

L U M B R E



Trampas del análisis experimental

Grado: 8

Materia: Diseño Experimental y A/B · Formación Transversal

11/10/2026

Antes de empezar

Aleatorizar no garantiza una conclusión válida: qué puede romper un experimento online y cómo detectar y corregir cada trampa antes de decidir.

QUÉ APRENDERÁS

- Identificar y controlar las trampas del A/B: mirar antes de tiempo, comparación múltiple, SRM e interferencia, y sus correcciones.

Trampas del análisis experimental

VIDEO

Trampas del análisis experimental

Video explicativo (inglés): los errores comunes en tests A/B — peeking, poder estadístico insuficiente y múltiples comparaciones.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=dLwH1kp03kE>

Un experimento mal leído engaña

Aleatorizar te da el derecho a hablar de causalidad. Leer mal el experimento te da la licencia para decidir con números que mienten.

Los ensayos controlados online son el estándar para medir el impacto de un cambio, pero esconden trampas que producen falsos positivos o conclusiones inválidas. Reconocerlas es tan importante como saber diseñar el experimento.

Mirar antes de tiempo (peeking)

Parar el experimento en cuanto el valor p cruza el umbral, sin corregir, dispara la tasa de falsos positivos: con suficientes miradas, casi cualquier efecto se vuelve "significativo" por azar. La salida son reglas de parada o análisis secuencial.

Comparación múltiple

Si mides veinte métricas o segmentos, es probable que alguna salga significativa solo por el número de pruebas. Controles como **Bonferroni** o el **FDR** ajustan el umbral para que el hallazgo no sea un artefacto de mirar mucho.



Las trampas aparecen tanto al diseñar como al leer; los supuestos se rompen en silencio.

Sample Ratio Mismatch (SRM)

Si el porcentaje real de unidades por variante difiere del esperado (por ejemplo, asignabas 50/50 y sale 52/48), hay un bug: el experimento está roto antes de interpretar nada. SRM invalida las comparaciones y hay que corregir el sistema, no "leer" el resultado.

Interferencia y efectos de red

La aleatorización asume que tratar a una unidad no afecta a las demás. En marketplaces o redes, un cambio en el grupo de control puede tocar al grupo de tratamiento (SUTVA violada). La solución es aleatorizar por clusters (ciudades, grupos).

Cada trampa tiene un nombre, un diagnóstico y una corrección concreta.

Trampa	Qué rompe	Cómo evitarla
Peeking	el control del error tipo I	parada secuencial corregida

Trampa	Qué rompe	Cómo evitarla
Comparación múltiple	la tasa de falsos positivos	Bonferroni o FDR
SRM	la validez del experimento	alertar y corregir el reparto
Interferencia	la independencia entre unidades	aleatorizar por clusters

Validez externa y novedad

- Efecto de novedad: al principio todo cambio sube la interacción; mide más allá.
- Efecto deprimacidad: los usuarios nuevos reaccionan distinto.
- Lo que gana en un experimento puede no generalizar a todo el año.

Errores frecuentes

- Decidir en cuanto "sale verde" sin corrección de miradas.
- Ignorar una alerta de SRM por no detener el lanzamiento.
- Extrapolar un resultado de una semana al año completo.

Ejemplo resuelto: un "ganador" sospechoso

1. Verificar SRM antes de mirar métricas de negocio.
2. Revisar cuántas veces y cuándo se consultó el resultado.
3. Corregir por el número real de métricas evaluadas.
4. Confirmar estabilidad del efecto tras la novedad inicial.

¿Para qué sirve en la realidad?

La cultura data-driven vive de experimentos; su credibilidad depende de no celebrar hallazgos frágiles. Evitar estas trampas protege tanto el resultado actual como la confianza en el método.

1. ¿Qué invalida un experimento por sí mismo, antes de leer cualquier métrica?

- ☐ Un tamaño de muestra algo pequeño
- ☐ Un sample ratio mismatch: el reparto real no coincide con el esperado
- ☐ Medir varias métricas
- ☐ Durar dos semanas

2. Detener un experimento en cuanto el valor p cruza el umbral, sin corrección, mantiene la tasa de falsos positivos bajo control.

☐ Verdadero

☐ Falso

3. Cuando una intervención sobre una unidad afecta a otra y se viola la independencia, se incumple el supuesto _____.

4. Clasifica cada situación según qué trampa representa.

Elementos:

- revisar el panel cada hora y parar al cruzar
- el reparto real 52/48 cuando tocaba 50/50
- un cambio en control afecta al tratamiento
- celebrar al primer día de resultado
- usuarios del mismo grupo influyen entre sí

Peeking	. . .
SRM	. . .
Interferencia	. . .

5. Une cada trampa con su corrección.

5.	Peeking	_____
5.	Comparación múltiple	_____
5.	SRM	_____
5.	Interferencia	_____

Opciones: parada secuencial corregida · Bonferroni o FDR · detectar y arreglar el reparto · aleatorizar por clusters

ACTIVIDAD INTERACTIVA DISPONIBLE EN
LUMBRE

6. Trampas experimentales



Trampas del análisis experimental

Trampas del A/B

Leer

peeking y parada

comparacion multiple FDR

Validar

SRM reparto

Supuestos

SUTVA interferencia

clusters

novedad y validez externa

El mapa mental interactivo es interactivo en la versión online.



Exp Platform (Kohavi, Deng, Tang) <https://exp-platform.com/> Referencia de los líderes del área sobre experimentación controlada online.

Problema de comparaciones múltiples (Wikipedia) https://en.wikipedia.org/wiki/Multiple_comparisons_problem Fundamento del control de errores al evaluar muchas hipótesis.

7. Tu equipo lanza un A/B a las 24 h porque "salió significativo" y con reparto 53/47 entre variantes. Diagnostica los problemas metodológicos y explica en qué orden los revisarías antes de tomar cualquier decisión.

Referencia curricular

Grado:	8
Materia:	Diseño Experimental y A/B · Formación Transversal

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



<https://univ.lumbre.lat/object/cdia-trans-dexp-004-trampas-del-analisis-experimental>

Material educativo generado desde Lumbre · lumbre.lat