Instructor's Guide to Stat Labs

The labs you find here in this text are case studies that serve to integrate the practice and theory of statistics. The instructor and students are expected to analyze the data provided with each lab in order to answer a scientific question posed by the original researchers who collected the data. To answer these questions, statistical methods are introduced, and the mathematical statistics underlying these methods are developed.

The Design of a Chapter

Each chapter is organized into five sections: Introduction, Data, Background, Investigations, and Theory. Sometimes we include a section called Extensions for more advanced topics.

Introduction.

Here a clear scientific question is stated, and motivation for answering it is given. The question is presented in the context of the scientific problem, and not as a request to perform a particular statistical method. We avoid questions suggested by the data, and attempt to orient the lab around the original questions raised by the researchers who collected the data.

The excerpt found at the beginning of a chapter relates the subject under investigation to a current news story, which helps convey the relevance of the question at hand.

Data.

Documentation for the data collected to address the question is provided in the Data section. Also, this section includes a description of the study protocol. The data can be found at the *Stat Labs* website: **www.stat.berkeley.edu/users/statlabs/**.

Background.

The Background section contains scientific material that helps put the problem in context. The information comes from a variety of sources, and

is presented in nontechnical language.

Investigations.

Suggestions for answering the question posed in the Introduction appear in the Investigations section. These suggestions are written in the language of the lab's subject matter, using very little statistical terminology. They can be used as an assignment for students to work on outside of the classroom, or as a guide for the instructor for discussing and presenting analyses to answer the question in class.

   The suggestions vary in difficulty, and are grouped to enable the assignment of subsets of investigations. Also included are suggestions on how to write up the results. Appendix 1 gives tips on how to write a good lab report.

Theory.

The theoretical development appears at the end of the chapter in the Theory section. It includes both material on general statistical topics, such as hypothesis testing and parameter estimation, and on topics specific to the lab, such as goodness–of–fit tests for the Poisson distribution and parameter estimation for the log-normal distribution. The exercises at the end of the Theory section are designed to give practice with the theoretical material introduced in the section. Some also extend ideas introduced in the section. These exercises can be used for paper-and-pencil homework assignments.

Statistical Topics

The table below lists the main statistical topics covered in each chapter. All of the basic topics found in most mathematical statistics texts are included here: descriptive statistics, experimental design, sampling, estimation, testing, contingency tables, regression, simple linear least squares, analysis of variance, and multiple linear least squares. We also list some of the additional specialized topics covered in each chapter.

Chapter	Main Topic	Some Additional Topics
1	descriptive statistics	quantile plots, normal approximation
2	simple random sampling	confidence intervals
3	stratified sampling	parametric bootstrap allocation
4	estimation and testing	goodness-of-fit tests, information, asymptotic variance
5	contingency tables	experimental design
6	Poisson counts and rates	Mantel-Hanzael test
7	regression	prediction
8	simple linear model	replicate measurements, transformations, inverse regression
9	ecological regression	weighted regression
10	multiple linear regression	model checking, projections
11	analysis of variance	unbalanced designs, indicator variables
12	response surface analysis	factorial design

Sample Courses

This book can be used as the main text for a course, or as a supplement to a more theoretical text. In a typical semester, we cover eight to ten chapters. We spend about one week on each of the chapters, with the exception of Chapters 4, 10, and 11, which require up to two weeks to cover.

The core chapters that we usually include in a course are Chapter 1 on descriptive statistics, Chapter 2 on simple random sampling, Chapter 4 on estimation and testing, and Chapter 7 on regression. Other chapters are chosen according to the interests of students. In a one semester course for engineers we may include Chapter 3 on stratified sampling, Chapter 5 on experimental design, Chapter 8 on calibration and inverse regression, Chapter 11 on analysis of variance, and Chapter 12 on response surface analysis. In a course designed for social and life science majors we tend to include Chapter 3 on stratified sampling, Chapter 6 on estimating mortality rates, Chapter 9 on ecological regression, Chapter 10 on multiple regression, and Chapter 11 on analysis of variance.

Lab Assignments

We have found that our course is most successful when we incorporate the labs into the class room, and not leave them as exercises for students to work on solely outside of class. Often, we have students work on four or

five of the ten labs covered in course. They have as their assignment, to address the suggestions in the Investigations section. Students analyze the data and write a report on their findings. We give two to three weeks for them to complete the assignment, and we sometimes allow them to work in groups of two or three. An alternative to this approach had students work on final projects of their own choosing. In this case, we make fewer lab assignments.

Software

For most of the labs, statistical software is needed to analyze the data. The exceptions are Chapters 2, 5, and 12. For these three chapters a statistical calculator is sufficient.

We have had success in using the software S-plus and R in the course. For those unfamiliar with R, the syntax of the language is similar to that of S-plus, and it is free. Students can easily down load a copy from the worldwide web (`lib.stat.cmu.edu/R/`) to run on their PCs, and so be able to work on assignments at home.

We advise students to consult an introductory text on how to use the software. For S-plus, we recommend *An Introduction to S and S-plus*, P. Spector, Wadsworth, 1994. For R we recommend *An Introduction to R*, the R Development Core Team, which can be found at the R website (`lib.stat.cmu.edu/R/`) and copied at no cost.

In our experience, we have found it important to provide assistance outside of class time on how to use the statistical software. One place where we do this is in section, where we sometimes meet in a computer laboratory room to work on the assignment and provide advice as needed. We also build a Frequently Asked Questions (FAQ) web page for each lab assignment. The page contains sample code and answers to questions students have asked in office hours, class, and section. These FAQs are available at the *Stat Labs* website (`www.stat.berkeley.edu/users/statlabs/`).

Grading

It can be difficult to grade the lab reports, because the investigations allow students to be creative in their solutions to the problem. We often base our grading on four aspects of the report: composition and presentation, basic analyses, graphs and tables, and advanced analyses. Sometimes we also request an appendix to the report for technical material.