Hartford Hospital Research Program Research Methods Lecture Series Part II :

Concepts of Inferential Statistics

Nov 2, 2009

Yes! We're going to talk about research methods!

OVERVIEW:

- October: Basic concepts of research design
- November: Concepts of inferential statistics
- December: Choosing the right statistic Part I
- January: Choosing the right statistic Part II
- February: Meta analysis and clinical trials
- March: Grant-writing

Presenters:

- Tara McLaughlin, PhD, MPE

Senior Scientist/ Grantwriter

(tmclaug@harthosp.org)

- Jyoti Chhabra, PhD

Senior Scientist (jchhabr@harthosp.org)

- Ilene Staff, PhD

Senior Scientist (istaff@harthosp.org)

- David O'Sullivan, PhD

Senior Scientist (dosulli@harthosp.org)

To Review – Our last presentation

- Characteristics of systematic design
- Developing background/significance
- Experimental and non-experimental approaches
- Common study designs
- Sampling methods
- Levels of measurement

Today's Presentation

- Descriptive vs Inferential Statistics
- Hypothesis Testing
- Statistical significance and power
- Sample Size calculation

Descriptive Statistics

- Describes the characteristics of the sample
  - Frequency Distributions
  - Measures of Central Tendency – mean, mode, median
  - Measures of Variability – range, variance, standard deviation
  - Measures of Association – correlation

Inferential Statistics

- Gathers data about the sample and ***infers*** information about the larger population.
- Uses principles of probability to decide if what is observed about the sample reflects the population reality or occurs by chance.
- Process referred to as **hypothesis testing** and forms the basis of what most people think of as “statistics”

Hypothesis Testing

- Set up alternative possibilities to test
- Can be looking for differences in two populations, the existence of a relationship, etc.
- A Null Hypothesis (H_O) – suggests that there are no differences, relationships, or equivalences
- Alternative Hypothesis (H_1) suggests that there are

Hypothesis testing: How decisions made

- Based on data collected and statistical findings, you take one of two actions – either reject the null hypothesis or not reject
- Basic rule -- Never accept null hypothesis, you haven't proven it, you just have failed to prove it wrong (this makes more sense later)

Hypothesis testing: How decisions made

- When doing analysis, you have data and make decisions about sample and try to understand the population
- But. . . if you could "know all" and know about the population that would set up four possible combinations

Hypothesis testing: Correct and Incorrect Inferences

	In The Population	
In the sample	**Differences Exist**	**No Differences Exist**
Differences observed	Correct Decision	Incorrect Inference
No differences observed	Incorrect Inference	Correct Decision

Hypothesis testing: Statistical significance

Decision re: Sample	In The Population: Differences Exist	In The Population: No Differences Exist
Reject H_O	Correct Decision	Type I Error (p = Alpha) "Signficance level"
Fail to Reject H_O	Type II Error (p = Beta)	Correct Decision

Statistical Significance Level

- Probability of a Type I error is the probability that an effect seen in sample is due to chance and does not reflect a true effect in population
- Want the probability of error to be as low as possible
- How would you do that????

Keeping Type I error low

- Be less likely to "Reject the Null Hypothesis"
- But then what happens??
- More likely to have Type II error
- Have to balance Type I error and Type II error

Statistical Significance Level

- Actually set by the researcher prior to the analysis – despite ***everyone calculating p values***
- Convention is to set alpha at .05
- But other factors can be in play to change what alpha should be
- How do you decide what the correct balance is?

Analogy to hypothesis testing: jury decisions

- Start with assumption of innocence (null hypothesis)
- Juries are presented with evidence (data)
- Reach a verdict – guilty or not guilty (reject or not reject null hypothesis)
- Verdict is meant to represent the actuality of guilt or innocence.

Analogy to hypothesis testing: jury decisions

Jury Decision	**Actual Guilt or Innocence**	
	Guilty	**Innocent**
Guilty	Correct Decision	Type I error **Wrongful sentence**
Not Guilty	Type II error Someone guilty goes free	Correct Decision

Balancing Types of Error - Analogy

- What type of case/court – criminal vs family
  - "beyond a reasonable doubt"
  - "preponderance of evidence"
- Back to statistics
  - Alpha is 'reasonable doubt'
  - When do we go to preponderance of evidence?
- Best aim is to maximize the likelihood of finding guilty party guilty – fair, safe for society, etc.

Hypothesis testing: Statistical significance and power

	In The Population	
	Differences Exist	**No Differences Exist**
Reject H_o	Correct Decision (p = 1 - Beta) "Power"	Type I Error (p = Alpha) "Signficance level"
Fail to Reject H_o	Type II Error (p = Beta)	Correct Decision (p = 1 – Alpha)

Hypothesis testing: How decisions made

- Based on data collected and statistical findings, you take one of two actions – either reject the null hypothesis or not reject
- Basic rule -- Never accept null hypothesis, you haven't proven it, you just have failed to prove it wrong (think of the analogy)

Keeping Type I and Type II error low

- You can decrease either at the expense of the other but how do you minimize both of them?
- Let's go back to our analogy
  - Only try cases that are really obvious
  - Get more evidence – more in depth information
  - Get more compelling evidence (clearer, less confusion)

What is statistical power?

- Power is the inverse of Type II error – it is the probability that a true effect in the population will be found;
- What determines power?
  - alpha level (need to balance between the two errors)
  - how substantial a population effect does one want to be able to identify
  - how much variance exists in the sample
  - **sample size – this is the one most easily changed**

Why is a sample size calculation needed?

- Increasing sample size is the best way to increase power without sacrificing significance level
- Determining proper balance between error types and providing large enough sample
  - insures best interpretation
  - prevents too many patients included unnecessarily and/or too few to make useful conclusions

Calculating power/sample size

Preparation

- What is your research question, design?
- What will be your primary measure? Is powering on that alone sufficient – what are implications of 'indefinite' interpretation
- What do you want (expect) to find?
  - How large an effect (difference) do you expect
  - How large is 'zone of indifference'
  - How wide a confidence interval is appropriate?

Calculating power/sample size

Estimating effect size or "How do you expect me to know what the differences will be, that's why I am doing the study? !!"

Base estimates on:

- Prior empirical findings – Are there studies that are similar to use as a guide? (Best)
- How large a difference has clinical significance? What are smallest differences in population that if exist, study should definitely be able to identify?
- Standard effect sizes small (.2), medium (.5) or large (.8) effect size (see Cohen, 1988)

Calculating power/sample size

- Best approach is to estimate effect size or precision desired and calculate required sample size
- Practical issues can limit available sample; can start with sample size and calculate what effect size/level of precision is adequately powered
- Even if power less than desired, knowing this will improve interpretation of any negative (no difference) findings

Calculating power/sample size

Resources

- Purchased software
  - (e.g.) nQuery, PASS, Sample Power
- Free internet software (e.g.) G-Power
  http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/
- Applets, published tables, specialized web sites
  - Various: http://www.stat.uiowa.edu/~rlenth/Power/
  - ROC: http://department.obg.cuhk.edu.hk/index.asp?scr=1024
- **Ask for help from the Research Program**

What about sample size for study that is 'just descriptive'?

- If objective is describing an event or process and its outcomes rather than testing an hypothesis – speak of precision rather than power
- Estimating means – confidence interval
- Estimating proportions – charts available

Interpreting the data:

- Describe the findings
- Were they statistically significant? Were they clinically significant?
  - Statistical significance is only about probability and generalizing from sample to population; it does not speak to importance to care
- How do the findings relate back to hypotheses, to findings in other studies outlined in background?
- What have we learned?

Summary: A Well Thought Out Research Process

- Clearly defined study objectives using and extending prior research findings
- A clearly defined target population and representative sample of sufficient size to yield good statistical power
- An appropriate research mode and design

Summary: A Well Thought Out Research Process (cont'd)

- Operationally defined measures that can be proven reliable and are as free from barriers to validity as possible
- Null and alternative hypotheses stated; alpha set
- Appropriate statistical tests selected
- Findings properly interpreted and generalized

Coming Attractions – next month. . .

Choosing the appropriate statistics

Join us for 3rd Annual Research Day Symposium, November 9th

Featured Speaker on Bioinformatics
Also-5 Hartford Hospital Researchers

CMEs/CEUs available w registration

Research Program Information

All day (7:00 – 2:15) - $25 including breakfast and lunch

OR, don't have all day??
Just stop by for 1 or more speakers – No food/no fee

Hartford Hospital Research Program

Questions??