

Efecto de la Falta de Aleatoriedad sobre la Cobertura de Intervalos de Confianza

Diego B. Meza Bogado

10 de junio de 2025

Índice

1. Diseño del experimento	1
2. Cobertura observada	1
3. Visualización de los IC	2
4. Discusión	2
5. Conclusiones	2

1. Diseño del experimento

Se simulan $R = 10\,000$ réplicas para una proporción poblacional $p = 0.5$ bajo dos esquemas:

1. **Muestreo aleatorio simple (MAS)**: cada réplica extrae $n = 50$ observaciones i.i.d. Bernoulli(0.5).
2. **Muestra determinista**: los valores se generan mediante $y_i = \mathbf{1}\{m i + b > 0\}$ con $m \sim U(-0.05, 0.05)$ y $b \sim U(-2.5, 2.5)$ sorteados por réplica, de modo que una vez fijados (m, b) la muestra carece de variabilidad interna.

Para cada muestra se estima \hat{p} y se construye un IC de Wald al 95 %

$$IC_{95\%} = \hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

2. Cobertura observada

Cuadro 1. Proporción de IC que contienen el valor verdadero $p = 0.5$ (R=10 000).

Diseño de muestreo	Cobertura
Muestreo aleatorio simple (MAS)	0.9336
Determinista (lineal)	0.0671

La cobertura bajo MAS se aproxima al 95 % nominal, mientras que la muestra determinista apenas cubre el 6.7 %, evidenciando la pérdida de validez al romper la aleatoriedad.

3. Visualización de los IC

La Figura 1 superpone los 10 000 intervalos de confianza de cada método. Las barras grises cubren $p = 0.5$; las rojas no lo hacen.

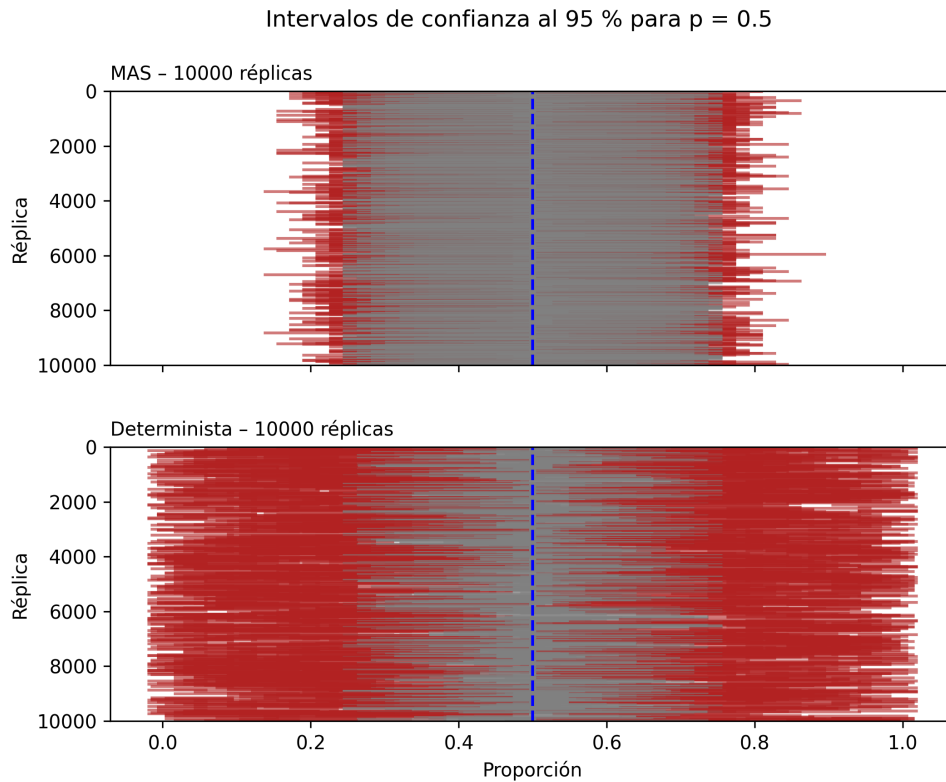


Figura 1. Intervalos de confianza al 95 % para $p = 0.5$; comparación entre MAS y muestra determinista. La línea azul discontinua marca el valor verdadero.

4. Discusión

El experimento demuestra que la fórmula clásica de Wald funciona cuando el supuesto de muestreo aleatorio se cumple; de lo contrario, el estimador \hat{p} puede quedar sesgado y la varianza interna subestimar la incertidumbre real. Como resultado, los IC calculados en la muestra determinista son demasiado estrechos y rara vez incluyen $p = 0.5$.

5. Conclusiones

- El diseño de muestreo es determinante para la validez de los IC: sin aleatoriedad, la cobertura nominal queda anulada.
- Aumentar el tamaño muestral no corrige el sesgo de selección ni restaura la cobertura.

- Estudios con muestras no probabilísticas deben aplicar ponderaciones o modelos de corrección para recuperar la inferencia.