

# Distribuciones muestrales y TLC

Versión PDF

II-1123 Estadística para Ingeniería Industrial II

Steven García Goñi

[steven.garciagoni@ucr.ac.cr](mailto:steven.garciagoni@ucr.ac.cr)

10 de agosto de 2026



# Agenda

- Preguntas generadoras
- Introducción
- Distribuciones muestrales
- Ley de los grandes números
- Teorema del límite central (TLC)
- Resultados importantes



# Preguntas generadoras

- ¿Por qué es necesario muestrear?
- ¿Qué es el sesgo estadístico y cómo evitarlo?
- ¿Cuáles métodos de selección de muestra existen y cuándo se usan?
- ¿Qué es el error de muestreo y qué consecuencias prácticas tiene?
- ¿Qué es una distribución muestral?
- ¿Para qué es útil el TLC?
- ¿Qué ventajas y desventajas tiene el TLC?

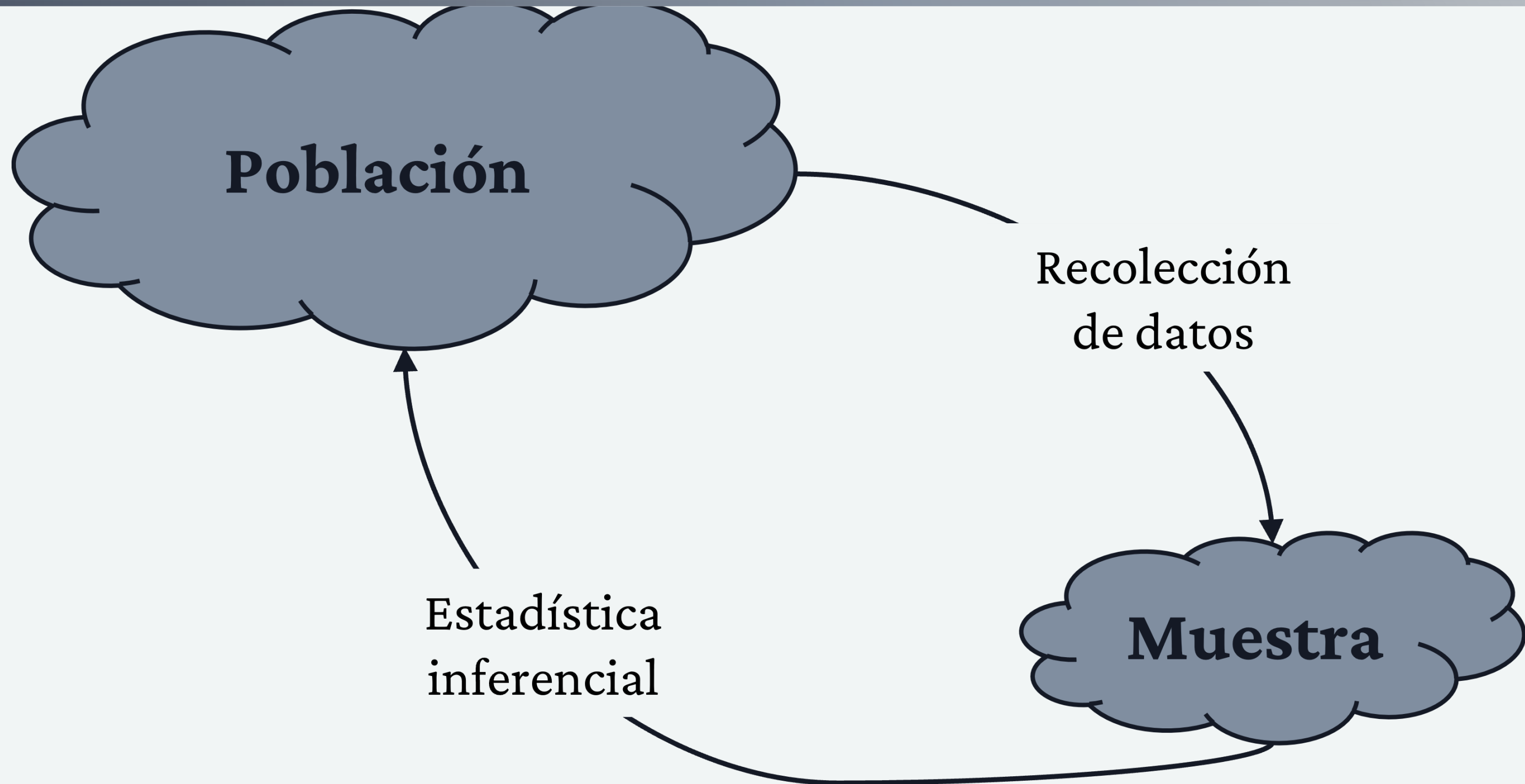


# Introducción

- ¿Por qué muestrear?
- Los métodos estadísticos que son usados para tomar decisiones y arribar a conclusiones acerca de las **poblaciones** son generalmente conocidos como **inferencia estadística**.
- En esta técnica se usa la información de una muestra para inferir dichas conclusiones.
  - Esto desde luego implica que la calidad de la muestra impacta la calidad de las conclusiones.
  - Recuerde lo visto en **Fundamentos de muestreo y manejo de datos**.



# Introducción



# Introducción

- La inferencia estadística se divide en dos grandes áreas:
  - Estimación de parámetros
  - Pruebas de hipótesis.
- Ambas nociones serán abordadas en otras lecciones, aunque la primera es una vieja conocida en esta área de conocimiento. Por ejemplo, el estimador:
  - De la media  $\mu$  es  $\bar{X}$ ,
  - De  $\sigma$  es  $S$
  - De  $p$  es  $\hat{P}$



# Introducción

- Los estadísticos como  $\bar{X}$ ,  $S$  y  $\hat{P}$  (así como otros:  $\hat{\Theta}$ ) son **variables aleatorias**.
  - Como las que estudiamos previo, que dieron origen a las distribuciones de probabilidad. De ser necesario repase este concepto.
- Estas son variables aleatorias porque dependen de la **muestras** que también son o deben ser **aleatorios**.



# Introducción

A la inferencia estadística le corresponde la toma de decisiones acerca de una población, basada en la información contenida en una **muestra aleatoria** de esa población.



# Distribuciones muestrales



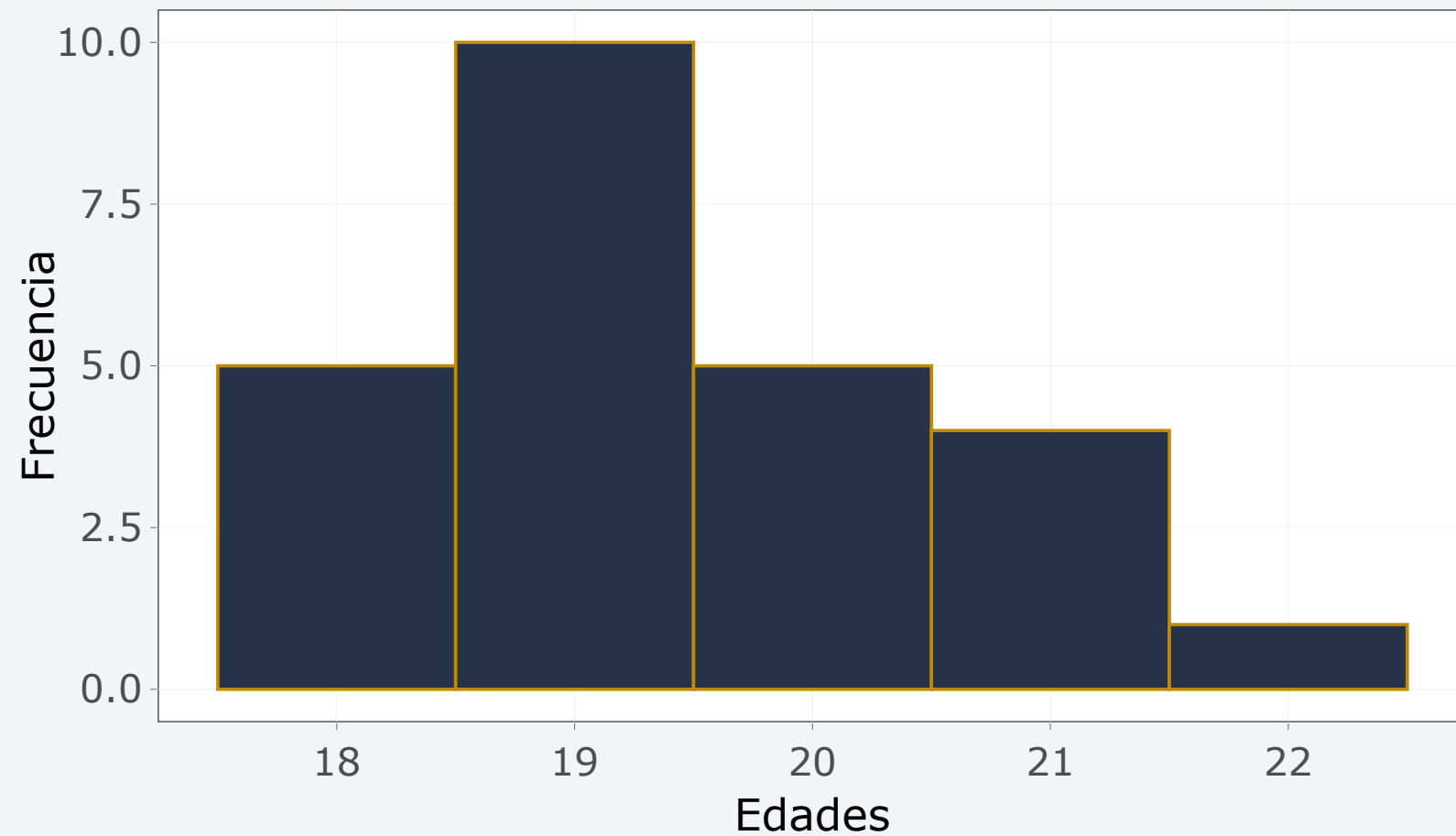
# Distribuciones muestrales

- Supóngase que estamos interesados en la edad promedio (en años) de los estudiantes de este curso, pues ese dato es relevante para la toma de decisiones respecto a las actividades por planear.
- Los siguientes datos corresponden a la edad de las personas presentes en la primera sesión de laboratorio del curso en el I-S 2025 grupo 003 de RRF.
  - A esto lo vamos a llamar la población.

|    |    |    |    |    |
|----|----|----|----|----|
| 19 | 20 | 18 | 19 | 18 |
| 18 | 22 | 21 | 20 | 19 |
| 18 | 19 | 21 | 19 | 19 |
| 21 | 21 | 20 | 19 | 18 |
| 20 | 19 | 19 | 20 | 19 |



# Distribuciones muestrales



- Nótese cómo la población NO se parece a una distribución normal, por el contrario, se parece más a una lognormal.
- Tome en cuenta esto para el desarrollo del resto de la sección.



# Distribuciones muestrales

- Cada valor numérico en los datos (**en este caso la edad**) es el valor observado de una variable aleatoria.
- Al subconjunto de valores numéricos se le conoce como muestra aleatoria.
  - Este es un supuesto **IMPORTANTE**, puesto que si la muestra no es aleatoria los métodos estadísticos por estudiar pueden conducir a conclusiones erróneas.
- El propósito de las muestras aleatorias es obtener información acerca de los parámetros desconocidos de la población.
- La media de la población, por ejemplo, se puede estimar a partir de una muestra aleatoria.



# Distribuciones muestrales

- Por tanto  $\bar{X}$  es una función de los valores observados en la muestra aleatoria.
- Entonces  $\bar{X}$  es un valor que variaría de muestra a muestra, lo que lo convierte en una **variable aleatoria** (como las que estudiamos antes).
- Esta variable aleatoria se llama estadístico
  - Un estadístico es una función de las observaciones en una muestra aleatoria.
- Ya hemos trabajado con estadísticos antes, por ejemplo, en el tema de descripción de datos.
- Entonces:
  - **La distribución de probabilidad de un estadístico es llamado como distribución muestral.**



# Sigamos con el ejemplo de la edad

- Se van a extraer, aleatoriamente, 20 muestras de tamaño 4 y se van a calcular los estadísticos  $\bar{X}$  y  $S$ .
- Cada uno de ellos, como variable aleatoria, va a seguir una distribución particular que vamos a averiguar paso a paso.



# Simulador de muestreo aleatorio

Simulador Didáctico e Interactivo

## Distribuciones Muestrales y Teorema del Límite Central

Por Steven García Goñi — II-1123 Estadística para Ingeniería Industrial II

Población de Edades (N = 25)

Tamaño n: 4

|    |    |    |    |    |
|----|----|----|----|----|
| 19 | 20 | 18 | 19 | 18 |
| 18 | 22 | 21 | 20 | 19 |
| 18 | 19 | 21 | 19 | 19 |
| 21 | 21 | 20 | 19 | 18 |
| 20 | 19 | 19 | 20 | 19 |

Registro Paso a Paso de Muestras

| Muestra   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| $\bar{x}$ | - | - | - | - | - | - | - | - | - | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  |
| s         | - | - | - | - | - | - | - | - | - | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  |

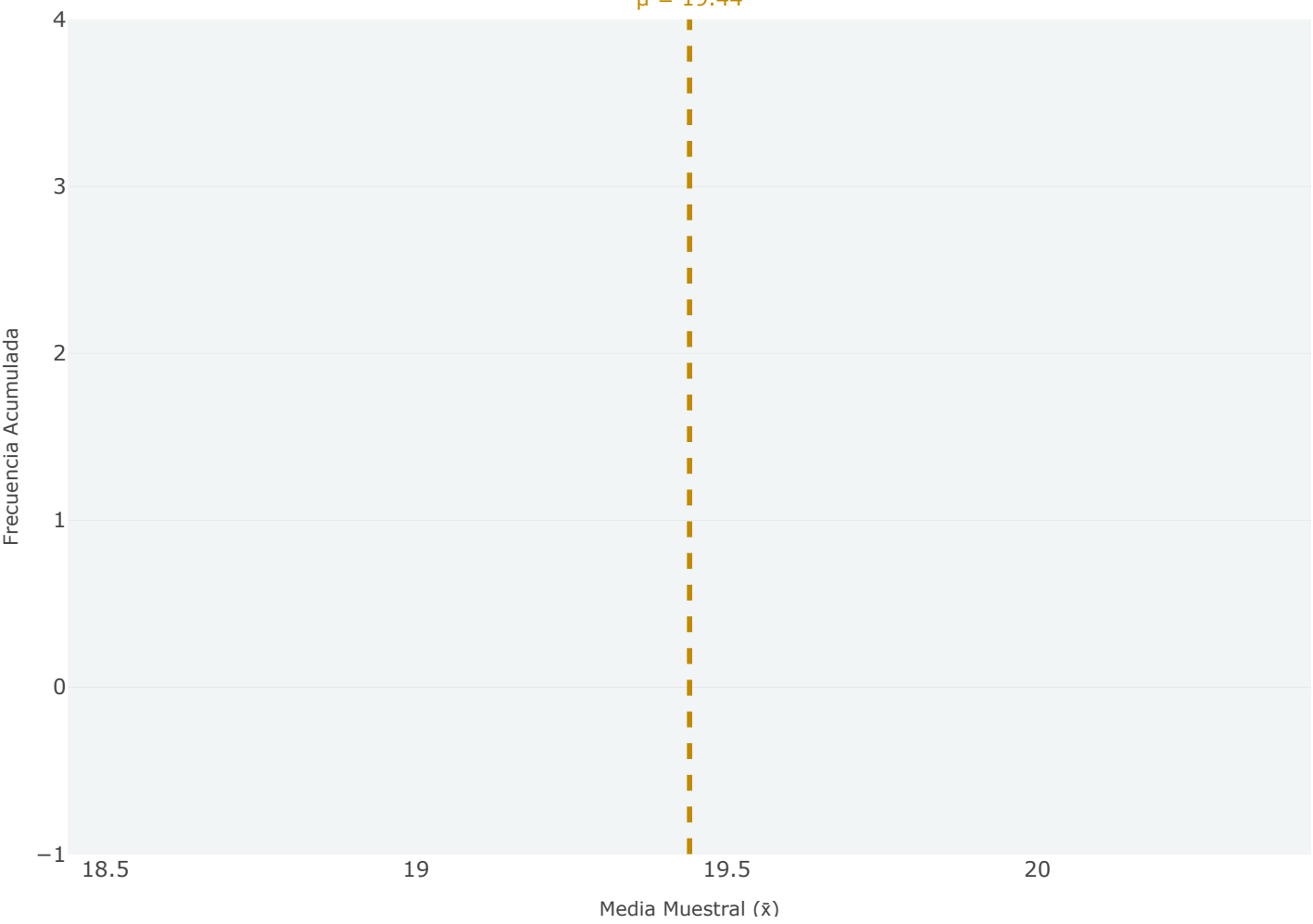
Muestra Paso a Paso

Simular +200

Reiniciar

Observe la convergencia empírica de  $\bar{x} \rightarrow \mu$  y la reducción de  $s_{\bar{x}} \rightarrow \sigma/\sqrt{n}$ .

Distribución Muestral de  $\bar{x}$  (n = 4)



Población (Teórica)

$\mu = 19.44$  |  $\sigma/\sqrt{n} = 0.55$

Empírica (Muestras k = 0)

$\bar{x} = -$  |  $s_{\bar{x}} = -$



# Resumen

- Como acabamos de ver, tanto  $\bar{X}$  como  $S$  son variables aleatorias en función del muestreo.
- Es decir, son distribuciones muestrales de la media y de la desviación respectivamente.
- Las distribuciones muestrales suelen ser muy útiles para hacer inferencia estadística cuando:
  - La población es demasiado grande, no es viable obtenerla toda.
  - La distribución que siguen los datos originales, no es una distribución normal.
    - Mucho de lo que vamos abordar depende del supuesto de normalidad de los datos.



# Ley de los grandes números



# Ley de los grandes números

- La componen teoremas que describen el comportamiento de las medias de muestras de variables aleatorias. Asegura que en una sucesión de variables aleatorias independientes idénticamente distribuidas con varianza finita, al aumentar el número de ensayos, el promedio de estas converge a la esperanza (media teórica) de las variables involucradas.
- En otras palabras, este permite explicar también como a medida que aumenta el tamaño de la muestra, la media muestral  $\bar{X}$  se aproxima a la media poblacional  $\mu$ .
- Se divide en dos:
  - Ley débil
    - En la que nos vamos a enfocar
  - Ley fuerte



# Ley débil

- Establece que se puede encontrar un tamaño de muestra  $n$  que garantice con una probabilidad tan cercana a uno como se quiera, que el promedio que se obtendría en esa muestra no difiera del promedio poblacional en menos de una cantidad tan pequeña como se desee  $\varepsilon$ .
- $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1$
- Usando la desigualdad de Chebyshev, podemos acotar esta probabilidad:
- $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$



# Ejemplo

- Para una población con media desconocida y  $\sigma = 2$ , determine el tamaño de muestra  $n$  que garantice una probabilidad de al menos 0.95 de que la media muestral no se aleje de  $\mu$  en más de 0.5 ( $\varepsilon = 0.5$ ).

- Aplicando la cota de Chebyshev:

$$- P(|\bar{X} - \mu| < 0.5) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \geq 0.95$$

- Despejando  $n$ :

$$- n \geq \frac{\sigma^2}{(1-0.95) \cdot \varepsilon^2} = \frac{2^2}{0.05 \cdot 0.5^2} = 320$$



# Ley fuerte

- La frecuencia relativa con que ocurre un hecho en pruebas independientes y en las mismas condiciones converge a la probabilidad del hecho observado con probabilidad 1. Su demostración excede el alcance del curso.



# Ley de los grandes números

- Consulte el siguiente [enlace](#), propuesto por la profesora Inga. Valentina Campos Aguilar como medio de apoyo para comprender de mejor manera la ley de los grandes números.



# Teorema del Límite Central (TLC)

- La distribución normal se puede usar frecuentemente para aproximar las distribuciones de **otras variables aleatorias**, como consecuencia del Teorema del Límite Central (TLC).
- El TLC dice, de **manera simplificada**, que la distribución de las **medias muestrales** (aleatorias e independientes) sigue una distribución normal, siempre que el tamaño de las muestras sea lo suficientemente grande.
- **¿Qué es suficientemente grande?** Muchos libros de texto enuncian que siempre que  $n \geq 30$ , no obstante, estudios relativamente “recientes” (de más de 15 años) dicen que  $n$  depende de, entre otras cosas, de la **asimetría de la distribución** que siguen los datos de la población. En resumen, a mayor asimetría, mayor tamaño de muestra es requerido.



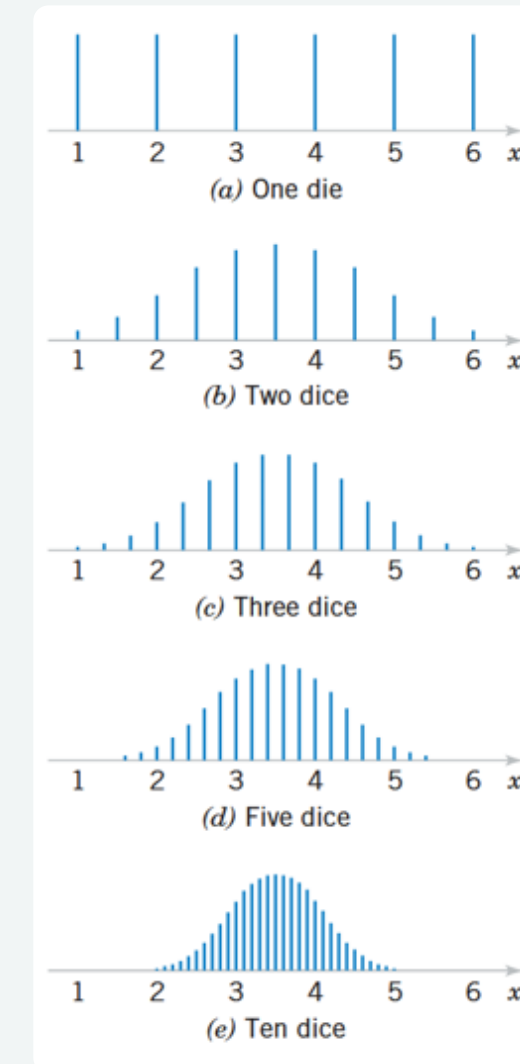
# Teorema del Límite Central (TLC)

- El siguiente [enlace](#) contiene una serie de demostraciones computacionales que demuestran la dependencia del TLC de la asimetría. La línea rosada en los gráficos representa a partir de donde se considera que los datos siguen la distribución normal.
- Note cómo a mayor asimetría, mayor tamaño de muestra se necesita para que se siga la distribución normal.
- En resumen, tamaño de muestra pequeños ( $n = 4$  o  $n = 5$ ) funcionan bien en la mayoría de los casos, particularmente cuando la población es continua, unimodal y simétrica.
- Tamaños de muestra grandes son requeridos en otras situaciones, dependiendo de la forma de la población.



# Teorema del Límite Central (TLC)

- Entonces,  $n \geq 30$  es una recomendación si no se conoce la forma de la población, pues con este tamaño de muestra se cumpliría satisfactoriamente la aproximación a la normal en una gran mayoría de los casos.
- Observe la siguiente **simulación**
- Cambie el tamaño muestral y note cómo la variabilidad de la distribución teórica varía, entre más grande, menos variabilidad
  - Nota: procure tener marcado el checkbox de “Distribución teórica”



# Teorema del Límite Central (TLC)

- En teoría, esa distribución muestral estará definida por la **misma media que la poblacional** y con una desviación estándar igual al cociente entre la desviación estándar poblacional dividida por la raíz cuadrada del tamaño de muestra ( $\sigma / \sqrt{n}$ ).
- Note de la fórmula que a mayor  $n$ , menor será la desviación de esta distribución normal
- $\bar{X} \sim N \left( \mu, \frac{\sigma}{\sqrt{n}} \right)$



# Teorema del Límite Central (TLC)

- Gracias al TLC, podemos acercarnos a conocer las características de la población, mediante la toma de muestras.
- Por ejemplo, la estimación de la media es directa.
- Para obtener la desviación se despeja la fórmula anterior.
- Finalmente

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



# Teorema del Límite Central (TLC)

- Formalmente:
- Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con media  $\mu$  y varianza finita  $\sigma^2$ , entonces la distribución de la variable estandarizada  $Z$  converge a una normal estándar  $N(0, 1)$  conforme  $n \rightarrow \infty$ :
- $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$
- Mientras más grande el valor de  $n$ , mejor es la aproximación.
- Repase esto con la simulación del muestreo de las edades.



# Finalmente

- A partir de los datos de una muestra, se puede aproximar razonablemente el comportamiento de la población.
- $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$



# Distribución muestral de la media ( $\bar{X}$ )

| Estadístico | Distribución muestral esperada                                                                                                                                      |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\bar{X}$   | <p>Si la población es normal o la muestra suficientemente grande</p> <p><b>Distribución normal</b></p> $\bar{X} \sim N \left( \mu, \frac{\sigma}{\sqrt{n}} \right)$ |

# Distribución muestral de la varianza ( $S^2$ )

| Estadístico | Distribución muestral esperada                                                                                                                                                                                                                                                                                                 |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $S^2$       | <div>Si la población es normal</div> <div>Si es una varianza</div> <div><math display="block">\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}</math></div> <div>Si son dos varianzas (bajo el supuesto de igualdad)</div> <div><math display="block">\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{(n_1-1,n_2-1)}</math></div> |

# Distribución muestral de la proporción ( $\hat{P}$ )

| Estadístico | Distribución muestral esperada                                                                                                                      |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| $\hat{P}$   | <p>Siempre que <math>n \cdot p \geq 5</math> y <math>n \cdot (1 - p) \geq 5</math></p> $\hat{P} \sim N \left( p, \sqrt{\frac{p(1 - p)}{n}} \right)$ |

# Bibliografía

- Walpole, R.; Myers, R.; Myers, S. y Ye, K. Probabilidad y estadística para ingeniería y ciencias (9na Edición).  
— Capítulo 8
- Montgomery, D; Runger, G. Applied Statistics and Probability for Engineers (5th Edition)  
— Capítulo 7.2



# **Distribuciones muestrales y TLC**

## **II-1123 Estadística para Ingeniería Industrial II**

**Gracias por su atención**

**Steven García Goñi**

[steven.garciagoni@ucr.ac.cr](mailto:steven.garciagoni@ucr.ac.cr)

**Dudas o correcciones requeridas pueden solicitarse al correo**

