Using Simulation Methods to Introduce Inference

Kari Lock Morgan
Duke University

In collaboration with
Robin Lock, Patti Frazer Lock, Eric Lock, Dennis Lock

CAUSE Webinar
12/13/11

Simulation Methods

- Inference: confidence intervals and hypothesis tests
- Simulation methods: bootstrapping and randomization
- I'll focus on randomization tests, because Chris Wild just gave a great CAUSE webinar on bootstrapping:

http://www.causeweb.org/webinar/activity/2011-11/

Hypothesis Testing

To generate a distribution assuming H_0 is true:

- *Traditional Approach*: Calculate a test statistic which should follow a known distribution if the null hypothesis is true (under some conditions)
- *Randomization Approach*: Decide on a statistic of interest. Simulate many randomizations assuming the null hypothesis is true, and calculate this statistic for each randomization

Traditional Hypothesis Testing

- Why not?
- With a different formula for each test, students often get mired in the details and fail to see the big picture
- Plugging numbers into formulas does little to help reinforce conceptual understanding

Paul the Octopus

http://www.youtube.com/watch?v=3ESGpRUMj9E

Paul the Octopus

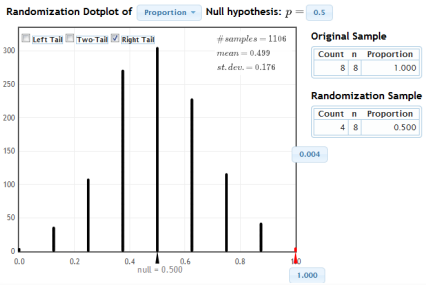
- Paul the Octopus predicted 8 World Cup games, and predicted them all correctly
- Is this evidence that Paul actually has psychic powers?
- How unusual would this be if he was just randomly guessing (with a 50% chance of guessing correctly)?
- How could we figure this out?

Simulate with Students

- Students each flip a coin 8 times, and count the number of heads
- Count the number of students with all 8 heads by a show of hands (will probably be 0)
- If Paul was just guessing, it would be *very unlikely* for him to get all 8 correct!
- How unlikely? *Simulate many times!!!*

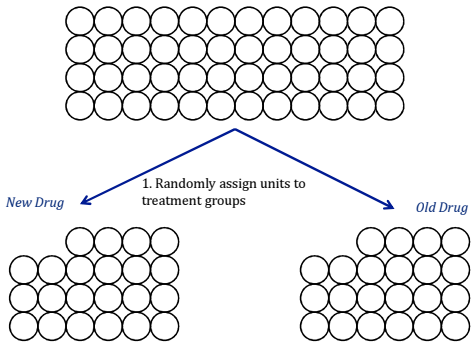
Simulate with StatKey

www.lock5stat.com/statkey

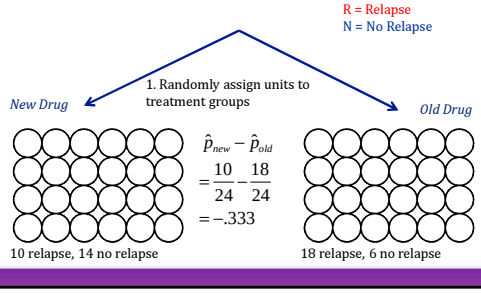


Cocaine Addiction

- In a randomized experiment on treating cocaine addiction, 48 people were randomly assigned to take either Desipramine (a new drug), or Lithium (an existing drug)
- The outcome variable is whether or not a patient relapsed
- Is Desipramine significantly better than Lithium at treating cocaine addiction?

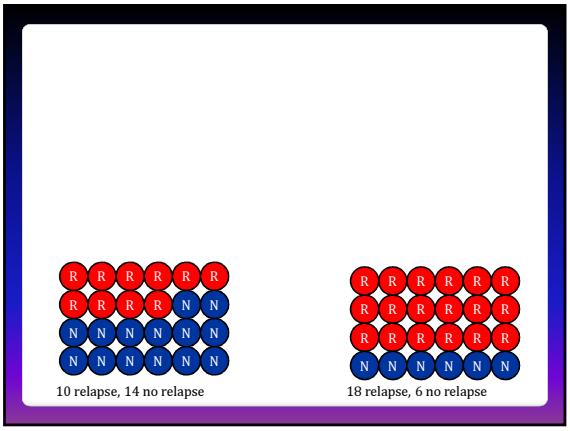


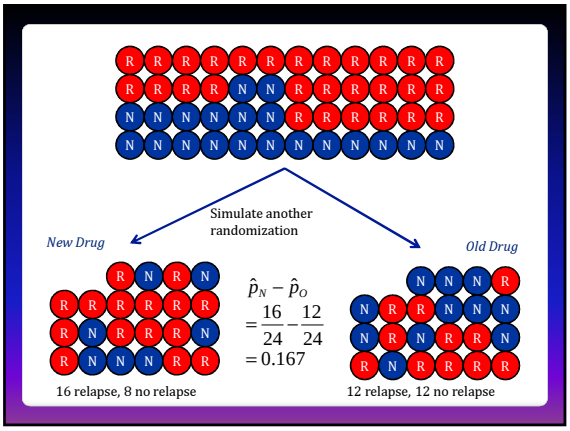
2. Conduct experiment
3. Observe relapse counts in each group

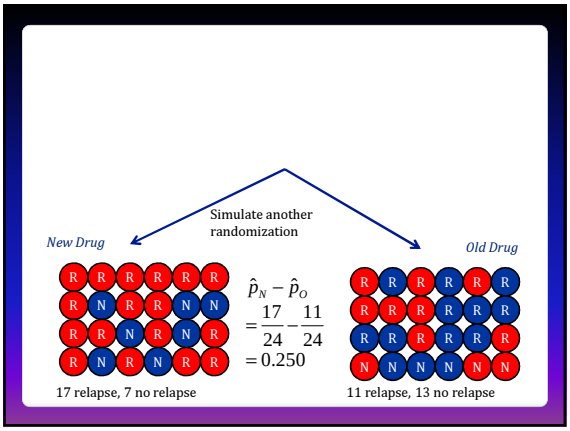


Randomization Test

- If the null hypothesis is true (if there is no difference in treatments), then the *outcomes would not change* under a different randomization
- Simulate a new randomization, keeping the outcomes fixed (as if the null were true!)
- For each simulated randomization, calculate the statistic of interest
- Find the proportion of these simulated statistics that are as extreme (or more extreme) than your observed statistic





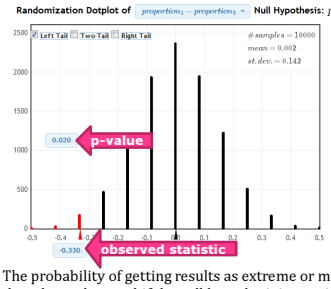


Simulate with Students

- Give students index cards labeled **R** (28 cards) and **N** (20 cards)
- Have them deal the cards into 2 groups
- Have them contribute to a class dotplot for the randomization distribution

Simulate with StatKey

www.lock5stat.com/statkey



The probability of getting results as extreme or more extreme than those observed *if the null hypothesis is true,* is about .02. p-value

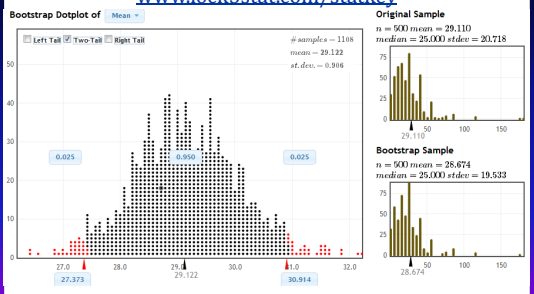
Cocaine Addiction

You want to know what would happen

- Why did you re-deal your cards?
  - by random chance (the random allocation to treatment groups)
- Why did you leave the outcomes (relapse or no relapse) unchanged on each card?
  - if the null hypothesis is true (there is no difference between the drugs)

Bootstrapping

www.lock5stat.com/statkey



Middle 95% of bootstrap statistics: (27.4, 30.9)

Simulation vs Traditional

- Simulation methods
  - intrinsically connected to concepts
  - same procedure applies to all statistics
  - no conditions to check
  - minimal background knowledge needed
- Traditional methods (normal and t based)
  - Familiarity expected after intro stats
  - Needed for future statistics classes
  - Only summary statistics are needed
  - Insight from standard error

Simulation AND Traditional?

- Currently, we *introduce inference* with simulation methods, then cover the traditional methods
- Students have seen the normal distribution appear repeatedly via simulation; use this common shape to motivate traditional inference
- "Shortcut" formulas give the standard error, avoiding the need for thousands of simulations
- Students already know the concepts, so can go relatively fast through the mechanics

Student Preferences

Which way do you prefer to do inference (confidence intervals and hypothesis tests)?

Bootstrapping and Randomization	Formulas and Theoretical Distributions
42	16
72%	28%

Student Preferences

Which way did you prefer to learn inference?

Bootstrapping and Randomization	Formulas and Theoretical Distributions
39	19
67%	33%

Student Preferences

Which way of doing inference gave you a better conceptual understanding of confidence intervals and hypothesis tests?

Bootstrapping and Randomization	Formulas and Theoretical Distributions
42	16
72%	27%

Student Preferences

DO inference	Simulation	Traditional
AP Stat	18	10
No AP Stat	24	6

LEARN inference	Simulation	Traditional
AP Stat	13	15
No AP Stat	26	4

UNDERSTAND	Simulation	Traditional
AP Stat	17	11
No AP Stat	25	5

Student Preferences

Student Preferences		*LEARN inference*	
		Simulation	Traditional
DO inference	Simulation	33	9
	Traditional	6	10

Student Preferences		*UNDERSTAND inference*	
		Simulation	Traditional
DO inference	Simulation	34	8
	Traditional	8	8

Student Preferences		*UNDERSTAND inference*	
		Simulation	Traditional
LEARN inference	Simulation	34	5
	Traditional	8	11

Simulation methods are useful for *teaching* statistics…

- The methods reinforce the concepts
  - A randomization test is based on the definition of a p-value
  - A bootstrap confidence interval is based on the idea that statistics vary over repeated samples
- Very little background is needed, so the core ideas of inference can be introduced early in the course, and remain central throughout the course

… and for *doing* statistics!

- Introductory statistics courses now (especially AP Statistics) place a lot of emphasis on checking the conditions for traditional hypothesis tests
- However, students aren't given any tools to use if the conditions aren't satisfied!
- Randomization-based inference has no conditions, and always applies (even with non-normal data and small samples!)

It is the way of the *past*…

"Actually, the statistician does not carry out this very simple and very tedious process [the randomization test], but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method."

-- Sir R. A. Fisher, 1936

… and the way of the *future*

"… the consensus curriculum is still an unwitting prisoner of history. What we teach is largely the technical machinery of numerical approximations based on the normal distribution and its many subsidiary cogs. This machinery was once necessary, because the conceptually simpler alternative based on permutations was computationally beyond our reach. Before computers statisticians had no choice. These days we have no excuse. Randomization-based inference makes a direct connection between data production and the logic of inference that deserves to be at the core of every introductory course."

-- Professor George Cobb, 2007

Thank you!

kari@stat.duke.edu