

**Biost 517**  
**Applied Biostatistics I**  
.....

Scott S. Emerson, M.D., Ph.D.  
Professor of Biostatistics  
University of Washington

1

© 2002, 2003, 2005 Scott S. Emerson, M.D., Ph.D.

**Biost 514**  
**Biostatistics I**  
.....

Scott S. Emerson, M.D., Ph.D.  
Professor of Biostatistics  
University of Washington

2

**The Use of Statistics to Answer  
Scientific Questions**  
.....

Scott S. Emerson, M.D., Ph.D.  
Professor of Biostatistics  
University of Washington

3

**General Philosophy**  
.....

"Everything should be as simple as possible, but no  
simpler."

- A. Einstein (paraphrased)

4

**Biost 517 / 514**  
**Applied Biostatistics I**  
.....  
Scott S. Emerson, M.D., Ph.D.  
Professor of Biostatistics  
University of Washington

Lecture 1:  
Course Structure; Overview

September 28, 2011

5

**Lecture Outline**  
.....

- Course Structure
- Overview of Setting
  - Scientific method
  - Case study

6

**Course Overview**  
.....

7

**Course Structure**  
.....

- Instructor: Scott S. Emerson, M.D., Ph.D.
  - » Fair warning
- TAs:
  - Greg Levin
  - Elisa Sheng
- Time and Place:
  - Lectures: 9:30 - 10:20 am MWF HST 439
  - Data Analysis:
    - 8:30 - 9:20 am M HST 473
    - 8:30 - 9:20 am Tu HST 498
    - 8:30 - 9:20 am W HST 473
    - 8:30 - 9:20 am F HST 473

8

## Assumed Prior Knowledge

.....

- Statistical coursework
  - None
  - (If you have had prior courses, unlearn
    - Need for normal data to test means
    - P value as entire summary of analysis
    - Significance testing to detect confounding
    - Overlapping CI signify not statistically significant
    - Wilcoxon as test of median
    - Fisher's exact test preferred in small samples
    - ...)

9

## Old Dogs, New Tricks

.....

- Recording of Lectures: Camtasia
  - Audio and computer video on web
  - Posted approximately 24 hours after class
- No guarantees
  - “Mistakes happen”

10

## Textbook: Rosner

.....

- Fundamentals of Biostatistics (kth ed.)
  - Classical organization
  - Optional: Used primarily as a reference
  - (Lectures follow a different organization)

11

## Computer Software

.....

- Extensively used for data analysis
- Students may use any program that will do what is required, however
  - Stata is used heavily in Biostat 536, 537, 540
  - Stata commands will be provided in lecture
  - R is free, but harder to use
    - Some R functions will be provided on the website
  - Help will presume the use of Stata or (for the brave) R
    - I am conversant in S-Plus, R (very) and SPSS (enough for this class)
  - Other packages may not compute robust standard errors

12

## Stata

.....

- Extremely flexible statistical package
  - Interactive
  - Excellent complement of biostatistical methods
- Graphical, report capabilities suboptimal
- Available in microcomputer lab
- Supplementary info on web page
- Syntax introduced in lectures as needed

13

## R

.....

- The ultimate in flexible statistical languages
  - Interactive
  - Many user-supplied functions
- Graphical functions generally very good
- Open source and free
- Supplementary info on web page
- Syntax provided (sometimes) as supplemental materials on webpages
- A collection of functions that I have written will be made available on the webpages
  - Caveat emptor

14

## Computer Software: Comments

.....

- Designed for people who know statistics, but do not want to write basic functions
  - Tries to be all things to all people
    - Much output that you will not want
    - Much output that I will recommend against

15

## Guiding Principles

.....

- This is a course in biostatistics, not statistical software
  - I will tell you how you can get the statistics I teach you to use
    - There are often multiple ways
    - I tend to teach one of them
  - I will not explain every number that appears on the printout

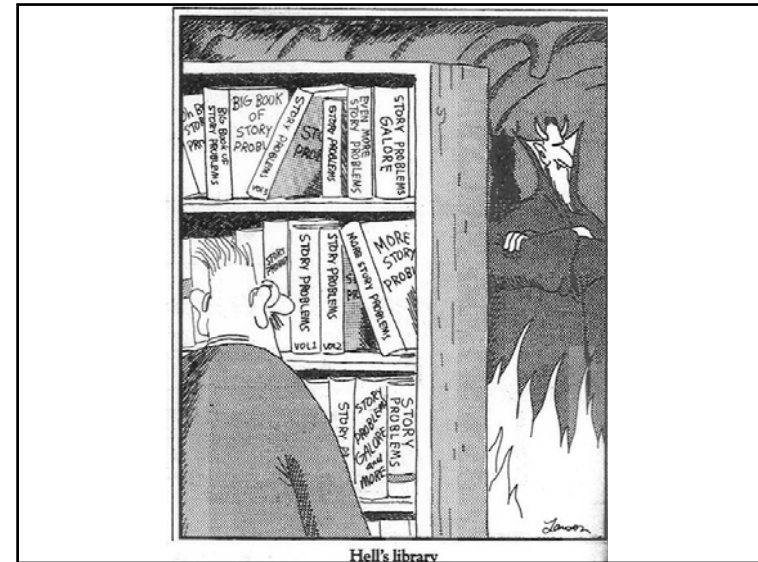
16

## Written Homeworks

.....

- Weekly homeworks: analysis of real data
  - Questions directed toward specific analyses
    - But questions will still be stated in as scientific terms (as opposed to statistical) terms as possible
  - Work handed in is expected to be organized scientifically
    - I expect nicely formatted tables, figures
    - Unedited Stata output is totally unacceptable
  - Biost 514: supplemental problems on occasion

17



## Errors to Avoid

Unedited Stata / R output is  
**TOTALLY**  
 unacceptable

## Errors to Avoid

Any assignments that  
 are handed in should  
 be only your work

## Homework Keys

.....

- Keys to the homeworks will be generally be available on the web pages
  - Annotated Stata output will typically be included
  - My answers will typically go beyond what I expected you to do
  - You are responsible for any new information that I provide in the homework keys, even if that information is not otherwise presented in class

21

## Quizzes and Lecture Discussion

.....

- Approximately once per week you will be asked to view a lecture prior to class
- Lecture will be used to
  - Take a brief quiz on the subject matter
    - Basic knowledge
    - Judgement
  - We will then discuss the reasoning that should be used to answer the quiz questions and current homework assignments
    - Participation in the discussion is required

22

## Discussion Section

.....

- Data analysis to answer scientific questions
- You will be given a scientific question and a data set which was collected to try to answer that question
  - Setting is more realistic than that which is given on written homeworks
- We will discuss the approach to the whole problem
- Nothing to hand in, but participation in discussion is required
  - You must inform me if you are attending a different discussion section

23

## Grading

.....

- 20% Homeworks (approx 8 - 9)
- 10% Quizzes and discussion
- 25% One Midterm (in class, closed book)
- 15% Data Analysis and Report
- 30% Final Exam (in class, closed book)

24

## Course Web Pages

.....

- Address: [www.emersonstatistics.com/b517/](http://www.emersonstatistics.com/b517/)
- Content
  - Syllabus
  - Lecture handouts
  - Recordings of lectures (and discussions?)
  - Homework assignments and keys
  - Datasets
  - Supplemental materials not discussed in class
    - Handouts, noteworthy emails
  - Prior years' webpages
    - Homework keys, exams, exam keys, projects

25

## Course Structure

.....

- Biost 517 / 514
  - One response variable; one grouping variable
    - One-, two-, K-sample description and inference
    - Simple regression
  - Stratified description and inference
    - Adjustment for confounding, precision
- Biost 518 / 515
  - Multivariable regression

26

## Biost 517 / 514: Topics

.....

- Scientific setting
  - Scientific questions
  - Study structures
  - Statistical role

27

## Biost 517 / 514: Topics

.....

- Descriptive statistics
  - Motivation
  - Types of measurements
  - Univariate summary statistics
  - Univariate depictions of distributions
  - Censored data descriptive statistics
  - Bivariate descriptive statistics
- Note: The time spent on descriptive statistics is meant to also guide the choice of summary measure to answer inferential questions

28

### Biost 517 / 514: Topics

.....

- Inferential statistics for two variables
  - Relevant probability
  - Types of statistical inference
    - Bayesian posterior distributions
    - Frequentist sampling distributions
  - Comparing means, geometric means, medians, proportions, odds, hazards, ...
  - Point and interval estimates
  - Hypothesis tests
    - T, chi squared, Fisher's exact, logrank, Wilcoxon
    - Simple regression

29

### Biost 517 / 514: Topics

.....

- Introduction to stratified analyses
  - Confounding, precision, effect modification
  - Descriptive statistics
  - Stratified analyses

30

### Overview of Setting

.....

Scientific Method

31

### Purpose of Statistics

.....

- Statistics is about science
  - (Science in the broadest sense of the word)
- Science is about proving things to people
  - (The validity of any proof rests solely on the willingness of the audience to believe it)

32

## First Stage of Scientific Investigation

.....

- Hypothesis generation
  - Observation
    - Measurement of existing populations
    - Disadvantages:
      - Confounding
      - Limited ability to establish cause and effect

33

## Further Stages of Investigation

.....

- Refinement and confirmation of hypotheses
- Experiment
  - Intervention
  - Elements of experiment
    - Overall goal
    - Specific aims (hypotheses)
    - Materials and methods
    - Collection of data
    - Analysis
    - Interpretation; Refinement of hypotheses

34

## Regulation of New Drugs / Biologics

.....

- Labeling
  - Wiley Act (1906)
- Safety
  - Food, Drug, and Cosmetics Act of 1938
- Efficacy
  - Kefauver – Harris Amendment (1962)
    - " [If] a lack of substantial evidence that the drug will have the effect ... shall issue an order refusing to approve the application. "
    - "...The term 'substantial evidence' means evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts..."
- Effectiveness
  - FDA Amendments Act (2007)
    - RCT registration, Pediatrics, Risk Evaluation and Mitigation Strategies<sub>35</sub> (REMS)

36

## Do You Need Statistics?

.....

- Two question test (Both must be YES)
  - In a deterministic world, do YOU know how to answer your question?
    - Is the question answerable in the real world?
    - How do you use a number to answer the scientific question?
  - In a world subject to variation, do YOU know how you would answer your question if you had the entire population?

36

### Ex: Smoking Effect on Lungs

.....

- Association between smoking and lung function in children
  - Long term smoking is associated with lower lung function
  - Are similar effects observed in short term smoking in children?

37

### Do You Need Statistics?

.....

- Two question test (Both must be YES)
  - In a deterministic world, do YOU know how to answer your question?
    - Is the question answerable in the real world?
    - How do you use a number to answer the scientific question?
  - In a world subject to variation, do YOU know how you would answer your question if you had the entire population?

38

### Ex: Smoking Effect on FEV

.....

- Scientific question
  - Does smoking lead to lower lung function in kids?
- Study design
  - 654 healthy children
  - Measure smoking by self report
  - Measure lung function by FEV
    - Forced expiratory volume: maximum volume of air that can be exhaled in 1 second

39

### The Data

**SMOKERS**

1.953 2.236 3.428 1.694 3.957 4.789 2.384 3.074 2.387 3.835 2.599 4.756 3.086 4.309 3.413 2.975 3.169 3.343 3.751 2.216 3.078 3.186 3.297 2.304 3.102 2.677 3.297 3.498 2.759 2.953 3.785 2.276 4.637 3.038 3.120 3.339 3.152 3.104 4.045 4.763 3.069 4.506 3.519 3.688 2.679 2.198 3.345 3.082 2.903 3.004 3.406 3.122 3.330 2.636 3.799 4.086 4.070 2.264 4.404 2.278 4.872 4.270 3.727 2.795

**NONSMOKERS**

1.708 1.724 1.720 1.568 1.895 2.336 1.919 1.415 1.987 1.942 1.602 1.735 2.193 2.118 2.258 1.932 1.472 1.878 2.352 2.604 1.400 1.256 0.839 2.578 2.988 1.404 2.948 1.755 2.980 2.100 1.292 3.000 2.673 2.053 1.812 2.175 2.725 2.071 1.547 2.004 3.195 2.420 1.776 1.931 1.343 2.076 1.624 1.344 1.650 2.752 2.017 2.797 3.556 1.703 1.634 2.570 3.016 2.419 1.569 1.698 2.123 2.481 1.481 1.577 1.940 1.747 2.069 1.631 1.536 2.560 1.962 2.531 2.715 2.457 2.090 1.789 1.858 1.452 3.842 1.719 2.111 1.695 2.211 1.794 1.917 2.144 1.253 2.659 1.580 2.126 3.029 2.064 1.611 2.215 2.388 2.196 1.751 2.165 1.682 1.523 1.292 1.649 2.588 0.796 2.574 1.979 2.354 1.718 1.742 1.603 2.639 1.829 2.084 2.220 1.473 2.341 1.689 1.196 1.872 2.219 2.420 1.827 1.461 1.338 2.090 1.697 1.562 2.040 1.609 2.458 2.650 1.429 1.675 1.947 2.069 1.572 1.348 2.288 1.773 0.791 1.905 2.483 1.431 2.631 3.114 2.135 1.527 2.293 3.042 2.927 2.665 2.301 2.460 2.592 1.750 1.759 1.536 2.259 2.048 2.571 2.046 1.780 1.552 1.953 2.893 1.713 2.851 1.624 2.631 1.819 1.658 2.158 1.789 3.004 2.503 1.933 2.091 2.316 1.704 1.606 1.165 2.102 2.320 2.230 1.716 1.790 1.146 2.187 2.717 1.796 1.335 2.119 1.666 1.826 2.709 2.871 1.092 2.262 2.104 2.166 1.690 2.973 2.145 1.971 2.095 1.697 2.455 1.920 2.164 2.130 2.993 2.529 1.726 2.442 1.102 2.056 1.808 2.305 1.969 1.556 1.072 2.042 1.512 1.423 3.681 1.991 1.897 1.370 1.338 2.016 2.639 1.389 1.612 2.135 2.681 3.223 1.796 2.010 1.523 1.744 2.485 2.335 1.415 2.076 2.435 1.728 2.850 1.844 1.754 1.343 2.303 2.246 2.476 3.239 2.457 2.382 1.640 1.589 2.056 2.226 1.886 2.833 1.715 2.631 2.550 1.912 1.877 1.935 1.539 2.803 2.923 2.358 2.094 1.855 1.535 2.135 1.930 2.182 1.359 2.002 1.699 2.500 2.366 2.069 1.418 2.333 1.514 1.758 2.535 2.664 2.487 1.591 1.624 2.798 1.891 1.999 1.869 1.004 1.427 1.826 2.688 1.657 1.872 2.015 2.371 2.115 2.338 1.495 2.894 3.328 3.381 2.170 3.470 3.068 1.811 2.524 2.642 3.741 4.336 4.842 4.550 2.841 3.166 3.816 2.561 1.654 2.481 2.665 3.203 3.549 3.222 3.111 3.490 3.147 2.520 2.292 2.889 2.246 1.937 2.646 2.957 4.007 2.386 3.251 2.762 3.011 4.305 3.906 3.583 3.236 3.436 3.056 3.007 3.489 2.864 2.819 2.250 4.683 2.352 3.106 3.994 4.393 2.592 3.193 2.946 3.515 2.754 2.720 2.463 2.633 3.048 3.111 3.745 2.094 3.183 3.977 3.354 3.411 3.171 3.887 2.646 2.504 3.587 3.945 2.971 2.891 1.823 2.472 2.175 2.735 4.273 2.976 4.065 2.318 3.596 3.395 2.751 2.673 2.556 2.542 2.608 2.354 1.458 3.795 2.491 3.060 2.545 2.993 3.305 3.774 2.855 2.988 2.498 1.169 2.887 2.704 3.515 3.425 2.287 2.434 2.365 2.696 2.868 2.813 3.255 4.593 4.111 1.916 1.858 3.350 2.901 2.241 4.225 3.223 2.224 4.073 4.090 2.696 4.411 3.791 3.089 2.465 3.200 2.913 4.877 2.358 3.279 2.581 2.347 2.691 2.827 1.873 2.538 2.758 3.050 0.079 2.201 1.858 3.403 3.501 2.576 1.665 2.081 2.974 4.073 4.448 3.984 2.250 2.752 3.680 2.862 3.023 3.681 3.255 3.692 3.356 4.591 3.082 3.258 2.216 3.247 4.324 2.362 2.563 3.206 3.585 4.720 3.331 5.083 2.417 2.364 2.341 3.231 3.078 3.369 3.529 2.866 2.891 3.022 3.127 2.866 2.605 3.056 2.569 2.501 3.320 2.123 3.780 3.847 3.924 2.132 2.752 2.449 3.456 3.073 2.688 3.329 4.271 3.530 2.928 2.689 2.332 2.934 3.110 2.894 2.435 2.838 3.035 4.831 2.812 2.714 3.096 3.519 2.232 2.770 3.341 3.090 2.531 2.622 2.936 2.568 2.387 2.499 4.130 3.001 3.132 3.877 3.223 2.290 2.859 2.622 2.140 4.203 2.997 2.662 3.082 3.806 2.458 2.391 3.141 2.579 2.100 2.785 4.294 2.906 5.102 4.429 4.270 4.500 2.635 3.082 3.387 5.793 3.985 4.220 4.724 3.731 3.500 3.674 5.633 3.645 2.887 3.960 4.299 2.981 4.504 5.638 2.853 3.211