What is involved (generally)

- Formulating a hypothesis
- Designing a test plan
  - Picking a method
  - Selecting users
  - Writing out procedure
- Getting IRB permission
- Dealing with users
- Dealing with data
Hypothesis Testing

What is a Hypothesis?

Why do hypothesis testing?

• What kinds of hypotheses do we usually test?
  – X is better/larger/faster than Y
  – X improved more than Y
Goals of Good Study Design

• Testing a hypothesis
  – Clearly state a research question
  – Identify a method for measurement of a single dependent variable as a number of independent variables are manipulated
  – Control for all other extraneous variables if possible

• Issues to deal with:
  – Selectivity of study method
  – Fidelity of research method
  – Quantitative vs. Qualitative data
  – Control vs Ecological validity
  – Cost vs. Relevance

Hypothesis Testing

1. Specify null hypothesis (H0) and the alternative hypothesis (H1). For example H0: \( \mu_1 - \mu_2 = 0 \) and H1: \( \mu_1 =/\neq \mu_2 \).
   
   Remember, define so that H1=true iff H0=false

2. Select a significance level. Typically \( P=0.05 \) or \( P=0.01 \)
3. Sample population and calculate statistics.
4. Calculate the probability (p-value) of obtaining a statistic as different or more from the norm.
5. Compared the p-value with the chosen significance level. If \( P<=\text{Significance level} \), reject null hypothesis

That \( P \) is greater does not mean H0 is true! Just means we can't tell that it is false.

When the null hypothesis is rejected, the result is "statistically significant." When we can't reject null hypothesis, result is not statistically significant.
**Within-subject vs. Between-subject Design**

- Repeated measures vs. single sample (or low number of samples)

- Are we testing whether two groups are different (between subjects), or whether a treatment had an effect (within subject)?
  - Between subjects we typically look at population averages
  - Within subjects we typically look at the average change in subjects (analysis of variance)

**Within-subject or Between-subject Design**

- Within-subject design
  - Cheap, fewer subjects, more data
  - Removes individual differences
  - Introduces learning and carryover effects
  - Can’t use the same tests as on between subjects because the observations are no longer independent
Controlled Studies

Controlled Experiments

You’re not paranoid, they really are watching you

- Isolate confounding variables, level playing field
- Removes “noise” from data
- Thoroughly documented
- Focus only on your question
What can we observe?

• Reaction time
• Completion time
• Work patterns
• Success/Failure rate
• Errors
• Emotional responses
• Satisfaction
• Reasoning

Equipment

• Multiple cameras
• Multiple monitors
• Recording systems
• Screen recorder
• Event logger
• Audio