

## Módulo 4 – Inferencia Estadística

# Distribución Muestral

Ciencia de Datos

# Objetivos de Aprendizaje



- Utiliza los conceptos básicos de estadística Inferencial.
- Realizar cálculos de probabilidad utilizando la distribución muestral para resolver un problema.

# Contenido:

1. Distribución muestral de medias.
2. Distribución muestral de varianzas.



# Distribución Muestral

La distribución muestral es una distribución de probabilidad que describe los posibles valores que puede tomar un estadístico (como la media, la mediana, la desviación estándar, etc.) al tomar muestras aleatorias de una población.

En otras palabras, la distribución muestral es una distribución teórica que muestra cómo se distribuyen los estadísticos de una población a través de todas las posibles muestras que se pueden extraer de esa población.

Por ejemplo, si queremos estimar la media de una población, podemos tomar muestras aleatorias de esa población y calcular la media de cada muestra. Si graficamos todas las medias de las muestras que se pueden extraer de la población, obtendremos una distribución de medias muestrales. Esta distribución de medias muestrales es la distribución muestral de la media.

En estadística, la distribución muestral sirve para **calcular la probabilidad que se tiene de acercarse al valor del parámetro poblacional al estudiar una sola muestra**. Asimismo, la distribución muestral permite estimar el error muestral para un tamaño de muestra dado.

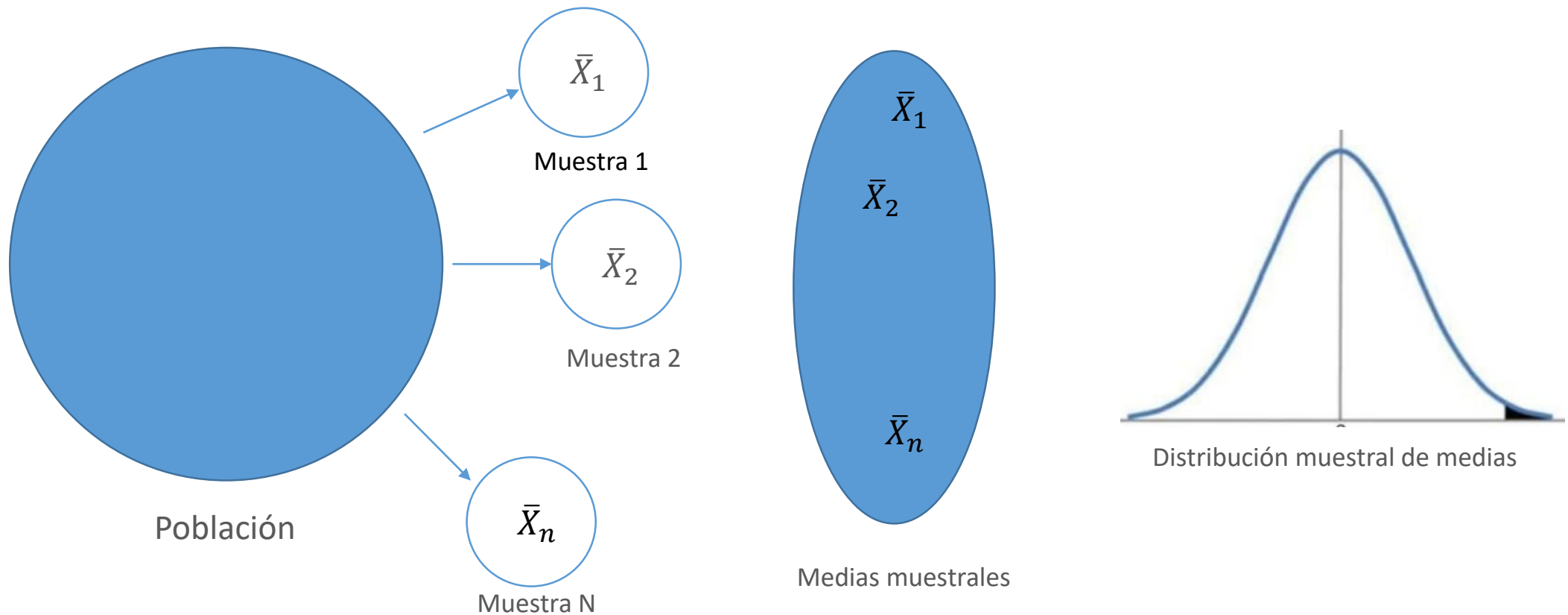
# Distribución Muestral

La distribución muestral es importante porque nos permite **hacer inferencias sobre la población utilizando información de la muestra**. Al conocer la distribución muestral de un estadístico, podemos calcular intervalos de confianza, realizar pruebas de hipótesis y hacer predicciones sobre la población.

Es importante destacar que la distribución muestral depende del tamaño de la muestra y de las propiedades de la población de donde se extrae la muestra. Además, la distribución muestral de algunos estadísticos, como la media, sigue una distribución normal según el teorema del límite central, lo que facilita el análisis estadístico.

# Distribución Muestral de Medias

Suponga que se han seleccionado  $n$  muestras aleatorias de tamaño  $m$  en una población grande. Se calcula la media muestral  $\bar{X}$  para cada muestra; la colección de todas estas medias muestrales recibe el nombre de **distribución muestral de medias**.



# Distribución Muestral de Medias

Respecto a la distribución muestral de medias:

- Su media coincide con la media de la población.

$$\mu_{\bar{X}} = \mu$$

- La desviación típica, Error Típico de la Media, es:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- La forma de la distribución tiende a la curva normal a medida que crece el tamaño de la muestra ( $n > 30$ ), si la muestra es pequeña, seguirá una distribución t con  $n-1$  grados de libertad.

# Distribución Muestral de Medias

Por lo tanto, como la distribución muestral de la media sigue una distribución normal, la fórmula para cualquier probabilidad relacionada con la media de una muestra es la siguiente:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

En donde,

$\bar{x}$  : media de la muestra.

$\mu$  : media de la población.

$\sigma$  : desviación típica de la población.

$n$  : tamaño de la muestra.

$Z$  : variable definida por la distribución normal estándar  $N(0,1)$ .

# Ejemplo

El peso de los estudiantes de una universidad sigue una distribución normal de media 68 kg y desviación estándar 9 kg. Determine:

- ¿Cuál es la probabilidad de que la media de una muestra aleatoria de 25 alumnos esté por debajo de 66 kg?
- Si se extraen 300 muestras con un tamaño de 25 alumnos cada una, ¿cuántas medias muestrales tendrán un valor por debajo de 66 kg?

# Ejemplo

Primero, calculamos el valor del estadístico correspondiente, según la siguiente fórmula:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{66 - 68}{\frac{9}{\sqrt{25}}} = -1.11$$

La probabilidad que buscamos corresponde al valor  $Z = -1.11$ , de la cola izquierda de la distribución normal estándar, que se puede obtener de la tabla de probabilidades o bien desde código Python. Por lo tanto,

$$P[\bar{x} \leq 66] = P[Z \leq -1.11] = 0.1335$$

La probabilidad de que la media de una muestra aleatoria de 25 alumnos esté por debajo de 66 kg es de 13.35%.

# Ejemplo

Ahora que ya sabemos la probabilidad de que la media de una muestra aleatoria esté por debajo de 66 kg, para saber el número de medias muestrales que están por debajo de 66 kg al sacar 300 muestras iguales tenemos que multiplicar la probabilidad calculada por el número total de muestras tomadas:

$$0.1335 \times 300 = 40.05 \approx 40$$

Por lo tanto, 40 de las muestras extraídas tendrán una media por debajo de 66 kg.

# Distribución Muestral de la Varianza

La distribución muestral de la varianza está definida por la distribución de probabilidad chi-cuadrado. Por lo tanto, la fórmula del estadístico de la distribución muestral de la varianza es la siguiente:

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

En donde,

$\chi^2$  : estadístico de la distribución muestral de la varianza, el cual sigue una distribución chi-cuadrado.

$s^2$  : varianza de la muestra

$\sigma^2$  : varianza de la población.

$n$  : tamaño de la muestra.

## Ejemplo

De una población con varianza conocida  $\sigma^2=5$  se escoge una muestra aleatoria de 17 observaciones. ¿Cuál es la probabilidad de obtener una varianza muestral mayor que 10?

# Ejemplo

Primero que todo, tenemos que sacar el estadístico de la distribución muestral de la varianza. Así que aplicamos la siguiente fórmula:

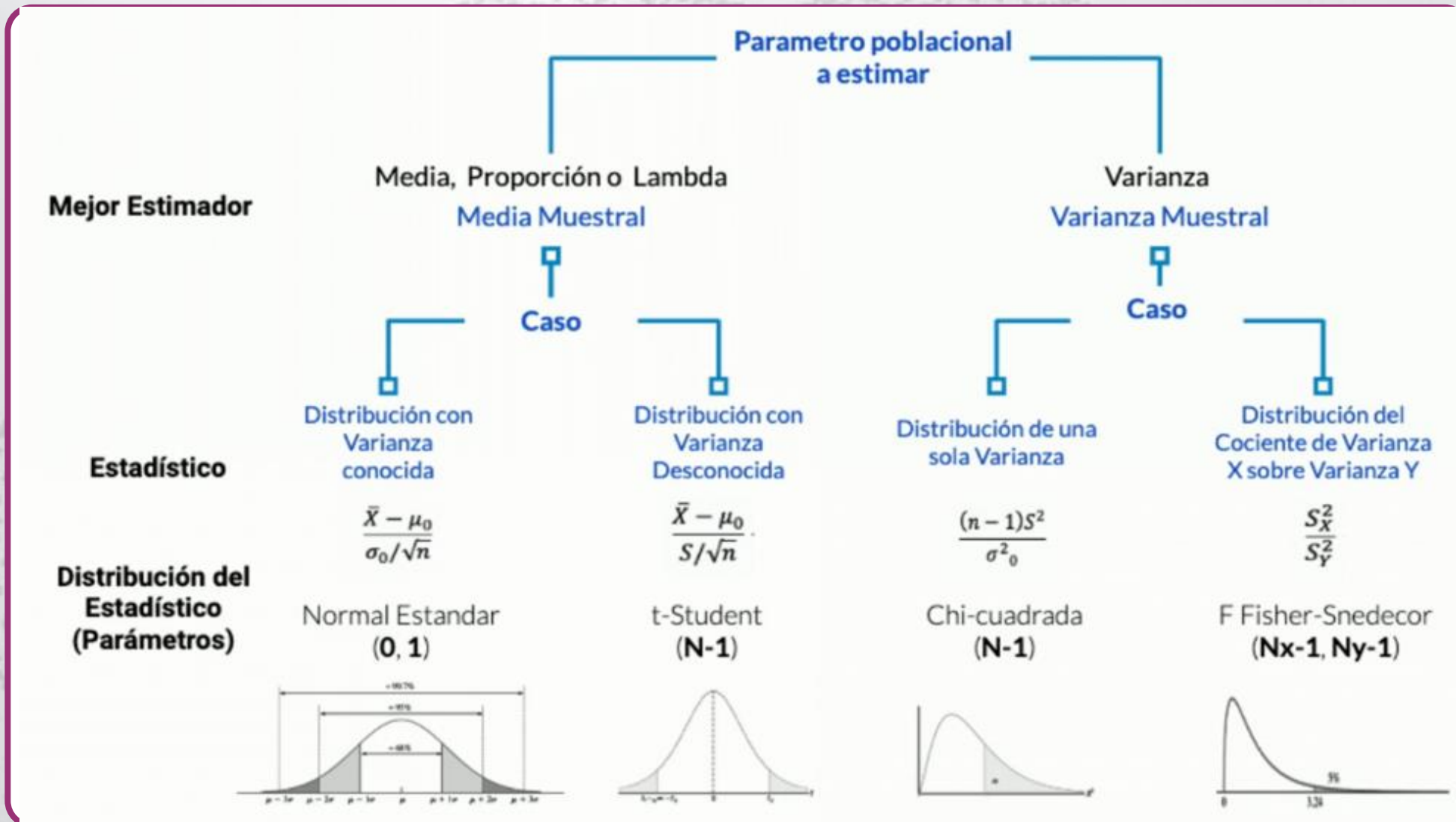
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(17-1) \cdot 10}{5} = 32$$

Como el tamaño de la muestra es  $n=17$ , la distribución chi-cuadrado tendrá 16 grados de libertad ( $n-1$ ). Por lo tanto, la probabilidad de que la varianza muestral sea mayor que 10 es equivalente a la probabilidad de tomar un valor de una distribución chi-cuadrado con 16 grados de libertad que sea mayor que 32. Revisamos la tabla de probabilidades o bien consultamos Python.

$$P[s^2 > 10] = P[\chi_{16}^2 > 32] = 0.01$$

En definitiva, la probabilidad de sacar una muestra con una varianza mayor que 10 es del 1%.

# Distribuciones Muestrales



# Dudas y consultas

Fin presentación