

Statistics as a Distillation of Everyday Experience

Gerald van Belle

Department Biostatistics,
Department of Occupational and
Environmental Health Sciences
University of Washington,
Seattle, WA.

Where Are We Going?

Statistics as distillation of everyday experience

1. Variation
2. Causation

Experience can benefit from everyday statistics

A. Variation in everyday experience

1. Describing and classifying variation
2. Selection in the face of variation
3. Controlling variation
4. Inducing variation
5. “Missing data”

1. Describing and classifying variation

- We tell stories of ab“normality”
 - * Air travel horror stories, laptop disasters,...
- We sort into genres: art, biology, literature
 - * Concept of “population”
 - * Characteristics of population and “sample”
- Variation in time, space, social structures, ...
 - * Waves on beach (non-stationarity)
 - * Hierarchy, “social class”
- We make inferences based on limited data
 - * And often get the wrong population
 - * Basis for a great deal of humor
 - * Switch in “expectation”



2. Selection in the face of variation

- Need to know selection mechanism
Spend very little time on this
Assumption of “missing at random”
- Error of thinking that current observation is representative
- Unintended intentional selection
- Two examples:



Chicago Daily Tribune

HOME

PHILADELPHIA, PA., 1944

CHICAGO, ILL., 1944

WEDNESDAY, MARCH 29, 1944

10 CENTS PER COPY (PAID UP)

DEWEY DEFEATS TRUMAN

G. O. P. Sweep Indicated in State; Boyle Leads in City

REPUBLICAN TICKET AHEAD OF 1944 VOTE

Tops Coghlan in Hot Race for Attorney

Probable Winners

Town Balloting Gives Trend

Probable Winners

RECORD CITY VOTE SEEN IN LATE TALLIES

Suburban Ballot Near 575,000

BULLETINS ON ELECTIONS

Illinois

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Chicago, Ill., March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Early Count Gives G.O.P. Senate Edge

Washington, March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Washington, March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Washington, March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

PUTS G.O.P. BACK IN THE WHITE HOUSE

Stable Electoral Margin Seen

Washington, March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Washington, March 29.—(AP)—The Republican ticket headed by Governor Dewey and Attorney General Coghlan, led in the early returns from the Illinois primary election today.

Seattle Times

THURSDAY

MAY 26, 2005

Metro Edition



INDEPENDENT AND LOCALLY OWNED SINCE 1896 | seattletimes.com

unscreen on:
ch some rays



Numbers speak for themselves, GOP says

GOVERNOR'S
ELECTION TRIAL



2. Selection in the face of variation--2

- Need to know selection mechanism
Random selection as gold standard
- “Representativeness”
 - * Kruskal and Mosteller papers
 - * Slippery concept
 - * Large sample vs small sample

3. Controlling variation

- Clearest examples in sports:
Divisions, “junior”, ...
- Societal examples
Min, max speed limits
Occupational (noise limits, flying hours)
Vergunningen, vergunningen, ...
- “Blocking” in statistics

4. Inducing variation

- Antitrust laws
 - * Increase competition, i.e. variability
- Draft system in sports
 - * Teams more equal, $P(\text{win})$ near $1/2$
- Societal
 - * Admission to medical school in Holland
 - * Representativeness (slippery concept)
 - * Key to clinical trials

5. “Missing” data

1. Serious problem, obviously
2. Anatomy of missingness
 - Normal (e.g. pediatrician chart)
 - Transcription error
 - Just not there (Murphy was here)
 - Deliberately missing (e.g. extended testing on subset of patients)
- Impacts population of inference
- Example:

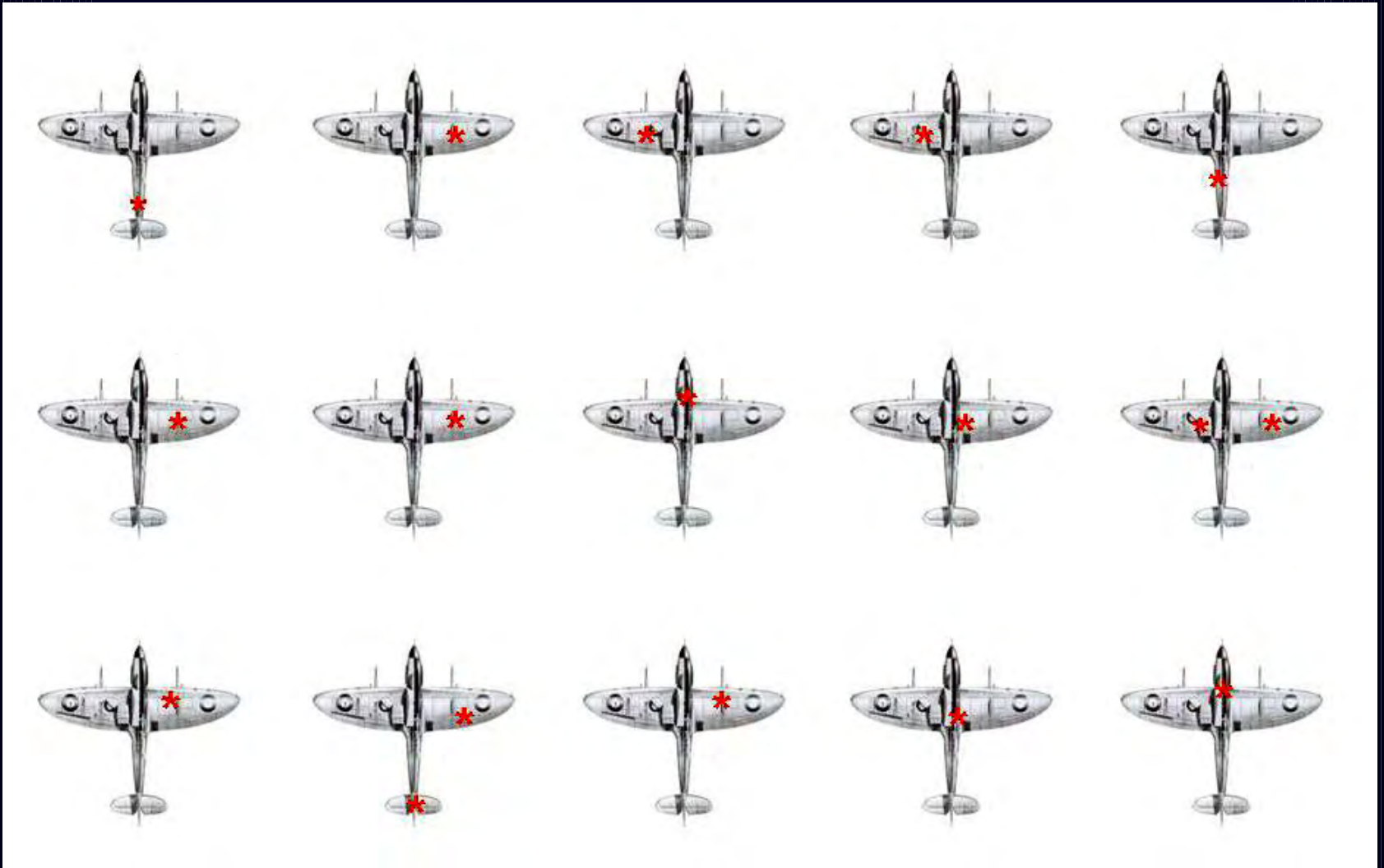


March 29, 2007

KNAW Lecture

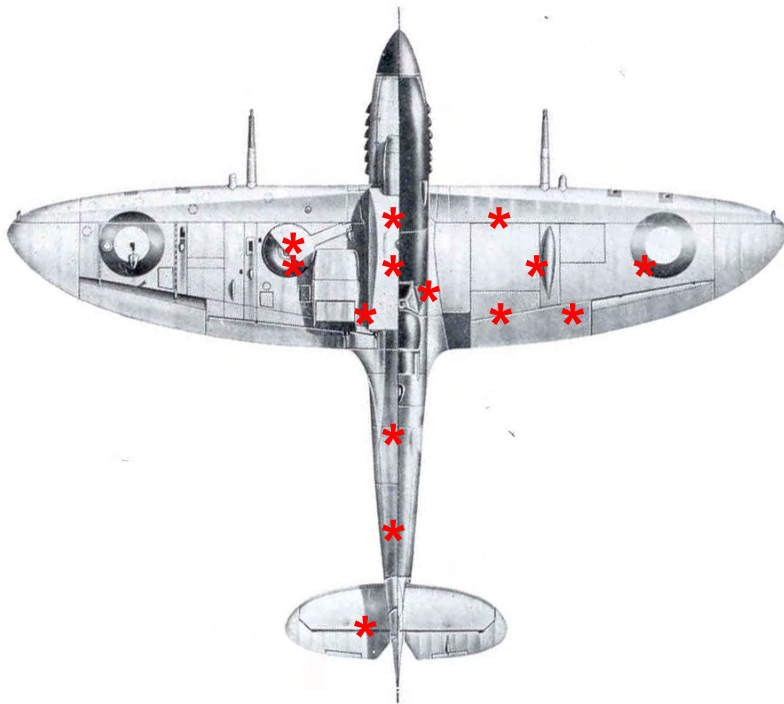
13

Vulnerability Analysis of Spitfires (sample: 15/400)



Composite of hits

Abraham Wald
Advice:



Another anatomy of missingness



...as we know,
there are known knowns;
there are things we know we know.
We also know there are known unknowns;
that is to say,
we know there are some things
we do not know.
But there are also unknown unknowns—
the ones we don't know we don't know.

Donald Rumsfeld

(set to music, see NPR website)

Translation into modern statistics

Non-missing

...as we know,
there are known knowns;
there are things we know we know.

**MCAR/
MAR**

We also know there are known unknowns;
that is to say,
we know there are some things
we do not know.

**Non-
ignorable**

But there are also unknown unknowns—
the ones we don't know we don't know.

Donald Rumsfeld

B: Causation in everyday experience

1. Aristotle's four causes
2. Hardwired to look for causation
3. Hardwired to assume association is causation
4. Hardwired to assign blame (secondary causes)

(The Dutch: "oorzaak" is closer to Aristotle's αιτιον)

1. Aristotle's four causes

- Material cause

(table made of wood)

- Formal cause

(four legs and flat top make this a table)

- Efficient cause

(carpenter makes a table)

- Final cause

(surface for eating or writing makes this a table)

(From S.M. Cohen, U Washington)

Four questions

- What **is** the question? **was**
- **Is** it testable? **Was**
- Where **will** you get the data? **did**

- What **will** the data tell you? **do**

The great divide

1. What is the Question? Is it testable?

**Not everything that can be counted
counts and not everything that counts
can be counted.**

Albert Einstein

"All the News
That's Fit to Print"

The New York Times

National Edition

Washington and Oregon: Periods of rain along the coast. Snow in the Cascades. Cloudy in the interior. Highs will be generally in the 30s and 40s. Weather map appears on Page C13.

VOL. CLVI, No. 53,813

Copyright © 2007 The New York Times

WEDNESDAY, JANUARY 3, 2007

Printed in Seattle ONE DOLLAR

PANEL PROPOSES THAT NEW JERSEY END EXECUTIONS

CALL FOR LIFE SENTENCES

State Would Be the First in
Over 35 Years to Drop
Capital Punishment

By LAURA MANSNERUS

TRENTON, Jan. 2 — A legislative commission recommended on Tuesday that New Jersey become the first state in more than 35 years to abolish the death penalty, finding "no compelling evidence" that capital punishment serves a legitimate purpose, and increasing evidence that it "is inconsistent with evolving standards of decency."

The report, whose lone dissenter was the original author of the state's modern death penalty statute, came a year after New Jersey joined Illinois and Maryland in imposing moratoriums on executions, and amid growing unease among politicians and the public about capital punishment. Nine other states, including New York, have also suspended executions in recent years, most because of court decisions. Maryland had lifted its moratorium in 2003, after a year, but a court essentially reinstated it last month.

Death penalty experts said that New Jersey was the first state to receive an official recommendation that capital punishment be abandoned, and it lands in a state where legislators have a Democratic majority along with a Democratic governor who supports repeal of the statute.

The governor, Jon S. Corzine, em-



Stephen Crowley/The New York Times

A Farewell to a President

Gerald R. Ford was remembered at his funeral yesterday in Washington as bringing decency, integrity and humility to government. Betty Ford, lower left, was escorted by President Bush as three former presidents, George Bush, Bill Clinton and Jimmy Carter, looked on. Pages A14-15.

Google Answer To Filling Jobs Is an Algorithm

A Surprising Secret to a Long Life: Stay in School

By GINA KOLATA

James Smith, a health economist at the RAND Corporation, has heard a variety of hypotheses about what it

THE NEW AGE

Piece in a Puzzle

education and even churchgoing. But the questions for researchers

ple more Social Security income, or less for that matter, will not really affect people's health. It is a good thing to do for other reasons but not for health."

Health insurance, too, he says, "is vastly overrated in the policy de-

IRAQ TO EXAMINE ABUSIVE CONDUCT TOWARD HUSSEIN

SUNNI PROTESTS INCREASE

Premier Pledges Inquiry
Into Taunts and Video
Taken at the Gallows

By JOHN F. BURNS
and JAMES GLANZ

BAGHDAD, Jan. 2 — Iraq's Shiite-led government said on Tuesday that it had ordered an investigation into the abusive behavior at the execution of Saddam Hussein, who was subjected to a battery of taunts by official Shiite witnesses and guards as he stood on the gallows.

Officials said a three-man Interior Ministry committee would look into the scenes that have caused outrage and public demonstrations among Mr. Hussein's Sunni Arab loyalists in Iraq, and widespread dismay elsewhere, especially in the Middle East. In an unofficial cellphone video recording broadcast around the world and posted on countless Web sites, Mr. Hussein is shown standing on the gallows platform with the noose around his neck at dawn on Saturday, facing a barrage of mockery and derision from unseen tormentors below the gallows.

As the shock of those scenes reached a new crescendo in Iraq, American officials revealed that they had worked until the last hours of Mr. Hussein's life to persuade Prime Minister Nuri Kamal al-Maliki to delay the execution. The officials, who spoke on condition that they not be named, said they appealed to Mr. Maliki not to proceed with the hanging at dawn on Saturday because of the onset of a m-

2. Frequent Consulting Scenario

Testable hypothesis

but

What was the question?

Example from *Science* (February 23, 2007)

| | |
|----------------|------------------------------|
| Title | Redefining the age of Clovis |
| Front page | Flints (pictures) |
| Page 1045 | Summary paragraph |
| Page 1067 | News story |
| Page 1122—1126 | Article (numbers) |

Very different “flavor” for each section

Question

Testable version

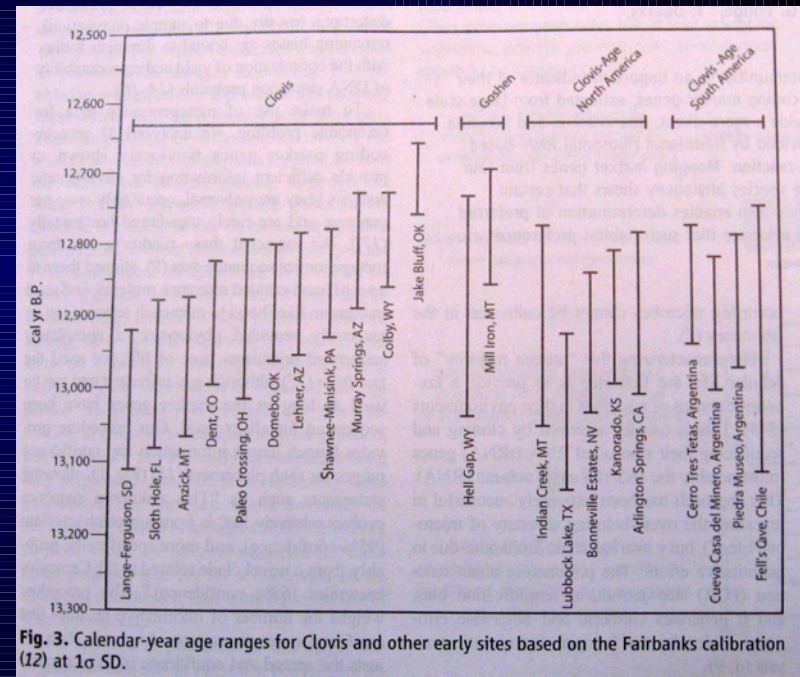


Fig. 3. Calendar-year age ranges for Clovis and other early sites based on the Fairbanks calibration (I2) at 1σ SD.

3. Hardwired to look for causation

- Story
- Instinctive looking for causes
- Challenging in courts
crime in search of criminal, stock market (“if only I had bought Microsoft in 1980”), and science (global warming)
- Life forces us to do this ex post facto
Whitehead quote

4. Hardwired to assume association is causation

1. Story
2. Criteria for causation help
3. Causation in observational studies is a great challenge
4. R.A. Fisher's design of experiments introduced randomization

5. Challenges to Causation in Observational Studies

1. Selection bias

“Where did you get the data?”

- Confounding

“What do you think the data are telling you?”

4. Interplay of selection bias and confounding

Effect modification needs to be considered

6. Hardwired to look for secondary causes

1. Assists in soccer, hockey, basketball
2. Tendency to blame i.e. move from efficient cause to material or formal cause; that is, change the “universe of discourse”
3. Surrogate outcomes in science (great work by Ross Prentice)

6. Observational vs Experimental Studies

| <u>Characteristic</u> | <u>Observational</u> | <u>Experiment</u> |
|-----------------------|----------------------|-------------------|
| Ethical issues | Fewer | More |
| Orientation | Retrospective | Prospective |
| Inference | Weaker | Stronger |
| Selection bias | Big problem | Less |
| Confounding | Present | Absent |
| Realism | More | Less |
| Causal plausibility | Weaker | Stronger |
| Researcher control | Less | More |
| Analysis | More complicated | Less |

The speech of numbers

If numbers could talk we could discern
the liars from the truth tellers

Numbers are a severe reduction of the
real world—the process is as important
as the product

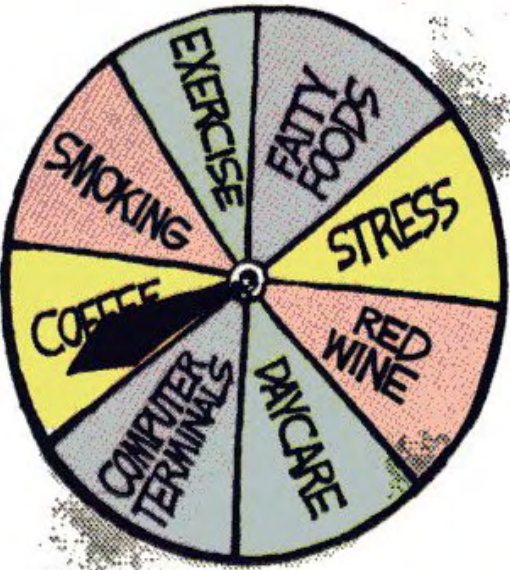
Consequence

In view of variation and hard-wired tendencies we have a lethal mixture.

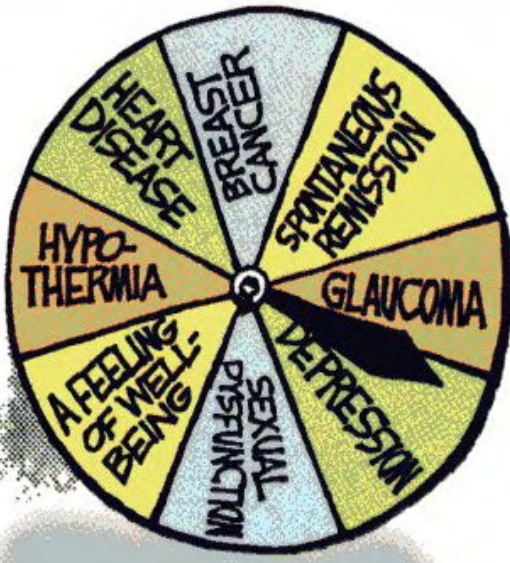
Today's Random Medical News

from the New England
Journal of
Panic-Inducing
Gobbledygook

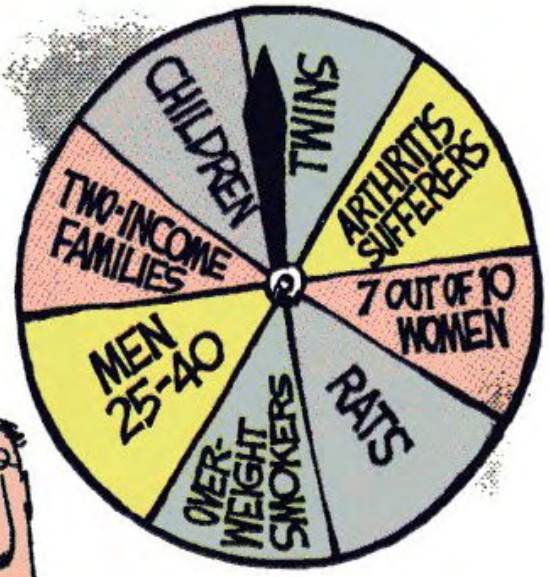
VINEYERMAN



CAN CAUSE



IN



ACCORDING TO A REPORT RELEASED TODAY....



Recapitulation:

1. What is the question?
2. Is it testable?
3. Where will you get the data?
4. What do you think the data are telling you?

Experience and everyday statistics

Variation

1. Variation is fact of life
2. “Population” as model
3. Representativeness
4. Regression to the mean

Causation

5. What is the question?
6. Testable question?
7. Association and causation
8. Causation through randomization