Planning a statistically sound research project.

Rick White

October 28, 2015

Outline

- Types of research studies
- Planning your research study
- Sample size calculations
- Experimental design
- Types of analysis
- Multiple comparisons.

Types of research studies

- Sample Surveys
  - Information about a population
- Observational Studies
  - Interested in comparing two or more groups
  - Assignment of groups is not controlled
- Controlled Experiments
  - Interested in comparing two or more groups
  - Group membership is assigned randomly

Sample Surveys

- Characteristics of interest
- Target Population of size N (usually finite)
- Random sample of size n from the population
  - yields data $x_1, \ldots x_n$
- Use data to make inference about the characteristics of interest
  - compute sample (point) estimates
  - compute error associated with those estimate
  - compute interval estimates with desired probability coverage
- sampling fraction could be important n/N

Issues with Sample Surveys

- Establishing the sampling frame.
- Over-coverage or Under-coverage.
  - Is the entire population in the sampling frame?
  - Are subjects outside the population in the sampling frame?
- Selection bias.
  - is the sample representative of the population?
- Non-response bias.
  - Do non-responders differ from responders?
- Wording and order of the questions

Roosevelt vs Landon: 1936

- Election Poll by Literary Digest during the Great Depression
- 10 million sampled from their readers, telephone users, and automobile owners
- n = 2.4 million responded (huge sample size)
- Poll result:
  - 57% for Landon, 36% for Roosevelt
- Actual Election Result:
  - 36.5% for Landon, 60.8% for Roosevelt
- Discussion point: Why do you think Literary Digest was so inaccurate?

Observational Studies

- Characteristics of interest compared between 2 or more distinct populations
  - Male vs Female, exposed vs unexposed
- Random sample from each population
  - often based on who's available
- Use data to compare the characteristics of interest between populations
  - compute sample estimates and their errors
  - compare the sample estimates
- Matching is sometimes used if one population is substantially larger than the other.

Issues with Observational Studies

- Confounding is a major issue
  - Differences could be due to another factor that differs across the groups
  - Analyses must adjust for such confounders
  - Extreme cases can lead to Simpsons paradox
- Analyses can suggest, but not establish causality
  - Bradford Hill's criteria for causality

Confounding Example

This is a real-life example from a medical study comparing the success rates of two treatments for kidney stones. The table below shows the success rates and numbers of treatments for treatments involving both small and large kidney stones, where Treatment A includes all open surgical procedures and Treatment B is percutaneous nephrolithotomy. (Wikipedia)

Size	Treatment A	Treatment B
Small Stones	**93% (81/87)**	87% (234/270)
Large Stones	**73% (192/263)**	69% (55/80)
Combined	78% (273/350)	**83% (289/350)**

Controlled Experiments

- Characteristics of Interest compared between 2 or more groups.
- Random sample is drawn from the study population.
  - sometimes from several populations if some observational factors are to be controlled.
- Subjects are randomized to groups and data is collected.
  - must be able to control the group assignment.
- Use data to compare the characteristics of interest between groups
  - compute sample estimates and their errors
  - compare the sample estimates

Issues with Controlled Experiments

- Ethics of experimentation.
  - Cannot force someone to smoke
- Choice of control: nothing, placebo or active.
  - the placebo effect
  - ethically must treat if one already exists
- Blinding of subjects and researchers.
  - subconscious effects

The discipline of Statistics

- Wikipedia: Statistics is the study of the collection, organization, and interpretation of data.
- American Heritage Science Dictionary: The branch of mathematics that deals with the collection, organization, analysis, and interpretation of numerical data.
- Merriam-Webster Dictionary: A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.
- Dictionary.com: the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data, . . .

Planning your Study

Before you collect any data

- Define the questions of interest.
- Determine the appropriate populations that will allow the questions to be answered.
- Create a plan to sample the populations and a randomization scheme if required.
- Determine what information (data) is needed from the sample to answer the questions.
- Create an appropriate analysis plan.
- Determine the sample size needed.

Defining the Question

What do you what to know

- Make sure questions are clear and focused.
- All questions should be based on the same populations or perhaps a subset.
- Each question should define a single hypothesis to test or quantity to measure.
- "Does Betaseron decrease the relapse rate in relapsing-remitting MS patients compared to Placebo, on average?"

The Population

Who or what you are going to study.

- Can you establish a proper sampling frame?
- Populations are hard to deal with in surveys without a proper sampling frame.
- Observational studies usually compare more than 1 self selecting population.
- Controlled experiments randomly assigns treatments to the sample to create distinct groups within the population. Any member of the population is equally likely to be assigned to a group.

The Sampling Plan

How you select units from the population.

- For surveys a poor sample means unreliable results.
- Experiments and observational studies usually rely on who is available.
- Controlled experiments require the sample be randomized to groups.
- Does the sample represent the population
  - This is who is sampled not how many.
  - Also known as accuracy or bias.
- Does the sample give enough precision?
  - This is how many are sampled.
  - For testing this is called power.

Randomization

Assigning units to groups

- Places sampled units into several groups.
- Results cannot be predicted or anticipated.
- A sampled unit can be assigned to any group.
- Eliminates bias from the study.
- Differences can still occur by chance.
- Randomization can be done within the levels of observational factors.
  - Randomize within gender.

The Data

The information needed to answer your questions

- Is the data collection procedure reliable?
- Is the data entry accurate?
- Are the data complete?
- Can you confirm or correct suspected problems in the data?
- Can you fill in missing data after the fact?
- The data issues should be resolved before analysis begins.

The Analysis Plan

How you answer your questions

- Should include graphics to examine the data.
- Analysis is mainly determined by the type of dependent variable.
  - continuous, count, binary, categorical
- Hypothesis testing or parameter estimation.
- Are the data truly independent
  - repeated measures or cluster effects
- Do you need to adjust for any confounders?
- Do you have a multiple comparison problem?

Sample Size Calculation

Matching statistical results with what's important

- As sample size increases
  - statistical power of a hypothesis test increases
  - precision of estimates increase
  - distribution of statistics become more predictable
  - statistical accuracy does not change
- averages eventually follows a normal distribution
- **On line demonstration**

Statisticians can help

- Focus and clarify the objectives
- Design an appropriate sampling plan
- Provide a randomization scheme
- Design an appropriate analysis plan
- Calculate the sample size or power.

Talk to a statistician before you collect your data !!!

Sample Size Calculation

- For a statistical test, the test statistic follows a specific statistical distribution under the Null Hypothesis. This is used to calculate the p-value.
  - z, t_n, χ^2_n, $F_{n,m}$
- Under the alternative Hypothesis, the test statistics a different distribution. This is used to calculate the power.
- As sample size increases
  - for testing, the power increases.
  - for estimation, the confidence interval shrinks
- Select the sample size that yields the desired power or interval size.

On-line Demonstration

Wikipedia -> **Statistical Power**

- 7. Software for Power and Sample Size Calculations

Russ Lenth on-line java applets

Power and Error Limits

R has built in functions for anova, proportions and t-test plus packages like *pwr*, *longpower* and many others.

Experimental Design

- Balance the data across the levels of the categorical predictors
- Correlated predictors can cause confounding and a loss of statistical power.
- Consider blocking factors to control variation. Make sure groups vary within each block.
- Make sure all combinations of factors are observed otherwise confounding and evaluating interactions can be a problem.
- Repeated measurements within a subject has a diminishing return except when evaluating predictors that vary within each subject.
- Be aware of any hierarchical structure in your data. Make sure your analysis controls for any such structure (mixed models).

Power of balanced designs

- Two group comparison with equal variances
  - 50/50 is as powerful as a 100/33 assignment.
- 4 treatments versus control: total N = 40
  - A: assign 8 to each treatment and 8 to control
  - B: assign 20 to treatment (5 each) and 20 to control
  - A & B are equally efficient for testing treatment versus control
  - A is 60% more efficient when comparing between treatments.
  - A is an overall better design.

Within Subject designs

A subject is exposed to more than one treatment

- It eliminates the between subject variability.
- treatment order needs to vary between different subjects
- With many treatments a latin square design can be used for counterbalancing.
- Even with counterbalancing, there can be carryover effects.
- If carryover effects exist then the study results can be unreliable.

Experimental Unit

- It is usually the level of sampling or treatment application.
- These observations are considered to be statistically independent of each other.
- Pseudo replication are observations within an experimental unit.
- Pseudo-replicates can be used but require a more complicated analysis to obtain accurate results.

Types of variables

- Data can play the role of the response variable (dependent, Y) or the explanatory variable (independent, X)
- Data can be numeric (continuous or discrete). This is a real number with a real interpretation.
- Data can be categorical , which can be ordinal (small, medium, large) or nominal (red, green, blue).
- The type of analysis depends on which type of variable is in each role.
  - The response variable is most important.

Common Analyses

- Continuous response variable
  - t-test, regression, ANOVA, ANCOVA
- Response variable is a count
  - Poisson or negative binomial regression
- Binary response variable
  - contingency tables, logistic regression
- Categorical responses are more complicated
  - ordinal or nominal logistic regression

Multiple Comparisons

If several tests are conducted then the chance of at least one false positive result will be much larger than α.

- Bonferroni correction (α/n) is very conservative.
  - guarantees chance of any false positives is α.
- False Discovery Rate (FDR) is less conservative.
  - guarantees that with n positive results, at most αn are false positives.
- Recommendations: limit the number of questions or select a few questions as primary

Questions?

www.stat.ubc.ca/SCARL

STAT 551 - Stat grad students taking this course offer free statistical advice. Fall semester every academic year.

SOS Program - An hour of free consulting to UBC graduate students. Funded by the Provost and VP Research.

Short Term Consulting Service - Advice from Stat grad students. Fee-for-service on small projects (less than 15 hours).

Hourly Projects - SCARL professional staff. Fee-for-service consulting.

The End