

I miti sui dati da sfatare

(Fonte: <https://twitter.com/KirkDBorne/status/1280708602236461061>)

Cherry Picking: Selezionare solamente i risultati che si adattano alla tua affermazione ed escludendo quelli che non lo sono.

Data Dredging: Testare ripetutamente nuove ipotesi rispetto allo stesso insieme di dati, senza riconoscere che la maggior parte delle correlazioni sarà il risultato del caso

Survivorship Bias: Trarre conclusioni da un insieme incompleto di dati perché i dati sono conformi ad alcuni criteri di selezione.

Cobra Effect: Fissare un incentivo che produce per sbaglio il risultato opposto a quello previsto. Conosciuto anche come “incentivo perverso”.

False Causality: Presupporre erroneamente quando due eventi sembrano correlati che uno deve aver causato l'altro.

Gerrymandering: Manipolazione dei confini geografici utilizzati per raggruppare i dati al fine di modificare il risultato.

Sampling Bias: Trarre conclusioni da un insieme di dati che non sono rappresentativi della popolazione che stai cercando di capire.

Gambler's Fallacy: Credere erroneamente che, poiché qualcosa è accaduto più frequentemente del solito, ora è meno probabile che accada in futuro (e viceversa).

Hawthorne Effect: L'atto di monitorare qualcuno può influenzare il suo comportamento, portando a risultati spuri. Conosciuto anche come “effetto osservatore”.

Regression Fallacy: Quando accade qualcosa di insolitamente buono o cattivo, nel tempo tornerà alla normalità.

Simpson's Paradox: Quando un trend appare in diversi sottoinsiemi di dati ma scompare o si inverte quando i gruppi vengono mischiati.

Mcnamara Fallacy: Affidarsi esclusivamente alle metriche in situazioni complesse e perdere di vista il quadro più ampio.

Overfitting: Creare un modello che sia eccessivamente su misura per i dati che hai e non rappresentativo della tendenza generale.

Publication Bias: È più probabile che i risultati di ricerche interessanti vengano pubblicati, distorcendo la nostra impressione della realtà.

Danger of Summary Metrics: Guardare soltanto le metriche di riepilogo e non tenere in considerazione le differenze nei dati grezzi.