Applicable Statistics for the Clinical Research Professional

San Francisco Bay Area SOCRA Chapter Meeting
July 9th

Shashi Sathyanarayana, Ph.D. CSSBB

Numeric Insight, Inc
www.numericinsight.com

practical and exceptional number crunching, and scientific programming

Let us hit the ground running

- The time taken for cessation of bleeding was recorded for a large number of persons whose fingers had been pricked. The mean time was found to be 1.407 minutes and the standard deviation was 0.588 min.
- In order to stably measure bleeding time, a standard pressure must be applied to the upper arm
- In an effort to determine whether pressure applied to the upper arm increases bleeding time, six persons had pressure equal to 20 mmHg applied to their upper arms and had their fingers pricked.
- For these six persons, the time taken (in minutes) for bleeding to stop were

  1.15

  1.75

  1.32

  1.28

  1.39

  2.50

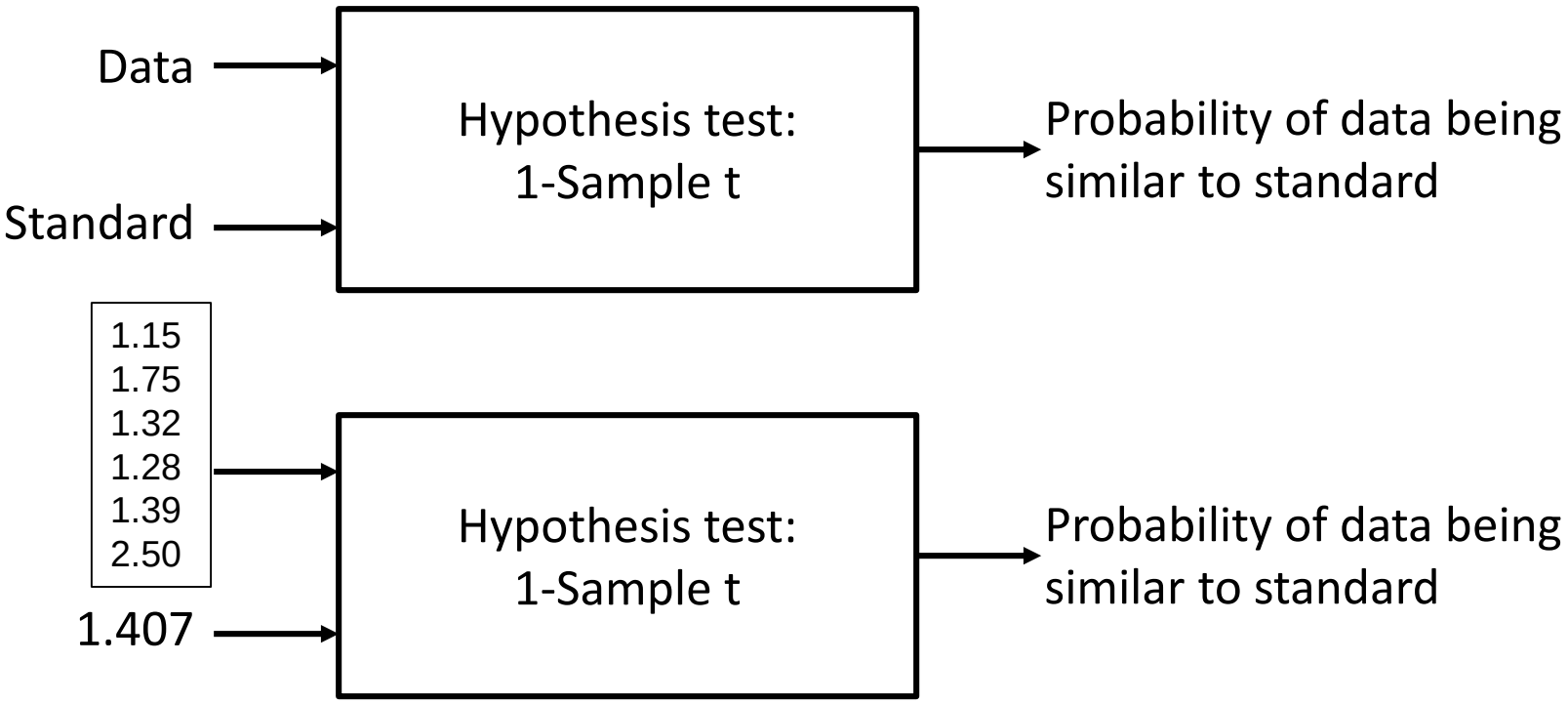
Clinical question: Does pressure increase bleeding time?

Basic Statistics: A Primer for the Biomedical Sciences (Wiley Series in Probability and Statistics) by Olive Jean Dunn and Virginia A. Clark (Jul 27, 2009). Fourth Edition pp.91

Exploratory

Hypothesis Test: 1-Sample t

Clinical question: Does pressure increase bleeding time? [w.r.t established standard]



Do statistical analysis

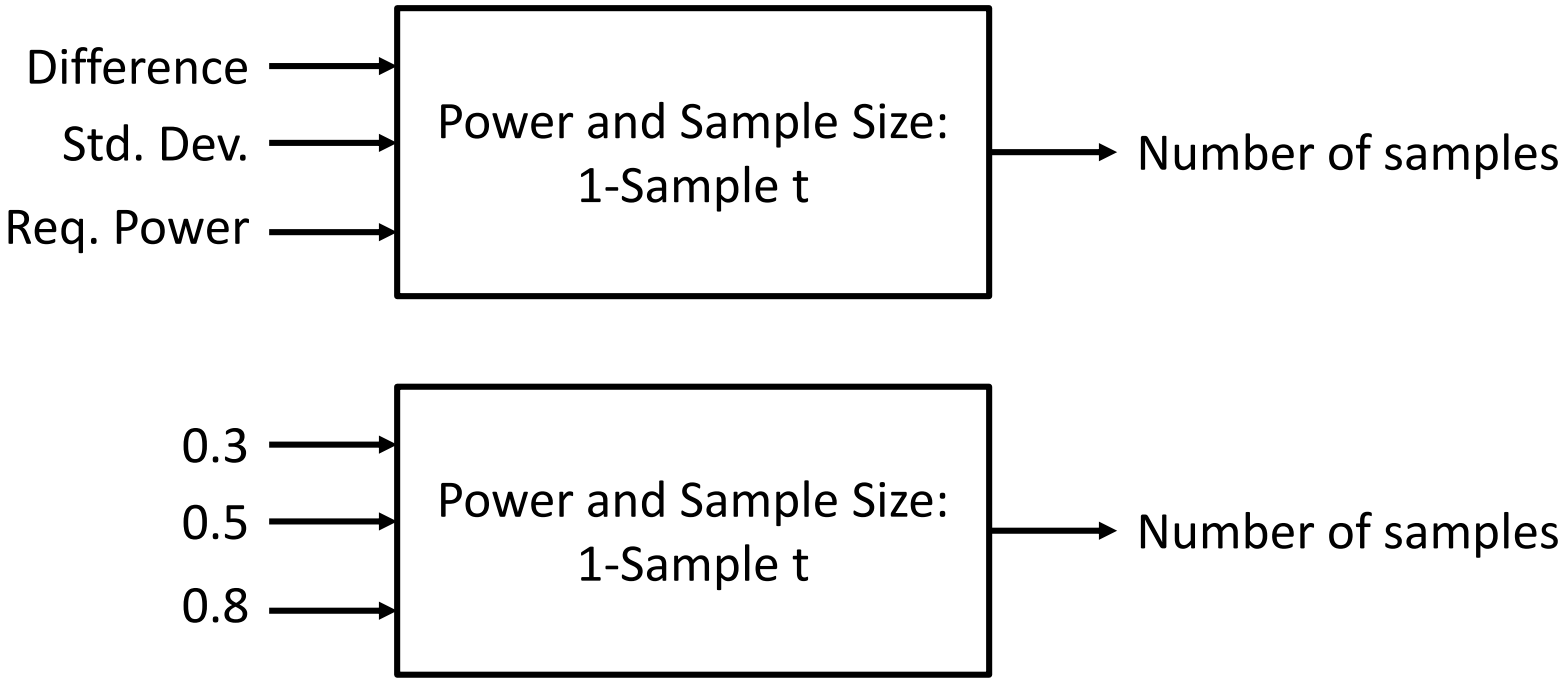
Conclusion: There is a good chance (47%) that data=standard

CAUTION! This test has very low power (10%)

Exploratory

Power and Sample Size: 1-Sample t

Statistical question: How many samples are needed to detect a clinically significant difference? [w.r.t established standard]



Do statistical analysis

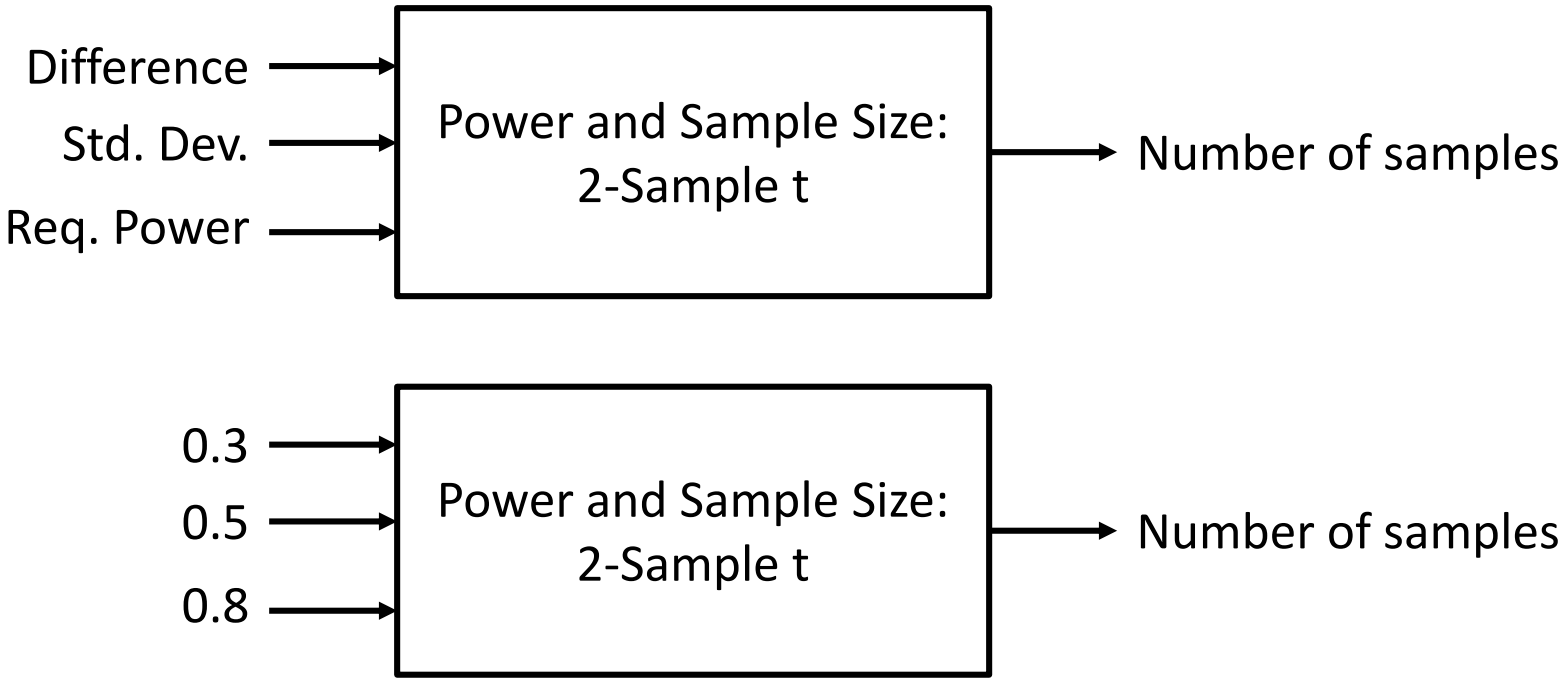
Conclusion: Need 24 samples to detect desired effect

"I know so little! Thorough investigation needed!"

"Clinical Trial"

Power and Sample Size: 2-Sample t

Statistical question: How many samples are needed to detect a clinically significant difference ***between two groups***?



Do statistical analysis

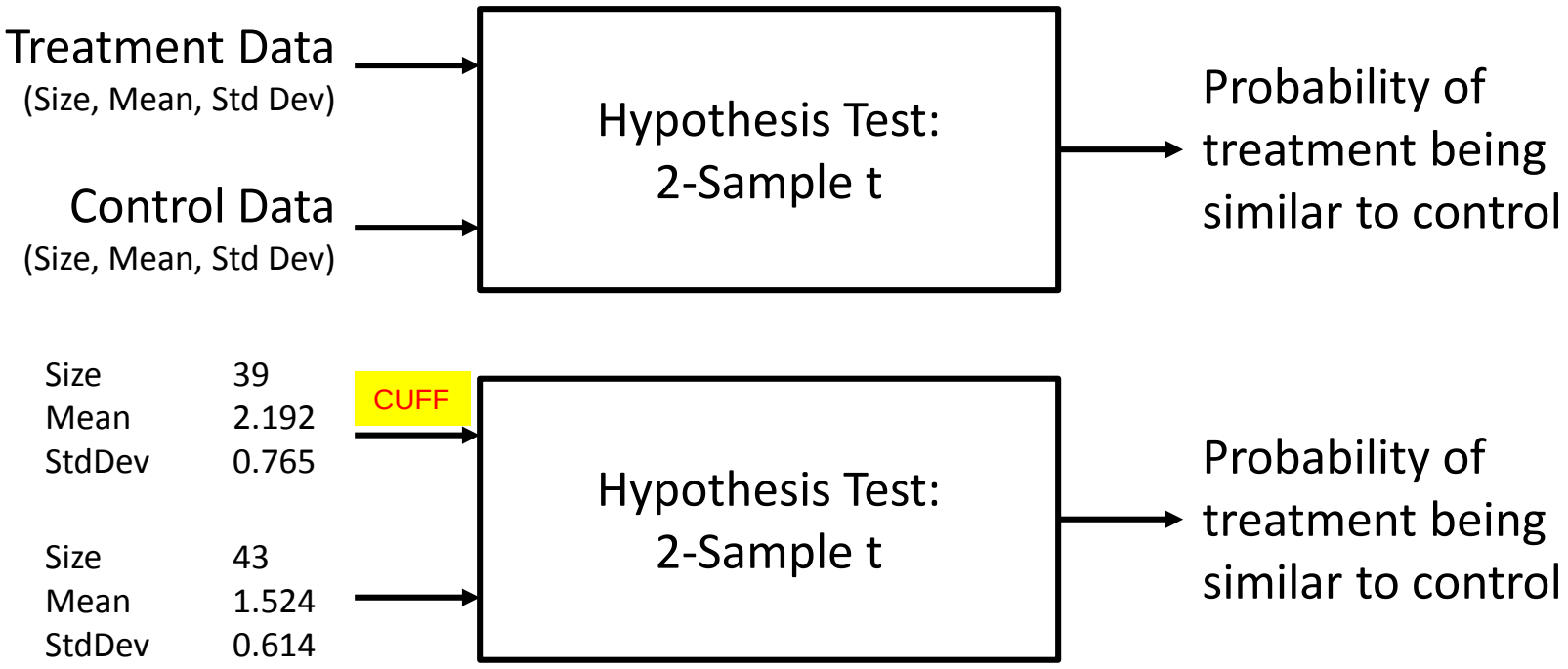
Conclusion: Need 45 samples of each group to detect effect

"To be safe, let me get 50 samples of each group."

"Clinical Trial"

Hypothesis Test: 2-Sample t

Clinical question: Does pressure increase bleeding time ***between two groups***?



Do statistical analysis

Conclusion: Cuffing definitely (p=0) increases bleeding time.

"I *MUST* standardize cuffing procedures in my study."

Wake Up!

- To do the simplest 1-sample t test you need data and hypothesized mean.
- The folks in the lab next door are also engaged in a similar experiment, except that their data has a smaller standard deviation. Is it more likely or less likely that they would reject the null hypothesis?
- If you are claiming to have discovered a better drug, the FDA would like you to produce solid data that accepts / rejects the null hypothesis in favor of a one-sided/two-sided alternative hypothesis.
- If you have too less data, you are likely to accept / reject the null hypothesis. This is called a Type I / II error. Also called false negative .

About Shashi Sathyanarayana, Ph.D CSSBB

- Numeric Insight, Inc www.numericinsight.com
- DOMAINS:
  - TRAINING
  - Bioinformatics
  - Biotechnology
  - Algorithms/Scientific Computing
  - Healthcare
  - Medical Devices
- Formerly
  - Boston Scientific, Cleveland Clinic Foundation

Another example: Insulin Binding Capacity

- The insulin binding capacity (pmol/mg protein) was measured for a sample of diabetic rats treated with a low dose of insulin and another sample of diabetic rats treated with a high dose
- The follow data was collected

Dose	N	Mean	SD
Low dose	8	1.98	0.51
High dose	12	1.30	0.35

Clinical question: Does the data indicate that there is any difference in the true average insulin binding capacity due to dosage level? Assume effect size of 0.65 pmol/mg

J. Clinical Investigation, 1978 552-560

Insulin Binding Capacity (cont)

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	8	1.980	0.510	0.18
2	12	1.300	0.350	0.10

Difference = mu (1) - mu (2)
Estimate for difference: 0.680
95% CI for difference: (0.278, 1.082)
T-Test of difference = 0 (vs not =): T-Value = 3.55 P-Value = 0.002 DF = 18
Both use Pooled StDev = 0.4195

There is strong evidence of significant difference between treatments. The probability that the two streams of data came from the same source is 0.002.

A Power and Sample Size calculation shows that this is a sufficiently powerful test.

Power and Sample Size

2-Sample t Test
Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05 Assumed standard deviation = 0.4195

Difference	Sample Size	Target Power	Actual Power
0.65	8	0.8	0.821441

The sample size is for each group.

Related test: Salk Polio Vaccine Experiment

- Salk Polio Vaccine Experiment (1954)
- The follow data was collected

Treatment	N	No of cases of paralytic polio
Placebo	201229	110
Vaccine	200745	33

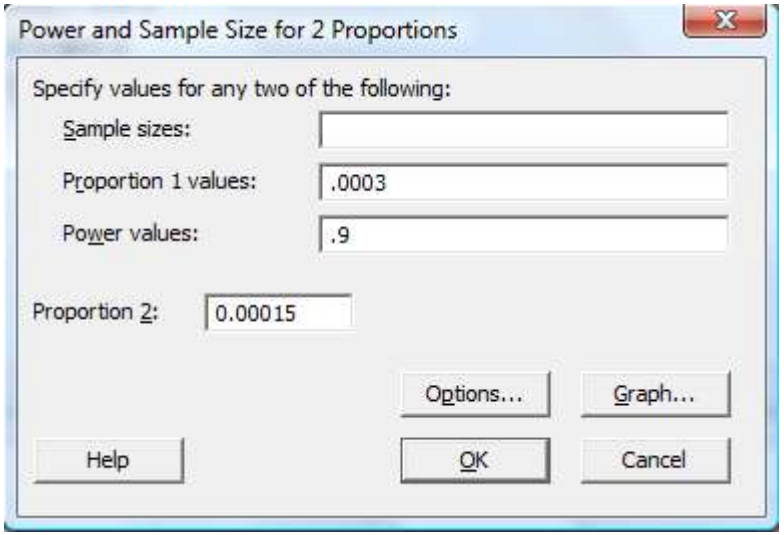
Clinical question: Does the data indicate that there is any difference in the proportion of children who contract polio? Is a vaccinated child less likely to contract polio than an unvaccinated child?

Thomas Francis, Robert Korn, et al. "An Evaluation of the 1954 Poliomyelitis Vaccine Trials."
American Journal of Public Health 45 (1955)

Salk Polio Vaccine Experiment (cont)

If this trial was yet to be done, this is how we would estimate sample size

- Note that 30 unvaccinated children per 100000 contract the disease
- ***Require*** that an effective vaccine should reduce the incidence to at least 15 per 100000 children
- This is how you enter these specifications into Minitab
- The sample size needed turns out to be 210100



Power and Sample Size

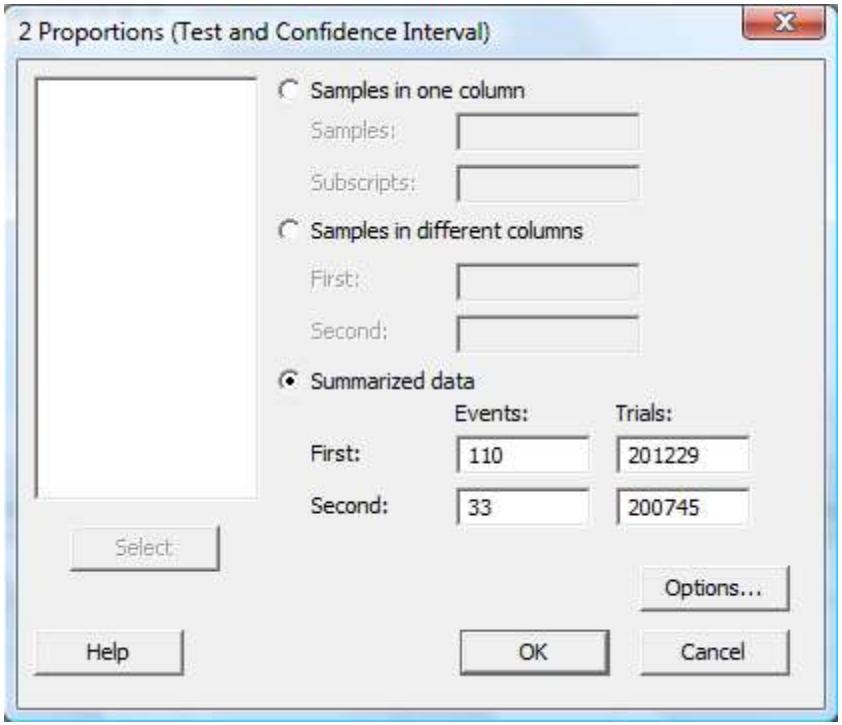
```
Test for Two Proportions
Testing proportion 1 = proportion 2 (versus not =)
Calculating power for proportion 2 = 0.00015
Alpha = 0.05

                Sample     Target      Actual
Proportion 1    Size       Power       Power
      0.0003    210100     0.9         0.900001
The sample size is for each group.
```

Close to the numbers actually used in the trial

Salk Polio Vaccine Experiment (cont)

Running the test is EASY

- This is how you enter the clinical trial data into Minitab
- The test answers the question *How likely is it that the two methods of "treatment" are similar?*
- Answer: VERY unlikely, with immeasurably small probability.



Test and CI for Two Proportions

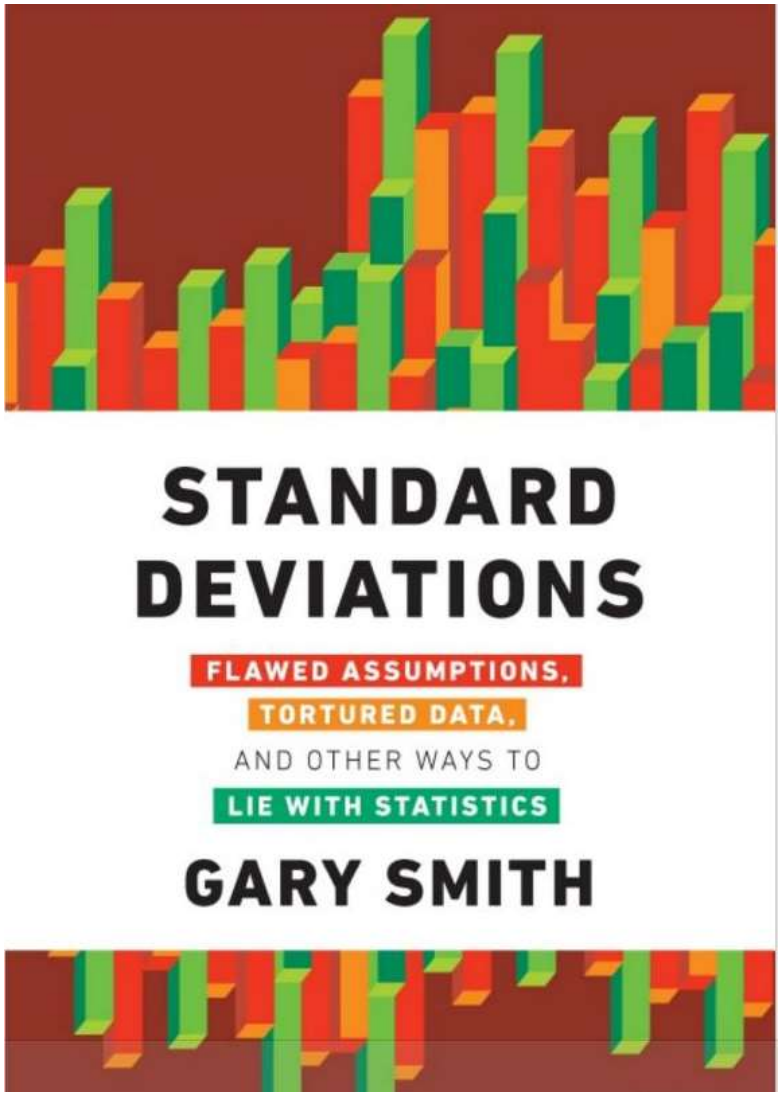
Sample	X	N	Sample p
1	110	201229	0.000547
2	33	200745	0.000164

Difference = p (1) - p (2)
Estimate for difference: 0.000382253
95% CI for difference: (0.000265742, 0.000498764)
Test for difference = 0 (vs not = 0): Z = 6.43 P-Value = 0.000

Statistics

- Statistics is used to summarize data so that non-statisticians can understand it
- Although clinical investigations usually involve collecting and generating a lot of data, at the end of the day, what you really need is a “punch-line”
  - Did the new drug/device/therapy work?
  - Are the two groups being compared the same or different?
  - Is the new more precise than the old?

A great read!


STANDARD
DEVIATIONS
FLAWED ASSUMPTIONS,
TORTURED DATA,
AND OTHER WAYS TO
LIE WITH STATISTICS
GARY SMITH

Concluding Remarks

- **Get your statistician involved before planning your study**
- Things statisticians do not like:
  - This grant/protocol/proposal is due tomorrow. Could you please go over it …
  - Could you please clean up the statistical sections in this paper
  - Request to wring meaning from data that has arisen from a poorly designed trial
- Statisticians have a lot to add
  - Rigor and new perspective on your study
  - Study will be more efficient and credible

Homework

- Many women claim they knew the sex of their child well before it was revealed to them on a medical examination. A researcher decided to examine this more closely.
- She asked 104 pregnant women to predict the sex of their unborn child, and 57 guessed correctly.
- Use Minitab to guide the conclusion the researcher can make.
- Get my email address from www.numericinsight.com and write to me. THANK YOU.

Legal Notice