



ADVANCE
Consortium



ATENEA 2025 BOOTCAMPS



Identificación Material

Bootcamp	Soluciones de Futuro con IA y Datos - Intermedio
Módulo	Desarrollo de soluciones basadas en datos
Eje Temático	Muestreo de Datos

Muestreo de Datos

MAPA DE CONTENIDOS

MUESTREO DE LOS DATOS

01



MARCO CONCEPTUAL

02



MUESTREO DE
RENGLONES

03



MUESTREO DE
ELEMENTOS POR
COLUMNAS

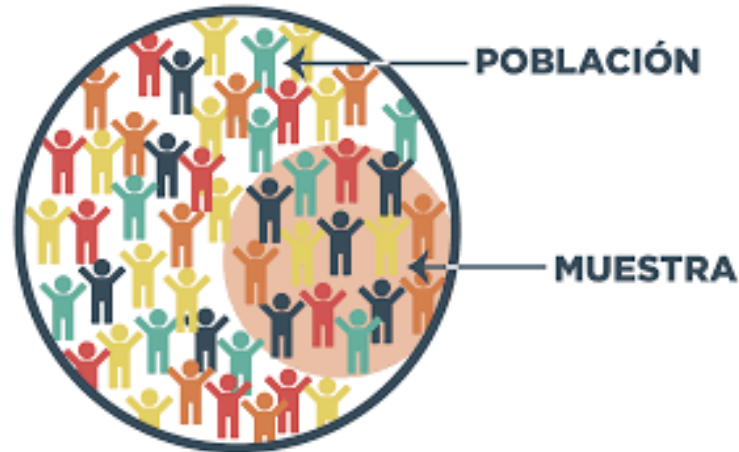


Universidad Tecnológica de Bolívar

MARCO CONCEPTUAL

Marco Conceptual

El **muestreo de datos** es una herramienta de la investigación científica. Su función básica es determinar qué parte de una realidad en estudio (población o universo) debe examinarse con la finalidad de hacer inferencias sobre dicha población. Además, es un proceso fundamental en la ciencia de datos **para seleccionar subconjuntos representativos de un conjunto de datos más grande**. En Python, hay varias formas de realizar muestreo, dependiendo de las bibliotecas que utilices.



Tipos de Muestreo

El **muestreo aleatorio simple** es un procedimiento de muestreo probabilístico que da a cada elemento de la población objetivo y a cada posible muestra de un tamaño determinado, la misma probabilidad de ser seleccionado.

Este sistema funciona a través de un método de «sorteo» entre un universo de individuos, en el que se le asigna a cada integrante un número para ser escogido



El **muestreo estratificado** es una forma de representación estadística que muestra cómo se comporta una característica o variable en una población a través de hacer evidente el cambio de dicha variable en subpoblaciones o estratos en los que se ha dividido. Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a característica a estudiar y que no se solapan.

El **muestreo sistemático** es un tipo de muestreo probabilístico que se basa en enlistar a toda una población, elegir de forma aleatoria al primer individuo para la muestra y luego, a partir de un intervalo definido por el investigador, seleccionar al resto de los individuos que conformarán la muestra.

El **muestreo por conglomerados** es un procedimiento de muestreo probabilístico en el que se seleccionan aleatoriamente varios grupos (llamados conglomerados, cúmulos o áreas) conformados por elementos heterogéneos de la población, pero que tienen algo en común

Fuente de Datos para práctica



```
#Muestreo aleatorio Simple
import numpy as np
import pandas as pd

#Leer archivo excel e imprimirlo
df = pd.read_csv('C:\\Otros\\UTB\\Talentotech\\AnálisisDatos\\Material\\Semana_5\\Ejercicios\\diabetes.csv')
print("\nArchivo CSV")
print(df)
```

```
Archivo CSV
   imc  glucosa
0  33.6     148
1  26.6      85
2  23.3     183
3  28.1      89
4  43.1     137
..   ...     ...
763 32.9     101
764 36.8     122
765 26.2     121
766 30.1     126
767 30.4      93

[768 rows x 2 columns]
```


MUESTREO POR RENGLONES

Muestreo por Renglones

```
# Método: Muestreo de Renglones  
df_muestra=df.sample(frac=0.10)  
print("\nMuestreo de Renglones (10%)")  
print(df_muestra)
```

La muestra es aleatoria y se toman, a través del `frac=0.10`, el 10% de la población, para este caso.

Muestreo de Renglones (10%)

	imc	glucosa
557	27.8	110
560	33.8	125
338	34.2	152
490	36.8	83
643	28.0	90
..
454	37.8	100
569	34.3	121
430	22.2	99
395	27.7	127
489	26.1	194

[77 rows x 2 columns]

MUESTREO DE ELEMENTOS POR COLUMNAS

Muestreo de elementos por columnas

```
# Muestreo de elemntos por columnas (Numpy)
df_muestra=np.random.choice(df["glucosa"],10)
print("\nMuestreo de elementos por columna (10 elementos)")
print(df_muestra)
```

```
Muestreo de elementos por columna (10 elementos)
[120  88 105 147 168 106 142 138 174 146]
```

La muestra es aleatoria y se toman 10 datos de la columna glucosa, para este caso



ADVANCE
Consortium

GRACIAS

