Next “Mission” to Test In the App

“It keeps me from looking at my phone every two seconds.”

Liam Walsh / The New Yorker

The Basics of A/B Testing

Prof. Morten Hansen
MIIC 2014

A/B (Split) Testing

- Inductive vs. Deductive
    - A-B is deductive (hypothesis-driven)
- Major benefit: causation
- Iterative => evolutionary theory based
- Challenge: historically difficult to do, web opened up a whole new possibility

A/B Testing: Key Design Parameters

1. What is your objective? Ideally, some outcome metric (not just website clicks). For example, “maximize donations” or “increase online sales per visitor by 20%.”

2. What are the key means to get to that objective (“bottlenecks”)? For example, “visit to e-mail address to donation” or “website-to-click on product category” or “time to checkout.”

3. What is the key metric? Must be a concrete, measurable, clear outcome! (Example: website visits-to-purchase)

4. What is the population you want to perform the test on?

5. What are the key hypotheses you want to start out with? A = baseline, B = “treatment.” Do you have a theory for “B”? (Example: behavioral principles in this course)

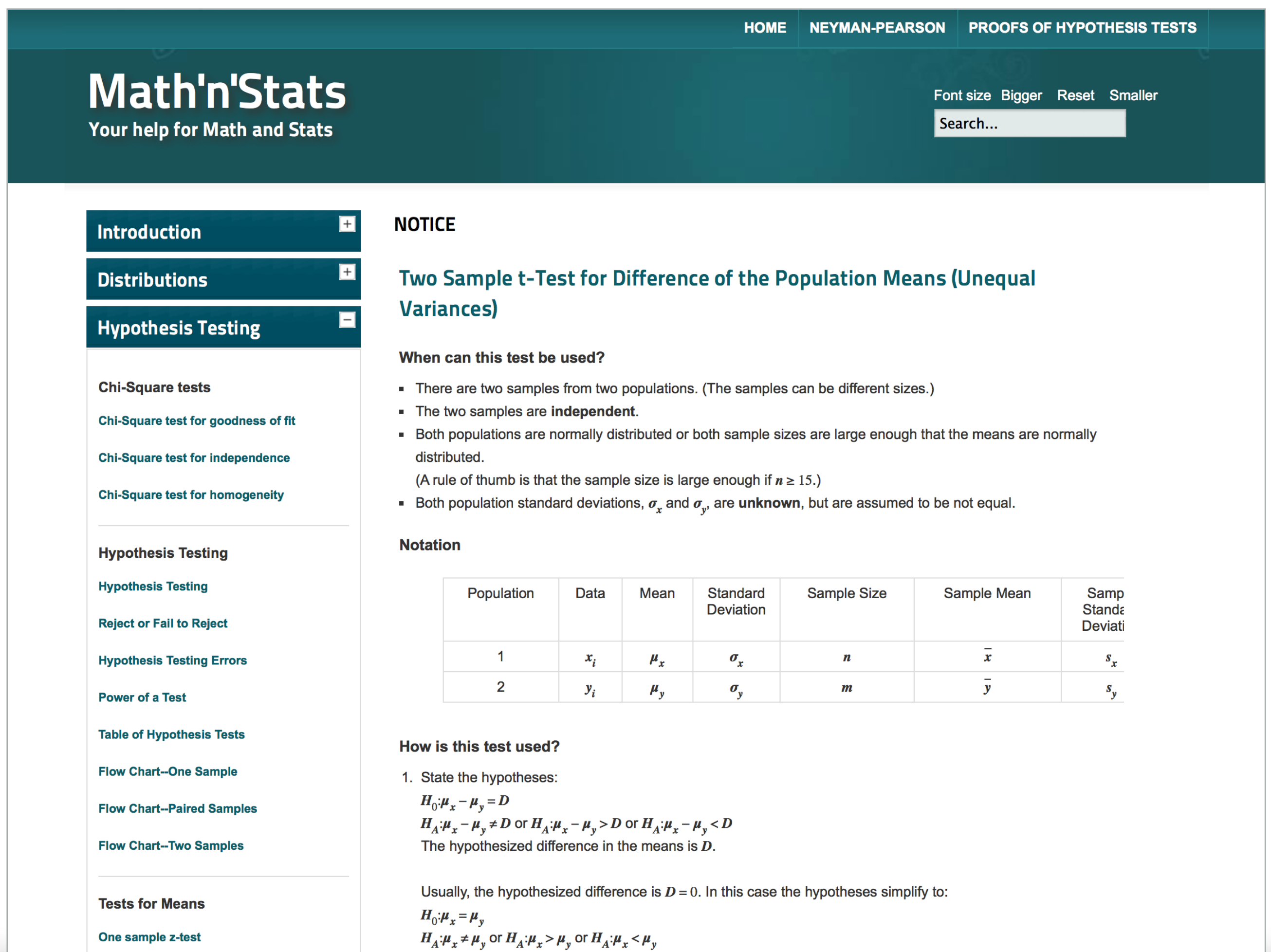
6. You need to have pre- and post-design (you can’t just run “B” - that is, just try a new version)

7. <u>Randomly</u> allocate population members to A and B. Very important!

8. Collect and compare outcome measure: mean and variance.

9. Retain the best one, design new hypothesis (step 5). Keep iterating.

Two Sample t-Test for Mean Difference (Unequal Variance)



http://mathnstats.com/index.php/hypothesis-testing/82-tests-for-means/122-two-sample-t-test-unequal-variances.html

Evolutionary: Variation, Selection and Retention Model

Variation		Selection		Retention
Blind?		Consumer pref?		Data wins?
Guided?	→	Biased?	→	Boss?
Incremental?		Know?		Change?
Big step?				Local optimum?

A/B: Pros and Cons

- Strengths:
  - Double-blind: testing causation, not correlation
  - Data-driven: data better than opinion (of genius' too?)
  - Iterative: can be fast, toward objective
- Limitations
  - No theory for what to test
  - Expensive / impossible outside of the Web
  - Can be very incremental, no major leaps (local optima)