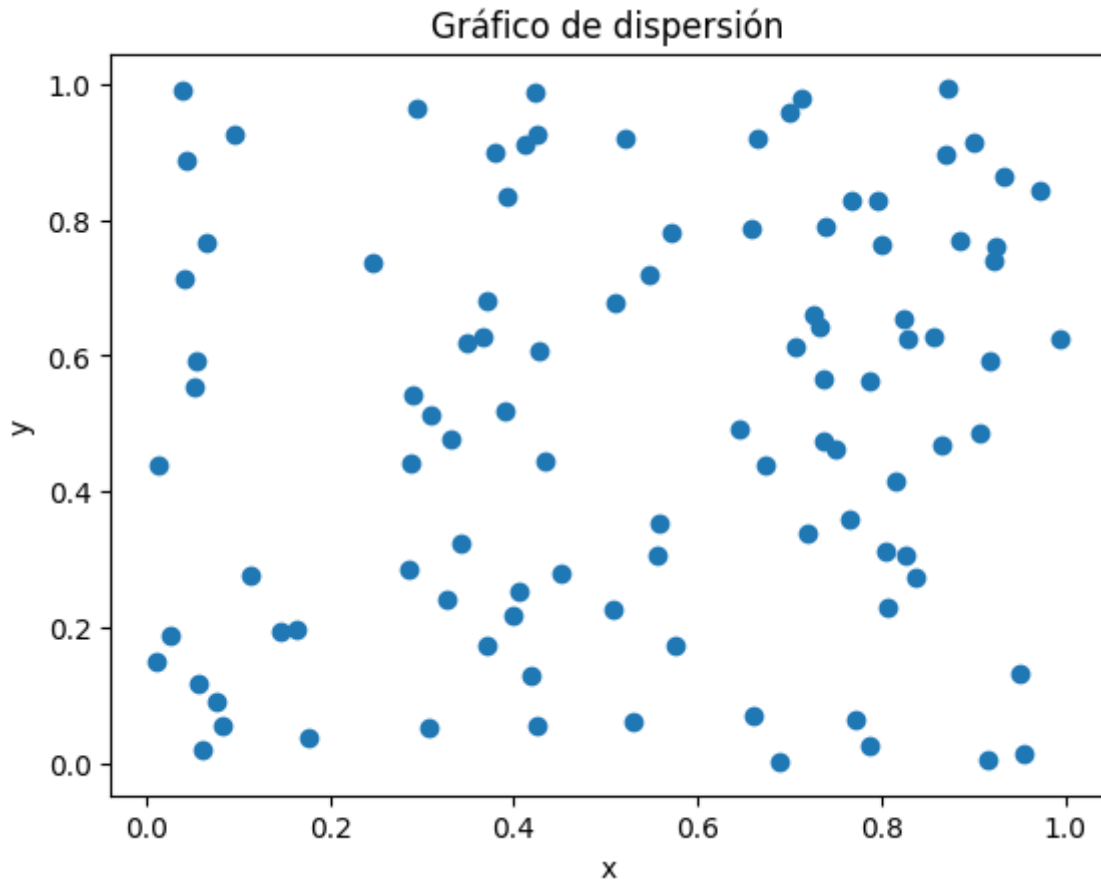
▼ Gráfico de dispersión

En esta celda, se crea un array 'x' con valores aleatorios y un array 'y' con valores aleatorios. Luego, se utiliza la biblioteca Matplotlib para trazar un gráfico de dispersión con 'x' en el eje x y 'y' en el eje y. También se agrega etiquetas y un título al gráfico.

```
x = np.random.rand(100)
y = np.random.rand(100)

plt.scatter(x, y)
```

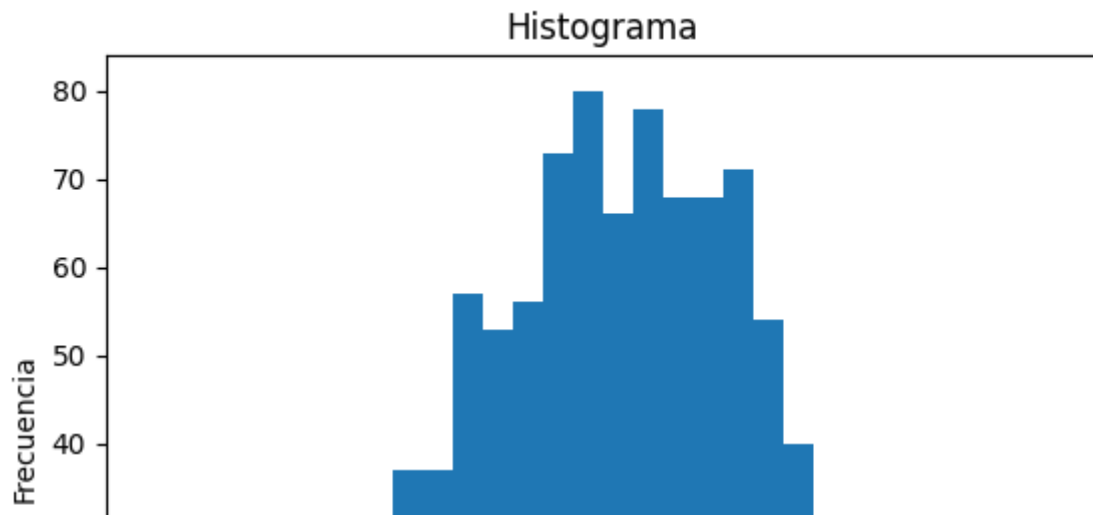
```
plt.xlabel('x')
plt.ylabel('y')
plt.title('Gráfico de dispersión')
plt.show()
```



En esta celda, se crea un array 'data' con valores aleatorios siguiendo una distribución normal. Luego, se utiliza la biblioteca Matplotlib para trazar un histograma de 'data'. También se agrega etiquetas y un título al gráfico.

```
data = np.random.normal(0, 1, 1000)

plt.hist(data, bins=30)
plt.xlabel('Valor')
plt.ylabel('Frecuencia')
plt.title('Histograma')
plt.show()
```



▼ Mediana

La función `np.median()` se utiliza para calcular la mediana de un array. Ejemplo de uso:

```

10 |-----|
array = np.array([1, 2, 3, 4, 5])
mediana = np.median(array)
print("Mediana:", mediana)

Mediana: 3.0

```

▼ Percentiles

La función `np.percentile()` se utiliza para calcular los percentiles de un array. Ejemplo de uso:

```

array = np.array([1, 2, 3, 4, 5])
percentil_50 = np.percentile(array, 50)
print("Percentil 50:", percentil_50)

Percentil 50: 3.0

```

▼ EJEMPLOS Y APLICACIONES

APLICACIÓN DE PROBABILIDAD

Se Realiza el lanzamiento de un dado EQUILIBRADO JUSTO, de manera reiterada con las mismas condiciones, y se anotan sus resultados= CARA O CRUZ.

Se le pide al usuario que defina la cantidad N de lanzamientos y se calcula su frecuencia relativa= 'PROBABILIDAD' de obtener cara o cruz. Se obtiene los siguientes resultados:

```

import numpy as np
n=int(input('ingrese la cantidad de simulación o lanzamientos'))

```

```
#===== FORMA 1 DE SIMULAR LANZAMIENTO DE UN DADO=====
```

```
dado=[]
cara=0
cruz=0
for i in range(n):
    dado.append(np.random.rand(1))
    dado1=np.array(dado)
    if dado1[i]<0.5:
        #print('salio cara')
        #print(dado[i])
        cara +=1
    else:
        #print('salio cruz')
        #print(dado[i])
        cruz +=1
```

ingrese la cantidad de simulación o lanzamientos100

```
# mostrar la probabilidad frecuencista de el lanzamiento de un dado
prom_cara=cara/n
prom_cruz=cruz/n
print('la probabilidad aproximada en '+str(n)+' lanzamientos de sacar CARA es:', prom_cara)
print('la probabilidad aproximada en '+str(n)+' lanzamientos de sacar CRUZ es:', prom_cruz)
```

la probabilidad aproximada en 100 lanzamientos de sacar CARA es: 0.47
la probabilidad aproximada en 100 lanzamientos de sacar CRUZ es: 0.53

```
#===== FORMA COMPACTA DE SIMULAR USANDO NUMPY=====
```

```
m=int(input('ingrese la cantidad de simulación o lanzamientos'))
lanzamientos = np.random.randint(0, 2, m)
```

```
# Calcular las probabilidades
```

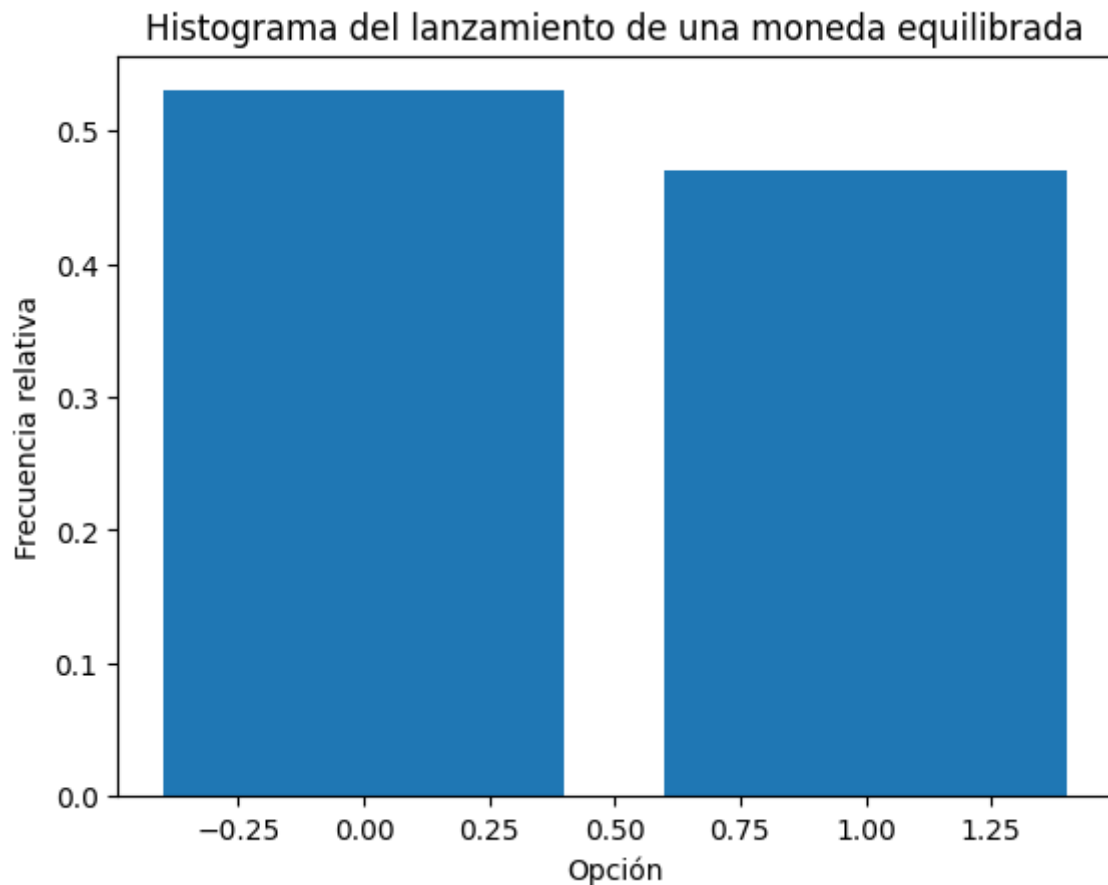
```
probabilidad_cara = np.sum(lanzamientos ==0 ) / len(lanzamientos)
probabilidad_sello = np.sum(lanzamientos== 1) / len(lanzamientos)
```

```
print(f"Probabilidad de cara: {probabilidad_cara:.2f}")
print(f"Probabilidad de sello: {probabilidad_sello:.2f}")
```

ingrese la cantidad de simulación o lanzamientos100
Probabilidad de cara: 0.53
Probabilidad de sello: 0.47

```
import matplotlib.pyplot as plt
x=np.array(range(2))
y=np.array([probabilidad_cara, probabilidad_sello])

plt.bar(x,y)
plt.xlabel('Opción')
plt.ylabel('Frecuencia relativa')
plt.title('Histograma del lanzamiento de una moneda equilibrada')
plt.show()
```



▼ EJERCICIOS

EJERCICIO 1

Simular el lanzamiento de un dado equilibrado de 6 caras.

- Mostrar la probabilidad aproximada de obtener el número 2, 3 y 5
- Graficar el histograma de las probabilidades

EJERCICIO 2

Simular el lanzamiento de un dado NO equilibrado o cargado de 6 caras, de manera que las chances o probabilidades de salir 1 o 2 sean el doble de obtener otro resultado (3,4,5 o 6)

- Mostrar la probabilidad aproximada de obtener el número 2, 3 y 5
- Graficar el histograma de las probabilidades

EJERCICIO 3

Realizar la simulación del juego {piedra,papel , tijera} donde el resultado de esta simulación deberá ser lo que mi oponente elije. No debe considerarse la simulación de mi elección.

- Calcular la probabilidad aproximada que mi oponente elija, 'piedra', 'papel', o 'tijera', en N jugadas.
- Realizar un gráfico de barra mostrando estas probabilidades. Concluir algo sobre las probabilidades de estos resultados, cuando el número de jugadas se incrementa bastante.

EJERCICIO 4

Dada la base de datos de ventas diarias durante una semana, de un negocio con 4 productos.-
 VER EJERCICIO PARTE 2 PRACTICO N 2- Realizar el siguiente análisis de datos para poder obtener información y planificar una estrategia de ventas.

1. Calcular la media en la ventas por cada producto y la mediana. Plotear estos nuevos datos en un gráfico de barra.
2. Calcular la media en la ventas por DÍA y la mediana. Plotear estos nuevos datos en un gráfico de TRAZO DE LINEAS.
3. Realizar una conclusión en función las graficas y sugerir al menos 5 estrategias de venta o inversión de manera de potenciar este negocio.
4. Calcular el desvío estandar de los datos por cada producto y concluir para cada producto como es su rango de variabilidad en las ventas de los mismos. Plotear los datos de ventas de la semana de los productos con mayor y menor desvío estandar y evidenciar la variabilidad indicada en el parametro '*DESVIO ESTANDAR*'

EJERCICIO 5

Solicitar al usuario que ingrese la cantidad N de datos a cargar. Luego generar de manera aleatoria los siguientes datos:

- SEXO= 'MASCULINO', 'FEMENINO', 'NO ESPECIFICA'
- EDAD= entre 28 y 60 años
- ALTURA= entre un rango de 165.0 mts hasta 210 mts
- PESO= entre un rango de 50 y 110 kg
- HORAS DE DESCANSO diario= un número decimal
- HORAS DE TRABAJO/ ESTUDIO diario= un número decimal
- HORAS DE EJERCICIOS diario= un número decimal

Realizar las siguientes indicaciones y extraer conclusiones sobre la información y la población muestreada.

1. Generar un array de dimensión 2 (matriz), con todos los datos ordenados, permitiendo visualizarlos como tabla.
2. Suponiendo que la distribución de probabilidad de los datos muestrados de las N personas, se distribuyen en forma NORMAL. Encontrar la media en los datos de altura,

peso, y horas de descanso, trabajo y ejercicios.

3. Graficar la muestra o distribución de los datos de pesos , y alturas en graficas separadas. Indicar el grado de variabilidad de sus datos
4. Plotear de manera conjunta y encontrar si presentan correlación los datos de peso vs altura,
5. Plotear de manera conjunta los datos de horas de descanso, estudio y ejercicio

EJERCICIO 6

Se realiza una encuesta a N usuario donde se registran los siguientes datos.

['DNI','GASTOS MENSUALES DE HOGAR','INGRESOS MENSUALES POR HOGAR', 'CANTIDAD DE MIEMBROS QUE VIVEN POR HOGAR','GASTO MENSUAL EN LINEA TELEFÓNICA', 'GASTO MENSUAL POR PAGO DE INTERNET']

Realizar la simulación de la carga de datos utilizando la siguiente información adicional:

1. El 70 % de los encuestados son jóvenes entre 15 y 21 años, el 10% son mayores de 60 años. Sugerencia, utilizar esto para generar los DNI
2. Los encuestados son de la provincia de Cordoba. Además el 80 % de la población de esa provincia tiene un grupo familiar de 5 personas. Con ingresos mensuales promedio de \$258.000
3. Cerca del 80 % de los encuestados Gastan el 90% de sus ingresos, el resto gasta el 80% de sus ingresos. Ayuda, para simular la cantidad de miembros utilizar la distribución normal.
4. Del gasto Total por hogar se sabe que se abona un 15 % en telefonía y un 7.8% en servicio de internet.

Con la información proporcionada, realizar la simulación de las encuestas de 100 y 1000 personas.

- Calcular el valor medio y desvio estandar de los gastos mensuales, ingresos mensuales y pagos de servicios.
- Plotear los datos de los ingresos , y gastos en gráficos de barra por separado y evidenciar la media y desvio estandar.
- Calcular la proporción de jóvenes menores a 21 encuestados. Mostrar sus datos de grupos familiares, gastos en servicios de internet y telefonía. Concluir algo en base a esta información.
- Realizar un histograma con el dato del grupo familiar, y concluir algo acerca de la distribución de probabilidad.

