Statistics & the Art of Deception

Déborah Berebichez, Ph.D.
Chief Data Scientist, Metis
"All models are wrong, but some are useful"

George Box (statistician)
SAMPLE SIZE

50% of Data Scientists Have Red Hair

* Survey Size: 2 Data Scientists

* Graphics adapted from VerticalMeasures.com
"The average of the results obtained from a large number of trials should be close to the expected (‘true’) value, and will tend to become closer as more trials are performed."

The larger the sample, the better the estimate
Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith
MMR vaccine controversy

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit
Misleading Election Results

“You know what I’m really happy about, because we’ve saying it for a long time? Forty-six percent with the Hispanics! Forty-six percent! Number one with Hispanics!”

—Donald Trump, Feb. 23, 2016 after winning Republican caucuses in Nevada

- Entrance poll of 1,573 Republican voters (margin of error ±2.5%)
- 126 Hispanic Republican voters (8%)
- 46% of Hispanic Republicans voted for Trump (margin of error ±2.5%)

Trump got 58 votes

National polling results for 300 Hispanic voters (margin of error is ±6%):
Hillary Clinton: 65% , Donald Trump: 17%, Gary Johnson: 9% and Jill Stein: 2%
SAMPLE SIZE REVISITED

- TOO BIG: Unwieldy, Expensive
- JUST RIGHT: Sample not representative (bias), Effect due to chance
- TOO SMALL: Effect due to chance
SAMPLE SIZE REVISITED

What’s the appropriate sample size?

SIZE of sampled population

VARIABILITY of your data (STD)
SAMPLE SIZE REVISITED

MARGIN OF ERROR you can accept in your estimate

How CONFIDENT you want to be that the true value falls within estimate
SIZE ISN’T EVERYTHING
Focus for success?

Study of 1,854 companies:
78% of all high-performance firms focused on one set of core activities

Problem: only included firms that survived study period
Excluded companies that focused and failed

When failed companies are included, reach opposite conclusion
HYPOTHESIS TESTING

Can the effect observed be explained by chance?

When chance is an unlikely explanation, we say the effect is “statistically significant”

Note: statistically significant does NOT necessarily mean that the effect is important or useful
1. $H_0$: Astrology = Coin Toss
2. $H_1$: Astrology > Coin Toss

Current Science, Vol. 96, No. 5, 10 March 2009, pp 641-643
HYPOTHESIS TESTING

• Coin Toss: on average match 20 of the 40 charts with group type A or B

• Statistically significant effect: at least 28 matches out of 40. Either fluke or not by chance
RESULTS

• Best astrologer: 24 matches out of 40

• Average of all astrologers: 17.25 matches out of 40

• “Many astrologers looked upon the success they had achieved...as a testimonial to their predictive ability.”
A/B TESTING

Has there been enough traffic (enough test subjects)?
Have there been enough conversions (enough test goals)?
Is the data skewed in any way?
Google: $200M increase in ad revenue
Microsoft: $80M increase in ad revenue
MULTIPLE COMPARISON PROBLEM

The more variations you test, the more likely you will see an effect caused just by chance

Null hypothesis:
purchase is independent of jelly bean color

Significance level: 5%

In \( m \) tests, the probability that at least one of the tests show an effect due to chance is

\[
\begin{align*}
  m &= 1, p_m = 0.05 \\
  m &= 20, p_m = 0.64 \\
  m &= 41, p_m = 0.88
\end{align*}
\]
Y dominates N...positive effect!
SPURIOUS CORRELATIONS

Divorce rate in Maine correlates with Per capita consumption of margarine

Correlation: 99.26% (r=0.992558)

Data sources: National Vital Statistics Reports and U.S. Department of Agriculture
The top tax rate will soon skyrocket to 39.6%.

4X or 4% Increase?
• Is the sample large enough and random?
• Is the correlation big enough to mean anything?
• Correlations without a measure of reliability (standard error) are not to be taken seriously.
• Does it make sense?
PARTING WORDS

“The first principle is that you must not fool yourself — and you are the easiest person to fool.”

Richard Feynman
Thank You!

Déborah Berebichez, Ph.D.

Chief Data Scientist, Metis