

Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#112 Observations and Biases

été 2020

Biases

"On two occasions I have been asked, 'Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?'" wrote computing pioneer Charles Babbage in 1864. "I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question." in **Bias Is AI's Achilles Heel**.
Here's How To Fix It

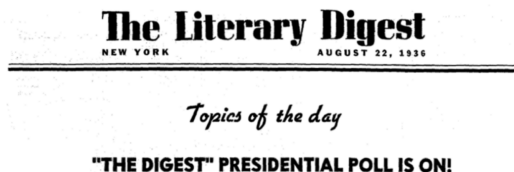


source : **Dilbert**

Sampling Bias

Sampling bias is a bias in which a sample is collected in such a way that some members of the intended population have a lower or higher sampling probability than others

see [Two Lessons of Sampling Bias from 1936 US Elections](#), Alf Landon vs. Franklin Delano Roosevelt

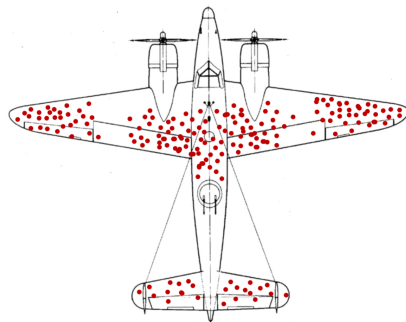


See [A Brief Introduction to Selection Bias](#) for more examples (and also [Berkson's paradox](#)). See also credit scoring...

Survival Bias

Study of damage done to aircraft that had returned from missions (World War II)

Recommendation that armor should be added to the areas that showed the most damage. Wald noted that the study only considered the aircraft that had survived their missions. Wald proposed that the Navy reinforce areas where the returning aircraft were unscathed since those were the areas that, if hit, would cause the plane to be lost. **Abraham Wald's Work on Aircraft Survivability**



see also **Do cats always land unharmed on their feet, no matter how far they fall?**, or Titanic dataset **#351**

Confirmation Bias

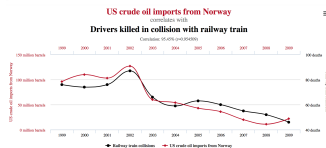
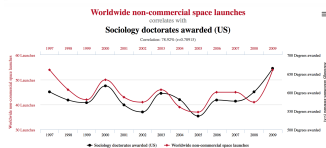
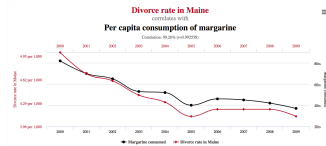
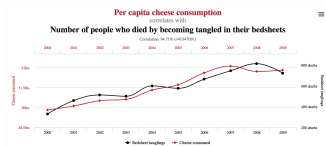
Can be seen as a cognitive bias [wikipedia](#) or [Are things getting worse – or does it just feel that way?](#), or a statistical one...

Many academic or research scientists run thousands of computer simulations where all fail to confirm or verify the hypothesis. Then they tweak the data, assumptions or models until confirmatory evidence appears to confirm the hypothesis. They proceed to publish the one successful result without mentioning the failures! (via [The Deadly Data Science Sin of Confirmation Bias](#))

see also [p-hacking \(#241\)](#) and [Why Most Published Research Findings Are False](#)

(Spurious) Correlation Bias ("Coincidence? I guess not")

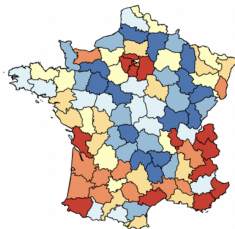
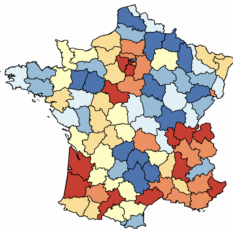
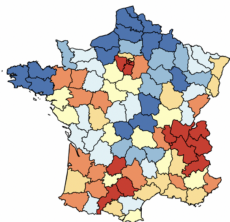
See <https://tylervigen.com/>



Esperance de vie

Licences de Tennis

Agences de publicité



Measurement, Stereotype (Endogeneity) & Omitted-Variable Bias

Measurement bias: poorly measuring the outcome you are measuring see [Measurement error and bias](#)

- ▶ difficult to assess some variables (e.g. self reported)
- ▶ difficult to assess for evasive questions in surveys ([#201](#))

Stereotype bias: Artificial intelligence is supposed to make life easier for us all – but it is also prone to amplify sexist and racist biases from the real world

Omitted-Variable bias: see [#121](#) (Simpson's paradox)