

Casos Históricos para Reflexión Estadística

Prof. José Neville Díaz Caraballo
Universidad de Puerto Rico en Aguadilla
Departamento de Matemáticas

Caso 1: El Sesgo del Superviviente — Aviones de la WWII

El Problema

Durante la Segunda Guerra Mundial, la Marina de los Estados Unidos quería reducir las bajas de aviones. Analizaron los aviones que regresaban de combate y mapearon los impactos de bala.

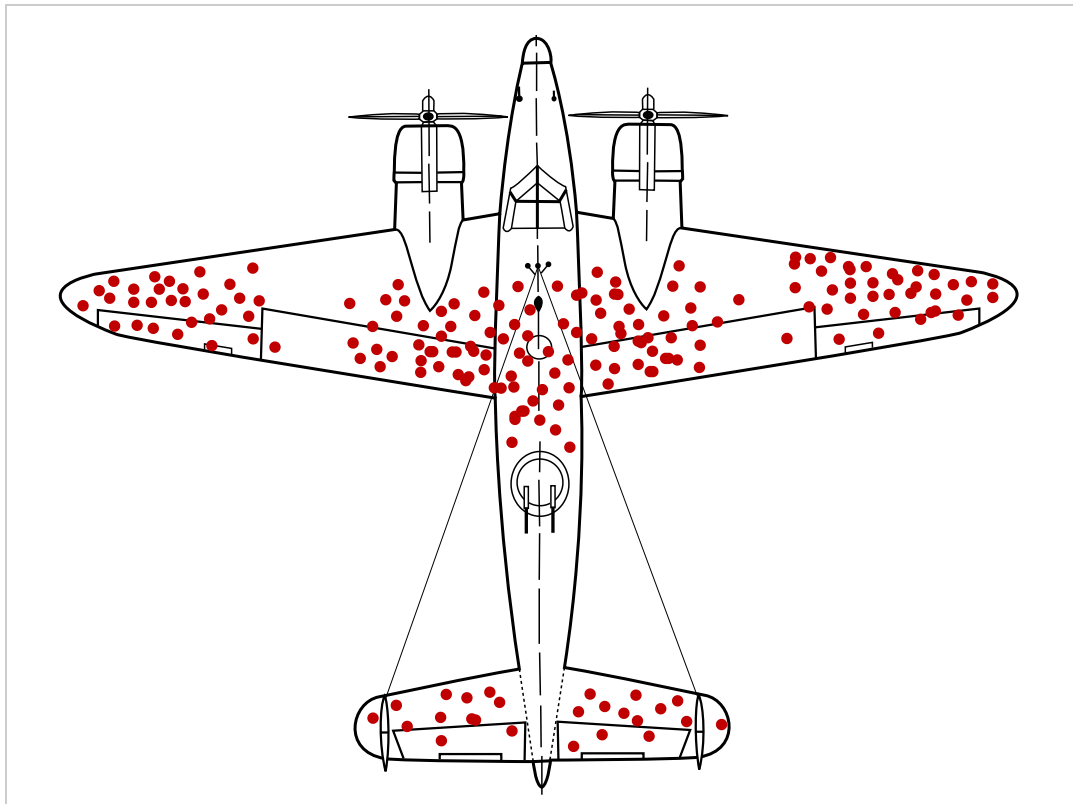


Diagrama que muestra la distribución de impactos en aviones que regresaron de combate.

Fuente: Wikimedia Commons

El Error Intuitivo

Los oficiales de la Marina querían reforzar las áreas **con más impactos** (alas, fuselaje central, cola).

La Genialidad de Abraham Wald

El estadístico Abraham Wald, del Statistical Research Group de Columbia, argumentó exactamente lo contrario:

*"Los agujeros representan áreas donde un avión **puede** recibir daño y **aún así regresar**. Las áreas **sin** agujeros (motores, cabina, líneas de combustible) son donde los aviones fueron impactados y **no regresaron**."*

Interpretación Visual

Área del Avión	Impactos Observados	Interpretación de Wald
Alas	Muchos	Puede recibir daño y sobrevivir
Fuselaje	Muchos	Puede recibir daño y sobrevivir
Cola	Muchos	Puede recibir daño y sobrevivir
Motores	Pocos/Ninguno	ZONA FATAL - No regresan
Cabina	Pocos/Ninguno	ZONA FATAL - No regresan
Tanques de combustible	Pocos/Ninguno	ZONA FATAL - No regresan

Lección Estadística

Solo estamos viendo los datos de los que sobrevivieron. Los aviones derribados no están en nuestra muestra. Este es el sesgo del superviviente: sacar conclusiones basadas únicamente en los casos que "sobrevivieron" al proceso de selección.

Aplicaciones Modernas del Sesgo del Superviviente

Contexto	Lo que vemos	Lo que ignoramos
Empresarios exitosos	"Abandoné la universidad y triunfé"	Miles que abandonaron y fracasaron
Edificios antiguos	"Antes construían mejor"	Edificios viejos que colapsaron

Música clásica	"La música de antes era mejor"	Miles de canciones olvidadas
Fondos de inversión	"Este fondo tiene 20 años de éxito"	Fondos que cerraron por pérdidas

Referencia: Wald, A. (1943). A Method of Estimating Plane Vulnerability Based on Damage of Survivors. Statistical Research Group, Columbia University.

Caso 2: John Snow y el Cólera en Londres (1854)

El Problema

Se creía que el cólera se transmitía por "miasmas" (aire malo). Esta teoría dominante impedía encontrar la verdadera causa de las epidemias.

El Método

El Dr. John Snow mapeó los casos de cólera en el distrito de Soho, Londres. Descubrió que los casos se agrupaban geográficamente alrededor de una bomba de agua pública ubicada en Broad Street.

El Mapa Original de John Snow (1854)



Mapa original de John Snow mostrando los casos de cólera en Soho, Londres.
Las barras negras representan muertes; la bomba de Broad Street está en el centro.

Fuente: Wikimedia Commons

Elementos Clave del Mapa

Elemento	Significado
Barras negras	Muertes por cólera (cada barra = 1 muerte)
Ubicación central	Bomba de agua en Broad Street
Patrón radial	Las muertes se concentran alrededor de la bomba

Evidencia Adicional

Snow notó que una cervecería cercana **no tuvo casos** de cólera. La explicación: los trabajadores bebían cerveza, no agua de la bomba contaminada.

El Resultado

Snow convenció a las autoridades locales de remover la manija de la bomba. La epidemia terminó poco después. Este evento marca el nacimiento de la **epidemiología moderna**.

Lección Estadística

La visualización geográfica de datos reveló una verdad que la teoría dominante ocultaba.
Un simple mapa con puntos de datos fue más poderoso que décadas de teoría médica incorrecta.

Referencia: UCLA School of Public Health - John Snow Site (<https://www.ph.ucla.edu/epi/snow.html>)

Caso 3: La Paradoja de Simpson — El Caso de Berkeley (1973)

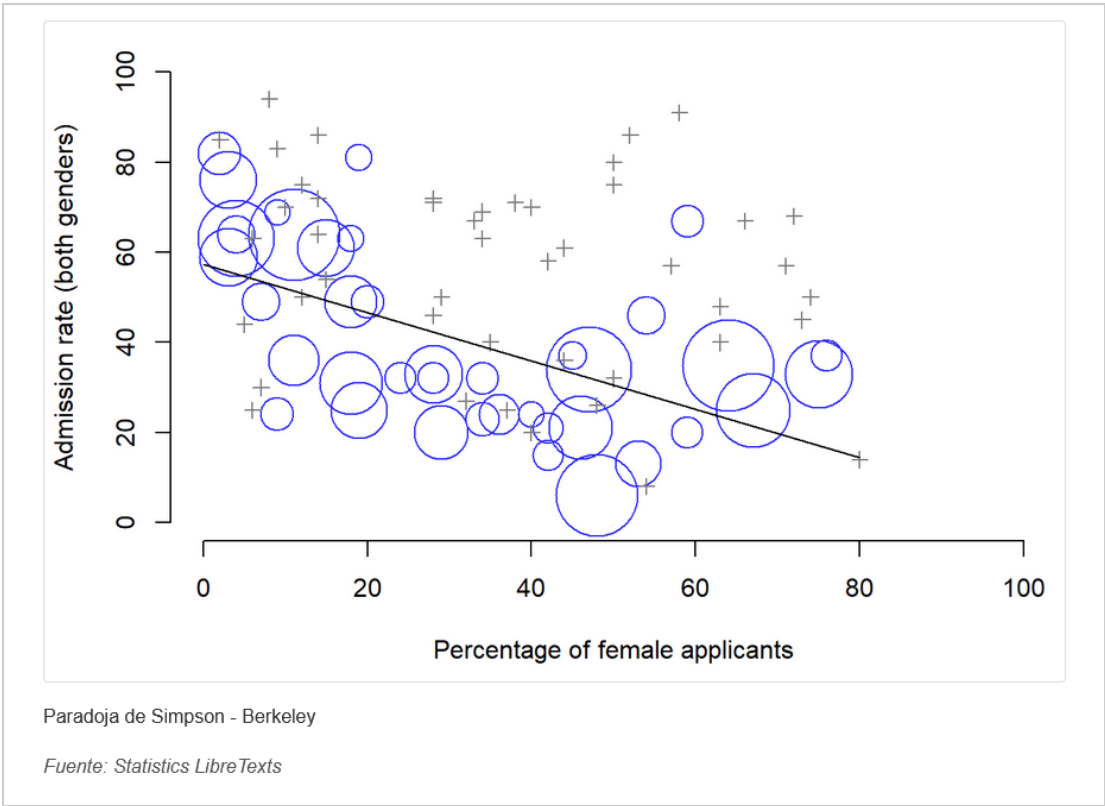
El Problema

La Universidad de California en Berkeley fue demandada por discriminación contra mujeres en admisiones de posgrado. Los datos agregados parecían contundentes:

Género	Solicitantes	Tasa de Admisión
Hombres	8,442	44%
Mujeres	4,321	35%

Una diferencia de 9 puntos porcentuales parecía evidencia clara de discriminación.

Visualización de la Paradoja



Paradoja de Simpson aplicada a Berkeley: Eje X = Porcentaje de solicitantes mujeres por departamento. Eje Y = Tasa de admisión (ambos géneros). Los círculos azules representan departamentos (tamaño = volumen).

Las cruces (+) son datos individuales. La línea de regresión muestra correlación negativa.

Fuente: Statistics LibreTexts

Interpretación de la Gráfica

La gráfica revela el mecanismo de la paradoja:

- **Eje X:** Porcentaje de solicitantes que son mujeres en cada departamento
- **Eje Y:** Tasa de admisión del departamento (para ambos géneros)
- **Círculos azules:** Cada círculo es un departamento; el tamaño indica el volumen de solicitantes
- **Línea de regresión:** Pendiente negativa — los departamentos con más mujeres solicitantes tienen tasas de admisión más bajas

Conclusión visual: Las mujeres no fueron discriminadas; simplemente aplicaron más a departamentos difíciles (parte derecha de la gráfica, con tasas bajas).

El Análisis Estratificado

Cuando los estadísticos Peter Bickel, Eugene Hammel y J. William O'Connell analizaron los datos **por departamento**, encontraron algo sorprendente:

Departamento	Hombres (Admitidos)	Mujeres (Admitidas)	¿Quién tuvo mejor tasa?
A	62%	82%	Mujeres
B	63%	68%	Mujeres
C	37%	34%	Hombres
D	33%	35%	Mujeres
E	28%	24%	Hombres
F	6%	7%	Mujeres

La Realidad

Dentro de cada departamento, las mujeres tenían tasas de admisión **iguales o superiores** a los hombres en 4 de 6 departamentos.

¿Cómo es Posible?

El sesgo aparente ocurría porque:

1. **Las mujeres aplicaban proporcionalmente más** a departamentos altamente competitivos (Departamentos C, D, E, F) con tasas de admisión bajas para TODOS.
2. **Los hombres aplicaban más** a departamentos con tasas de admisión altas (Departamentos A, B).
3. **Resultado agregado engañoso:** Al combinar los datos, parecía que las mujeres eran discriminadas, cuando en realidad el patrón de aplicación era el factor determinante.

Lección Estadística

Los datos agregados pueden mentir. El análisis estratificado es esencial antes de sacar conclusiones. Este fenómeno, conocido como la Paradoja de Simpson, demuestra que una tendencia que aparece en grupos de datos combinados puede **revertirse completamente** cuando los grupos se analizan por separado.

Referencia: Bickel, P.J., Hammel, E.A., & O'Connell, J.W. (1975). Sex Bias in Graduate Admissions: Data from Berkeley. *Science*, 187(4175), 398-404.

Resumen: Tres Lecciones Fundamentales

Caso	Año	Lección Principal
Sesgo del Superviviente	1943	Solo vemos a los que sobrevivieron; los datos faltantes son críticos
John Snow	1854	La visualización geográfica revela patrones ocultos
Paradoja de Simpson	1973	Los datos agregados pueden invertir la realidad

Moraleja

La estadística no es solo números — es pensamiento crítico con datos.

*Material suplementario para el Capítulo I: ¿Qué es Estadística?
Universidad de Puerto Rico en Aguadilla
Prof. José Neville Díaz Caraballo*