

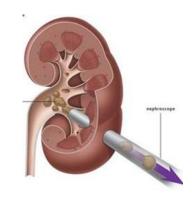
# Falacias con los datos



En un estudio se compararon los porcentajes de éxito de la extracción de cálculos renales según el método utilizado.



Cirugía	Nefrolitotomía
Abierta	percutánea
78%	83%





El estudio concluía que lo mejor era la nefrolitotomía percutánea.

Sin embargo, partiendo el dataset de otra forma lo que se puede ver es distinto.



	Cirugía Abierta	Nefrolitotomía percutánea
Piedra <2cm	93%	83%
Piedra >=2cm	73%	69%

Lo que sucede es que los datos que se tienen son de tratamientos reales, en donde a más grande la piedra más chances hay de ir a cirugía.

https://www.bmj.com/content/309/6967/1480.full



	Cirugía Abierta	Nefrolitotomía percutánea
Piedra <2cm	93%	83%
Piedra >=2cm	73%	69%

No es que la cirugía sea peor, sino que se le suelen asignar los casos más complejos.

## Paradoja de simpson: Solución?



Podemos agarrar a la mitad de los pacientes y arbitrariamente elegirlos al azar un tratamiento.

En medicina esto sería bastante polémico, pero sirve para muchos otros casos en donde hacer este tipo de pruebas no mata a nadie. Esto se llama **validación cruzada.** 





Somos wikipedia, tenemos un cartelito que dice donar.





Somos wikipedia, tenemos un cartelito que dice donar.

A alguien se le ocurre hacerlo rojo.





Somos wikipedia, tenemos un cartelito que dice donar.

A alguien se le ocurre hacerlo rojo.

Las donaciones aumentan 5%.





¿Cómo sabemos si fue gracias al cambio del botón o por otra razón?

### A/B testing



#### Usuarios "A", 50% del tráfico



#### Usuarios "B", 50% del tráfico



### A/B testing



#### Usuarios "A", 50% del tráfico

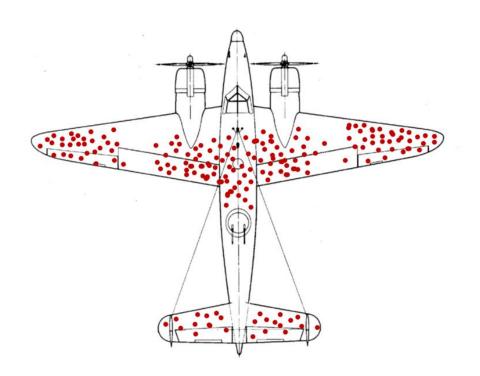


#### Usuarios "B", 50% del tráfico



## Sesgo de supervivencia



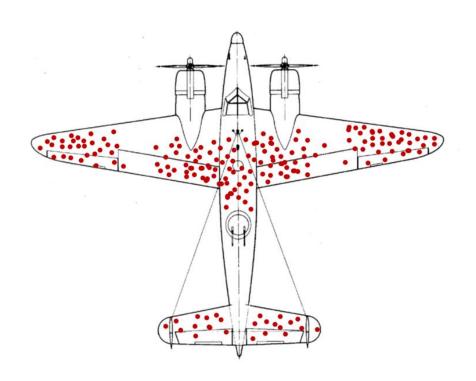


Estados Unidos de cara a la segunda guerra mundial analiza los aviones que vuelven del combate.

Concluye que debe reforzarlos en donde más se dañaron.

## Sesgo de supervivencia



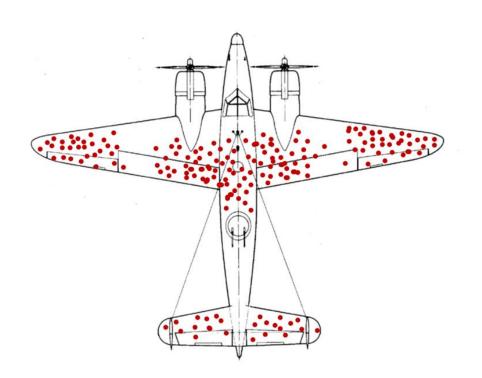


Resulta que estaban dañados en todos los lugares que no son críticos, porque pudieron volver.

Alguien nota esto y indica que en realidad se debe reforzar los lugares en donde no vemos daños de los que volvieron.

## Sesgo de supervivencia: Fix





Siempre hay que preguntarnos, ¿cuál es el origen de nuestros datos?