

Preface

This book began as a collection of lecture notes and has evolved over the last several years into its current form. My goal in putting the book together was to remedy some problems I perceive with extant introductory statistics texts. To that end, this book has several features that differentiate it from standard texts. First, typical introductory statistics texts are now commonly over 500 pages long. To be blunt, I don't believe that students actually read texts when they are so lengthy. So, although a 500+-page book may be chock-full of examples, illustrations, and interesting side information, such detail is offset by the fact that students will not read it. Here, I have tried to provide limited, but sufficient, examples of concepts and to keep the actual text limited to about 200 pages. In doing so, I have *not* covered a number of statistical methods that are not commonly used, like Fisher's exact test, Kruskal's Tau, and the Spearman correlation. While these, and similar, methods are certainly still used, they have limited application in contemporary social research. Thus, they can be investigated in detail by students when needed in the future (e.g., for a thesis or other research paper) via other sources.

Second, in this book, I integrate statistics into the research process, both philosophically and practically. From a philosophical perspective, I discuss the difficulties with concepts like "proof" and "causality" in a deductive scientific and statistical framework. From a practical perspective, I discuss how research papers are constructed, including how statistical material should be presented in tables and figures, and how results and discussion sections should be written.

Third, I emphasize probability theory and its importance to the process of inference to a greater extent than many books do. Most books discuss probability theory, but I feel its relationship with statistics is underdeveloped. In this book, probability is painstakingly shown to be the basis for statistical reasoning.

Fourth, I have included a number of homework problems at the end of almost every chapter, along with answers to the first half of them in each chapter in an appendix. There are several differences between the homework problems and answers I provide and those found in typical texts. First, I do not provide as many exercises as is commonly found in texts. In most chapters, there are approximately 10 exercises. The exceptions are Chap. 5, with 50 items, and Chap. 6,

with 40 items. My rationale for this is that for developing an understanding of most concepts, I do not believe students need to perform dozens of exercises: once they have done one or two, especially ones that are word based and involve more thought in implementation, they understand the process. With problems involving more complicated (or at least tedious) calculations—like the chi-square test of independence, ANOVA, correlation, and simple regression—performing repetitive calculations may even be detrimental, fostering a “losing the forest for the trees” problem.

Thus, a second difference between my exercises and those often found in texts is that I have tried to construct exercises that are interesting, that often may be solved in different ways, that are realistic in the sense that they often require thinking like a full-fledged researcher would, and that often involve information that cannot simply be dropped into a formula in a rote fashion. For example, in the exercises for one-way ANOVA modeling, I have three exercises without solutions (and three with solutions). In one, the raw data are provided, and the student must compute sums of squares and complete the ANOVA table. This problem tests the student’s ability to perform the basic computations. In another, a partial ANOVA table is provided, and all the student must do is understand the structure of the ANOVA table to complete it (i.e., $SST = SSB + SSW$ and $df(T) = df(B) + df(W)$). Finally, in another, the student is provided with group-level statistics like the mean and variance and must recover (two of) the sums of squares by “undoing” the variance calculations before constructing the ANOVA table. This problem tests the student’s deeper understanding of the connection between variances and sums of squares.

A third difference between my exercises and those found in other texts is that I use the exercises in some cases to introduce additional material not covered directly in the chapter and as further examples of concepts and methods discussed in the chapter. As an example of the former strategy, the last exercise in Chap. 6 introduces the basic ideas underlying capture-recapture methodology. The detailed answers to exercises exemplify the latter strategy: the answers are detailed much as a solution to an example problem would be within a chapter in a typical text. They are placed at the end of the chapter so that they do not disrupt the flow of the text and so that students can work the problems and check their thought and computation processes as they work them.

A final difference between this text and others is that I routinely use simulated or contrived data, rather than real data, to illustrate concepts. To be sure, I use real data from the General Social Survey heavily in both examples and exercises. However, while real data are useful for giving students concrete examples, I think simulated data are often more useful when trying to illustrate difficult ideas, especially when showing how unknowns can be estimated. In real data, the unknowns are unknown! Also, it is generally the case that real data are messy. It is easier, for example, to illustrate the notion of row and column percentages in cross-tabs using rounder numbers than may be found in real data. Perhaps as a better example, it is easier to illustrate the idea of within versus between sums of squares in ANOVA modeling by creating contrived data in which the within sum of squares is 0 and contrasting it with data in which the between sum of squares is 0.

Layout for a Course

The book is intended as a text for a semester-long course on introductory statistics for social science students. Princeton semesters are relatively short (12 weeks), and our courses have one-day-per-week “precepts” in which the large class is broken into smaller groups (of about 15 students each) that are led by preceptors (graduate teaching assistants or the course instructor). I use these precepts to teach students how to use statistical software and to help students work through homework problems. Given the precepts, there are two 1-h lectures per week. After discounting 3 days for exams, that leaves 21 lectures. Table 1 shows the organization of the lectures and the corresponding book chapters.

As the table shows, I give four exams in my course. The first covers the basics of the research process and descriptive statistics. The midterm exam covers probability theory and one sample inference. The third exam covers various two-sample statistical tests, like the independent samples t test, and the chi-square test of independence. The final exam covers ANOVA, the correlation, and regression modeling.

Lecture	Topic	Chapter
1	Overview of research	1,2
2	Data acquisition	3
3	continued	
4	Summarizing data	4
5	continued	
*	First exam	none
6	Probability theory	5
7	continued	
8	Central Limit Theorem	6
9	One sample tests	
10	Confidence intervals	
11	Review	none
*	Mid-term exam	
12	Two-sample tests/intervals	
13	continued	
14	Chi-square	7
15	ANOVA	8
16	Correlation & simple regression	9
17	continued	
*	Third exam	none
18	Multiple regression	10
19	continued	
20	continued	
21	Tables and figures	11

Table 1. Typical course structure based on book

<http://www.springer.com/978-1-4614-8572-8>

Using Statistics in Social Research

A Concise Approach

Lynch, S.M.

2013, XXIII, 229 p. 44 illus., Hardcover

ISBN: 978-1-4614-8572-8