# How to Do Experiments

David W. Martin, 2008

Goals

- Understand and discuss hypothesis testing

- Discuss random variables, statistical significance, and statistical regression

- Design an experiment as a class

Testing a Hypothesis

The world is noisy and no experiment will be perfect, so we must select what conditions to control and how.

X - Independent
Y - Dependent
M - Manipulation
CF - Confounding
C - Control
R - Random
CR - Constrained
      Random
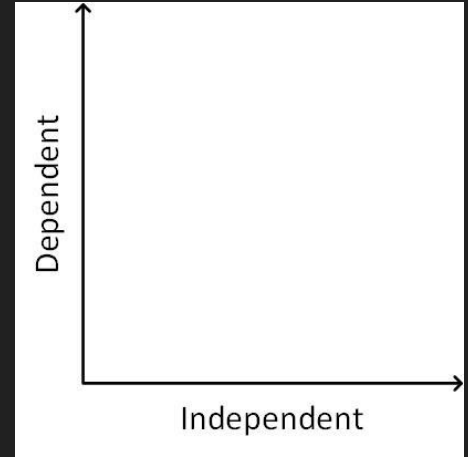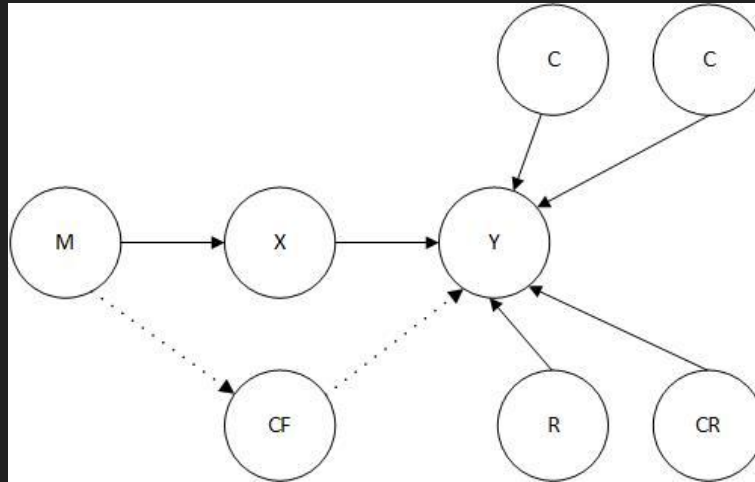
* Image Credit: Xinyu Zhang

Random Variables

Why wouldn't we want to deterministically control all of the parameters of an experiment?

Random Variables

Total control leads to poor generalization

"The author makes an interesting statement about producing random results… 'Humans are notoriously bad at producing random events'. There is considerable debate surrounding this topic. Are humans inherently bad at random event generation or have we evolved from children to adults with this negative ability?" -- Calvin Gomez

"The author mentions this in the closing discussion about random variables - "new experimenters commonly make mistakes in randomization". This again calls for a discussion on some sort of heuristics to pick experimental variables" -- Toushar Koul

Testing a Hypothesis

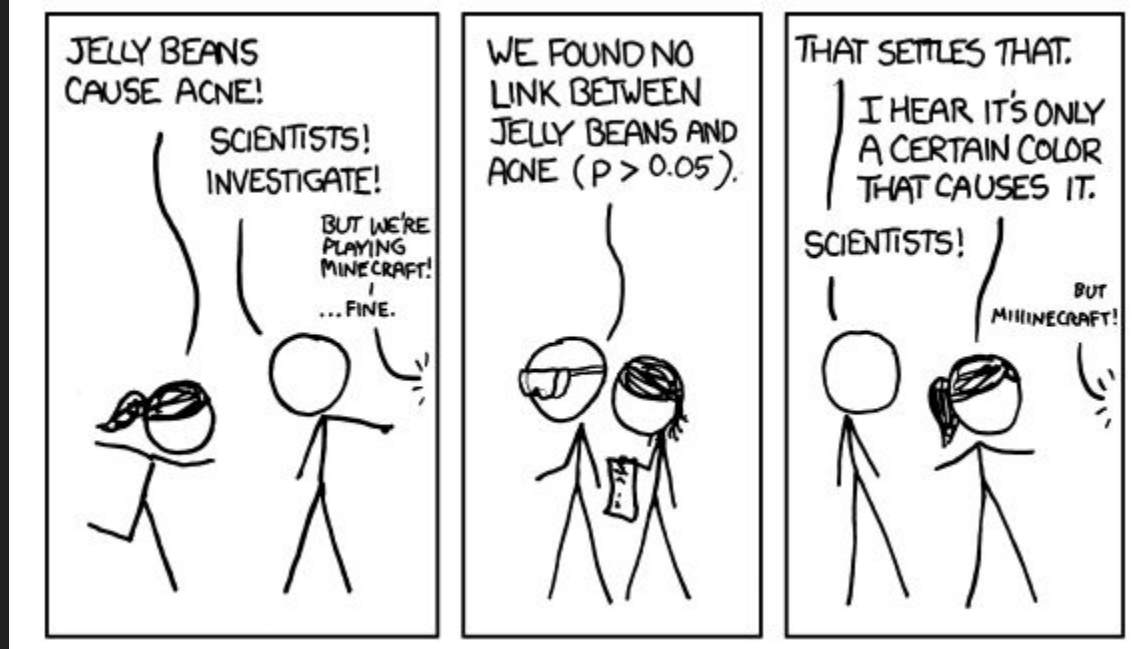How can we justify that our hypothesis is true? What might be some good criteria?

"...how does one make appropriate conclusions when there are so many possibilities? Given the time and resources available, it seems satisficing is the way to go, and I wonder how it can be done." -- Kandarp Khandwala

Statistical Significance

- Null hypothesis vs. Alternative hypothesis
    - Null hypothesis -- there is no relationship between X and Y
    - Alternative hypothesis -- the posed claim of a relationship (the research hypothesis)
        - Non-directional vs. directional
- Potential criteria:
    - p-test: The probability of obtaining your result given that the null hypothesis is true
    - α: The probability of your study rejecting the null hypothesis given that the null hypothesis is true
- Results can be considered "significant" if $p < \alpha$ (0.05, 0.01, 0.001, for lower tailed p-test)

What might be some potential problems with the p-test? Discuss ~1m

Statistical Significance

Credit: xkcd (https://xkcd.com/882/)

Statistical Significance

Statistical Significance

('So, uh, we did the green study again and got no link. It was probably a--' 'RESEARCH CONFLICTED ON GREEN JELLY BEAN/ACNE LINK; MORE STUDY RECOMMENDED!')

Credit: xkcd (https://xkcd.com/882/)

Statistical Significance

p-hacking:

http://io9.gizmodo.com/i-fooled-millions-into-thinking-chocolate-helps-weight-1707251800

"...the study was 100 percent authentic. My colleagues and I recruited actual human subjects in Germany. We ran an actual clinical trial, with subjects randomly assigned to different diet regimes. And the statistically significant benefits of chocolate that we reported are based on the actual data...And for reporters who don't have science chops, as soon as they tapped outside sources for their stories—really anyone with a science degree, let alone an actual nutrition scientist—they would discover that the study was laughably flimsy."

Threats to Internal Validity

History

Maturation

Selection

Statistical Regression

Mortality

Testing

Interactions with Selection

Discussion question: What is statistical regression? Give an example.

"Assume the sample in the study was selected exclusively from the low pretest scorers. Then, when we conduct posttest on the same sample, the resulting mean will appear to regress towards the mean of the population, namely higher than previous mean score in pretest. Same phenomenon happens in the other end, the high pretest scorers." -- Chenxi Zheng

"An example would be that say I went bowling with some friends last Saturday. It was an amazing day for me and I got 12 strikes, which is a 300 score, a perfect game. It was amazing, but that is no excuse for me to quit my job and become a Pro Bowler. It is likely that the next time we go bowling, I will only get an average 150 score. This is what we call regression to the mean." --Daniel Pan

Threats to Validity: Statistical Regression
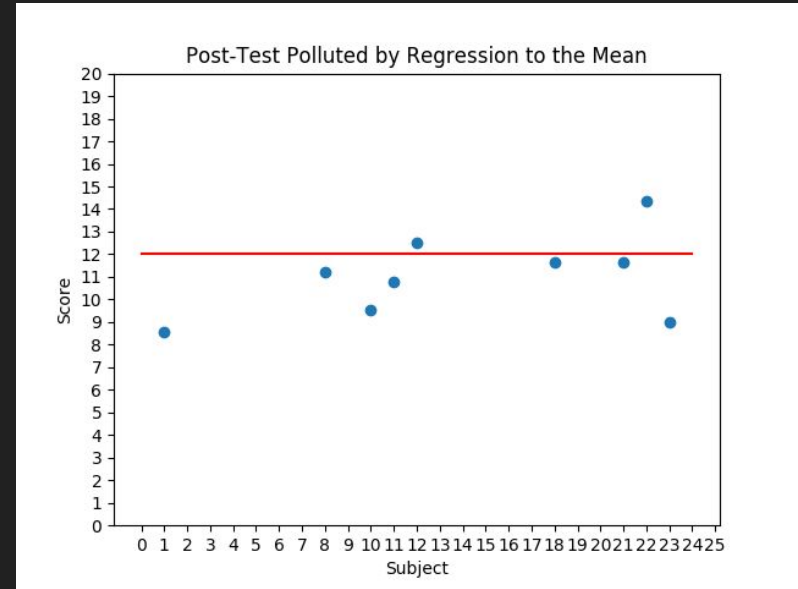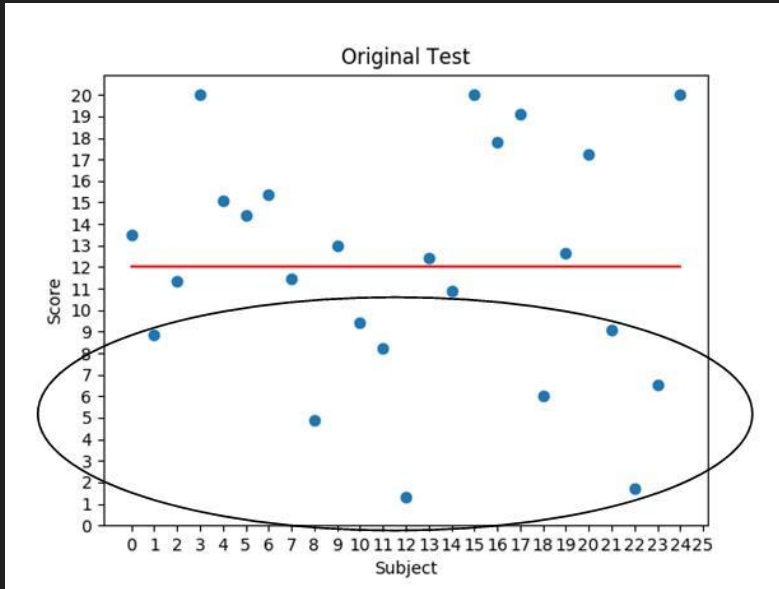
Exercise!

"Using font X in news articles leads to increased retention of material"

How might we test this hypothesis?

What are our variables (independent/dependent/control/random)?

What are some looming threats to our experiment (confounding variables/validity)?

Discuss with your partner ~2min

Prototyping
Tools and Techniques

Michel Beaudouin-Lafon, Wendy Mackay in Human Computer Interaction Development Process, *2003*

Objectives

- Understand prototypes
  - Definition
  - Analysis of the role played in design
  - Identify approaches

Prototypes

- Concrete (tangible artifact) representation for a concept
- Basis for future systems
- Allows for interaction in comparison with abstract representations that are open to interpretation

Frame taken from Mercedes-AMG youtube channel

Discussion Question

How would one make use of a prototype?

What sort of information becomes available through prototyping?

Analysis of Prototypes

- Four dimensions
  - Representation - the form (e.g. online, offline)
  - Precision - the level of detail (e.g. dialogue box containing words)
  - Interactivity - extent that a user can interact (e.g. fixed, open)
  - Evolution - life cycle (e.g. rapid, iterative)

Discussion Question

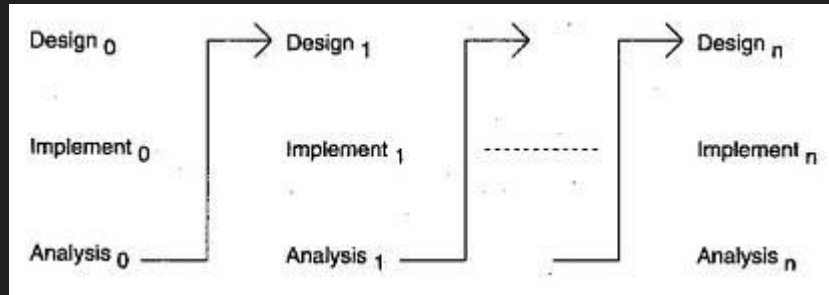What other dimensions, if any,

can be used to qualify prototypes?

Rapid Prototyping

- Develop prototypes with the goal of speed in mind
- Explore many different avenues of designs (examples?)
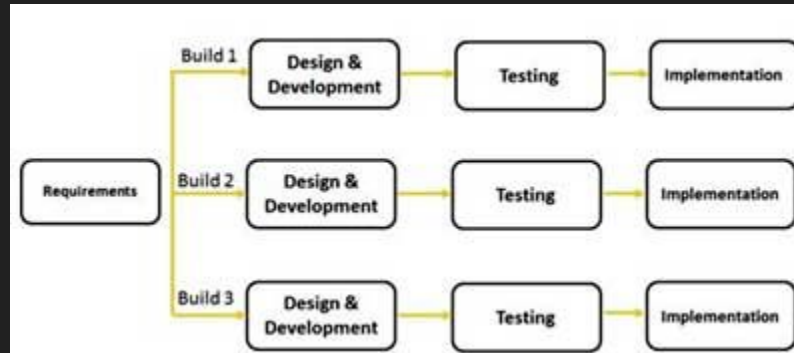
Iterative Prototyping

- Reflection of design in progress
- Explore variations and unforeseen occurrences
- Refine precision in details as well as interactions

Evolutionary Prototyping

- A form of iterative prototyping mainly applicable to software
- Prototype evolves into the next part of the system
- Pick working ideas to move into the final system
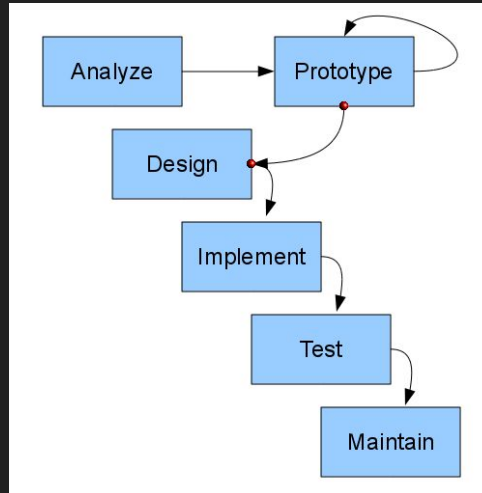
What is different between iterative and evolutionary prototyping? What is similar between the two?

"Iterative prototyping consists of constantly going back to improve on a prototype, while evolutionary prototyping is iterating on a prototype while simultaneously grooming it to become the final product."

"Evolutionary prototypes intended to evolve into the final product require more traditional software development tools"

"Evolutionary prototypes are a specific subset of iterative prototypes… applicable only to software prototypes"

"[They are] closely related in that they keep producing better and better prototypes"

Application

What sort of prototyping would be applicable to your research project? Could you define your prototype in another way?

In what ways might prototyping be useful for answering your research question?

Prototypes

- Provide insight to how a system would look and feel
- Rich for experimentation since they are concrete
- Part of a design process in which ideas are generated and expanded upon