In Data Science…

- ...Uncertainty is a fact of life
- Many sources...
  - — Measurement error
  - — Missing data
  - — Model selection
  - — Sample selection

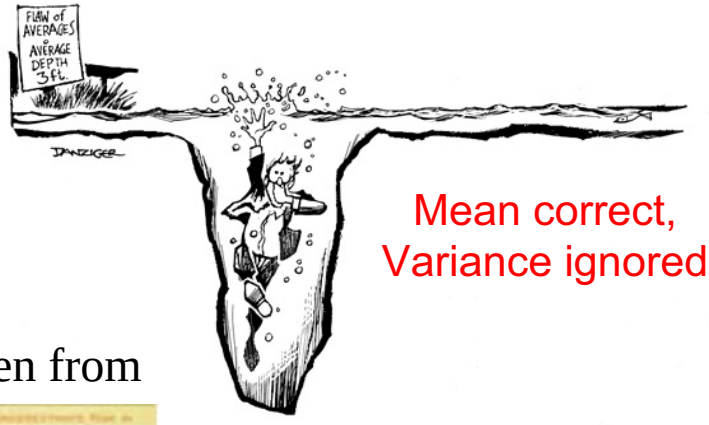
But It's Usually Just Ignored

- Data and models assumed to be precise
- Sometimes with dramatic consequences

But It's Usually Just Ignored

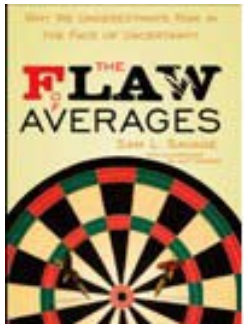
- Data and models assumed to be precise
- Sometimes with dramatic consequences

Flaw of averages (weak form):



Mean correct,
Variance ignored

taken from



Flaw of averages (strong form):

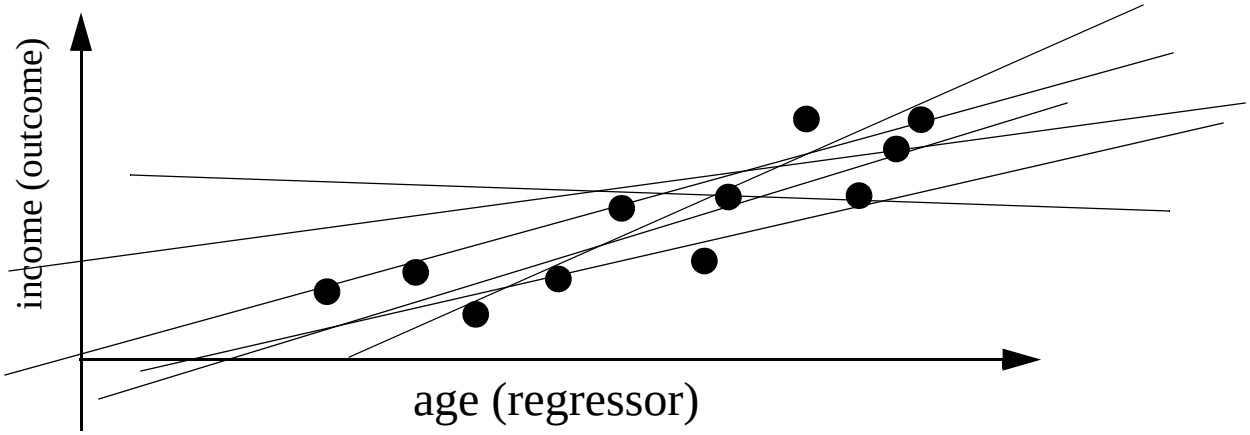
Wrong value of mean:
$f(E[X]) \neq E[f(X)]$

Solution: Embrace the Bayesian Approach

- Then regression (for example) is not just fitting a line...

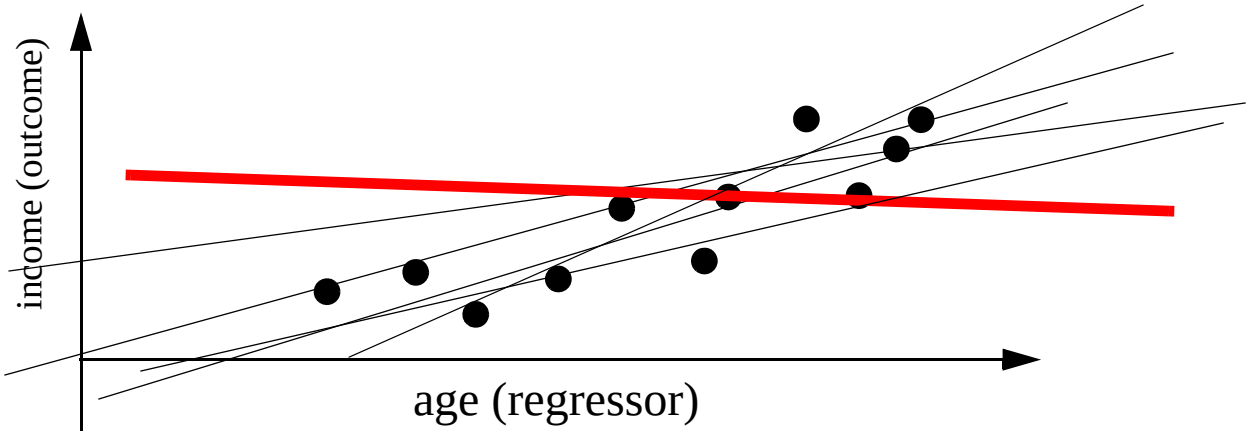
Solution: Embrace the Bayesian Approach

- It's fitting a **family** of lines...



Solution: Embrace the Bayesian Approach

- It's fitting a **family** of lines...



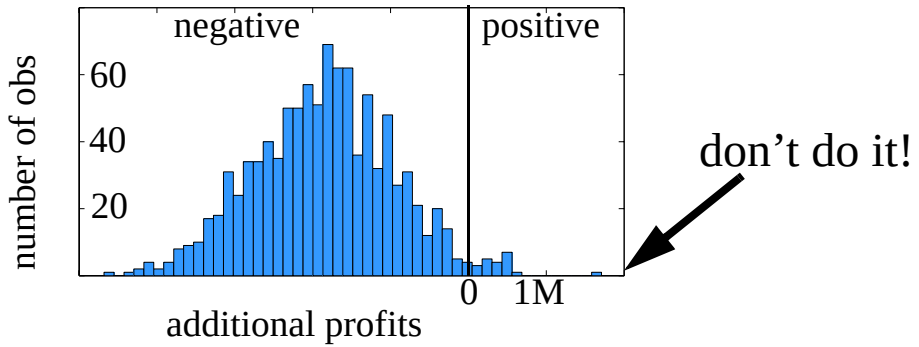
And understanding there's a chance we might have it all wrong

Data Processing Systems Should Support This

- By (for example) natively supporting Monte Carlo
- And natively supporting data and model uncertainty

Data Processing Systems Should Support This

- By (for example) natively supporting Monte Carlo
- And natively supporting data and model uncertainty
- If I ask: what will my profits be if I raise margins by 5%?
  - — You compute this once, you get one answer
  - — You do this again, you get another answer
  - — How to handle this?
  - — Redo the computation many times (Monte Carlo) to obtain a *distribution* of results



Thank You!