

TEOREMA CENTRAL DEL LÍMITE (TCL)

Es uno de los fundamentos más importantes de la estadística inferencial, ya que permite comprender por qué, bajo ciertas condiciones, las operaciones de muestreo generan resultados que pueden analizarse utilizando la distribución normal, incluso cuando la población original no sigue esta distribución. En términos simples, el TCL establece que cuando se extraen muestras aleatorias e independientes de tamaño suficientemente grande de una población con media finita y varianza finita, la distribución de las medias muestrales tenderá a aproximarse a una distribución normal conforme el tamaño de la muestra aumenta; esta propiedad convierte a la media muestral en un estimador altamente útil y confiable en estudios aplicados.

El TCL se fundamenta en el hecho de que la media muestral es una combinación lineal de variables aleatorias, y al promediar un número creciente de observaciones, las fluctuaciones extremas tienden a cancelarse. Por ello, aun si la población tiene una distribución altamente sesgada, multimodal o con colas no simétricas, las medias muestrales convergen gradualmente a una forma de campana. Este fenómeno, conocido como **convergencia en distribución**, es clave porque permite aplicar herramientas basadas en la normalidad sin necesidad de conocer la forma exacta de la distribución poblacional.

El Teorema Central del Límite establece que, cuando el tamaño de la muestra es suficientemente grande ($n \geq 30$), la distribución de las medias muestrales se aproxima a una distribución normal, independientemente de la forma de la distribución de la población.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ cuando } n \text{ es grande}$$

Un elemento fundamental asociado al TCL es el error estándar, que representa la variabilidad de las medias muestrales y se calcula como $\frac{\sigma}{\sqrt{n}}$. A medida que el tamaño de la muestra aumenta, el error estándar disminuye, lo que significa que las medias muestrales se agrupan más estrechamente alrededor del valor real de la media poblacional. Esto explica por qué, con muestras grandes, la media muestral no solo tiende a ser aproximadamente normal, sino también más precisa.

El TCL no solo aplica a medias, sino también a proporciones:

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

Con condiciones:

$$np \geq 5 \text{ y}$$

$$nq \geq 5.$$

Cuando estas condiciones se satisfacen, la distribución de la proporción muestral puede aproximarse a una normal, permitiendo la construcción de intervalos de confianza y la aplicación de **pruebas z** sin requerir transformaciones adicionales o supuestos estrictos sobre la población.

Las implicaciones del TCL en la inferencia estadística son profundas. Gracias a él, es posible emplear la distribución normal estándar (z) para construir intervalos de confianza y realizar pruebas de hipótesis en muestras grandes.

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

Sin el TCL, la estadística inferencial moderna (que abarca desde análisis de mercado hasta modelos econométricos) sería extremadamente limitada, especialmente en contextos reales donde las distribuciones poblacionales rara vez son normales. Además, el TCL justifica el uso de métodos asintóticos que fundamentan técnicas avanzadas de modelado, simulación y predicción utilizadas en economía, finanzas, operaciones y comercio internacional.

El Teorema Central del Límite explica por qué las medias muestrales convergen a la normalidad, reduce la importancia del conocimiento exacto de la distribución poblacional y proporciona la base matemática para los métodos de inferencia utilizados en prácticamente todas las disciplinas aplicadas. Su relevancia radica no solo en su elegancia teórica, sino en su utilidad práctica:

El TCL permite convertir datos complejos y desordenados en resultados confiables y comparables, haciendo posible la toma de decisiones informadas en entornos empresariales, económicos y globales.

Ejemplo práctico:

Supongamos que se estudian los **días que tarda en llegar un embarque marítimo** desde un puerto de Asia a un puerto en México.

- La distribución real de los tiempos **NO es normal**: hay muchos envíos que llegan en 18-22 días, pero algunos se retrasan mucho (clima, congestión, aduanas).
- A partir de datos históricos se conoce, de forma bastante confiable, que:
 - Media poblacional: $\mu = 20$ días
 - Desviación estándar poblacional: $\sigma = 6$ días

Una empresa de logística quiere saber: “Si se toma una muestra de 36 embarques recientes, ¿cuál es la probabilidad de que el promedio muestral de días sea menor o igual a 18 días?”

Como la distribución poblacional no es normal pero el tamaño de la muestra es grande ($n = 36 \geq 30$), entonces es posible utilizar el Teorema Central del Límite.

Identificación de datos

- $\mu = 20$ días
- $\sigma = 6$ días
- $n = 36$ embarques
- Queremos: $P(\bar{X} \leq 18)$

Cálculo del error estándar (SE)

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

Cálculo de $\sqrt{n}$:

- $n = 36$
- $\sqrt{36} = 6$

Entonces:

$$SE = \frac{6}{6} = 1$$

El error estándar es **1 día**, lo que significa que **las medias muestrales** (promedios de 36 embarques) varían alrededor de la media poblacional con una desviación de aproximadamente 1 día.

Aplicación del TCL: distribución de la media muestral

Por el Teorema Central del Límite:

$$\bar{X} \sim N\left(\mu = 20, \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{36}{36} = 1\right)$$

Es decir:

$$\bar{X} \sim N(20,1)$$

Estandarización (cálculo del valor z)

Se busca que:

$$P(\bar{X} \leq 18)$$

Convertir a la distribución normal (z):

$$z = \frac{\bar{X} - \mu}{SE} = \frac{18 - 20}{1} = \frac{-2}{1} = -2$$

Resultando que:

$$P(\bar{X} \leq 18) = P(Z \leq -2)$$

Buscando en la tabla de la normal estándar, se ubica el valor de z en:

$$P(Z \leq -2) \approx 0.0228$$

Interpretación:

Hay aproximadamente un 2.28 % de probabilidad de que, al tomar una muestra aleatoria de 36 embarques, el promedio de días de tránsito sea 18 días o menos.

Construcción de un intervalo de confianza

Supóngase ahora que la empresa tomó realmente una muestra de 36 embarques y obtuvo:

- $\bar{X}_{obs} = 19.2$ días
- Queremos un **intervalo de confianza del 95 %** para la media poblacional.

Utilizando:

$$IC = \bar{X} \pm z \cdot SE$$

Teniendo que:

- $z = 1.96$ (para 95 %)
- $SE = 1$ (ya calculado)

Entonces:

$$IC_{95\%} = 19.2 \pm 1.96(1) = 19.2 \pm 1.96$$

Se calculan los límites:

- Límite inferior: $19.2 - 1.96 = 17.24$
- Límite superior: $19.2 + 1.96 = 21.16$

$$IC_{95\%} = (17.24, 21.16) \text{ días}$$

Interpretación:

Con un intervalo de confianza del 95 %, la media real de días que tardan los embarques está entre 17.24 y 21.16 días.

Referencia:

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2021). Statistical Techniques in Business & Economics (18th ed.). McGraw-Hill Education.