Early Inference: Using Randomization to Introduce Hypothesis Tests

Kari Lock, Harvard University
Eric Lock, UNC Chapel Hill
Dennis Lock, Iowa State

Joint Mathematics Meetings
New Orleans, 1/9/11
Traditional Hypothesis Testing

• In many introductory statistics classes now, too many students may see hypothesis tests as a series of steps and often meaningless formulas.

• With a different formula for each test (proportions, means, etc.), students often get mired in the details and fail to see the big picture.

• Following formulas and looking up a p-value in a table does nothing to help reinforce conceptual understanding.
**p-value**

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

- To calculate a p-value, we need a *distribution* for results we would observe if the null hypothesis were true.

- The only difference between traditional and randomization based approaches to hypothesis testing is how this distribution is obtained.
Distribution Under $H_0$

- *Traditional Approach*: Calculate a test statistic which should follow a known distribution if the null hypothesis is true (under some assumptions).

- *Randomization Approach*: Decide on a statistic of interest. Simulate many randomizations assuming the null hypothesis is true, and calculate this statistic for each randomization.
Example: 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?
1. Randomly assign units to treatment groups

New Drug

Old Drug
2. Conduction Experiment

3. Observe Outcome Data

1. Randomly assign units to treatment groups

\[
\hat{p}_{\text{new}} - \hat{p}_{\text{old}} = \frac{10}{24} - \frac{18}{24} = -0.333
\]

R = Relapse
N = No Relapse

Old Drug

New Drug

10 relapse, 14 no relapse

18 relapse, 6 no relapse
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
\[ \hat{p}_{\text{new}} - \hat{p}_{\text{old}} = \frac{10}{24} - \frac{18}{24} = -0.333 \]

10 relapse, 14 no relapse

18 relapse, 6 no relapse
Simulate another randomization

New Drug

16 relapse, 8 no relapse

\[ \hat{p}_{\text{new}} - \hat{p}_{\text{old}} = \frac{16}{24} - \frac{12}{24} = .167 \]

Old Drug

12 relapse, 12 no relapse
Simulate another randomization

New Drug

- 17 relapse, 7 no relapse

Old Drug

- 11 relapse, 13 no relapse

\[
\hat{p}_{\text{new}} - \hat{p}_{\text{old}} = \frac{17}{24} - \frac{11}{24} = 0.25
\]
The probability of getting results as extreme or more extreme than those observed \textit{if the null hypothesis is true}, is about .0193.
Flexibility

- I just illustrated the randomization test for a difference in proportions, but the exact same idea holds for other parameters!
In-Class Activity

Does 5 seconds of exercise increase pulse rate?

1. Randomly assign half the students to exercise for 5 seconds, then measure everyone’s pulse
2. Have the students record all the pulse rates on their own sets of index cards
3. Calculate the observed difference in means
4. Have each student randomly split their cards into two groups, calculate the difference in means, and contribute to a class dotplot
5. Use a computer to continue building up the randomization distribution
6. Calculate the p-value
Randomization-Based Inference is useful for teaching statistics...

- The whole idea of a randomization test is centered around the definition of a p-value
  - How extreme would the observed results be if the null hypothesis were true?
  - Can they be explained just by random chance?

- 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!)
"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
“... 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!

lock@stat.harvard.edu

www.people.fas.harvard.edu/~klock