Research Methods
&
Experimental Design

16.422 Human Supervisory Control

April 2004

Research Methods

- Qualitative vs. quantitative
- Understanding the relationship between objectives (research question) and variables is critical
- Information $\neq$ Data
  - Information=data + analysis
- Planning in advance is a must
  - To include how data will be analyzed

Qualitative Research Methods

- Social & cultural phenomenon
- Case studies
- Focus groups
- Observations
- Usability testing
  - Can be quantitative
- Interviews
- Questionnaires

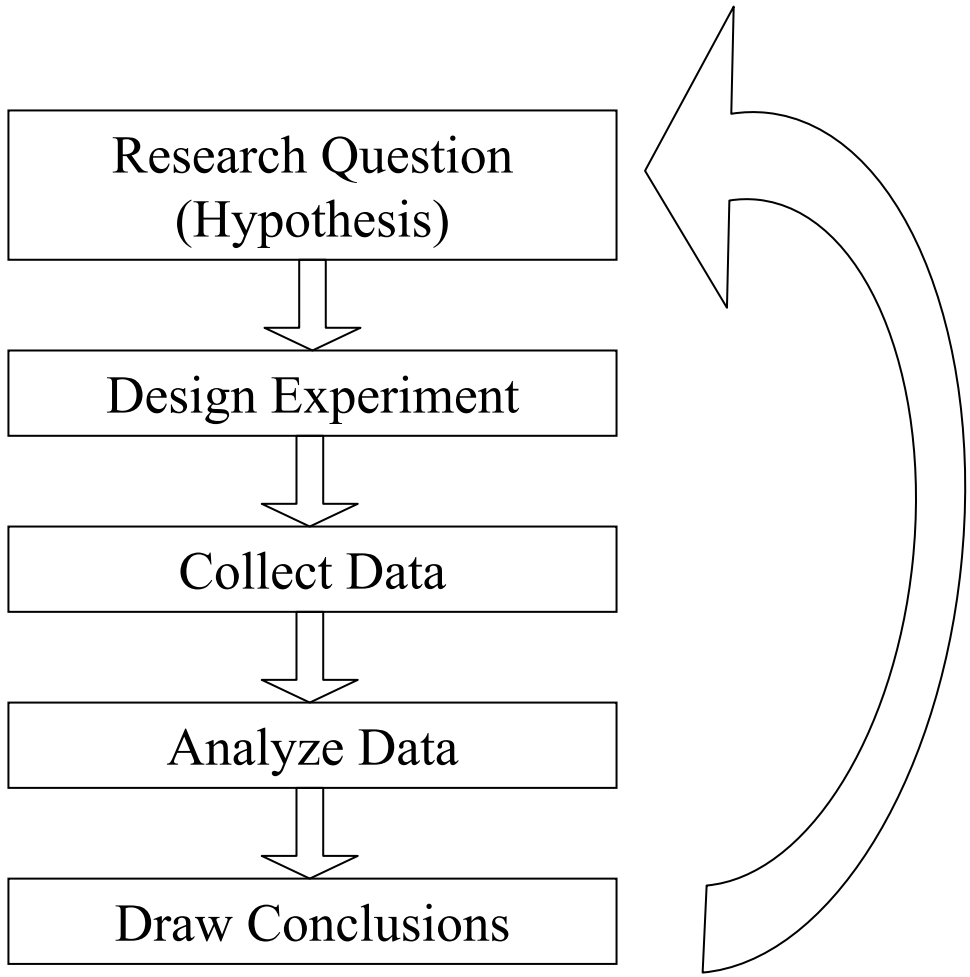
Quantitative Research Methods

- Natural phenomenon
- Mathematical modeling
- Experiments
- Optimization
- Game theory
- Surveys
- Bottom line – statistics are a must

Project Assignment

- Design and conduct an experiment in which you explore some measure of human performance through testing, analyze the results, and discuss the broader implications.
- Design an actual display that uses automation for decision support… While formal experimental testing is not required, a small group of users should be used to identify problems with the design to include functionality evaluation as well as recommendation for future improvements and systems integration.

The Experimental Design Process



Experimental Design

- Design of Experiments (DOE) defined:
  - A theory concerning the minimum number of experiments necessary to develop an *empirical* model of a research question and a methodology for setting up the necessary experiments.
  - A parsimony model
- Human subject vs. object experimentation
- Other DOE Constraints
  - Time
  - Money

Experimental Design Basics

- Two kinds of data gathering methodologies
  - Observation
    - Can't prove cause & effect but can establish associations.
    - Hawthorne effect, social facilitation
  - Experimental
    - Cause & effect
    - Variables of interest – factors vs. treatments
      - Independent variable
        - Treatment – manipulations of variables of interest
        - Treatment vs. control group
      - Dependent variable is what you are measuring

More Basics

- Confounds
- Randomization Concerns
  - Randomization prevents experimental bias
    - Assignment by experimenter
      - Counterbalancing
  - Statistical assumptions
    - A requirement for statistical tests of significance
- Why would you use the observation methodology instead of experiments?

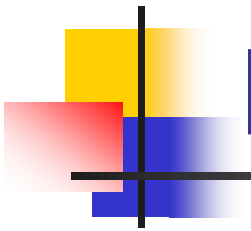
DOE Terminology

- Replications
  - Independent observations of a single treatment.
- Variance
  - The measuring stick that compares different treatments.
- Internal validity
  - The extent to which an experiment accomplishes its goal(s).
- Reproducibility
  - Given the appropriate information, the ability of others to replicate the experiment.

DOE Terminology (cont.)

- External validity
  - How representative of the target population is the sample?
    - Can the results be generalized?
    - Generalizations for field experiments are easier to justify than lab experiments because of artificialities.
- Medical Trials
  - Placebo
  - Double Blind

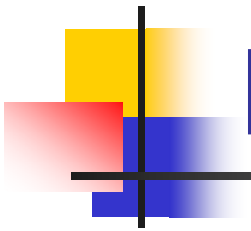
• If so, what is the population to which it can be generalized?
• Can the results be generalized to the real world?



Data Analysis

- Data Types
  - Variables
    - Categorical
    - Numerical
  - Scales of Measurement
    - Nominal
    - Ordinal
    - Interval
- Computer Programs
  - Excel, SAS, S+, SPSS

ANOVA
Within group variance is noise and between group variance is information we seek. ANOVA separates these out.



Basic Statistical Tests

- Assumptions for comparison of means
  - Independent & random
  - Normality
  - Variances roughly equal
- t-tests
  - One or two samples
- Chi-square tests
  - NID(0,1)
  - Categorical data, non-parametric

Chi square important because any sum of squares in normal random variables divided by the variance is chi-square distributed

Null Hypothesis: H_o

- Defined: The difference in two different populations parameters is 0.

$$H_o: \mu_1 = u_2$$

$$H_a: \mu_1 \neq u_2$$

- H_o: Always predicts absence of a relationship & assumed to be true.
- If the null hypothesis is NOT rejected, we CANNOT conclude that there is no difference, only that the method did not detect any difference.
- $p < .05$????

A Very Important Research Question

- Does drinking cappuccino one hour before a test improve results?
  - What is the metric (dependent variable)?
- Experimental Design
  - Treatment group vs. control group
    - A single comparison
  - Experimental efficiency
    - Perhaps we want to look at who makes the cappuccino (Seattle's, Starbucks, Pete's) as well as the difference between coffee and cappuccino.
    - 2X3 Factorial
    - Interaction effects

Caffeine/Performance Experiment

	GB	SB	ER
Capp			
Coffee			

We now know the general layout of the experiment – but what is missing?

Caffeine/Performance Experiment

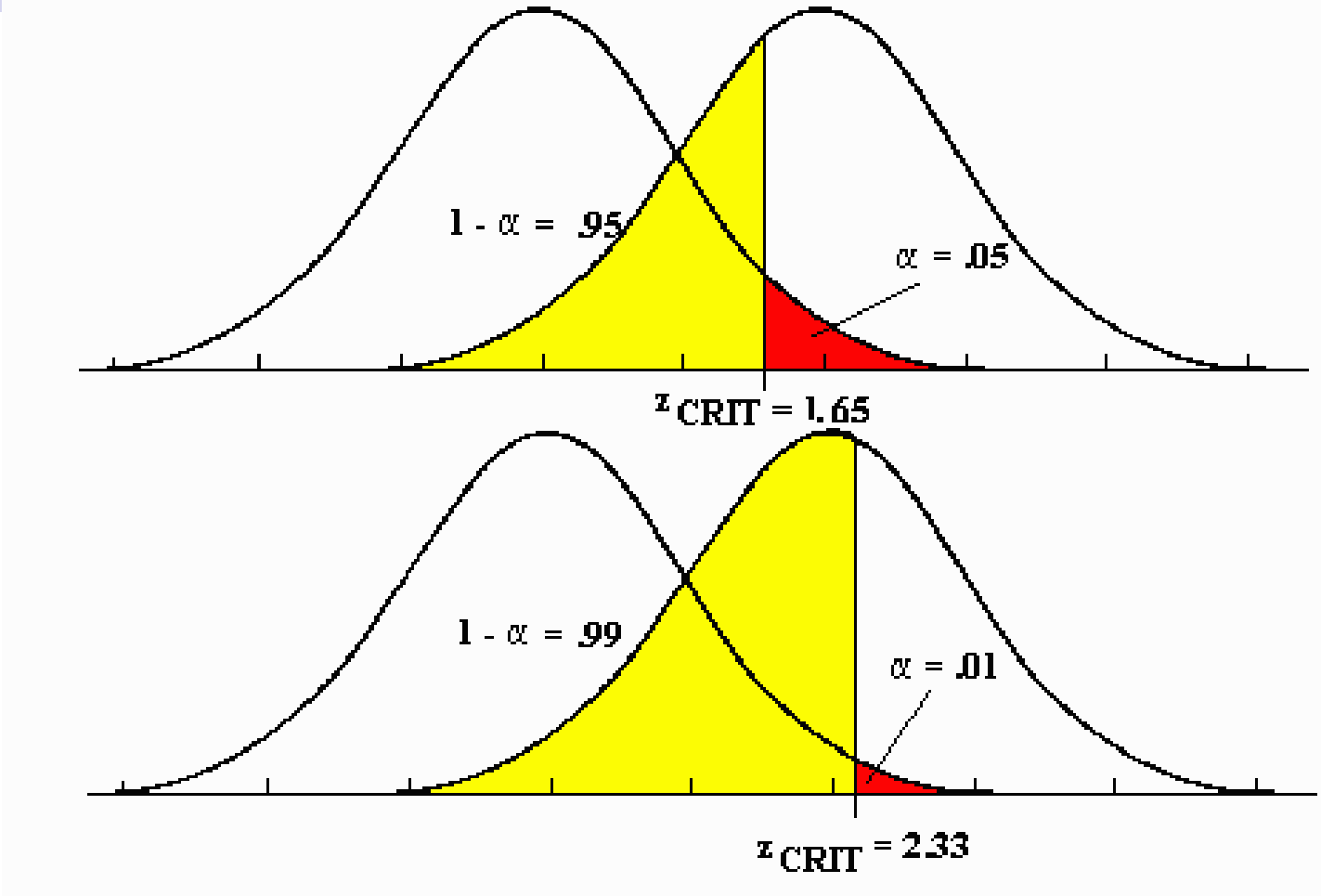
- How many subjects do we need?
  - Sample Size
    - Related to power – the complement of a Type II error…

Decision	H_o True	H_o False
Reject H_o	Type I error $p = \alpha$	Correct decision $p = 1 - \beta =$ **POWER**
Fail to reject H_o	Correct decision $p = 1 - \alpha$	Type II error $p = \beta$

Ask what Ho is?

Null hypothesis – no significant difference exists between experimental groups.

Don’t Panic…



Caffeine/Performance Experiment

- So how do you determine sample size?
  - http://members.aol.com/johnp71/javastat.html
  - Sensitivity is an issue
  - # of factors influences sample size
- Recruitment Issues
  - Population selection
  - How do we assign subjects to treatment categories?
- Confounds
  - Experience
  - Self-selection
  - Control techniques

Other Subject Considerations

- What is the most efficient way to use human subjects?
  - Between subjects
  - Within subjects
    - Repeated measures
    - Increases power but…
    - Confounds – practice & fatigue
    - Counterbalance
  - Mixed subjects
    - Pre-test/post-test
    - Tests over time

Pre/post Test Considerations

Between Subjects

	Pre-Test	Post-Test
Intervention A		
Intervention B		

Within Subjects

- Ideally pre-test scores will be equivalent
- You want to see a difference between the experimental and control group.

Statistical Tests (cont.)

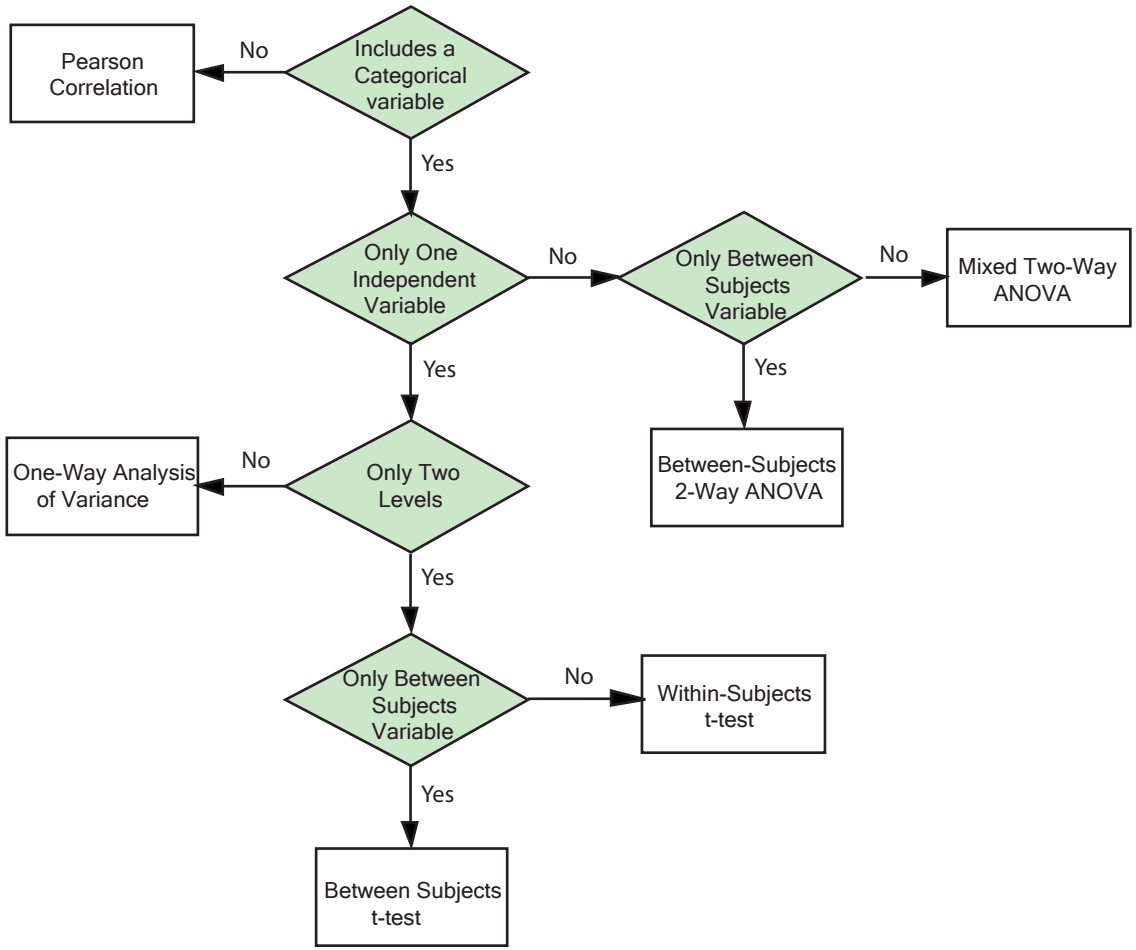
- Analysis of variances (ANOVA)
  - Testing the differences between two or more independent means (or groups) on one dependent measure (either a single or multiple independent variables).
  - One way vs. factorial
  - F test – ratio of variances
  - MANOVA

Other DOE considerations:

- Full Factorial
- Blocking
  - More homogenous grouping
    - Coffee of the day v. another kind
    - Starbuck's at the Marriott vs. Galleria
- Pairing
  - Increases precision by eliminating the variation between experimental units
  - Randomization still possible
- Many others…

• Full factorial – should be run twice
• Tennis shoe example – try to find out which sole is better for shoes so each boy wears two different shoes. Randomization comes in assigning which shoe to which foot.

What test to use?



Adapted from University of Maryland Psychology World.

Example Experiment

- Are web-based case studies better than print versions.
  - How can we test this?
- This question was tested with 2 classes with 2 different professors.
  - What are the independent & dependent variables?
  - Was it within/between/mixed?
  - What statistical test should we use?

Results

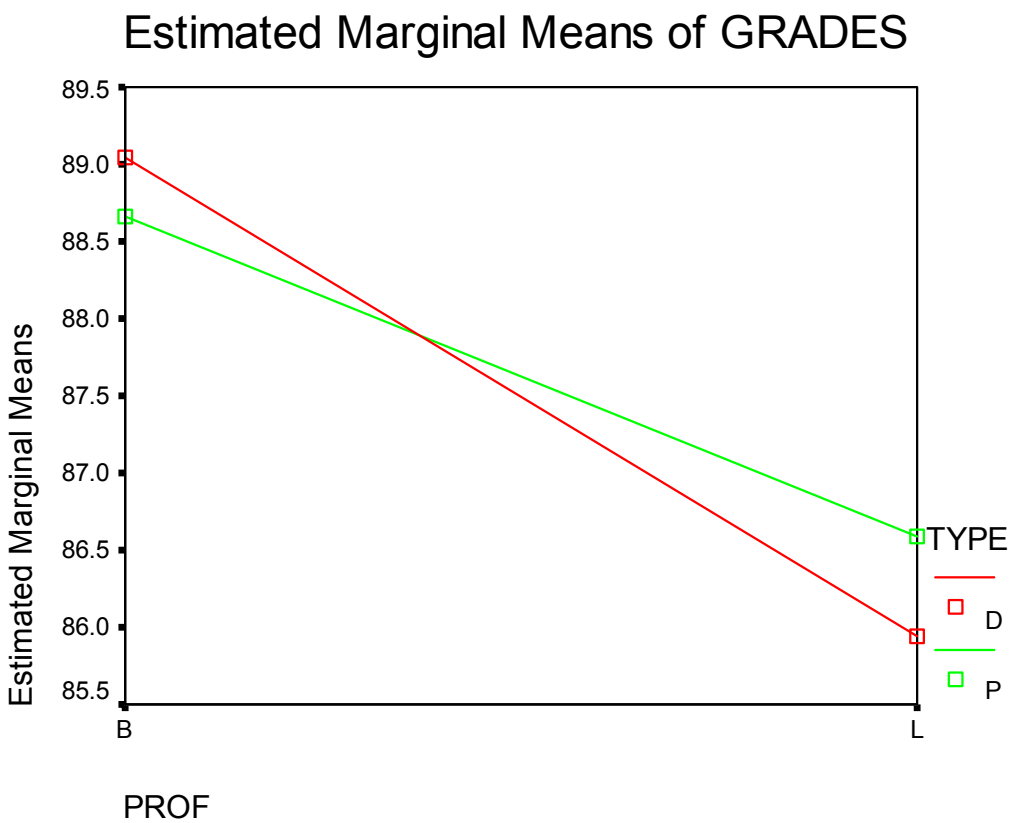
Tests of Between-Subjects Effects

Dependent Variable: GRADES

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	173.681[a]	4	43.420	.986	.420
Intercept	190832.489	1	190832.489	4333.757	.000
PROF	157.697	1	157.697	3.581	.062
TYPE	26.217	2	13.109	.298	.743
PROF * TYPE	11.840	1	11.840	.269	.605
Error	3654.818	83	44.034		
Total	673001.300	88			
Corrected Total	3828.499	87			

a. R Squared = .045 (Adjusted R Squared = -.001)

Interactions



- Interaction effect: the response of one variable depends on effect of another variable
- No interaction – parallel lines
- Significant interaction:
- Which professor would you rather have?

Non-Parametric Tests

- Use when you have no good information about an underlying distribution
- Parametric tests:
  - Parametric form - parameters either assumed to be known or estimated from the data
    - The mean and variance of a normal distribution
  - Null hypothesis can be stated in terms of parameters and the test statistic follows a known distribution.
- Non-parametric tests are still hypothesis tests, but they look at the overall distribution instead of a single parameter
- Particularly useful for small samples

All data is not normal….

Parametric

- Correlation & Association
  - Pearson
- T-tests
  - Independent & dependent
- ANOVA
  - Factorial
  - Repeated measures
  - MANOVA
- Linear regression

Non-parametric

- Association
  - Spearman
- Chi-Square
  - Contingency tables
- Kruskal-Wallis test
- Sign-test
- Friedman ANOVA
- Logistic regression