BIOS 312: Course Overview

Chris Slaughter

Department of Biostatistics, Vanderbilt University School of Medicine

January 2, 2013
Outline

1. Course Structure
2. Software
3. Grading
4. Course Materials
Contact Information

- Instructor: Chris Slaughter, DrPH
  - email: james.c.slaughter@vanderbilt.edu
  - phone: 322-0439 (but no voicemail)
  - office: MCN D-2207

- TA: Uche Sampson
  - email: u.sampson@vanderbilt.edu
Lectures: Tuesdays and Thursdays, 10:30 - 12:00,
Lab: Wed, 11:00 - 12:00
Most classes held in Light Hall 306A
Exceptions noted on course web page
Syllabus: See course web page
(http://biostat.mc.vanderbilt.edu/CourseBios312)
Guiding Principles

- Alternative course title: How to Use Statistics to Answer Scientific Questions (part 2)
- Put science before statistics
- Emphasize parameter estimates and confidence intervals (credible intervals, likelihood intervals) over hypothesis testing and p-values
  - The End of Statistical Significance (Jonathan Sterne)
  - What’s Wrong with P-Values (Bland, Altman)
  - Key difference between scientific/clinical significance and statistical significance
- This is a course in Biostatistics, not Stata
  - I will show you how to get and interpret the key statistics, but not interpret every number of the Stata output
  - Often there is more than one way to arrive at identical final answers
Example

- 5 clinical trials conducted to determine if drug A, B, or C lowers cholesterol
- Assume that a decrease of 10 mg/dl or more is important to clinicians
- Study design
  - Cholesterol measured at baseline
  - Subjects take drug for 1 month
  - Cholesterol measured at 1 month
  - Change in cholesterol is the outcome of interest
- Summary of results

<table>
<thead>
<tr>
<th>Trial</th>
<th>Drug</th>
<th>Pts</th>
<th>Mean diff</th>
<th>Std dev</th>
<th>Std error</th>
<th>95% CI for diff</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>30</td>
<td>-30</td>
<td>191.7</td>
<td>49.5</td>
<td>[-129, 69]</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>1000</td>
<td>-30</td>
<td>223.6</td>
<td>10</td>
<td>[-49.6, -10.4]</td>
<td>0.002</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>40</td>
<td>-20</td>
<td>147.6</td>
<td>33</td>
<td>[-85, 45]</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>4000</td>
<td>-2</td>
<td>147.6</td>
<td>3.3</td>
<td>[-8.5, 4.5]</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>5000</td>
<td>-6</td>
<td>100.0</td>
<td>2</td>
<td>[-9.9, -2.1]</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Outline

1. Course Structure
2. Software
3. Grading
4. Course Materials
Many packages will suffice (Stata, R, SPSS, SAS)
Students may use any package they prefer
  - Will show code to obtain results in both Stata and R
  - I focus on Stata for output
  - Will also use JAGS and R later in course for applied Bayesian analysis

Stata
  - My second favorite statistical package
  - Flexible, user-written addons
  - Relatively easy syntax, which will be introduced in lectures
  - Report writing for reproducible research sub-optimal
  - Graphics not as good as R
  - Designed for people who know statistics, but don’t want to write basic functions
• R
  • Free
  • Most adaptable statistical package
  • Steeper learning curve, but “R commander” or other addons can be helpful
  • Need to be able to compute robust standard errors (the “sandwich” estimator)
  • Used by the majority of academic Biostatisticians

• Reproducible research in R
  • packages: knitr, Sweave
  • Dynamic report generation with R
  • RStudio
Other Software

- **SAS**
  - Unique language
  - Does everything needed for this course
  - Great for data management
  - Provides too much statistical output (much of the output is useless)

- **SPSS**
  - Statistical Package for the Social Sciences
  - Mostly menu driven
  - Package I know the least about, but should do everything needed for the course
Software for Bayesian Analysis

- **JAGS**
  - Just Another Gibbs Sampler
  - Free
  - Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods
  - Not a complete statistical packages (can interface with R)
  - Limited use in this course

- Others: WinBUGS, OpenBUGS, R packages, SAS ...
Outline

1. Course Structure
2. Software
3. Grading
4. Course Materials
Grading

- Midterm (25%)
- Final Exam (25%)
- Homework and quizzes (25%)
- Project (25%)
- Class participation

This is a 4-credit course. Your lab and lecture grades will be the same.
Homework

- Up to 1 per week (probably 6 or 7 total)
- Will focus on real data analysis and interpretation with some mathematical derivations of important quantities
- Questions will focus on specific analyses, with questions stated in as scientific terms as possible
- Work handed in should address the scientific questions
  - Format Table and Figures
  - No raw, unformatted Stata output
- Keys will be provided shortly after the homework is turned in
  - No late homework accepted after the key is posted
  - Answers in keys may go beyond what is expected of your homework and present concepts in more detail. You are responsible for any material in the keys for exams.
- May discuss the homework with others in the class, but the work you turn in should be your own
Exams

- Exams: Midterm and Final in class
- Focus on understanding concepts, not memorizing formulas
- I will provide an example midterm and final
- For midterm, you will be allowed 1 page of your own notes
- For final, you will be allowed 2 pages of your own notes
- Quizzes
  - At my discretion, we will have a 10-minute quiz in lab (up to one per week)
Project

- Detailed analysis of a dataset
- Prepare a 4-10 page reporting of your results (excluding tables and figures)
- Will be assigned at mid-point of course; more details at that time
- Likely will be a collaborative project (about 3 people per group)
Outline

1. Course Structure
2. Software
3. Grading
4. Course Materials
Course notes will be the primary source

Available on main web page:
http://biostat.mc.vanderbilt.edu/CourseBios312

Daily class schedule including notes for that day

Schedule will be updated throughout semester
Recommended Books

Vittinghoff, Glidden, Shiboski, and McCulloch. *Regression Methods in Biostatistics*

- **Pros**
  - Covers applied linear, logistic, survival, and repeated measures regression
  - Biomedical orientation
  - Uses Stata for examples

- **Cons**
  - Does not discuss robust standard errors, which I will use almost exclusively
  - Emphasizes model diagnostics more than I will
  - Does not cover Bayesian approaches
  - Does not cover matrix algebra
  - Little theory and mathematical derivation

- I will give recommended readings throughout the semester
- If the course notes and book differ, go with the notes
Recommended Books

Weisberg. *Applied Linear Regression*

- **Pros**
  - Mathematical theory is at the level of Biostat qualifying exams
  - Uses matrix algebra notation

- **Cons**
  - Poor for applied data analysis
  - Not oriented towards biomedical applications
  - Only covers linear regression
  - Does not cover Bayesian approaches
Recommended Books

Kruschke. *Doing Bayesian Data Analysis*

- **Pros**
  - Excellent introductory applied Bayesian analysis
  - Makes comparisons to frequentist approaches

- **Cons**
  - Does not cover theory
One last thing...