

Segunda Clase Métodos Cuantitativos II

Promedios, Intervalos de Confianza y Teoría del Límite Central

Taller en R-studio

Docente: Francisco Meneses

Resultados ticket n° 0

Conceptos elementales

- Sobre los datos
 - **Población** (Grupo de interés de estudio)
 - **Muestra** (Subconjunto de la población utilizado en estudio)
- De los cálculos
 - **Parámetro** (Valor/cálculo en la población: ej μ)
 - **Estadístico** (Valor/cálculo en la muestra: ej \bar{x})
 - **Inferencia** (Inferir información de la población en base muestras)
 - **Diferencia significativa** (Diferencias en la muestra que asumimos existen en la población)
- De las ideas
 - **Hipótesis** (Resultado teórico o estadístico esperado)

Promedios: lo básico

- En investigación cuantitativa, **resumimos fenómenos** con medidas estadísticas.
- Una de las más comunes es el **promedio (media)**.

El promedio es el "centro" de una distribución:
suma de los valores dividida por la cantidad de observaciones.

Ejemplo:

Puntajes PSU:

500, 520, 480, 510, 490

Promedio: $(500 + 520 + 480 + 510 + 490) / 5 = 500$

¿Por qué usar promedios?

- **Simplifican** la información.
- Permiten comparar **grupos** o **momentos en el tiempo**.
- Son la base para el análisis **inferencial**.

¿Qué limitaciones tienen?

- Son sensibles a **valores extremos (outliers)**.
- No reflejan la **variabilidad** entre personas.

¿Qué tan seguro estoy del promedio?

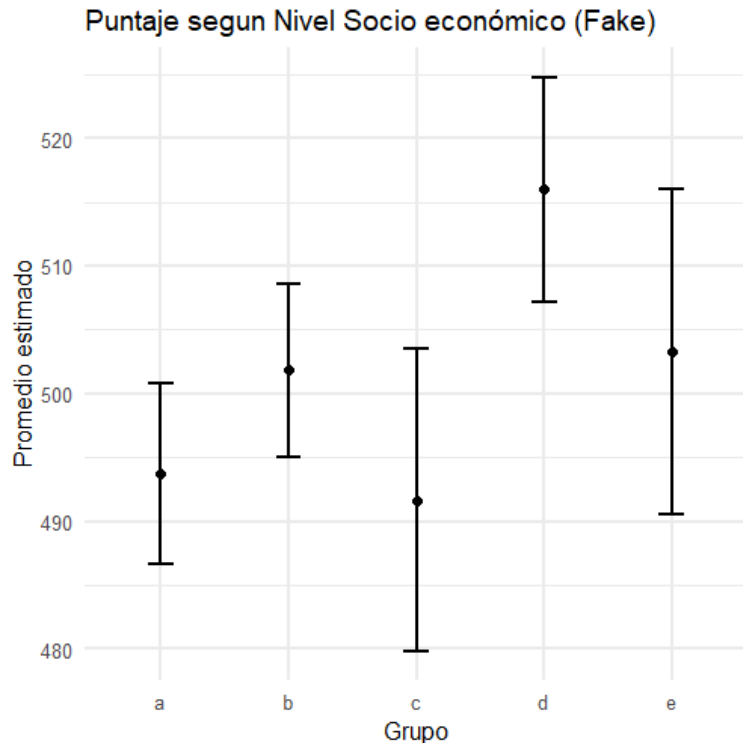
- Cuando calculamos un promedio a partir de una **muestra**, no es exacto.
- Hay **incertidumbre**: otra muestra daría otro promedio.
 - | Por eso usamos **intervalos de confianza (IC)**.

Intervalo de Confianza (IC)

Es un **rango** en torno al promedio muestral que contiene al verdadero promedio poblacional con cierta seguridad (confianza).

- Ej: “El promedio estimado es 500 puntos, con un IC del 95% entre 485 y 515”.
- Significa que **si repitiéramos el estudio muchas veces**, el 95% de los intervalos obtenidos contendría el valor real.

¿Cómo se ve un intervalo de confianza?



Nota: Apreciamos 5 grupos socioeconómicos con distintos promedios SIMCE y distintos intervalos de confianza. Algunos intervalos son más grandes que otros. Algunos grupos se solapan y otros no. El grupo c no se solapa con el grupo d

Si los intervalos no se solapan decimos que existen diferencias significativas.

Diferencias significativas:
Diferencias que existen más allá de la muestra.

¿Utilidad de los intervalos de confianza?

- Podemos **comparar grupos**: ¿las diferencias son reales o por el azar de la muestra?
- Podríamos evaluar si un programa **funcionó**.
- Ayuda a **comunicar resultados** con transparencia.

¿Cómo se calcula el intervalo de confianza para una media?

Para una **variable continua**, como por ejemplo el puntaje en una prueba, el **intervalo de confianza del 95%** para la media se calcula como:

$$\text{IC 95\%} = \text{media} \pm (\text{valor crítico}) \times (\text{error estándar})$$

- **media**: promedio muestral
- **valor crítico**: depende del nivel de confianza (ej: 1.96 para 95%)
- **error estándar**: $EE = \frac{s}{\sqrt{n}}$

donde:

- s : desviación estándar de la muestra
- n : tamaño de la muestra

Ejemplo:

Supongamos que en una muestra de estudiantes:

- Media = 70
- Desviación estándar = 10
- Tamaño de muestra = 25

$$IC = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$$

$$IC = 70 \pm 1.96 \cdot \frac{10}{\sqrt{25}}$$

Margen de error: 3,92; Intervalo de confianza: 66,08 - 73,92

Interpretación: Estamos un 95% confiados de que el promedio real en la población se encuentra entre **66.08 y 73.92 puntos**.

¿Qué intervalos queremos?

- Intervalos más pequeños son más precisos y útiles

¿De qué depende el tamaño del intervalo?

- Del tamaño de la muestra (n)
 - Muestras más grandes dan intervalos más pequeños (Precisos)
- Nivel de confianza (90%, 95%, 99%, 99.9%)
 - A mayor nivel de confianza más grande es el intervalo (Impreciso pero más confiable)
- Nivel de dispersión de los datos
 - Datos más dispersos implican intervalos más grandes.

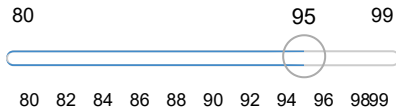
Intervalos de confianza para tres muestras

Promedio poblacional (μ):

**Desviación estándar
poblacional (σ):**

Tamaño de la muestra (n):

Nivel de confianza (%):



Vamos a calcularlo en R

Teoría del Límite Central (TLC)

- Es la **base teórico matemático** de los Intervalos de confianza
- Desde aquí nace la idea del 1.96 en la fórmula para un intervalo de confianza que utiliza el 95% de confianza

“Si graficamos la distribución del promedio de muchas muestras tiende a una normal al aumentar el tamaño de la muestra, **sin importar la distribución original.**”

En suma

Si sacamos una muestra al azar, es altamente improbable que el promedio de la muestra sea muy distinto al obtenido

Podemos calcular un intervalo en torno al promedio muestral en el cual tendremos un % de confianza de que el parámetro poblacional está incluido.

Simulación del Teorema del Límite Central

Selecciona la distribución:

Normal ▼

Media de la población:

6000

Desviación estándar de la población:

200

Tamaño de la muestra (n):

30

Número de muestras:

1000

Generar Muestras

Distribución de la Variable Original

Distribución del Promedio de las Muestras

¿Por qué es importante?

- Nos permite **estimar la incertidumbre** con herramientas conocidas.
- Hace posible usar la **distribución normal** (z, t, etc.) para construir intervalos y hacer pruebas.

Recapitulación

- El **promedio** resume nuestros datos, pero tiene incertidumbre debido a trabajar con muestras.
- Los **intervalos de confianza** permiten expresar esa incertidumbre.
- Gracias a la **Teoría del Límite Central**, podemos confiar en la forma en que estimamos esa incertidumbre.

Pregunta para pensar 🤔

| ¿Por qué es más útil reportar un promedio **con su intervalo de confianza**, en vez del promedio solo?

- ¿Qué responderías tú como investigador/a?

Actividad breve

- Forma grupos de 2-3 personas.
- Responde:
 - ¿Cómo explicarías el concepto de intervalo de confianza a alguien que no estudia estadísticas?
 - ¿En qué tipo de decisiones crees que sería útil usarlo?
 - ¿Puedes dar un ejemplo en educación, salud o política?