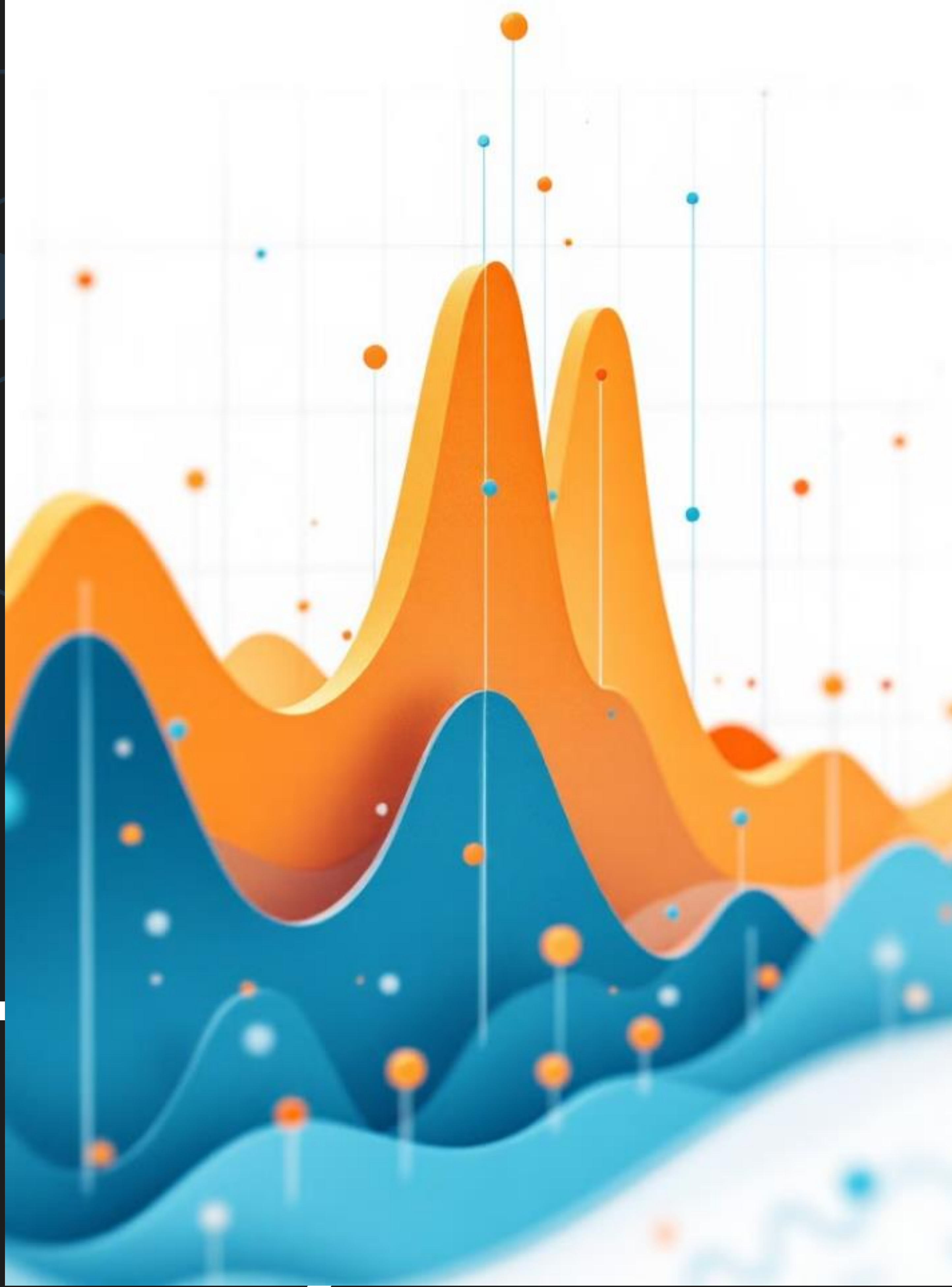
Inferencia

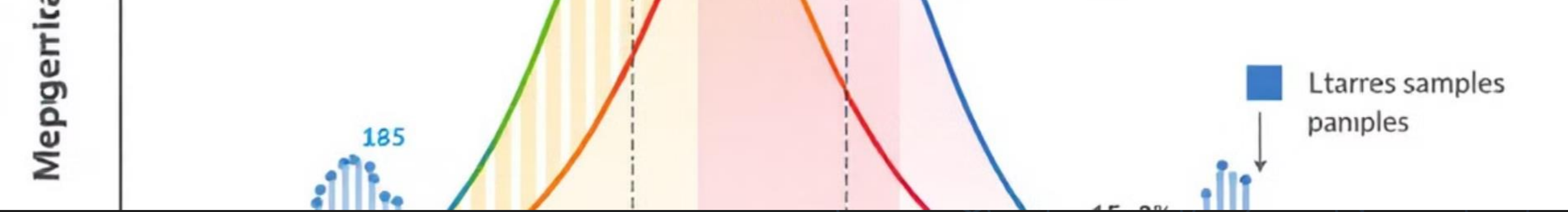
Estadística

Sesión 4

Distribución Muestral y Teorema del Límite Central

La distribución muestral y el teorema del límite central son conceptos fundamentales en estadística y ciencia de datos. Estos tópicos permiten entender cómo se comportan las muestras extraídas de una población y cómo inferir propiedades de la población a partir de muestras.





La Distribución de Muestreo

Definición

La distribución de muestreo es la distribución de probabilidad de un estadístico (como la media, la proporción o la varianza) calculado a partir de múltiples muestras de una población.

Media de la Distribución

La media de la distribución muestral es igual a la media de la población: $E(\bar{X}) = \mu$

$$\mathbb{E}(\bar{X}) = \mu$$

Error Estándar

La desviación estándar de la distribución muestral (error estándar) se calcula como: $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, donde σ es la desviación estándar poblacional y n el tamaño de la muestra.

Ejemplo de Distribución Muestral

Parámetros

Supongamos que la población tiene una media μ=50 y una desviación estándar σ=10. Si tomamos muestras de tamaño n = 25, la distribución de muestreo de la media tendrá:

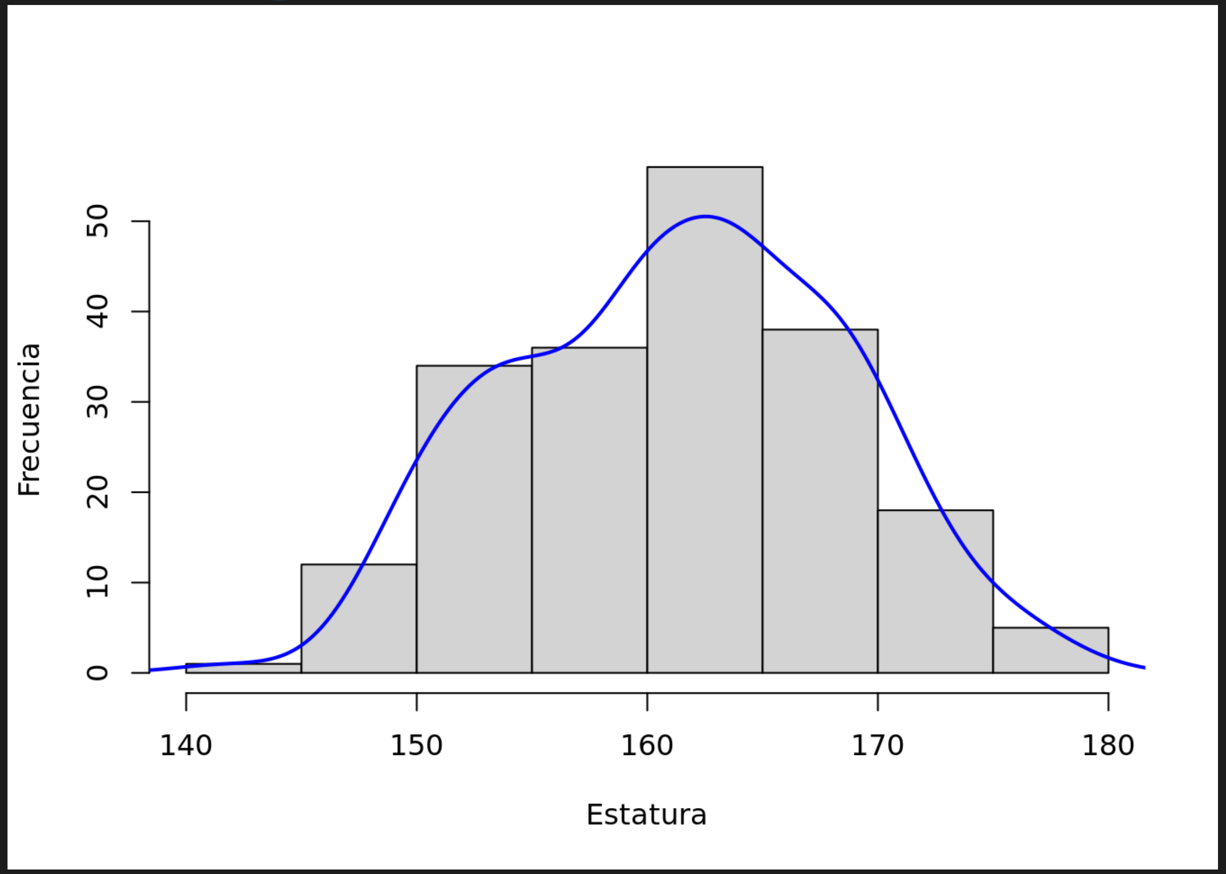
Media: E(X̄) = 50

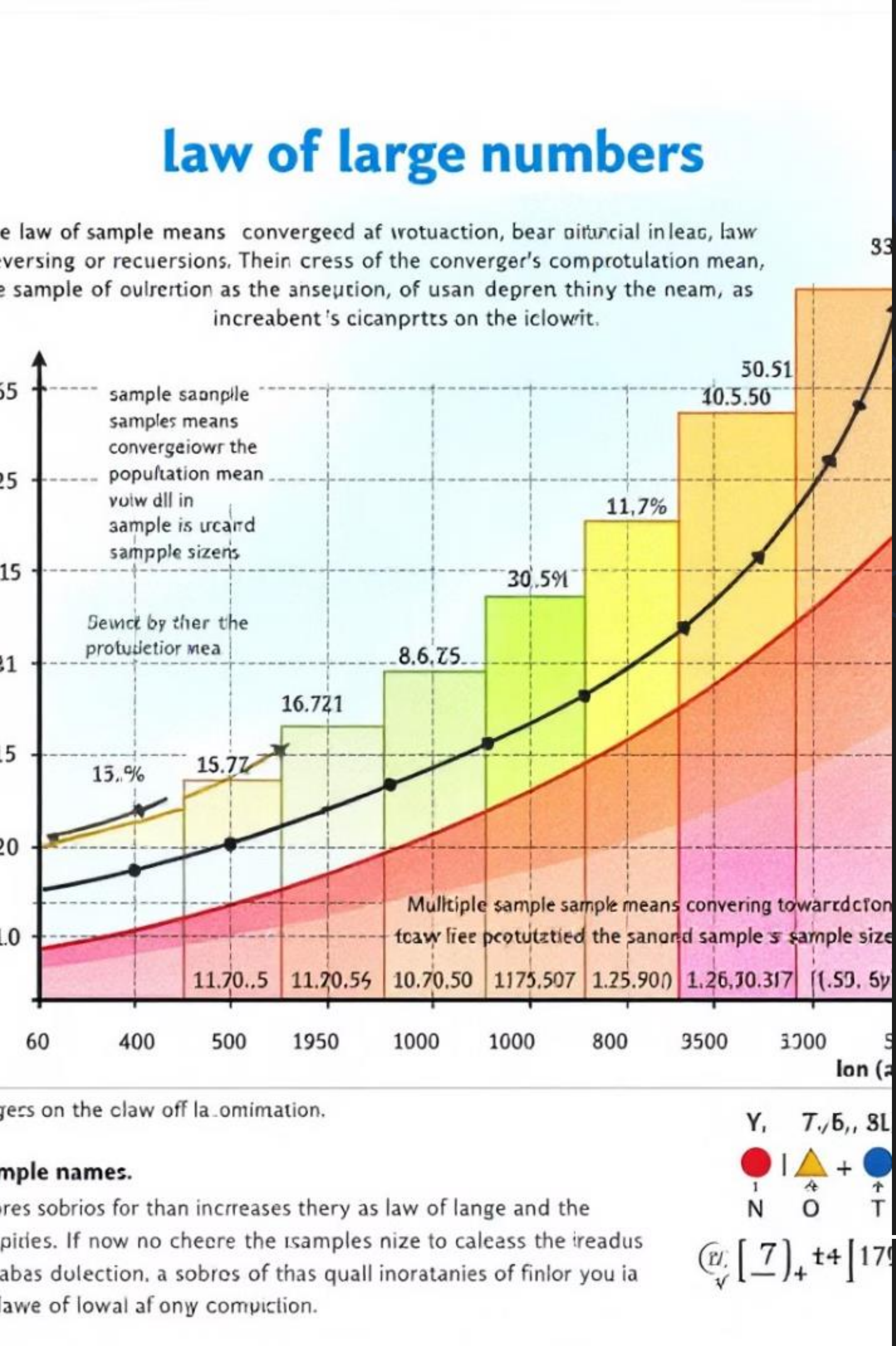
Error estándar: σX̄ = 10/√25 = 10/5 = 2

$$\sigma_{\bar{X}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2$$

Interpretación

Esto significa que la media de las muestras oscilará en torno a 50 con una variabilidad de 2 unidades. Cuanto mayor es el tamaño de la muestra n, menor es el error estándar, lo que significa que las medias muestrales estarán más concentradas alrededor de la media poblacional.





Ley de los Grandes Números

1

Definición

La Ley de los Grandes Números (LGN) establece que, a medida que el tamaño de la muestra n aumenta, la media muestral $\bar{X}$ se aproxima a la media poblacional μ.

2

Fórmula Matemática

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas con media μ, entonces: $X^{-}=1/n$ $n\sum i=1$ $Xi \to \mu$ cuando $n \to \infty$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad \text{cuando } n \to \infty.$$

3

Tipos

Ley Débil: Garantiza que la media muestral converge en probabilidad a la media poblacional.

Ley Fuerte: Afirma que la media muestral converge casi seguramente a la media poblacional.

Ejemplos de la Ley de los Grandes Números

Lanzamiento de Moneda

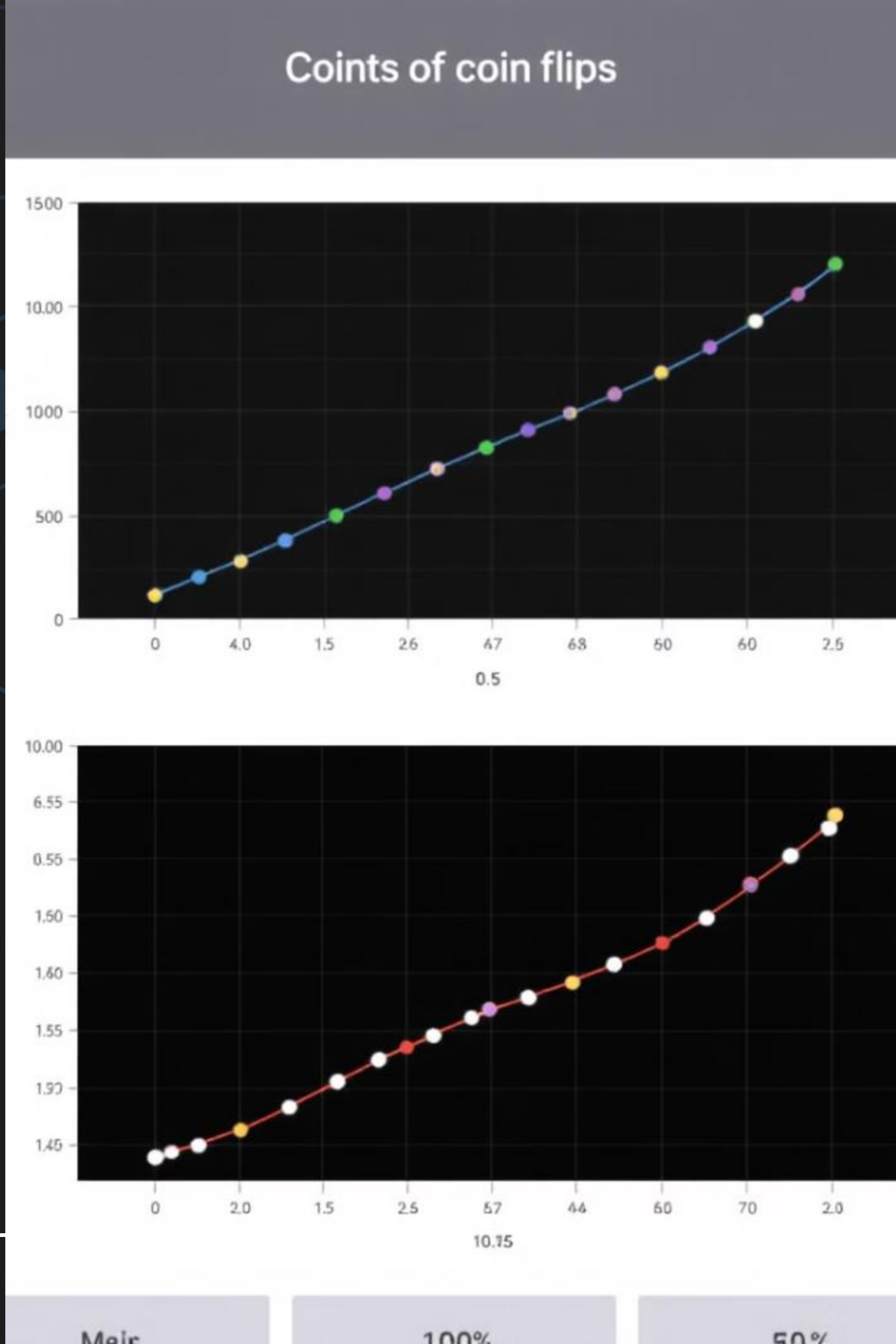
Si lanzamos una moneda 10 veces, la proporción de caras puede ser diferente de 0.5. Si lanzamos la moneda 10,000 veces, la proporción de caras estará muy cerca de 0.5.

Promedio de Alturas

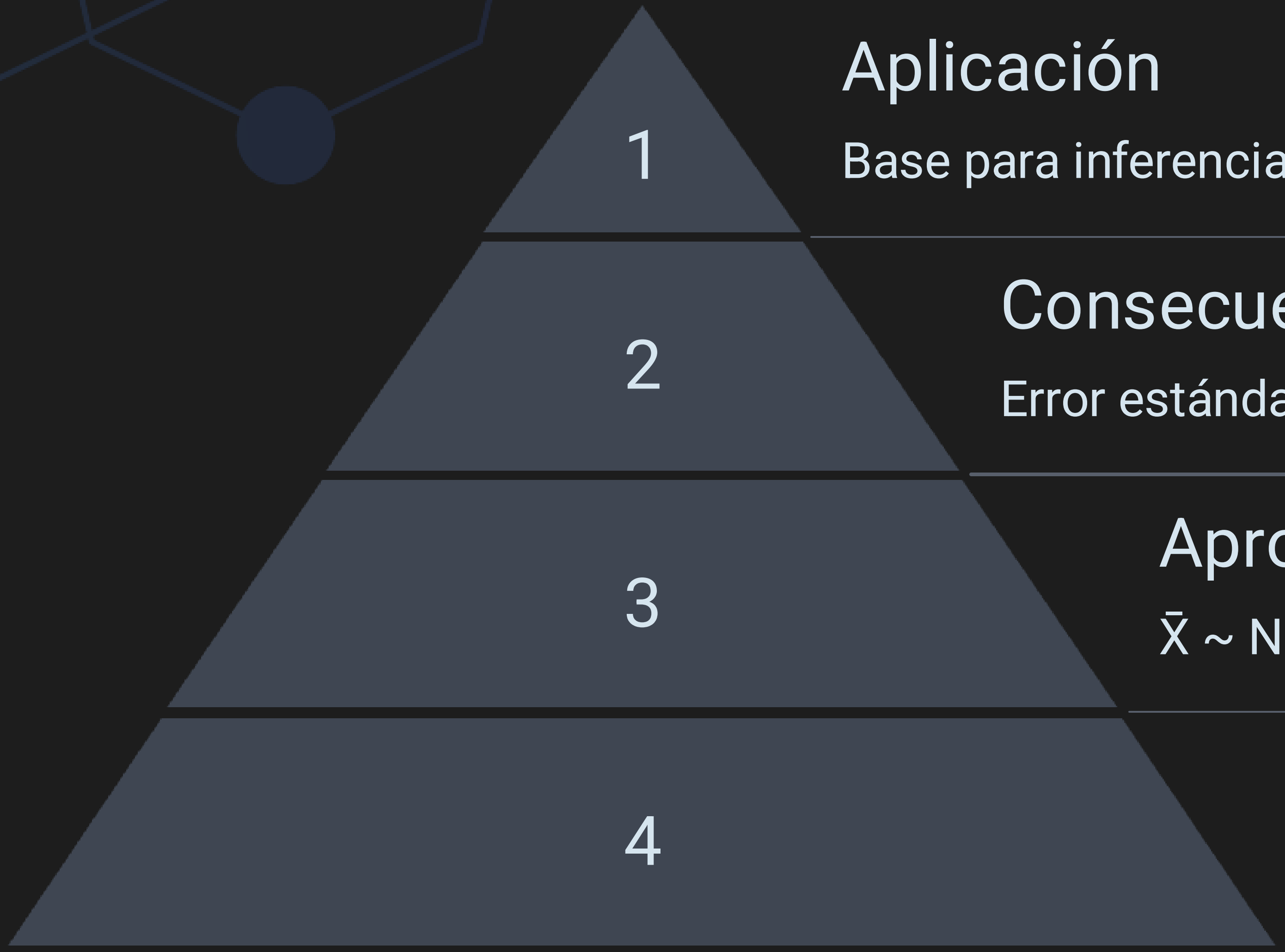
Si medimos la altura de 5 personas, la media puede diferir de la media poblacional. Si medimos la altura de 100,000 personas, el promedio se acercará a la media real de la población.

Simulación en Python

Podemos simular la Ley de los Grandes Números para observar cómo la media muestral converge hacia la media poblacional, como en el ejemplo de lanzamientos de moneda donde la proporción acumulada converge a 0.5.



Teorema del Límite Central



Aplicación

Base para inferencia estadística

Consecuencia

Error estándar: $\sigma\bar{X} = \sigma/\sqrt{n}$

$$\text{Error estándar} = \frac{\sigma}{\sqrt{n}}$$

Aproximación

$\bar{X} \sim N(\mu, \sigma^2/n)$

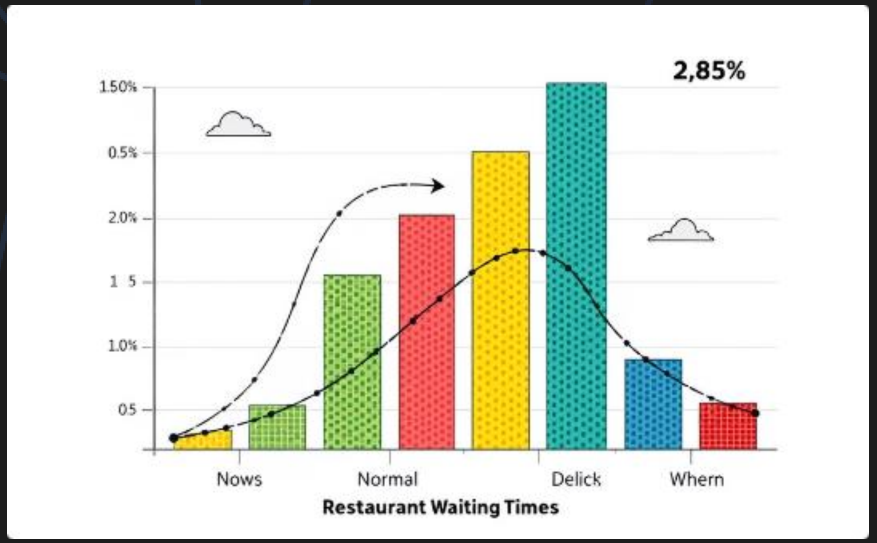
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{cuando } n \to \infty.$$

Definición

Para n grande, la media muestral sigue distribución normal

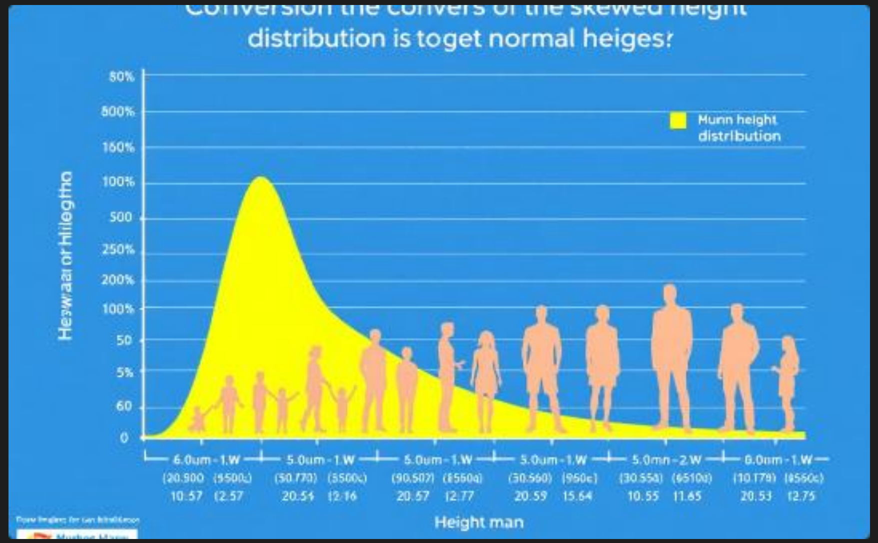
El Teorema del Límite Central (TLC) afirma que, bajo ciertas condiciones, la distribución de la media muestral se aproxima a una distribución normal, sin importar la forma original de la distribución de la población. Esto ocurre cuando las variables son independientes e idénticamente distribuidas y el tamaño de muestra es suficientemente grande.

Ejemplos del Teorema del Límite Central



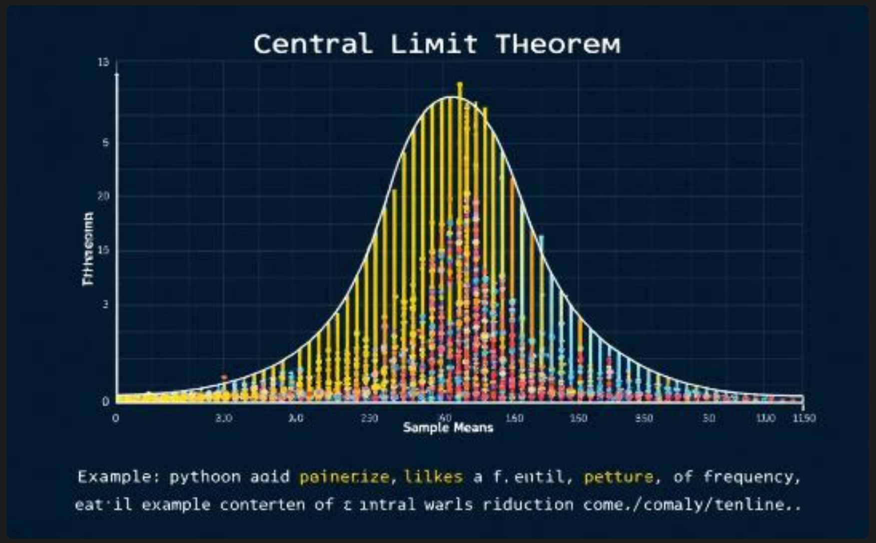
Tiempo de Espera en Restaurante

Si el tiempo de espera sigue una distribución desconocida, con media de 15 minutos y varianza de 9, al tomar muestras de n = 30 clientes, la distribución de medias muestrales se aproximará a N(15, 9/30) = N(15, 0.3).



Altura de Estudiantes

Si la distribución original de alturas es asimétrica y la media poblacional es $\mu=1.70$m con desviación estándar $\sigma=0.08$m, para muestras de n=50, la distribución de medias muestrales será $N(1.70, 0.08^2/50) = N(1.70, 0.00013)$.



Simulación en Python

Podemos simular el TLC generando muestras de tamaño n=30 a partir de una distribución exponencial (no normal), calculando las medias muestrales y observando cómo su distribución se aproxima a la normal.

Relación entre Distribuciones

1

Distribución de la Población

Representa la distribución de una variable en toda la población. Incluye todos los posibles valores que puede tomar la variable de interés. Sus parámetros (media μ y varianza σ^2) suelen desconocerse.

2

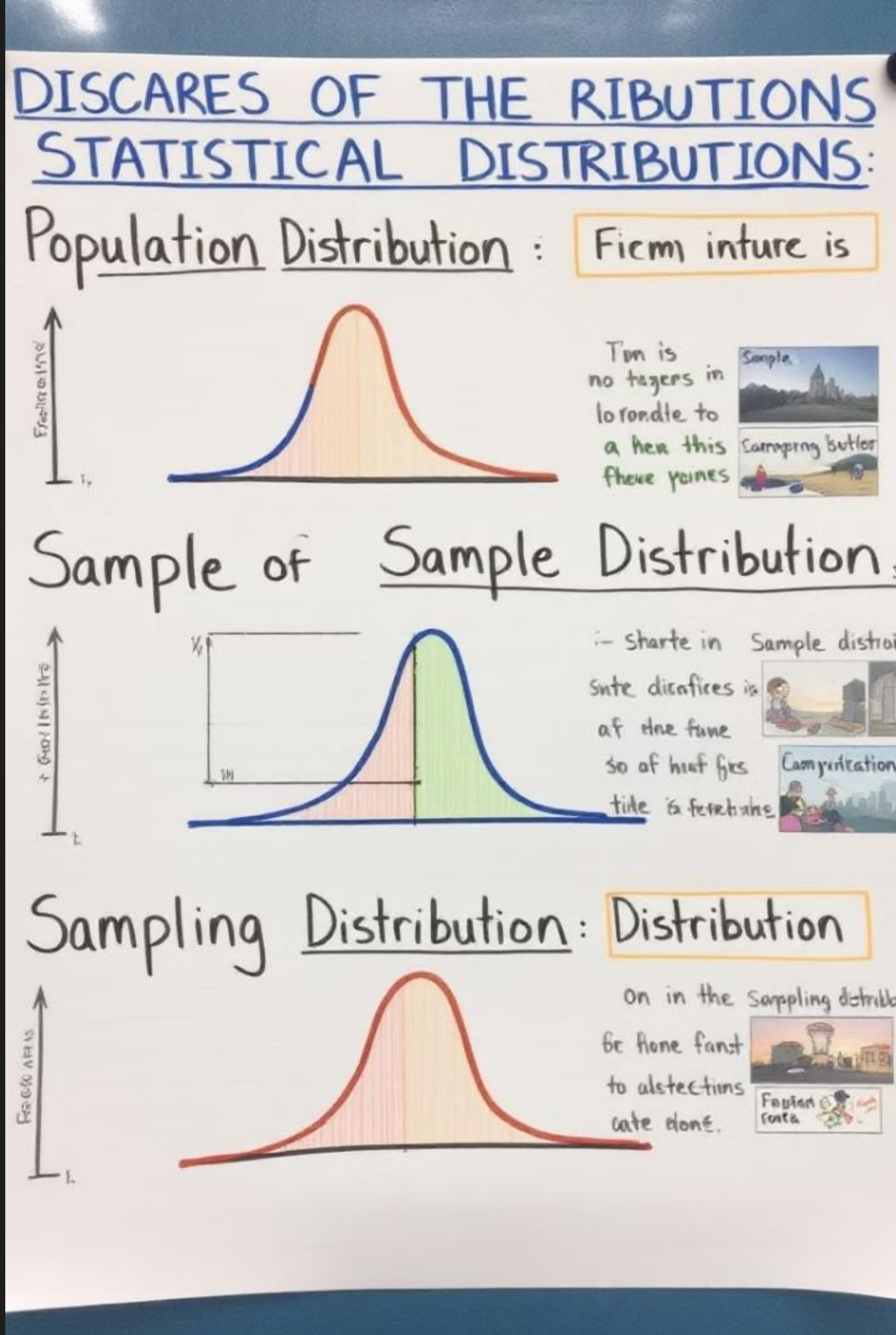
Distribución de la Muestra

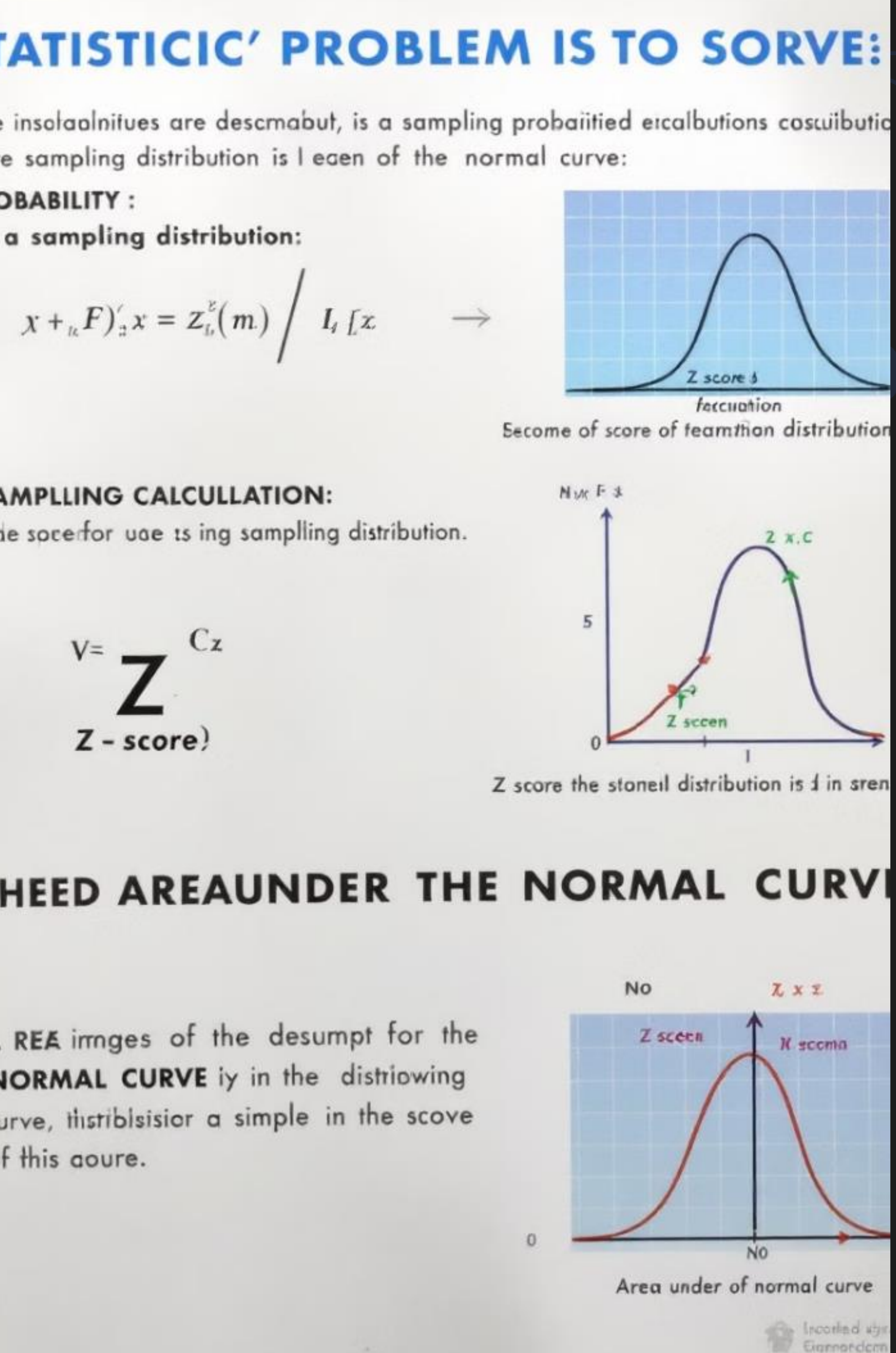
Describe la distribución de la variable en una muestra específica extraída de la población. Su media y varianza ($\bar{X}$, s^2) son estimaciones de los parámetros poblacionales.

3

Distribución Muestral

Es la distribución de un estadístico (como la media o la proporción) calculado a partir de muchas muestras de tamaño n. Si el tamaño de la muestra es suficientemente grande, la distribución muestral de la media se aproxima a una normal.





Cálculo de Probabilidades con la Distribución Muestral

Identificar el Estadístico

Puede ser la media muestral ($\bar{X}$), la proporción muestral ($\hat{p}$), entre otros.

Determinar la Distribución Muestral

Si la población sigue una distribución normal, la media muestral también lo hará. Si la población no es normal pero $n \geq 30$, podemos usar el TLC para aproximar a una normal.

Calcular el Error Estándar

$$SE = \frac{\sigma}{\sqrt{n}}$$

Para la media muestral: $SE = \sigma/\sqrt{n}$. Para la proporción muestral: $SE = \sqrt{p(1-p)/n}$.

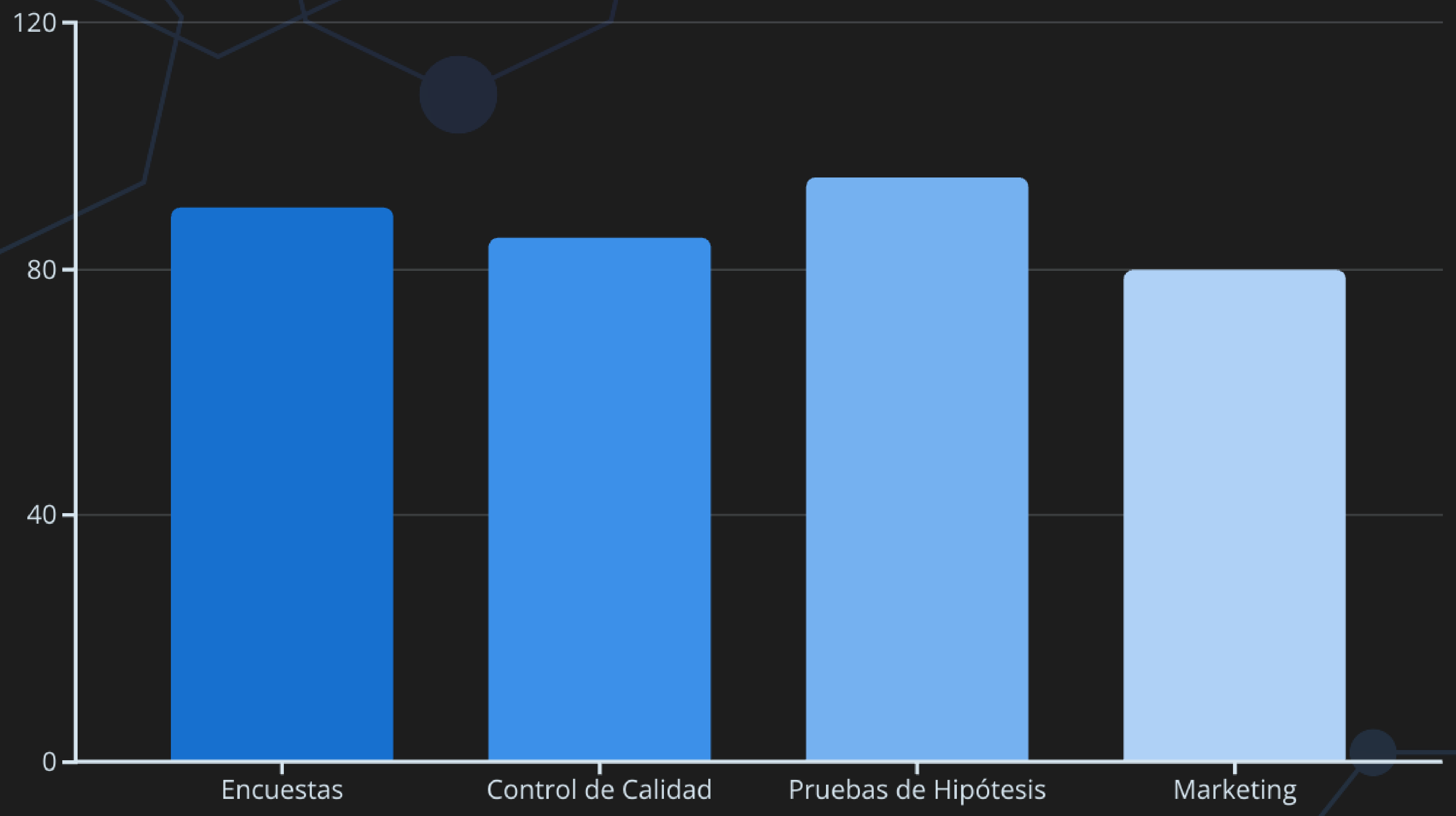
$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Estandarizar con Z

Convertimos el valor de interés a una puntuación Z: $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$. Luego usamos tablas Z o software estadístico para encontrar la probabilidad.

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

Distribución Muestral de Proporciones

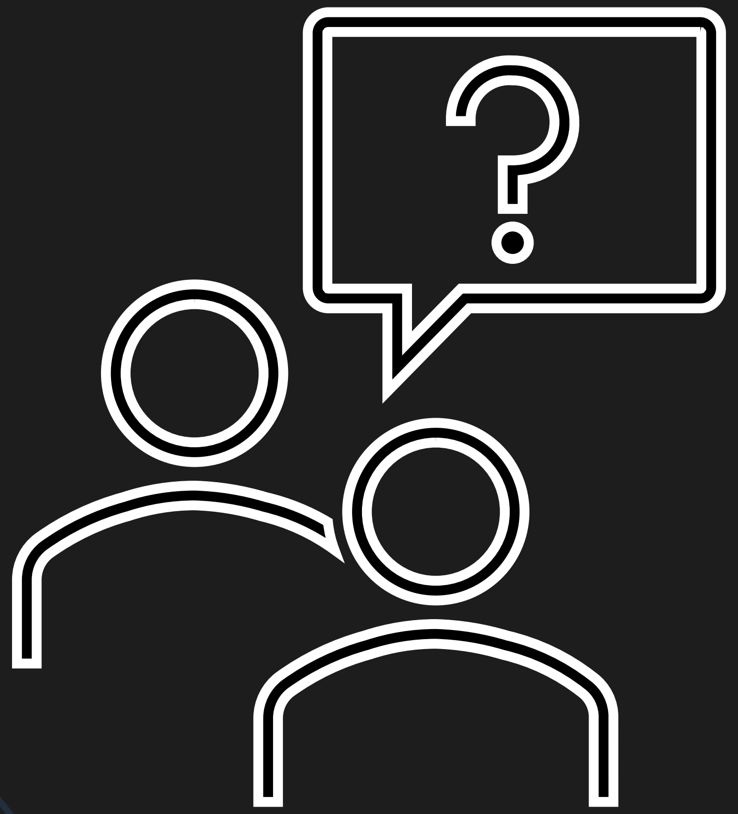


La distribución muestral de proporciones describe cómo varía la proporción muestral $\hat{p}$ en múltiples muestras. Si p es la proporción poblacional y se toman muestras de tamaño n, entonces $\hat{p}$ sigue aproximadamente una distribución normal cuando n es suficientemente grande ($np \geq 5$ y $n(1-p) \geq 5$).

La media de esta distribución es $E(\hat{p}) = p$ y el error estándar se calcula como $\sigma\hat{p} = \sqrt{p(1-p)/n}$. Esto permite calcular probabilidades como $P(\hat{p} > \text{valor})$ utilizando la distribución normal estándar, facilitando inferencias sobre proporciones poblacionales.

Preguntas

Sección de preguntas



Inferencia

Estadística

Continúe con las actividades