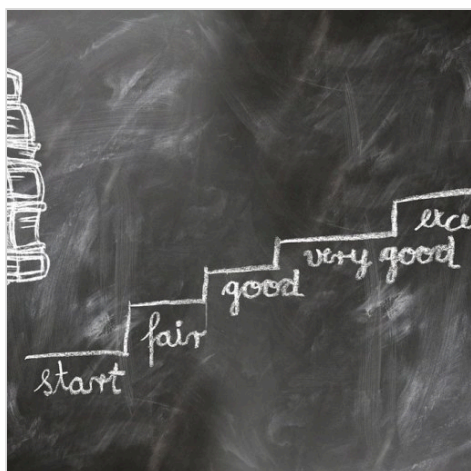


L U M B R E



Muestreo y error de selección

Grado: 3

Materia: Estadística Inferencial · Probabilidad y Estadística

11/10/2026

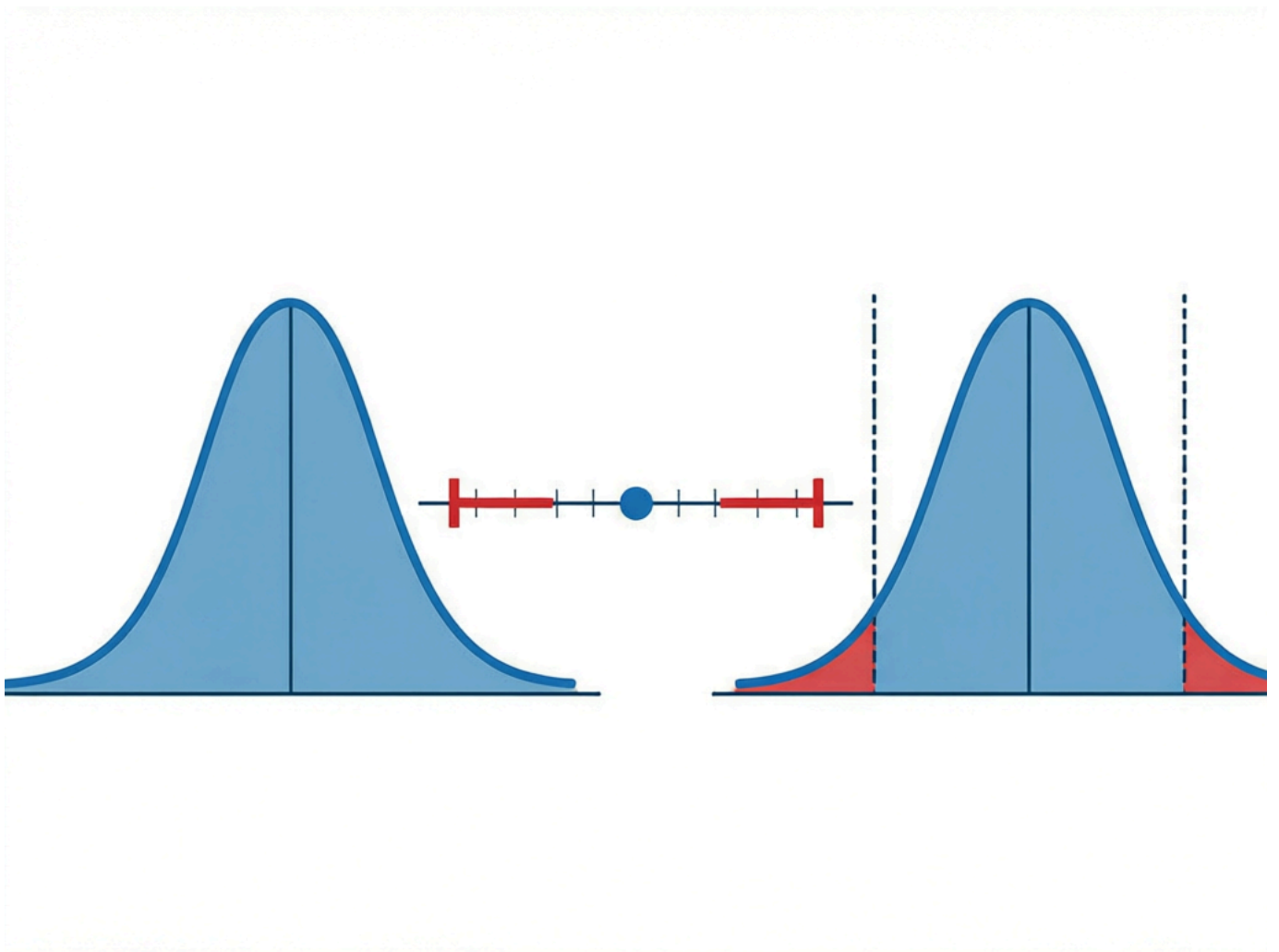
Antes de empezar

Casi toda conclusión estadística parte de una muestra. Si la muestra está mal elegida, ningún cálculo posterior la salva. Esta lección cubre cómo seleccionar bien y cómo reconocer el sesgo que lo contamina todo.

QUÉ APRENDERÁS

- Diseñar muestras representativas, distinguir tipos de muestreo y detectar los sesgos de selección que invalidan la inferencia.

Muestreo y error de selección



Estadística inferencial: estimacion, pruebas de hipotesis y regresion.

Mapa conceptual

- **Estadística Inferencial**
 - Muestreo
 - Aleatorio

- Estratificado
- Estimación
 - Puntual
 - Intervalos
- Hipótesis
 - H_0 y H_1
 - Valor p
 - Potencia
- Pruebas
 - t-Student
 - χ^2
 - ANOVA
- Regresión
 - Pendiente
 - Correlación

El mapa mental interactivo es interactivo en la versión online.



Mejor pocas y bien puestas

No hace falta medir a todos para conocer a muchos, pero sí medir a los indicados. Una muestra pequeña y bien elegida vale más que una enorme y sesgada. El error de selección no se corrige con más datos.

Población y muestra

La población es el conjunto completo sobre el que queremos concluir; la muestra, el subconjunto que realmente observamos. La inferencia consiste en usar la muestra para afirmar algo sobre la población, y solo funciona si la muestra la representa.

Marco muestral

El marco muestral es la lista desde la que se eligen las unidades: un censo, una base de clientes. Si el marco difiere de la población real (excluye a un grupo), aparece un sesgo de cobertura imposible de corregir con el método de selección. Definir bien el marco es el primer paso.

Muestreo aleatorio simple

Cada posible muestra de tamaño n tiene la misma probabilidad de ser elegida, y cada individuo la misma de entrar. Es el estándar de referencia: sin sesgo y con una teoría de la probabilidad limpia. Su limitación práctica es necesitar un marco completo y accesible.

Muestreo estratificado

Se divide la población en estratos homogéneos (edad, región) y se muestrea cada estrato, a veces en proporción a su tamaño. Garantiza que los grupos relevantes estén representados y reduce la varianza respecto al aleatorio simple cuando los estratos difieren entre sí.

Muestreo por conglomerados

Se eligen al azar grupos naturales (manzanas, aulas) y se observa a todos o varios dentro de ellos. Barato operativamente cuando la población está muy dispersa, a costa de mayor error porque los individuos dentro de un conglomerado se parecen entre sí.

Muestreo sistemático

Tomar un elemento cada k a partir de un arranque aleatorio. Simple y a menudo casi tan bueno como el aleatorio, pero peligroso si la lista tiene un ciclo oculto con periodo cercano a k , que introduciría un patrón sesgado.

Sesgo de selección

Ocurre cuando el proceso de elección hace que ciertos individuos tengan distinta probabilidad de entrar, de modo que la muestra no representa a la población. A diferencia del error de muestreo, el sesgo no desaparece aumentando n : se arrastra pase lo que pase.

Sesgo de no respuesta

Aunque la muestra esté bien elegida, si ciertas personas sistematizan no contestar (los más ocupados, los más afectados), los que responden forman una submuestra sesgada. Es un error de selección post-hoc, frecuentísimo en encuestas.

Muestreo por conveniencia

Encuestar a quien está a mano (voluntarios, clientes que pasan por la tienda) es barato pero peligroso: la muestra hereda el perfil de lo accesible. Los sondeos en línea con auto-selección son su caso emblemático, y su «volumen» no compensa el sesgo.

El legendario fracaso

Una encuesta gigantesca puede fallar estrepitosamente si su marco está sesgado. El tamaño no cura el sesgo: una muestra enorme de la gente equivocada da una estimación enorme y equivocada. La calidad de la selección manda sobre la cantidad.

1. Define la población y el objetivo de inferencia.
2. Consigue el marco muestral más fiel posible.
3. Elige el diseño (aleatorio, estratificado, por conglomerados) según coste y estructura.
4. Minimiza la no respuesta con seguimiento.
5. Documenta el método para poder juzgar su sesgo.

Tamaño de muestra y precisión

El error aleatorio decrece con la raíz cuadrada del tamaño: duplicar la precisión exige cuadruplicar la muestra. Hay rendimientos decrecientes, y ningún tamaño salva un diseño sesgado. El dimensionamiento supone muestreo sin sesgo.

No hay diseño perfecto: cada uno cambia coste por precisión por sesgo.

Diseño	Ventaja	Riesgo
Aleatorio simple	sin sesgo, teoría limpia	requiere marco completo
Estratificado	asegura grupos, menor varianza	exige conocer los estratos
Por conglomerados	barato y operativo	mayor error intra-grupo
Por conveniencia	rápido y económico	sesgo de selección alto

1. Una consulta en línea recibe un millón de respuestas de voluntarios auto-seleccionados. ¿Cuál es el principal problema?

- ☐ La muestra es demasiado pequeña
- ☐ Sesgo de selección: los voluntarios no representan a la población
- ☐ Falta varianza
- ☐ No se puede calcular la media

Ponderación correctiva

Cuando se conoce que ciertos grupos están sub o sobre representados, se pueden re-ponderar las respuestas para compensar. Es un parche útil pero imperfecto: corrige desequilibrios medibles, no el sesgo de los que nunca pudieron entrar.

Muestreo probabilístico frente a no probabilístico

Solo en el diseño probabilístico se conocen las probabilidades de selección, lo que permite cuantificar el error y justificar la inferencia. Los métodos no probabilísticos (cuotas, boule de nieve) sirven para explorar, pero sus márgenes de error no son defendibles.

Errores frecuentes

- Creer que una muestra grande elimina el sesgo de selección.
- Ignorar la tasa de no respuesta y asumir que contestan los que callaron.
- Usar un marco muestral que excluye sistemáticamente a parte de la población.

2. Aumentar el tamaño de una muestra sesgada reduce su sesgo.

☐ Verdadero

☐ Falso

Muestreo en datos digitales

En analítica web o de producto, el equivalente al sesgo de selección es analizar solo usuarios activos, o solo una plataforma. Los abandonos silenciosos (churn sin voto) son la no respuesta digital: ignorarlos sesga cualquier conclusión de negocio.

ACTIVIDAD INTERACTIVA DISPONIBLE EN
LUMBRE

3. Vocabulario de muestreo



4. Ordena el diseño de un muestreo riguroso:

Elementos:

- define a quién quieres conocer
- consigue la lista más fiel posible
- elige el método de selección
- evalúa y controla el sesgo restante

Poblacion	. . .
Marco	. . .
Diseno	. . .
Sesgo	. . .

Seeing Theory <https://seeingtheory.io/>Cómo la aleatoriedad del muestreo produce variabilidad en las estimaciones.

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=1v1ySMzTQzI>

5. Una empresa quiere conocer la satisfacción de TODOS sus clientes y encuesta solo a quienes escriben al soporte. Identifica el tipo de sesgo y propone un diseño que lo reduzca.

NIST Engineering Statistics Handbook <https://www.itl.nist.gov/div898/handbook/>
Referencia completa de métodos estadísticos.

Referencia curricular

Grado:	3
Materia:	Estadística Inferencial · Probabilidad y Estadística

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-esti-001-muestreo-y-error-de-seleccion>

Material educativo generado desde Lumbre · lumbre.lat