CSE 312

Foundations of Computing II

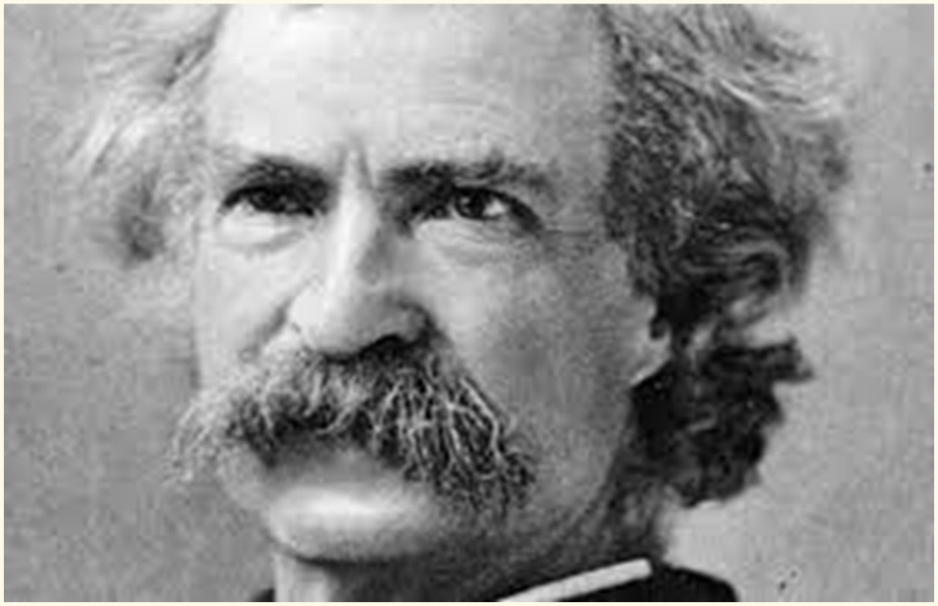
Lecture 28: How to Lie/Be Misled/Detect Lies with Statistics

Slide Credit: Based on Stefano Tessaro's slides for 312 19au incorporating ideas from Anna Karlin, Alex Tsun, & Maya Bar-Hillel

Random Quote

"There are three kinds of lies: lies, damned lies, and statistics."

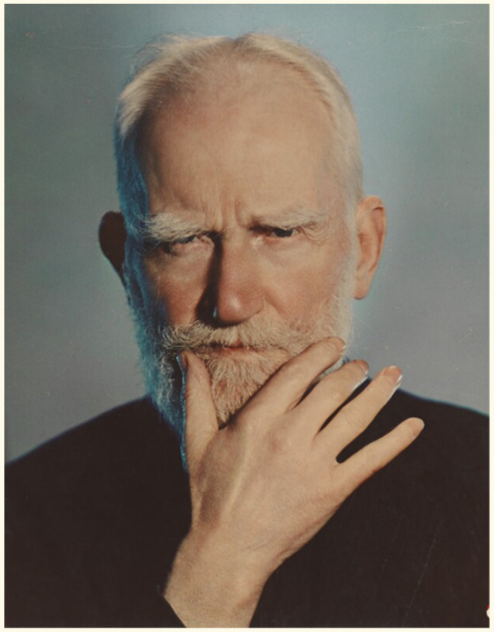
- Mark Twain



Random Quote

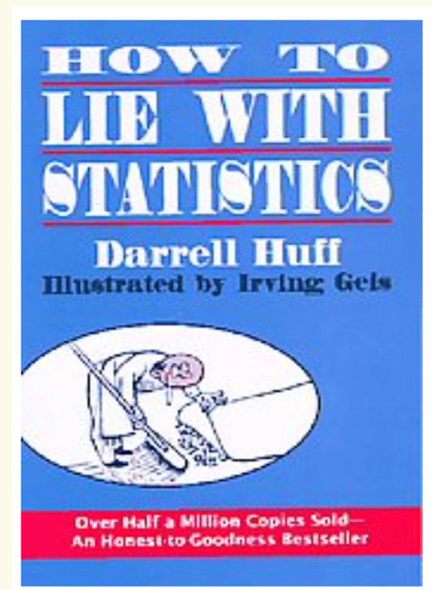
"It is the mark of a truly intelligent person to be moved by statistics"

- George Bernard Shaw



The Book

- Published in 1954, over 500,000 copies sold
- “A great introduction to the use of statistics, and a great refresher for anyone who’s already well versed in it” - Bill Gates.



The Book

- Published in 1954, over 500,000 copies sold
- “A great introduction to the use of statistics, and a great refresher for anyone who’s already well versed in it” - Bill Gates.
- Doesn’t teach how to lie with statistics, but how we are/can be lied to using statistics

The Book

- Published in 1954, over 500,000 copies sold
- “A great introduction to the use of statistics, and a great refresher for anyone who’s already well versed in it” - Bill Gates.
- Doesn’t teach how to lie with statistics, but how we are/can be lied to using statistics
- In the current age, we are lied to all the time, e.g., by **politicians**, and **marketers**.
    - Often make decisions based on these lies: “4 out of 5 dentists recommend….”

To be clear…

- Many lies are unintentional
- People passing on misinformation/bad information that they don’t even know is bad.
- People using bad data to make inferences
- People not understanding statistics well enough

Random Quote

"Statistical Thinking will one day be as necessary for efficient citizenship as the ability to read and write"

- H.G. Wells

Statistical Inference

- Making an estimate or prediction about a **population** based on a **sample**.

Statistical Inference

- Making an estimate or prediction about a **population** based on a **sample**.
  - Often very expensive/impossible to survey an entire population (all students at UW, all residents in the U.S)

Statistical Inference

- Making an estimate or prediction about a **population** based on a **sample**.
  - Often very expensive/impossible to survey an entire population (all students at UW, all residents in the U.S)
  - Need to use a **random unbiased** *sample* of *population* to draw conclusions (with some chance/margin of error)

1. Sampling Gone Wrong (Bias)

"The Literary Digest" Magazine wanted to predict 1936 election:

- Alfred Landon vs Franklin D Roosevelt
- Sent 10 million surveys and received 2.4 million responses
- From a "List" containing: their subscribers, owners of cars and telephones

Electoral Votes	**Prediction**	**Actual**
Landon		
Roosevelt		

1. Sampling Gone Wrong (Bias)

"The Literary Digest" Magazine wanted to predict 1936 election:

- Alfred Landon vs Franklin D Roosevelt
- Sent 10 million surveys and received 2.4 million responses
- From a "List" containing: their subscribers, owners of cars and telephones

Electoral Votes	**Prediction**	**Actual**
Landon	370	
Roosevelt	161	

1. Sampling Gone Wrong (Bias)

"The Literary Digest" Magazine wanted to predict 1936 election:

- Alfred Landon vs Franklin D Roosevelt
- Sent 10 million surveys and received 2.4 million responses
- From a "List" containing: their subscribers, owners of cars and telephones

Electoral Votes	**Prediction**	**Actual**
Landon	370	8
Roosevelt	161	523

1. Sampling Gone Wrong (Bias)

"The Literary Digest" Magazine wanted to predict 1936 election:

- Alfred Landon vs Franklin D Roosevelt
- Sent 10 million surveys and received 2.4 million responses
- From a "List" containing: their subscribers, owners of cars and telephones

Electoral Votes	**Prediction**	**Actual**
Landon	370	8
Roosevelt	161	523

1. Sampling Gone Wrong (Bias)

Let $x_1, x_2, \ldots, x_n$ be **iid** samples…

1. Sampling Gone Wrong (Bias)

Let $x_1, x_2, \ldots, x_n$ be **iid** samples…

- Not Representative
  - Voluntary Response Bias
    - Only 24% of respondents answered the poll.

1. Sampling Gone Wrong (Bias)

Let $x_1, x_2, \ldots, x_n$ be **iid** samples…

- Not Representative
  - Voluntary Response Bias
    - Only 24% of respondents answered the poll.
  - Not the Right Population
    - Was biased toward people with more money, education, information, alertness than average American

1. Sampling Gone Wrong (Bias)

Let $x_1, x_2, \ldots, x_n$ be **<u>iid</u>** samples…

- Not Representative
  - Voluntary Response Bias
    - Only 24% of respondents answered the poll.
  - Not the Right Population
    - Was biased toward people with more money, education, information, alertness than average American
- Not Random
  - Convenience Sampling
    - Only to people whose contact information they have.
    - Like standing outside a church and asking “Do you believe in God?”, using those samples to represent the US population.

1. Sampling Gone Wrong (Bias)

Let $x_1, x_2, \ldots, x_n$ be **<u>iid</u>** samples…

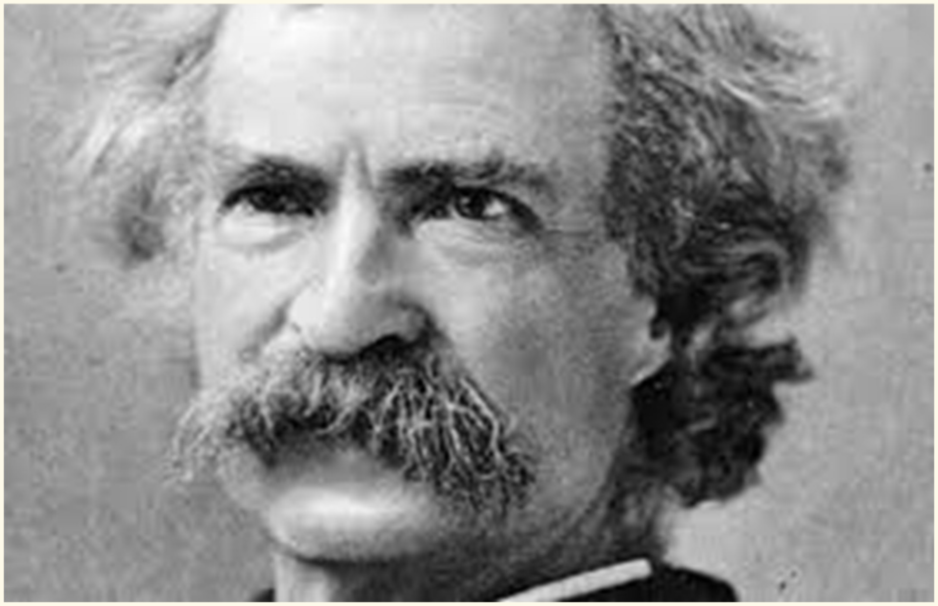
- Not Representative
  - Voluntary Response Bias
    - Only 24% of respondents answered the poll.
  - Not the Right Population
    - Was biased toward people with more money, education, information, alertness than average American
- Not Random
  - Convenience Sampling
    - Only to people whose contact information they have.
    - Like standing outside a church and asking "Do you believe in God?", using those samples to represent the US population.

<u>More samples is NOT a solution for bad sampling technique...</u>

Random Quote

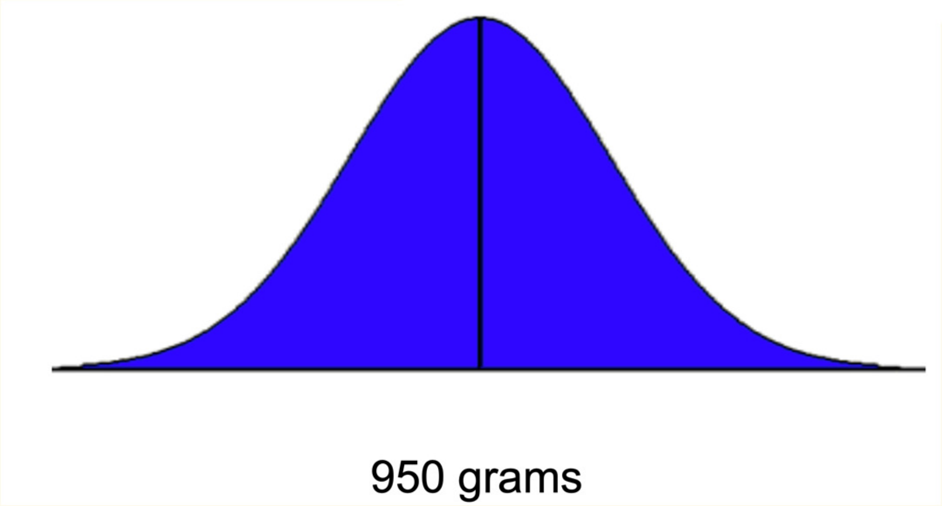
"Facts are stubborn, but statistics are more pliable."

- Mark Twain



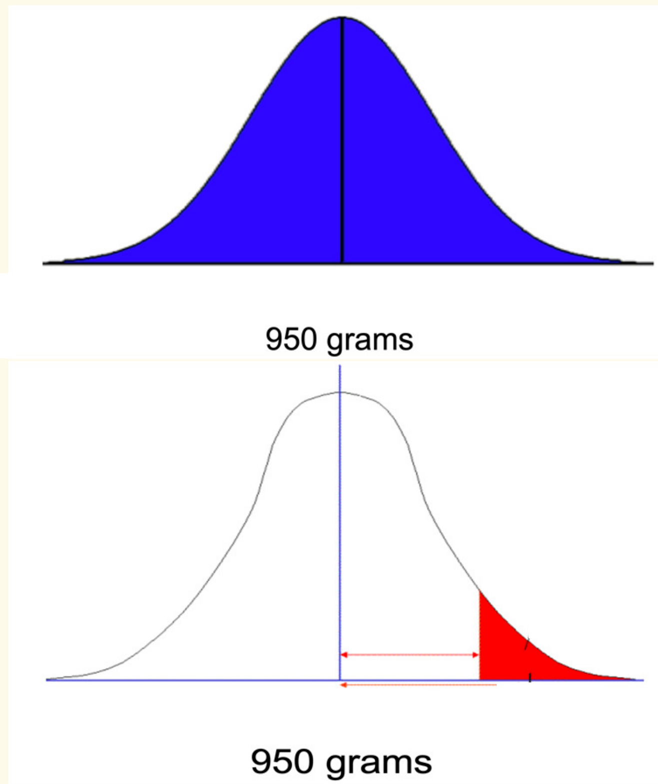
2. Detecting lies with statistics

A story about the famous French mathematician Henri Poincare



2. Detecting lies with statistics

A story about the famous French mathematician Henri Poincare



To fake a distribution…

You’d better know what it looks like….

People that are untrained in statistics often don’t.

For example, people are really bad at faking a sequence of fair coin tosses.

Random Quote

"It's easy to lie with statistics. It's hard to tell the truth without statistics."

- Andrejs Dunkels

First digit phenomenon

Suppose that I pick a random integer in the range 1..999

What's the chance that the first digit of the number I pick is a 1?

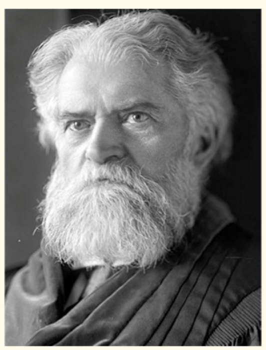
a). About 1/9
b). About 11%
c) 30%
d) I don't know.

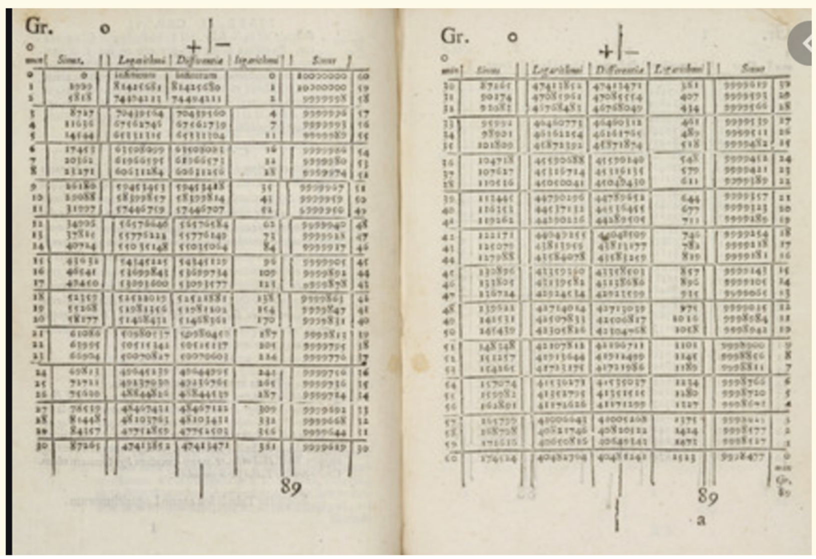
About 1/9, which is about 11%

Benford's Law

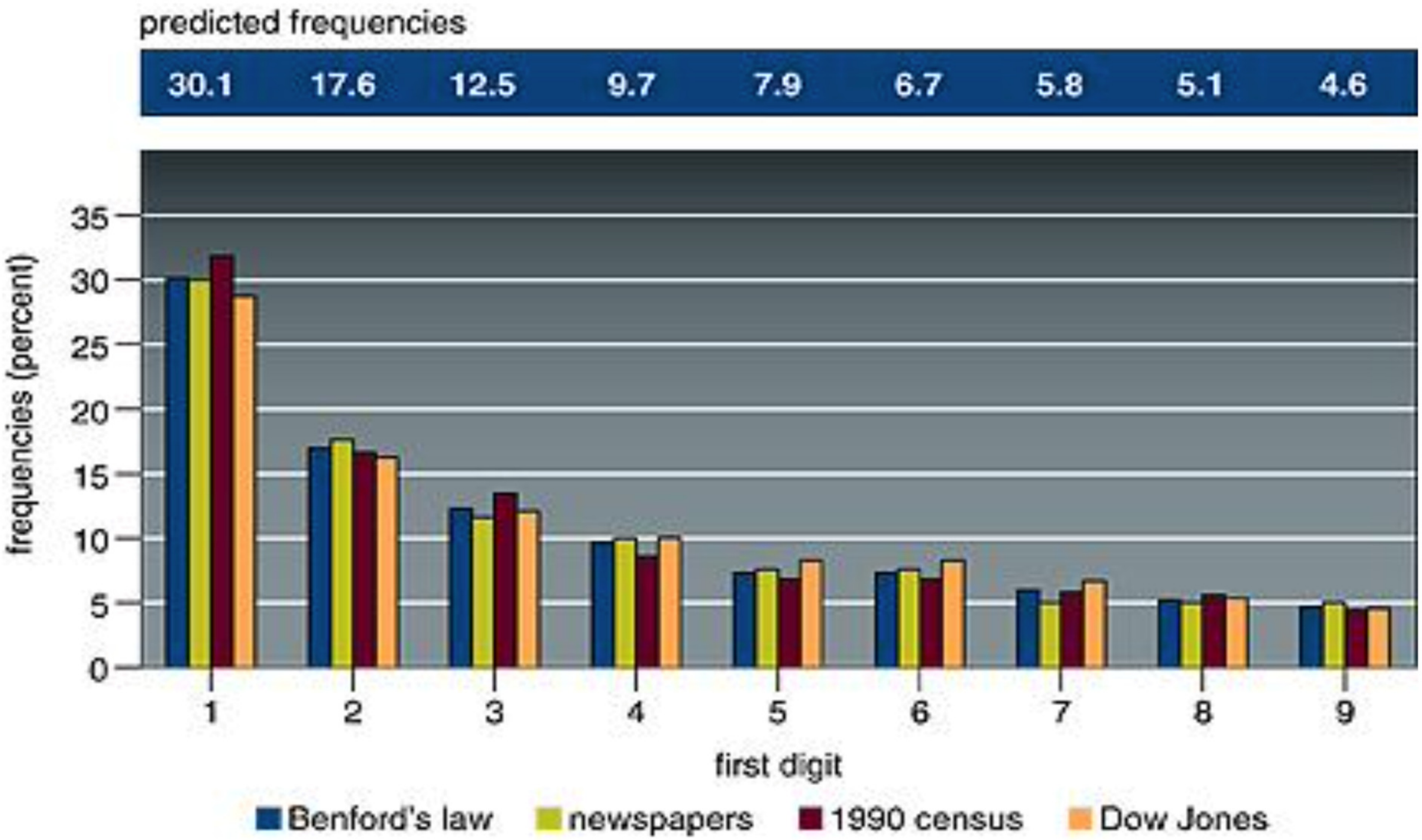
How about in real life? Do certain digits in numbers collected randomly from the front pages of the newspaper or census statistics or from stock-market prices occur more often than others?

Frequency with which first significant digit is d = log (1 + 1/d)





From "The First-Digit Phenomenon" by T. P. Hill, American Scientist, July-August 1998)



Long-term efforts to "prove" Benford's Law

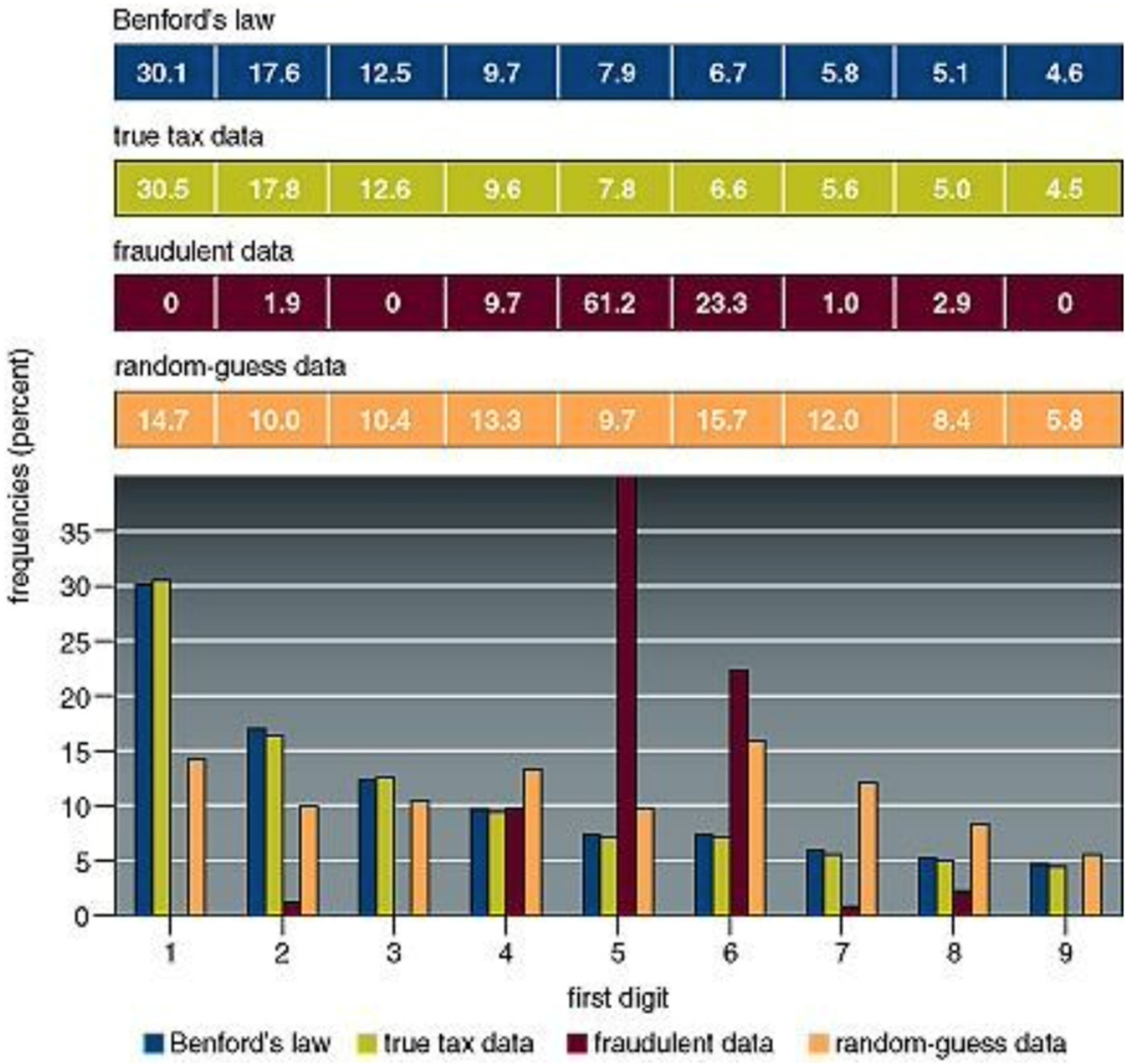
Properties of a random sample that result in such a distribution? e.g. not true for Unif {1,…999}

- Scale invariance: e.g. convert from dollars to pesos shouldn't change the first digit frequencies much
- Independent of base: Equally valid when numbers expressed in base 10, base 100, or others

The only distributions on numbers that satisfies these conditions satisfy
Pr(first significant digit = d) = log (1 + 1/d)

Modern Application

- Using Benford's law to detect fraud or fabrication of data in financial documents.



Random Quote

“It is easy to lie with statistics, but easier to lie without them”.

Fred Mosteller

“Too good to be true”

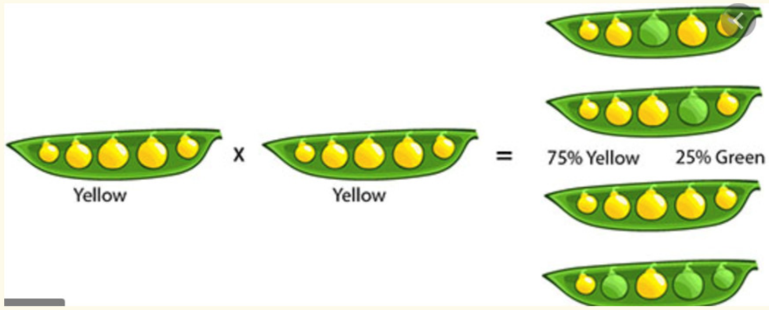
- The special case of not appreciating the expected magnitude of sampling error.

- Data comes out “too good to be true”, a telltale sign of having been tampered with, if not generated out of “whole cloth”.

Gregor Mendel's Sweet Peas

Postulated that self fertilization Of hybrid yellow-seeded sweet peas would yield offspring with

- 0.75 chance yellow-seeded
- 0.25 chance green seeded.



1865, reported results of 8023 experiments:

- 0.7505 yellow-seeded
- 0.2495 green-seeded.

Probability of observations as close to expected value as he reported is minute.

Some telltale signs of fakery….

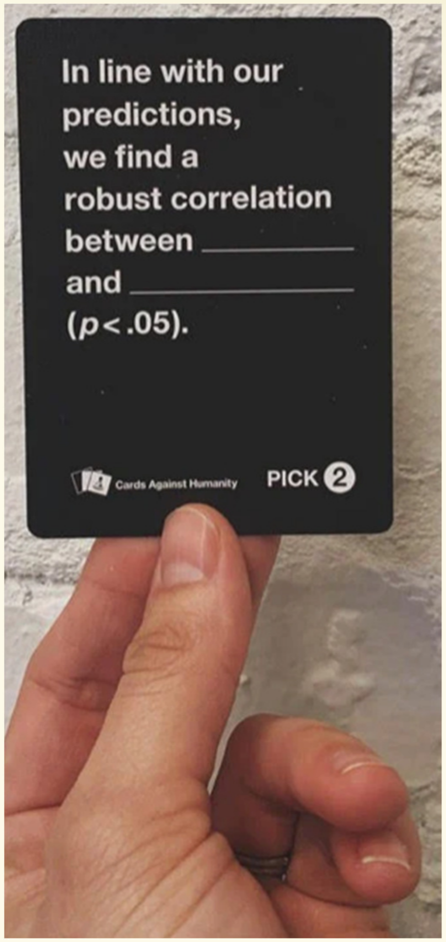
- Wrong shape
- Too close to expected value (especially replicated)
- Too far from expected value
- Replications too good to be true.

Another famous example: Sir Cyril Burt's Twins

3 data sets: IQ same to 3 decimal points.

3. p-Hacking

Manipulating data or statistical analyses to get **"significant p-values"**



First, a brief primer on hypothesis testing and p-values.

Suppose that I believe that jelly beans cause acne.
How might I provide evidence of this?
Approach – "probabilistic proof by contradiction"

Hypothesis Testing

Want to provide evidence that the null hypothesis can be rejected!

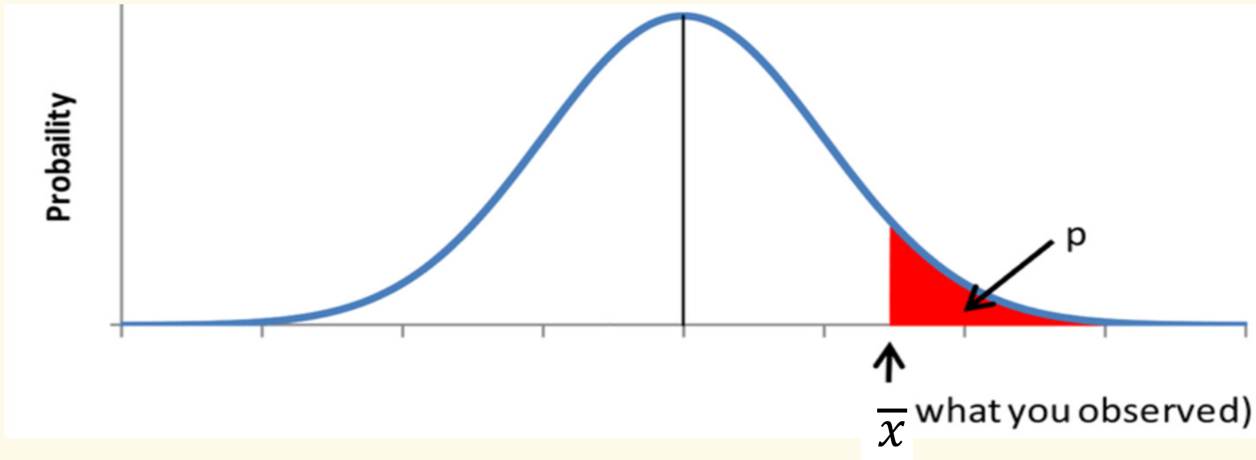
Average teenager has amount of acne with mean μ and variance σ^2

H_0 – null hypothesis (baseline): the mean amount of acne someone who eats jelly beans has is μ, i. e. , **jelly beans have no effect on acne**

H_A - Alternative hypothesis: the mean amount of acne someone who eats jelly beans has is $> \mu$

Choose ***significance level,*** say 0.05

Observe 100 jelly-bean-eating teenagers and measure their acne levels.
Suppose sample mean observed $\overline{x}$

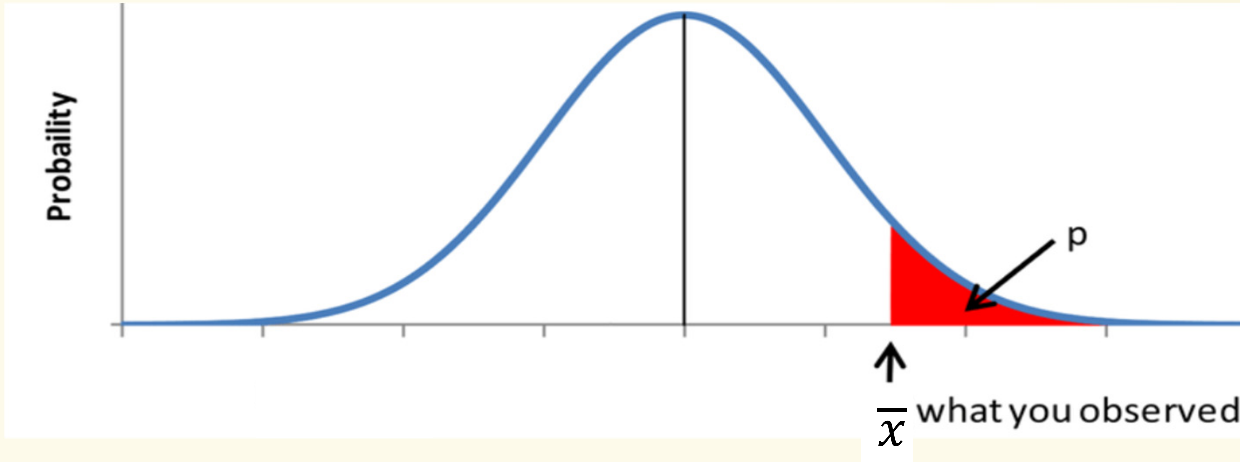


Hypothesis Testing

H_0 – null hypothesis (baseline):
jelly beans have no effect on acne

H_A - Alternative hypothesis:
Jelly beans increase acne

Suppose find that for measured $\overline{x}$

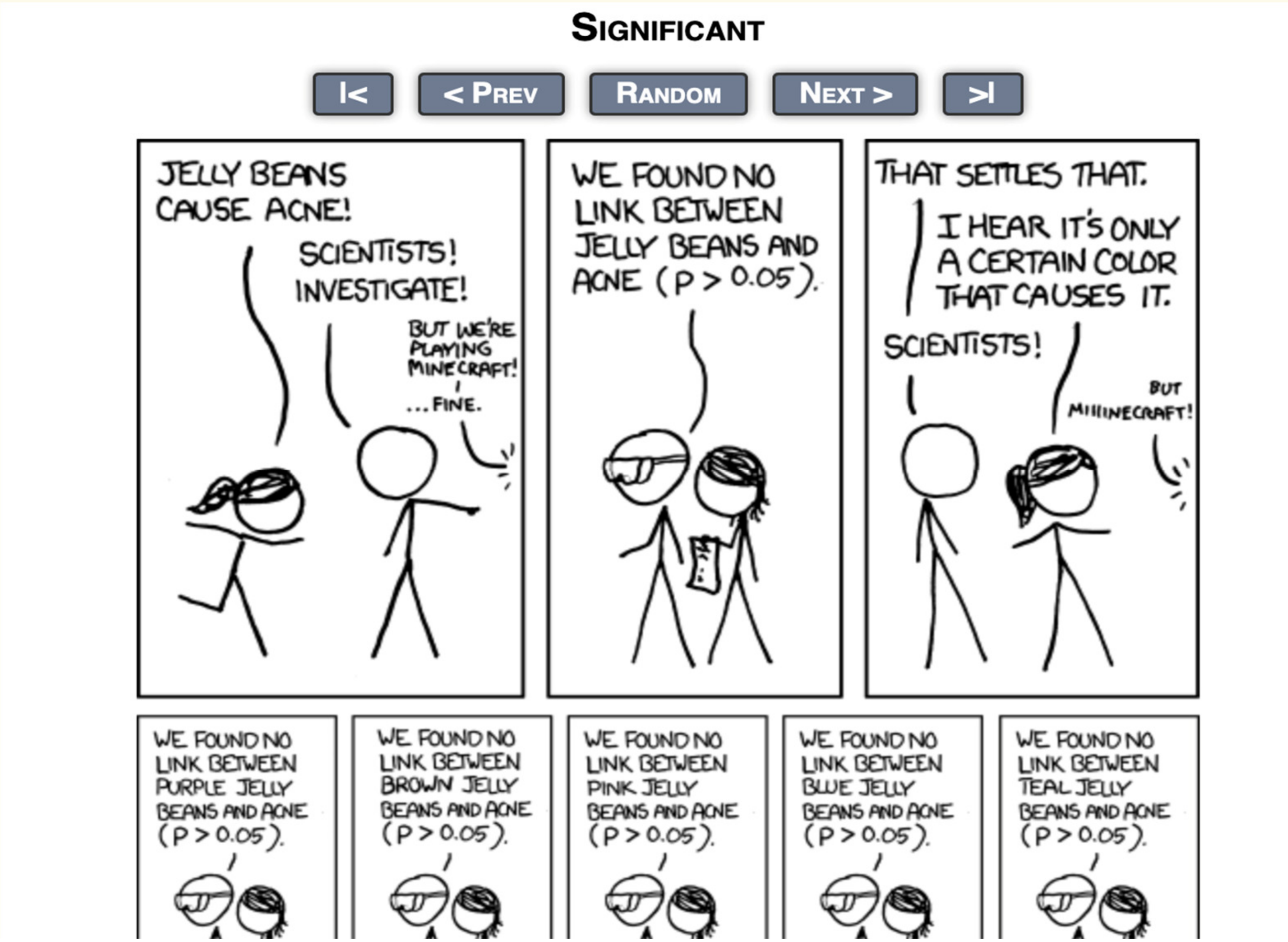


Pr (observing amount of acne this high if H_0 true) = **Pr ($\overline{X} \geq \overline{x}$) = 0.0162**. This is our **p-value.**

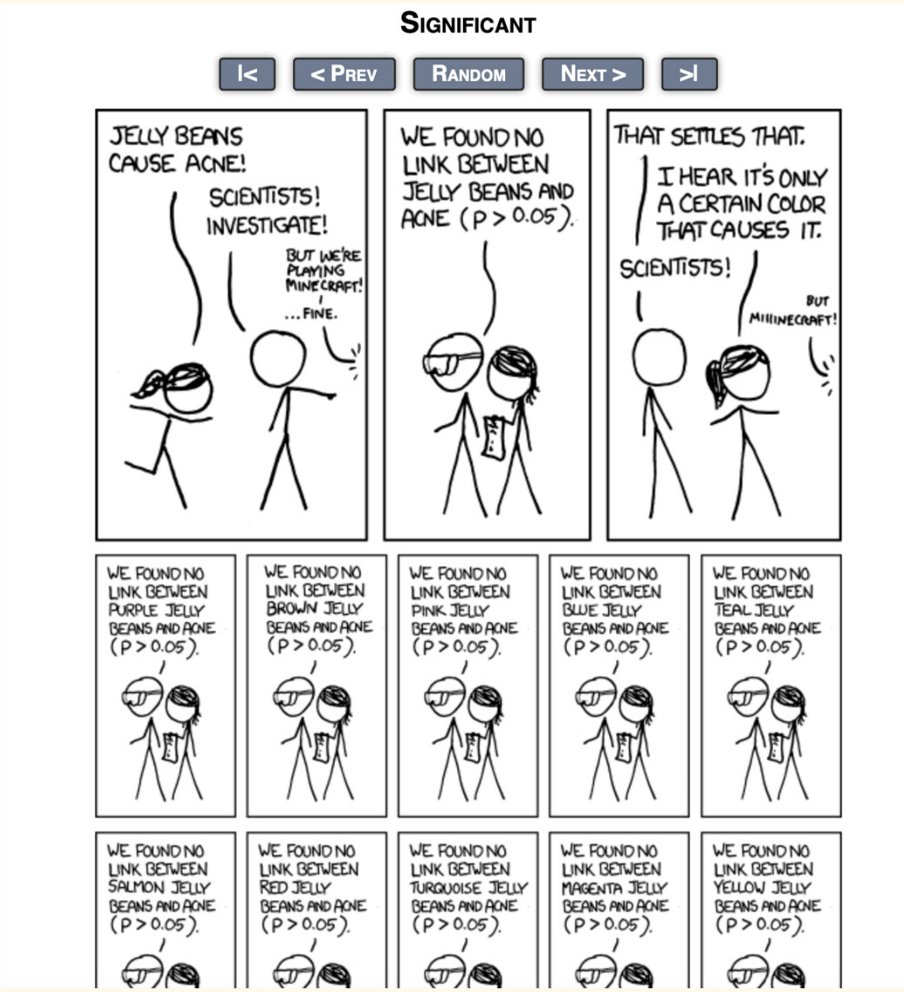
If $p < 0.05$ reject H_0 at the 0.05 significance level, i.e**., strong statistical evidence that jelly beans cause an increase in acne.** (If H_0 was true, this would be a very unlikely outcome).

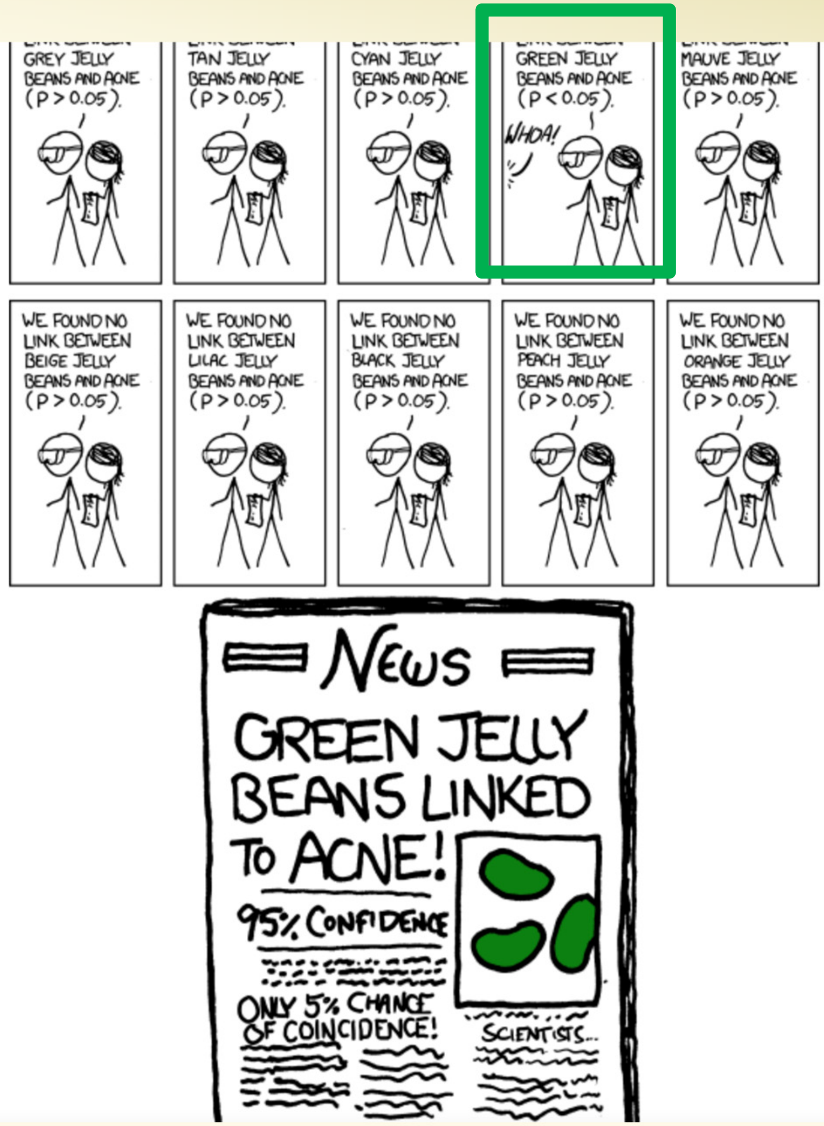
If $p > 0.05$, fail to reject $H_{0;}$
Not enough evidence to suggest the jelly bean effect on acne was significant.

3. p-Hacking


SIGNIFICANT
|<
< PREV
RANDOM
NEXT >
>|
JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).

3. p-Hacking


SIGNIFICANT
|<
< PREV
RANDOM
NEXT >
>|
JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).


GREY JELLY BEANS AND ACNE (P > 0.05).
TAN JELLY BEANS AND ACNE (P > 0.05).
CYAN JELLY BEANS AND ACNE (P > 0.05).
GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

3. p-Hacking

- Scientists concluded that “Eating green jelly beans gives you more acne” after testing that teenagers who ate green jelly beans have more acne than those who don’t, with a p-value of 0.05”.
  - **The p-value means:** if the null hypothesis is true (teens who eat green jelly beans and those who don’t have the **same** amount of acne), the probability of observing at least as extreme an outcome as we did is p.
  - **Putting it another way, a p-value of 0.05 means:** only a 5% chance of seeing this much acne if green jelly beans don’t cause acne
  - But what if I repeat similar experiments 20 times?
  - The chance that in 20 trials I will never get a p value < 0.05 is

$$0.95^{20} \approx 0.358$$

In other words 64% of the time one of these tests will be significant.

This result has no significance! Happened by random chance!

3. p-Hacking

- **Definition**: Performing the same hypothesis test multiple times in order to get a statistically significant result.

4. Malicious p-Hacking

- **Definition**: Performing the same hypothesis test multiple times in order to get a statistically significant result.
- The particularly evil thing: reporting only the significant tests, but not reporting the other 19 tests…..

Random Quote

"Torture numbers, and they'll confess to anything"

- George Easterbrook

5. p-Hacking due to Adaptive Data Analysis

Typical scenario:

1. Collect training data set $D = x_1, x_2, \ldots, x_n$ (expensive)
2. Compute $a = \mathbb{E}_D[g(x)]$ for some test function g (typically a polynomial)
3. Repeat for up to m steps or until (a, g) seems significant/surprising
   a. Update polynomial g based on the value of a found
   b. Compute $a = \mathbb{E}_D[g(x)]$

Problem: If m is too large compared to n, results are no longer valid

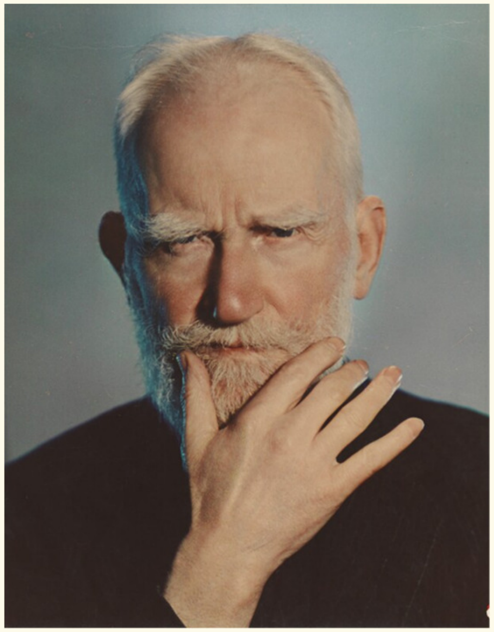
Differential privacy can help: Can compute a differentially private approximation to $a = \mathbb{E}_D[g(x)]$ that is just as good for significance and lets one use a much larger value of m

[Dwork, Feldman, Hardt, Pitassi, Reingold, Roth 2015]

Random Quote

"If at first you don't succeed, try two more times so your failure is statistically significant".

- George Bernard Shaw



Another interesting misuse of statistics

Attali/Bar-Hillel noticed that SAT answer keys are not randomized.

Keys are balanced rather than randomized.

Was easy for statisticians to detect by examining published tests.

This is a case of thinking “**randomization is too important to be left to chance** “!

Suggests a strategy for test-takers

- Answer all the questions you can.
- When guessing the rest, pick an answer position that occurs least frequently in your answers.

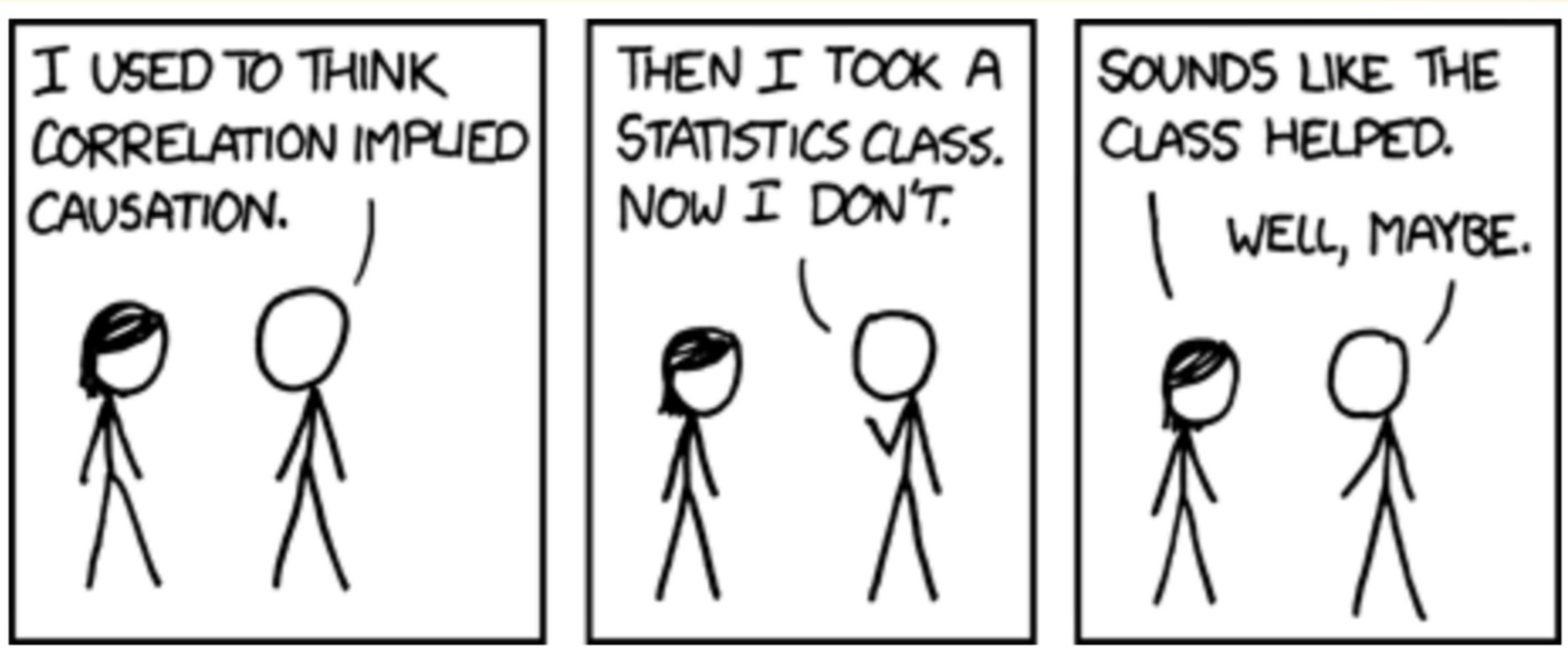
Simulations shows this adds 10-16 points over random guessing.

Claimed to be more gain than some very expensive SAT prep courses!

Conclusions

1. Determine if the samples are **random** and **representative**.
2. Ask for a confidence interval.
3. Be dubious. Be extremely dubious.
4. Don’t make up data or statistics. You’ll get caught.
5. Be wary of p-hacking (and don’t do it yourself)!
6. Be careful about seeing patterns where there are none.
7. Correlation does not imply causation.

Random Quote



Source: https://xkcd.com/552/

Conclusions

1. Determine if the samples are **random** and **representative**.
2. Ask for a confidence interval.
3. Be dubious. Be extremely dubious.
4. Don’t make up statistics. You’ll get caught.
5. Be wary of p-hacking (and don’t do it yourself)!
6. Be careful about seeing patterns where there are none.
7. Correlation does not imply causation.
8. Be careful with interpreting conditional probabilities. Intuition sometimes doesn’t work here!
9. Be wary of assuming things are independent that aren’t independent.

Random Quote

"Data is the sword of the 21st century, those who wield it well, the Samurai."

- Jonathan Rosenberg (ex-Google SVP)