STATISTICS FOR MEDICAL RESEARCH I:
The Big Picture

Rosey Zackula, MA
Senior Research Associate/Statistician
Office of Research
University of Kansas School of Medicine-Wichita
(316) 293-3810; rzackula@kumc.edu
WHAT YOU SHOULD KNOW AFTER THIS PRESENTATION:

STATISTICS FOR MEDICAL RESEARCH

You should know:

1. When to consult a statistician during your research project.
2. Where statistics fit in the research process.
3. How to use power analysis.
4. How statistical tests work.
5. What is statistical significance.
6. What is a confidence interval.
7. Where to go for help.
Statistics and Research
The BIG Picture

Rosey Zackula, MA
Senior Research Associate/Statistician
Office of the Associate Dean for Research
University of Kansas School of Medicine-Wichita

Statisticians and physicians have a lot in common

Physicians
- Patients
- Take samples
- Physical exams
- Run diagnostics
  - X-rays, CT scans, MRI’s
- Write up reports
  - Patient diagnosis

Statisticians
- Data
- Take samples
- Physical exams
- Run diagnostics
  - Charts, graphs, tables
- Write up reports
  - (results)
  - Data diagnosis

What you will learn

- Ten step research process
- Where statistics fit in
  - Hypotheses
  - Data distributions
  - Power analysis
  - Statistical tests
  - P-values
  - Confidence intervals
- When/where to get statistical help

Where do statistics fit in?

Watch for UPPER CASE, BOLDED items on the following slides

Step 1: Turn Ideas into Research Questions

- Is your idea novel?
- Why does it matter?
- How will patients (or service users) benefit from your research?
- Consult colleagues and other researchers
- In general terms, what is your aim?
- IS YOUR RESEARCH QUESTION ANSWERABLE?
Step 2: Review Literature

- What is generally accepted?
- What are important gaps?
- What are possible misconceptions?
- Identify current state of research
  - NULL HYPOTHESIS
    - gather data for power analysis
    - more about this later
- In specific terms, what is your hypothesis
  - ALTERNATIVE HYPOTHESIS
    - gather data for statistical test(s)

Step 3a: Establish Research Design

- Consider participant involvement
- Construct method for sampling from the population
  - SAMPLING DESIGN
  - Identify what to measure
  - DISTRIBUTIONS
    - Counts
    - dichotomous: two categories (yes or no)
    - nominal: unordered categories (race)
    - ordinal: ordered categories (number of injections)
    - Continuous measures (blood pressure)

Step 3b: Develop Statistical Methods

- Estimate number of subjects required to answer the research question(s)
  - POWER ANALYSIS
    - Select main statistical test (formula) to be conducted with your data
    - Formula includes sample size “n”
    - use algebra to solve for n
    - insert key information into formula
      - from the null hypothesis gathered during literature search

Step 4: Write Proposal

- Title
- Abstract/summary
- Background or rationale of the project
- Aims/objectives
- EXPERIMENTAL DESIGN AND METHODS
- Ethical considerations
- Benefits of the study
- Resources and costs

Step 5: Address Funding Issues

- Patient costs
  - Labs
  - Travel reimbursement
  - Incentives
- Research assistants
- Supplies
- Outside statistical consulting fees
- IRB fees

Step 6: Get Study Approval

- Departmental signatures
- Scientific review
- STATISTICAL REVIEW
- IRB approval
Step 7a: Collect data

- Follow HIPAA guidelines for protecting patient identification
- Format
  - Rows: patient information
  - Columns: measurement per patient
- Beware of biases and inaccuracies

Step 7b: Correct the data

- CLEAN AND ASSESS DATA
  - Data accuracy
  - Outliers
  - Missing data
  - Check statistical test rules (assumptions of the test)
    - Distribution (shape of data)
    - Sampling design
      - Independent or dependent (matching)
      - Random or convenience sample

Step 8: Analyze and Interpret Results

- STATISTICAL TEST(S)
  - Same test as the power analysis
  - Uses your data
- P-VALUE(S)
  - Strength of evidence against null hypothesis
    - The smaller the value, the more we doubt the null
    - Interpret through picture or graph
- CONFIDENCE INTERVAL(S)
  - "Guesstimates" true population value without measuring everyone

How do statistical tests work?

Distribution-based tests (parametric)

- Based on shape of data (and other rules)
  - Where is the most frequently occurring point?
  - Average: typically this is the value to test
  - How does the data vary?
    - What is the spread of the data
    - Select distribution with same shape as data
    - Conduct test associated with that distribution
      - input average from data, sample size, etc.
      - result of test marks the start of area under the curve
      - Measure area under the curve (p-value)

SINGLE VARIABLE EXAMPLE

Research question:
Is mean AGE significantly different from 40?

Null hypothesis
H₀: The mean age is 40

Alternative hypothesis
H₁: The mean age is not 40

Statistical test example: plot data
**Statistical test example: match data shape and conduct test**

- Match the shape of your data to known distribution
  - Bell shape from normal distribution
  - 40 is standardized (set to 0)

- Conduct test
  - Find area under the curve (to the right of test result)
  - Calculus determines the value
  - Also found on standard z or t tables

**What is statistical significance?**

- When a p-value is significant (i.e., very small) it simply means you are very sure it's reliable
- It doesn't mean
  - the finding is important
  - it has any decision-making utility
  - the difference is large
  - No information about size or direction
- Why do tests turn out to be significant?
  - sample size is large enough
  - test is powerful enough to detect a difference
  - sometimes it's significant just by chance

**More on confidence intervals**

- Measure of precision for study results
  - Range of values either side of an estimate of the population value
  - Strict definition of 95% CI (Altman, 2005)
  - 95% of CI carried out in a series of identical studies with different samples would contain the true population value
- Range of the interval will vary with the size of the sample
  - Small samples don't give much information about the whole population, so the range is wider

**Step 9: Evaluate Implications**

- APPLY STATISTICAL RESULTS TO POPULATION
  - Generalizability
    - Unbiased, random sampling
    - representative of the population
- Limitations of study
  - STATISTICAL LIMITATIONS
    - Sample selection bias
    - Sample size
    - Outliers and missing data
    - Failure of test assumptions

**Step 10: Report the Findings**

- Posters, Oral presentations, Publications
  - Introduction: HYPOTHESIS
  - Methods: POWER ANALYSIS, DESIGN, STATISTICAL ANALYSIS
  - Results: STATISTICAL TEST, P-VALUE, CONFIDENCE INTERVAL
  - Discussion: APPLY RESULTS, STATISTICAL LIMITATIONS
The 10 step research process
- Null and alternative hypotheses
- Power analysis
- Statistical tests
- Statistically significant p-value
- Confidence intervals
- When to consult a statistician

Free research/statistical help
- Research scientist within your department
- KUSM-W Office of the Associate Dean for Research
  - 293-3810, rzackula@kumc.edu
- Textbooks
- Journal articles on statistical methods
  - BMJ, Alman and Band
- CONSORT: CONsolidated Standards Of Reporting Trials
  - http://www.consort-statement.org/home/
- STROBE: STrengthening the Reporting of OBservational studies in Epidemiology
  - http://www.strobe-statement.org/
- Equator Network: Enhancing the QUAlity and Transparency Of health Research
  - http://www.equator-network.org/