

Data dredging (aka. p-hacking)

Description

"A well-wrapped statistic is better than Hitler's "big lie" it misleads, yet it cannot be pinned on you."

~ Darrell Huff, How to Lie with Statistics

Data dredging (also known as **data snooping** or **p-hacking**)[1][a] is the misuse of [data analysis](#) to find patterns in data that can be presented as [statistically significant](#), thus dramatically increasing and understating the risk of false positives. This is done by performing many [statistical tests](#) on the data and only reporting those that come back with significant results.[2]

The process of data dredging involves testing multiple hypotheses using a single [data set](#) by [exhaustively searching](#)—perhaps for combinations of variables that might show a [correlation](#), and perhaps for groups of cases or observations that show differences in their mean or in their breakdown by some other variable.

Category

1. General

Date Created

6. December 2022

Author

web45