7.1 DESIGNING STUDIES

DO YOU LIKE MY INTERFACE?

- WHAT is COMPARISON?
- WHAT is THE YARD STICK?

HOW TO IMPROVE

- BASERATES
  how often does Y occur?
- CORRELATIONS
  do X and Y co-vary?
- CAUSES
  does X cause Y?

SKETCHNOTESPACE.COM
ANNA IURCHENKO

MANIPULATION
- diff conditions
- independent variables

MEASURES
- dependent variables
- accuracy
- recall
- emotional response

DON'T DEPEND ON PARTICIPANTS, THEY ARE SET BY EXPERIMINTER

GENERALIZABILITY
- external validity
- does this apply to this particular users?

PRECISION
- INTERNAL validity
- if you ran this again will you see the same results
- # of people

„CONTROLLED COMPARISON ENABLES CASUAL INFFRENCE"

IS MY NEW APPROACH BETTER THAN THE INDUSTRY STANDARD?

STRADEGIES FOR FAIRER COMPARISONS

- insert your new approach into production setting
- scale things down so you're looking at a piece of a larger system
- train people up
- make a version of a production thing in the same style as new approach

← you can learn stuff ☺

7.2 ASSIGNING CONDITIONS

SKETCHNOTESPACE.COM
ANNA IURCHENKO

Vaccum type (manipulation)
Speed, cleanliness (measurement)

personal difference ??
(USE ONE OF THE DEVICES)

- Faster?
- Cleaner
- Fatigue
- Ergonomic
- Ease of use
- ...

Between Subject Design
Within Subject Design
ordering ??
(use both cleaning device)

THREE or MORE ALTERNATIVES

1	2	3
2	3	1
3	1	2

LATIN SQUARE

- each person will use all three conditions, order is changed

SHOULD EVERY PARTICIPANTS USE EVERY ALTERNATIVE

1. WITHIN SUBJECTS
(everyone tries all the options)
- not worry about learning

2. BETWEEN-SUBJECTS
(each person tries one)
- more people, attention to fair assign.

3. COUNTERBALANCING
- minimize variatio in a between subject design

ASSIGNMENT SHOULD BE RANDOM (HAWTHORNE EFFECT)

A	B
35	32
59	57
⋮	⋮

4

(TYPERS EXAMPLE)

REGRESSION is a DANGER

COUNTERBALANCE ASSIGNMENT
- USE PRE-TEST
- EACH PARTICIPANT HAS AN EQUAL CHANCE OF LANDING IN EITHER CONDITIONS

- DAVID MARTIN
DOING PSYCHOLOGY EXPERIMENTS

MAKE CLEAR GOALS
- narrow a scope to the purpose of your study
- WHO, WHERE WHEN
- SCENARIO (REALISTIC)

Question | Data to collect | Setup

7.3 IN-PERSON STUDIES

- Planning
- EXECUTING
- ANALYSING

CONCRETE TASKS
- WRITE THEM DOWN
- THINK OF ORDER
- WHAT TO DO IF USER CANNT ACCOMPLISH?
- TRAINING?

REMIND USERS
You are testing the site, not the users

PILOT
↓ 1 COLLEAGUES
↓ π REAL USERS (ONE)

REVEAL THE KINKS OF THE STUDY DESIGN

CAPTURING
THINKING ALOUD
GREETING P.
DEBRIEF WERE
- WHAT YOU DOING

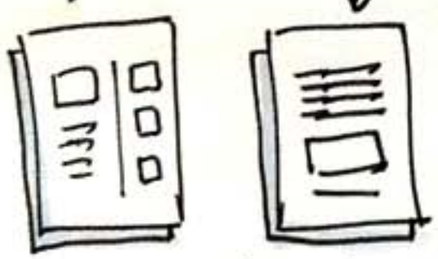
7.4 RUNNING WEB EXPERIMENTS

ROLL OUT DIFFERENT VERSIONS OF A USER INTERFACE, GET FEEDBACK, ITERATE QUICKLY

DUSTIN CURTIS [dcurt.is BLOG]

AB

50% 50%



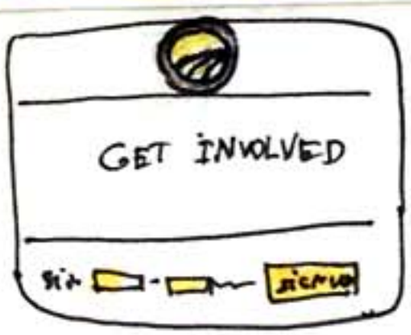
METRICS? CTR?

MULTIVARIATE TEST (variations of two diff. parts of the page)

WAYS DESIGN MAKES A DIFFEREN.

- position & color of call to action
- whitespace?
- position of testimonials
- position of heading
- number of columns
- # of visual evement competing for attract.
- the age, sex, appearance of people on the photo

CALL TO ACTION EXPERIM.



Baseline 8,2%

Variations 7,5% 8,91%

- button (sign up, learn more ...)
- MEDIA (family img, change img, obama img ...)

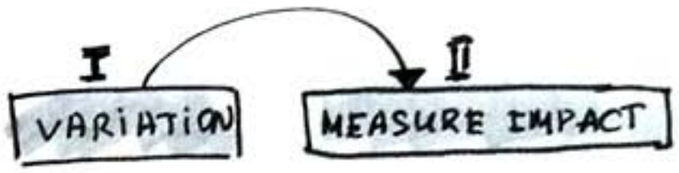
Small changes, HUGE IMPACT (text on the button)

TYPOGRAPHY EXPERIMENT

Ron Kohavi

color change on MSN search

↑3% (AD) conversion RATE?



SMALL DISTRACTIONS (extra fields) CAN YIELD BIG CHANGES

COMMITMENT ESCALATION

BY COMBINING ITERATIVE DESIGN AND CONTROLLED EXPERIM.

YOU CAN DIAL IN THESE PHENOMENA TO ↑ EFFECT

PRINCIPLES

0,1%

RAMP-UP & AUTO ABORT (simple analyses to find egregious problem)

EQUAL AMOUNT OF PEOPLE

MEASURE THINGS THAT MATTER

RUN IT LONG ENOUGH

1. CONSISTENT
2. DURABLE
3. INDEPEND.

rules of random assignments

DESIGNER ROLE SHIFTS TO BEING ABOUT CREATING MULTIPLE ALTERNATIVES

SKETCHNOTESPACE.COM
ANNA IURCHENKO

7.5 ANALYZING EXPERIMENTS

COMPARING RATES

ANALYZE DATA IN 3 STEPS

1. HOW IT LOOKS LIKE ? (plot your data
2. OVERALL NUMBERS (deviation, average)
3. IS THE DIFF. „REAL"? (compute significance)

PEARSON'S CHI-SQUARED TEST

compare the rater of an expected value to an observed value

$$X^2 = \frac{(\text{OBSERVED} - \text{EXPECTED})^2}{\text{EXPECTED}}$$

Sign Up

IMPROVED CLICK-THROUGS

[example]

- 10% CTR
- CHANGE TEXT
- 1 week, ~~119~~ clicked "Learn More" out of 1000

Learn More

YES

? CAN WE SAY WITH CONFIDENCE THAT „LearnMore" have ↑ CTR

$$\frac{(119-100)^2}{100} + \frac{(881-900)^2}{900} = 4{,}01$$

$df = 1$ (degree of freedom) $p \leq 0{,}05$

STATISTICAL TESTING

formalise "we're pretty sure"

helps generalize from small samples

1908, WILLIAM GOSSET „Guinness"

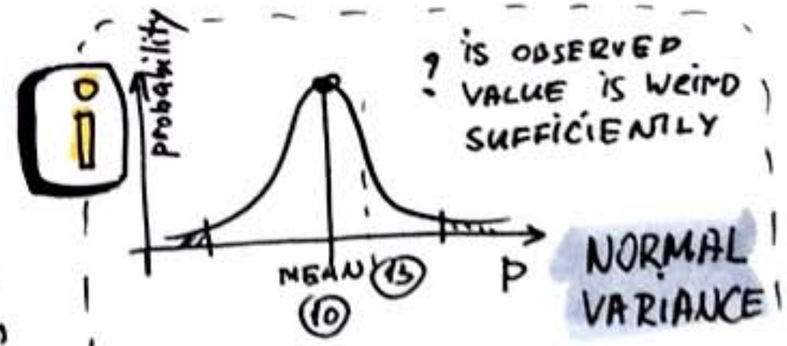
T-test (small sample) 2 conditions

ANOVA (compare > 2 conditions)

Data often ain't „Normal"

- A/A tests
- randomised testing

BIMODAL



NORMAL VARIANCE

THE NULL HYPOTHESIS

- our opening bid in any stat. test is that there is no relation between measured phenomena

- Practical statistic for HCI, Jacob Wobbrock, depts.washington.edu
- Doing Psychology Experiments, David Martin
- Statistics as Principled Argument, Robert P. Abelson
- Learning to use stat. test, Judith Green

SKETCHNOTESPACE.COM
ANNA IURCHENKO