

---

## **Clase 182 — Intervalos de confianza**

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 5 + Bruce & Bruce, cap. 2  
Confidence Intervals. Duración estimada: 70 min.

## Clase 182 — Intervalos de confianza

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 5 + Bruce & Bruce, cap. 2 Confidence Intervals. Duración estimada: 70 min.

### Objetivo

Construir e interpretar correctamente intervalos de confianza para media (t-based, z-based, bootstrap) y proporción (Wald, Wilson, Clopper-Pearson), entendiendo que un IC95 % NO significa "95 % de probabilidad de que el parámetro caiga en el intervalo" sino "si repitiéramos el experimento muchas veces, el 95 % de los intervalos construidos contendrían el parámetro". Saber elegir el método según n y la métrica.

### Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Construir un IC para la media usando la distribución t:  $\bar{x} \pm t_{\{\alpha/2, n-1\}} \cdot (s/\sqrt{n})$  con `scipy.stats.t.interval`.
- Construir un IC para la proporción con tres métodos y entender cuándo cada uno falla (Wald falla con p cerca de 0/1; Wilson y Clopper-Pearson son robustos).
- Usar `scipy.stats.bootstrap` ( $\geq 1.7$ ) para IC sin supuestos paramétricos (anticipa Clase 153).
- Interpretar correctamente la frase "intervalo de confianza al 95 %" (es una propiedad del procedimiento, no del intervalo específico).
- Relacionar IC y test de hipótesis: si el IC95 % de la diferencia no incluye 0, el test bilateral al  $\alpha=5$  % rechaza  $H_0$ .

### Temas

- IC para la media (varianza desconocida): t de Student con n-1 gl.
- IC para la media (varianza conocida o n grande): z.
- IC para proporción: Wald ( $\hat{p} \pm z \cdot \sqrt{\hat{p}(1-\hat{p})/n}$ ) vs Wilson score (recomendado por Agresti & Coull 1998) vs Clopper-Pearson (exacto, conservador).
- IC bootstrap percentil (anticipa Clase 153).
- IC del odds ratio, riesgo relativo (medicina/epidemiología).
- Margen de error (ME = z·SE) y cómo determina n.

### Definiciones y características

- Intervalo de confianza (IC) al  $1-\alpha$ : par (L, U) calculado a partir de la muestra tal que, sobre repeticiones del experimento,  $P(L \leq \theta \leq U) = 1-\alpha$ . El parámetro  $\theta$  es fijo (no aleatorio); lo aleatorio son L y U.
- Cobertura nominal vs real: la cobertura "nominal" es 95 %; la "real" depende del método y la distribución. Wald infla error con n chico o p extrema.
- Standard error (SE) de la media:  $s / \sqrt{n}$ . Se reduce con  $\sqrt{n}$  — para reducir el SE a la mitad necesitas 4× la muestra.
- t-distribution: para IC de la media cuando estimas  $\sigma$ . Tiene colas más anchas que la normal, lo que infla correctamente el IC con n chico.
- Wilson score interval: usa la fórmula  $(\hat{p} + z^2/(2n) \pm z \cdot \sqrt{(\hat{p}(1-\hat{p})/n + z^2/(4n^2))}) / (1 + z^2/n)$ . Mantiene

cobertura  $\approx$  nominal incluso con  $\hat{p}$  cerca de 0 o 1.

- Clopper-Pearson: intervalo exacto basado en la distribución binomial. Garantiza cobertura  $\geq$  nominal (puede ser conservador).
- Bootstrap percentile interval: cuantiles  $\alpha/2$  y  $1-\alpha/2$  de la distribución bootstrap del estadístico. No requiere supuestos paramétricos.
- Margen de error (ME): mitad del ancho del IC. Para n requerido al planificar un estudio:  $n = (z \cdot \sigma / ME)^2$ .

### Dataset / recursos

- seaborn.load\_dataset('tips'): IC de la propina media.
- Encuesta sintética: simular  $n=500$  respuestas binarias con  $p=0.03$  (proporción chica  $\rightarrow$  Wald falla).
- Librerías: scipy.stats, statsmodels.stats.proportion, pingouin.

### Ejercicios

1. IC t para la media: con tips.total\_bill, calculá el IC95 % con scipy.stats.t.interval(0.95, n-1, loc=mean, scale=sem). Verificá contra pingouin.compute\_bootci.
2. IC para proporción extrema: con rng.binomial(1, 0.03, 100) (proporción de eventos raros), calculá IC con statsmodels.stats.proportion.proportion\_confint(count, n, method='normal') (Wald), 'wilson' y 'beta' (Clopper-Pearson). Observá cómo Wald da límite inferior negativo (¡imposible!) y los otros dos no.
3. Cobertura empírica: simulá 5 000 muestras de tamaño 30 de  $N(50, 10)$ . Para cada una, construí IC95 % t. Contá qué % contiene  $\mu=50$ . Debería ser  $\approx 95$  %.
4. Bootstrap IC: con tips.total\_bill, aplicá scipy.stats.bootstrap((tips.total\_bill,), statistic=np.mean, n\_resamples=10\_000, method='percentile'). Compará con el IC t.
5. Sample size: querés estimar una proporción con margen de error de  $\pm 2$  %, asumiendo  $\hat{p} \approx 0.5$  (peor caso). Calculá el n requerido para 95 % de confianza.

### Homework verificable

Diseñar un estudio para estimar la proporción de clientes satisfechos en una tienda:

1. Determinar n requerido para margen de error  $\pm 3$  % con 95 % de confianza asumiendo  $\hat{p} \approx 0.5$ .
2. Simular el experimento con esa n y  $p_{verdadera}=0.78$ .
3. Construir los 3 IC (Wald, Wilson, Clopper-Pearson) y comparar anchos.
4. En 3 líneas: justificar cuál reportarías y por qué.

Criterio de aceptación:  $n \approx 1068$ . Los 3 IC contienen 0.78. La justificación debe mencionar que Wilson tiene buen comportamiento general (cobertura  $\approx$  nominal con menos ancho que Clopper-Pearson) y es el recomendado actual.

### Errores comunes

Síntoma / mensaje	Causa y cómo arreglar	
"Hay 95 % de probabilidad de que $\mu$ esté en	Interpretación frecuentista incorrecta. Fi	datos) = 0.95").
Wald da IC con límite negativo para propor	Pasa con $\hat{p}$ cerca de 0/1 o n chico. Fix: W	
IC del 95 % se interpreta como "el dato ca	No, eso sería un intervalo de predicción (	
Construyo IC asumiendo normalidad con $n=8$	t.interval requiere normalidad o n grande.	
Comparo dos IC: "se solapan, no hay difere	Solapamiento de ICs no implica $p > 0.05$ . P	

## Preguntas frecuentes

¿Por qué a veces uso z y a veces t?

z cuando conocés  $\sigma$  poblacional (raro) o  $n \geq 30$  (TCL hace que la diferencia sea trivial). t cuando estimás  $\sigma$  con la muestra y n es chico. En la práctica moderna, siempre t (con n grande coincide con z, así que no perdés nada).

¿Wilson o Clopper-Pearson para proporciones?

Wilson por default (Agresti & Coull 1998, Brown et al. 2001 lo recomiendan). Clopper-Pearson si necesitás garantizar cobertura  $\geq$  nominal (FDA, ensayos clínicos).

¿Bootstrap siempre es mejor?

No siempre. Si tus supuestos paramétricos se cumplen, t-based es más eficiente (intervalos un poco más cortos). Bootstrap brilla con n chico no normal, estadísticos no estándar (mediana, percentil,  $R^2$ ), o estimadores complejos donde no hay fórmula cerrada.

¿IC95 % es siempre simétrico?

Para la media t-based, sí. Para proporciones y bootstrap, no — sobre todo cerca de los bordes. Eso es una característica, no un bug: refleja la asimetría real de la distribución muestral.

¿Cómo le explico el IC al cliente sin entrar en frecuentismo?

"Si repitiéramos el experimento muchas veces con muestras del mismo tamaño, el 95 % de los rangos que produciríamos contendrían el valor real. Este rango es uno de esos 95 % en promedio." O directamente: "el valor real está plausiblemente entre L y U; rangos más angostos requieren más datos".

## Referencias

- ISLP, cap. 5 — Resampling Methods.
- Bruce & Bruce, cap. 2 — Confidence Intervals.
- Agresti, A. & Coull, B. (1998), Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions, American Statistician.
- Brown, Cai & DasGupta (2001), Interval Estimation for a Binomial Proportion, Statistical Science — review de métodos.
- statsmodels.stats.proportion.proportion\_confint.
- scipy.stats.bootstrap — API moderna.

## Siguiente clase

Clase 183 — Bootstrap y permutation tests

## Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

## Archivos complementarios

- notebook.ipynb