

Next “Mission” to Test In the App



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

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. Randomly 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)

HOME NEYMAN-PEARSON PROOFS OF HYPOTHESIS TESTS

Math'n'Stats
Your help for Math and Stats

Font size Bigger Reset Smaller
Search...

Introduction
Distributions
Hypothesis Testing

Chi-Square tests
Chi-Square test for goodness of fit
Chi-Square test for independence
Chi-Square test for homogeneity

Hypothesis Testing
Hypothesis Testing
Reject or Fail to Reject
Hypothesis Testing Errors
Power of a Test
Table of Hypothesis Tests
Flow Chart--One Sample
Flow Chart--Paired Samples
Flow Chart--Two Samples

Tests for Means
One sample z-test

NOTICE

Two Sample t-Test for Difference of the Population Means (Unequal Variances)

When can this test be used?

- There are two samples from two populations. (The samples can be different sizes.)
- The two samples are **independent**.
- Both populations are normally distributed or both sample sizes are large enough that the means are normally distributed.
(A rule of thumb is that the sample size is large enough if $n \geq 15$.)
- Both population standard deviations, σ_x and σ_y , are **unknown**, but are assumed to be not equal.

Notation

Population	Data	Mean	Standard Deviation	Sample Size	Sample Mean	Sample Standard Deviation
1	x_i	μ_x	σ_x	n	\bar{x}	s_x
2	y_i	μ_y	σ_y	m	\bar{y}	s_y

How is this test used?

- State the hypotheses:
 $H_0: \mu_x - \mu_y = D$
 $H_A: \mu_x - \mu_y \neq D$ or $H_A: \mu_x - \mu_y > D$ or $H_A: \mu_x - \mu_y < D$
The hypothesized difference in the means is D .

Usually, the hypothesized difference is $D = 0$. In this case the hypotheses simplify to:
 $H_0: \mu_x = \mu_y$
 $H_A: \mu_x \neq \mu_y$ or $H_A: \mu_x > \mu_y$ or $H_A: \mu_x < \mu_y$

- Pick a significance level, α .

Evolutionary: Variation, Selection and Retention Model

Variation

Blind?
Guided?
Incremental?
Big step?



Selection

Consumer pref?
Biased?
Know?



Retention

Data wins?
Boss?
Change?
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)