Springer Texts in Statistics

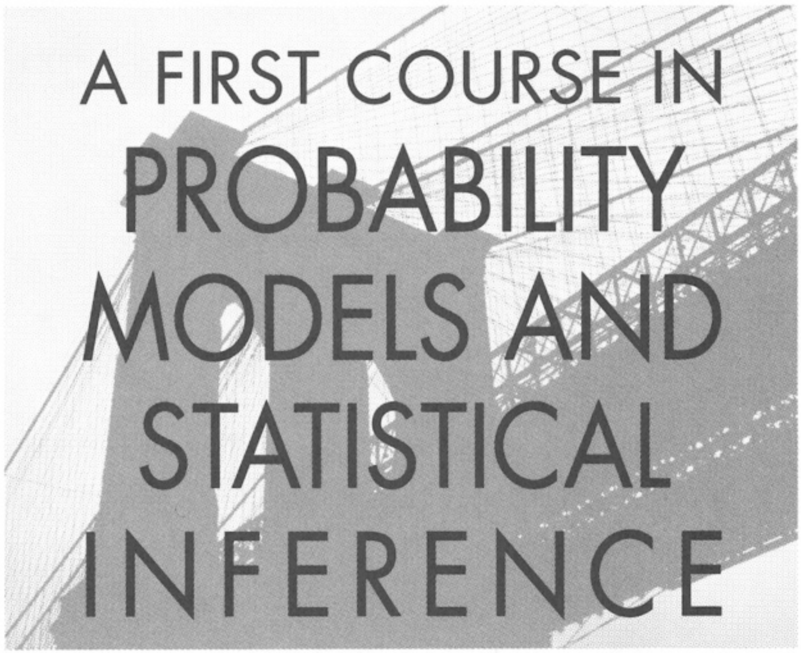
Advisors:
Stephen Fienberg   Ingram Olkin

Springer Texts in Statistics

Alfred	Elements of Statistics for the Life and Social Sciences
Berger	An Introduction to Probability and Stochastic Processes
Blom	Probability and Statistics: Theory and Applications
Chow and Teicher	Probability Theory: Independence, Interchangeability, Martingales, Second Edition
Christensen	Plane Answers to Complex Questions: The Theory of Linear Models
Christensen	Linear Models for Multivariate, Time Series, and Spatial Data
Christensen	Log-Linear Models
Creighton	A First Course in Probability Models and Statistical Inference
du Toit, Steyn and Stumpf	Graphical Exploratory Data Analysis
Finkelstein and Levin	Statistics for Lawyers
Jobson	Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design
Jobson	Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods
Kalbfleisch	Probability and Statistical Inference: Volume I: Probability, Second Edition
Kalbfleisch	Probability and Statistical Inference: Volume 2: Statistical Inference, Second Edition
Karr	Probability
Keyfitz	Applied Mathematical Demography, Second Edition
Kiefer	Introduction to Statistical Inference
Kokoska and Nevison	Statistical Tables and Formulae
Lindman	Analysis of Variance in Experimental Design

Continued at end of book

J H C   C R E I G H T O N



Springer Science+Business Media, LLC

J H C Creighton
San Francisco State University
1600 Holloway Avenue
San Francisco, CA 94132
USA

Mathematical Subject Classification (1991): 60-01, 62-01

Library of Congress Cataloging-in-Publication Data
Creighton, James H.C.
An first course in probability models and statistical inference / Creighton, James H.C. -- 1st ed.
p. cm. -- (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 978-1-4612-6431-6 ISBN 978-1-4419-8540-8 (eBook)
DOI 10.1007/978-1-4419-8540-8
1. Probabilities. 2. Mathematical statistics. I. title. II. Series.
QA273.C847 1994
519.5--dc20 93-25369

Printed on acid-free paper.

Production managed by Karen Phillips; manufacturing supervised by Jacqui Ashri.
Typeset by Digital Graphics Inc., East Lansing, MI.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4612-6431-6

आदिगुरवे नमः

Katherine Mayo Cowan
1883–1975

Student's Introduction

Welcome to new territory: A course in probability models and statistical inference. The concept of probability is not new to you of course. You've encountered it since childhood in games of chance—card games, for example, or games with dice or coins. And you know about the "90% chance of rain" from weather reports. But once you get beyond simple expressions of probability into more subtle analysis, it's new territory. And very foreign territory it is.

You must have encountered reports of statistical results in voter surveys, opinion polls, and other such studies, but how are conclusions from those studies obtained? How can you interview just a few voters the day before an election and still determine fairly closely how HUNDREDS of THOUSANDS of voters will vote? That's statistics. You'll find it very interesting during this first course to see how a properly designed statistical study can achieve so much knowledge from such drastically incomplete information. It really is possible—statistics works! But HOW does it work? By the end of this course you'll have understood that and much more. Welcome to the enchanted forest.

So now, let's think about the structure of this text. It's designed to engage you actively in an exploration of ideas and concepts, an exploration that leads to understanding. Once you begin to understand what's going on, statistics becomes interesting. And once you find it interesting, you don't mind working at it. It does require work. Statistics is hard; there's no getting around that. But it's interesting. Once it becomes interesting, the "hard" part is not too onerous.

The text is divided into three parts. The first part presents some basic information and definitions and leads you immediately to a set of exercises called "Try Your Hand." The second and third parts of the text are solutions to these exercises: "Solutions Level I" and "Solutions Level II." Leaf through the text and you'll see that the solutions comprise over half the book. They're much more than just "the right answer." They provide complete discussions for the issues raised in the exercises.

Over the first week or so you'll develop your own method for working with the Try Your Hand exercises. Different students learn in different ways, after all, and so you must find what works best for you. Still, a few words of orientation will help. First, it WON'T work to just read the problems and then read the solutions. As you read you'll be saying, "Yes, yes, yes. That sounds right. Yes, and that's right too. Yes, yes, yes. . . . " But afterward, when your instructor mentions some item from that list (on a quiz?), you'll swear you've never seen anything like it! In one eye and out the other.

Here's something else that won't work: Sometimes students are diligent in an unproductive way. One form of wrong diligence is to think you must master each step before going on to the next step. That's possible alright—if you have 30 or 40 hours a week to devote to this course. But it's not an efficient way to learn. The human mind is not a linear machine. Efficient learning is always grasping for the "total picture," grabbing something here and something there, leaving the details to be filled in later. Think how a small child learns. How she learns to talk, for example. Children are master learners, very efficient learners. They don't proceed in a logical step-by-step manner. They grab for everything at once. Of course, there does come a time when you're ready for careful detail and when you'll go over each step with a fine-tooth comb. But that's the last step in the learning process. In the beginning, trying for that kind of detail is counterproductive.

All of this tells you something about how to use this book. When you begin a new section, read over the text material quickly just to get an idea of what's there. Aim for the next set of exercises. Read the first question in the exercise set, reflect for a moment (30 seconds?) on what that question wants from you. Try to grasp the meaning of the question. Often you'll go back into the text for some detail. Realize that you may or may not be able to answer the question. Don't worry about that. Learning happens in the ATTEMPT. Whether or not you can answer the question right then is secondary. When you've given the question as much effort as seems appropriate, turn to the Level I solution. There you'll find help of some form—a hint, a clarification, the beginning of a solution, further information, and so on. Now, process this Level I information. Give a bit more time attempting to formulate a complete answer. You'll succeed if you've understood everything up to that point. But . . . if you're studying properly you WON'T have understood everything up to that point! That means even with the help of Level I, you may or may not be able to give a complete answer to the question. Try. Then turn to Level II for a complete solution.

So you see how the solutions are structured. Level I is help in some form, Level II provides a complete solution. Don't omit Level I! Level II will be meaningless without Level I if for example, Level I helps you, by providing a start on the solution, as sometimes it does. Or suppose Level I provides more information. If you skip Level I, you'll not have that information. Occasionally (not often) the Try Your Hand exercises ask a question that, as asked, CAN'T be answered. Ideally, you would respond by saying, "This question can't be answered." And you would explain: "It requires more information. We need to know this, this, and this." Or maybe you would say, "It's ambiguous, it could mean A or could mean B." As a practical matter, you'll struggle with questions like this, suffering a vague sense that something's wrong. And maybe—just maybe—after you've encountered a few such challenges, you'll have gained enough clarity to suspect something's missing or that there's an ambiguity. Great! You're learning. In any case, when you turn to the Level I solution, you'll get the clarification you need. But notice, it's in your struggle with the question that you learn, even though you don't "succeed" on your own in giving an answer or even in understanding the question!

Remember the principle: Grasp for everything at once. If you find you just can't understand some problem, GO ON TO THE NEXT PROBLEM! You can come back to this one later.

One more bit of advice: Find someone to study with! Better yet, form a study group with two or three other students. You'll be surprised how much you learn from each other. And it takes less time because the very point YOU get stuck on will be the thing that seems clear to someone else. Then, just around the next bend, what some other student finds totally mysterious seems clear to you. And both of you gain clarity trying to explain to the other person. Studying should be a social enterprise!

So success in answering questions is not your goal. The goal is to develop your understanding. And that happens as a result of your STRUGGLE to understand. That's why the exercises are called "Try Your Hand" instead of something like "Do It." Maybe we should have called the whole book "Try Your Hand," that's what it's all about.

Finally, a word about the technical terms: You will notice that technical terms are given in boldface italics in the very sentence where they first appear. Be sure you learn the definition for each such term PRECISELY, otherwise your thinking will necessarily be vague and confused everytime you encounter that term. Well, that's it—that's all the advice I can give. Good luck and have fun!

Instructor's Introduction

The concepts encountered in a first course in statistics are subtle, involving quite sophisticated logic. Understanding the statistical techniques presented in that course, understanding their application and the real-world meaning of the conclusions depends very critically on understanding those concepts. This text is my attempt to structure an exploration for the student leading to that kind of understanding.

The fundamental idea underlying the entire structure is the concept of a probability distribution as a model for real-world situations, a concept that is not in any sense elementary. Probability itself is hardly elementary. Still, even children grasp (naively!) the idea of probability as expressed in phrases such as "a fair coin" or, in card games, "the chance of drawing a spade." That being true, games of chance seem a good place to begin. That's Chapter 1. And on the first page of Chapter 1 comes the idea of a probability model. Thus the student has time—the entire course—to assimilate this concept with all its subtlety and later mathematical elaboration.

Before discussing chapter content in detail, a few general observations will be helpful. This "first course" covers the standard topics of introductory precalculus statistics, but with a very nonstandard presentation. The text itself is brief, leading the student quickly to "Try Your Hand" exercises in which the student actively explores concepts and techniques. Often, no more than a page or two passes before the student sees the next set of exercises. Much of the exposition usually given discursively in the text itself is presented here through these Try Your Hand exercises. This approach is possible only because the exercises have complete solutions with full discussions. The solutions are given in two levels. Level I gives hints, clarifications, further information, partial solutions, and so on. Having completed Level I, the student turns to Level II to find a full solution.[1] A number of problems with real data have been included to give some idea of the variety of applications of statistics and some feel for the unexpected questions which arise in specific situations.

By judicious use of the Try Your Hand exercises, the instructor can focus the course according to the students' needs and abilities. Some exercises might be omitted altogether; others might be presented by the instructor in class. For example, probability formulas for the discrete distributions of Chapter 3 are derived in the exercises. The student is led by the hand through the derivation with the help of Level I. An instructor who chooses to omit formula derivations can simply omit those exercises. The student will be totally unencumbered by the derivations. She won't even see them because they are not included in the main body of the text itself.

The presentation given in this text is more sophisticated with respect to the underlying logic of statistics than most introductory books. In Problem 6.2.21, for example, the student sees that two very different p-values could arise from the same objective data, depending on how that data is "modeled." The ramifications of that are discussed in Level II. In Problem 5.5.12, two different prediction intervals for the same problem—one parametric, the other nonparametric—are compared. The discussion of statistical testing in Chapter 6 is more thorough than in any book I know of at this level. The presentation of regression in Chapter 7 avoids the usual list of unmotivated assumptions. Instead, I give a natural characterization which reveals simple linear regression as the next logical increment in complexity beyond previous chapters. For further detail on all of this, please see the individual chapter discussions given below.

Finally, I give much in the way of informative heuristics, such as the "elementary errors" interpretation for the normal distribution introduced in Chapter 4. That criterion, which is just an intuitive formulation of the Central Limit Theorem, is used systematically throughout the course to explain many details otherwise left obscure. To give one example, the criterion explains why only in the small sample case for inferences about means we must assume we are sampling from a normal distribution (see Problems 5.3.2 and 5.3.3).

[1] The instructor should remind the students (often) not to neglect the Level I solutions. Level II will seldom be complete in itself.

This text should be appropriate for students requiring an elementary introduction to statistics, assuming little mathematical sophistication, who anticipate more than a passive involvement with statistics and who will be learning more sophisticated statistical techniques in later courses. It should be appropriate for engineering, economics, computer science, psychology, sociology, and education majors. It would probably be appropriate for students in fields such as geography, ecology, and so on. It certainly would be an ideal freshman introduction to statistics for mathematics majors with the expectation of later follow-up in a thorough Mathematical Statistics course. In my experience, it is difficult to teach a meaningful Mathematical Statistics course to students with no prior exposure to statistics and, consequently, with no intuitive orientation.

Chapter 1

The discussion above suggests that the probabililty distribution of a random variable is a central idea. Certainly it is. It's a sophisticated concept at the heart of virtually every technique of statistics. Even for exploratory and nonparametric techniques a probability distribution often lurks in the background (at least) as a standard of comparison. And, of course, the prior probability distribution is fundamental to Bayesian statistics.

To ease the student's initial exposure to this abstract concept, I introduce probability distributions at the very beginning through simple examples of real-world situations, namely, games of chance with coins, dice, cards, and so on. This term "real world" is used throughout the text in contrast to "abstract theory" (not in contrast to "artificial" or "contrived"). Random variables and linear functions of random variables are used in this chapter to model these simple games of chance. For example, if you receive two dollars for each dot on the uppermost face of a die and pay six dollars to play, your gain/loss random variable is $G = 2X - 6$, where X models the die. Thus, linear functions of random variables are also explored in some depth with the simple examples of Chapter 1. Most of this chapter relies on the student's intuitive idea of probability together with some simple ad hoc rules. After "reminding" them of what they already know, a more rigorous development of probability is presented in Section 1.4.

Betting games with dice, loaded in various ways, are effective for investigating the mean and variance of a random variable and for understanding the variance as a measure of risk, predictability, accuracy, and so on. The significance of the variance is much more readily appreciated by students in the dynamic context of random variables than in the static context of observed data, another reason for introducing random variables at the very beginning of the course. The die—in general, the random mechanism for the game—need not be fair. Loaded dice offer a variety of interesting situations and an opportunity to understand interesting concepts. For example, we can ask the student to load a given die differently so the mean stays the same, within a specified degree of accuracy, but the standard deviation is smaller or larger by a specified amount. Or again, we can ask which is more predictable, a die with a given loading or a fair die? Which game would you prefer to play, the one with the fair die or the one with the loaded die? The answer depends not only on how the die is loaded but also on one's motivation for playing. Students readily appreciate the relevance of this kind of analysis to more realistic situations such as portfolio analysis or variability in a manufactured product, or variety within a genotype, or any of many other situations of possible interest to the student.

Random variables are presented as providing a "bridge" from the real-world situation with all its complexity to the relatively simple world of theory (see the picture of this bridge on page 7). This metaphor is much more than a clever hook—it is a significant pedagogical device. I have the picture of this bridge on the blackboard every day for the first half of the course and am constantly surprised, and surprised again, how often misconceptions can be resolved by reference to this picture. Many errors arise from confusing the outcomes with the values of the random variable. I simply point to the picture and students catch their error immediately (well, almost)! For example, isn't a "constant random variable" a contradiction in terms? After all, "it" is predictable! Look at the picture: "It" (the value) is predictable, but "it" (the outcome) is not. Or, if we have a pair of fair dice, isn't the number of dots on the top faces uniformly distributed? Since the dice are fair, "they" are equally likely. Yes, "they" (the outcomes) are equally likely, but "they" (the values) are not. The bridge metaphor is particularly helpful in resolving the confusion of values with outcomes because it places them symmetrically on opposite sides of the River Enigma.

Chapter 2

This chapter covers topics of descriptive statistics, introducing frequency and relative frequency distributions, their histograms, and related ideas. In this chapter, we introduce random sampling: first sampling from a probability distribution, then sampling from numeric and dichotomous populations. After Chapter 1, sampling from a distribution seems quite natural to students. It's conceptually simpler than sampling from a population. The key question for sampling from a distribution is simply the independence for repetitions of the underlying random experiment. For example, if we're interested in monitoring "fill" for cups from a soft drink vending machine, 10 cups taken in succession will be a simple random sample from the distribution of "fill" provided only that the amount of drink dispensed is independent from one cup to the next. Students are readily able to suggest how that might or might not be true depending on the circumstances.

Most, if not all, introductory statistics texts seem to avoid sampling from distributions—a wise choice if the students don't understand what a distribution is—and, consequently, many examples throughout such a text force very artificial interpretations of sample data by reference to some hypothetical nonexistent population where certainly the data was NOT selected through any sampling plan. Did we use a random number table to select 10 cups from a population of "all possible cups?" In exactly what warehouse are "all possible cups" to be found? If that's not the procedure, what justifies calling those 10 cups a random sample? Interpreting such examples as "sampling from a population" not only does not help, it's a serious obstacle to clarity. You dare not ask the student if the assumption of randomness would, under the circumstances, seem justified. The definition before them is so artificial that any practical discussion of its relevance is impossible. An intelligent student can only conclude that one blindly assumes whatever one wants in order to make the theory work. I prefer the idea that one makes assumptions only where those assumptions seem reasonable and where they can later be verified.

Chapter 3

This chapter presents nine "models," nine classes of discrete probability distributions of varying degrees of concreteness, interrelated in various ways. I strongly urge that none of these models be omitted. Understanding a sophisticated, abstract concept—here, probability distributions—requires more than one or two examples. The goal of this chapter, at least from the point of view of the statistical material to be presented later, is to develop the student's skill in recognizing an appropriate model for a real-world problem. This requires experience with a number of different models. The instructor can mitigate the difficulty of this chapter without compromising its principal thrust by omitting some or all of the probability formulas, focusing instead on model recognition. Skill at model recognition can be developed and tested without probability questions per se, by simply restricting to questions about the mean and variance, questions such as, "How many cups should we get from this drink machine before the machine malfunctions?" (the mean of a geometric random variable).

The statistical topics of this course—random sampling, sampling distributions, estimation, statistical testing, and the regression model—cannot be understood if the underlying theoretical models are not understood in their roles as models, models for the sampling process or for the more complex situation of regression. In my experience, Chapter 1 and the nine models of this chapter will indeed bring the students to the requisite understanding of probability distributions as models for real-world situations.

The discussion from Chapter 1 is continued at a more sophisticated level with the nine models of this chapter. Constant and uniformly distributed random variables form two very simple classes, already familiar from Chapter 1. In particular, Chapter 1 has already shown how constant random variables arise very naturally through combinations of other random variables. For example: $X + Y = 7$, where X and Y are, respectively, the number of dots on the top and hidden faces of a six-sided die. These two simple examples—constant and uniformly distributed random variables—help us to establish what we mean by a "class" of random variables and set the pattern for the rest of this chapter.

The classes of this chapter are interrelated in various interesting ways. There are two sampling distributions: Sampling with or without replacement from a dichotomous population form one group. The binomial, geometric, and negative binomial form another group (with the geometric a special case of the negative binomial). The binomial has the previous "sampling with replacement" model as a special case. The Poisson model is the most abstract of the models in this chapter, having been derived abstractly through a purely mathematical process from the binomial. It becomes a model for real-world situations only after the fact and for that reason has a less concrete feel about it. The student is alerted to this "abstract" versus "concrete" consideration. That idea is picked up again in Chapter 4 where we introduce continuous distributions. The understanding that some models are more abstract than others is helpful in understanding the normal and chi-squared distributions which are indeed quite abstract.

An important challenge in this chapter (and again in Chapter 6) comes in the set of mixed review problems at the end. In these problems, the student is on her own to identify a correct model for a given problem. A few problems can be correctly modeled in more than one way. This review is very important and should not be omitted. A real difficulty for students, which will show up in these review problems, is their single-minded focus on the abstract part of the model. Students complain about abstraction, but, in fact, they love it—it's easier. The abstraction is precise and clear; equations and formulas can be learned. The real world, by unhappy contrast, is messy, ambiguous, and confusing. However, to identify an appropriate model for a real-world problem we have to look where the problem is—in the real world. That requires focusing on the real-world description of the random experiment, the real-world component of the model, and matching that description with what's going on in the problem. The skill to do this is developed through these review problems at the end of the chapter.

Chapter 4

This chapter extends the presentation of the previous chapter to continuous distributions. First are the uniform and exponential distributions; then the most abstract model so far encountered, the normal distribution, modeling random error or, by extension, any situation where the difference in two values "looks like" random error (see the criterion for normality on page 148. Finally, we see the chi-squared distribution, the most abstract of all among the distributions of this text.

Chapter 5

This chapter presents sampling distributions, the Central Limit Theorem, and interval estimates as a unified topic. We do three types of interval estimates: confidence intervals, prediction intervals, and tolerance intervals.

For understanding sampling distributions, variability from one sample to the next is not the really difficult concept. People with no knowledge at all of statistics see this variability very clearly. That's why they're so ready to criticize statistical surveys, complaining that "it's all just based on a sample!" This is the thinking students come to us with. We must show them that

They're right if they think one sample alone can't tell them anything.

BUT . . . They're wrong if they think sampling is useless or statistics a sham or if they think only very large samples are legitimate. And they're especially wrong if they think a very large sample carries any information by itself.

BECAUSE . . . The missing ingredient which makes sense of one sample is the *entire context of that sample*. That "entire context" is the sampling distribution, a sophisticated theoretical construct not easily understood.

For example, a sample mean by itself tells you nothing. On the other hand, a sample mean seen as just one of the many possible values of a normal distribution centered on the unknown true mean, with most of the probability concentrated there and with a standard deviation intimately related to the standard deviation of the original distribution, TELLS YOU A LOT!

When asked, "What assumption must you make about the data here?" (that it's a random sample) students will respond, "That it's more or less typical of the population." Well, if it's typical, you don't require statistics! But, in fact, you'll never know whether it's typical or not and YOU HAVE NO THEORY FOR THAT. There is, however, a theory for random sampling and that theory controls the error which could arise from a possibly atypical sample. Control of error is the theme, probability distributions the tool.

There may be an objection to the presentation in this chapter which treats only interval estimates and does not allow for point estimates. But point estimates are only appropriate when the estimator is in some sense "best" for the problem at hand. That more advanced discussion involves everything in the discussion above and more. For this reason, I present only interval estimates with the understanding that the point estimate is incomplete to the point of being meaningless if no further investigation is carried out.

There may also be an objection to interpretations of confidence intervals which begin "There's a 95% probability that. . . . " The usual argument tells the student to replace the wrong expression by another one where the offending term "probability" is replaced by the undefined term "confidence." This just replaces error by ignorance, hardly an improvement! No wonder highly intelligent people say they never could understand statistics. I prefer to use a natural probability expression, but acknowledge openly that it's *ambiguous*. One reading is wrong (with the 95% probability referring to the parameter), the other correct (it refers to the interval). The student is held responsible for understanding the two readings, understanding why the one is wrong and the other correct.

This approach is consistent with my exposition throughout the text, where I hold the student responsible for certain standard ambiguities or misstatements. For example, the question "What are the chances for a female on this committee?" almost certainly is asking for $P(X \geq 1)$, although technically it asks for $P(X = 1)$. Or I leave out the phrase "on average" in situations where it's clearly implied. Or, again, I ask for a count where a proportion is all that's possible. This significantly challenges the student's clarity of thought because she is on her own to discriminate among possible meanings. "How many," for example, might mean "how many on average" or "what proportion." There's no getting around it, she has to understand the context! In this way, the student of this text grows accustomed to dealing with ambiguity and the resolution of ambiguity as a part of the natural intellectual environment. All of this is possible through judicious use of the Level I answers to the Try Your Hand exercises.

Chapter 6

This chapter presents tests of statistical hypotheses. This vexed topic is presented with in-depth discussion of the possible misinterpretations, misuses, and limitations of the technique as well as a careful discussion of the correct interpretation of *p*-values, conclusions, and errors. The first section of the chapter gives an overview of two testing procedures, the "test of significance" (*p*-values) and the "hypothesis test," and introduces some terminology and some comparisons (the details of which are deferred until later sections). The second section presents tests of significance, including chi-squared tests. The third section of the chapter presents the "hypothesis test" as providing a decision procedure for a monitoring process as, for example, in quality control.

So the "test of significance" is presented first, with its formal answer (a *p*-value) and its not-so-straightforward real-world interpretation. The logic of statistical testing is quite subtle, involving considerable controversy. A significant amount of confusion has been introduced into the topic by not distinguishing between the "test of significance" and the "hypothesis test" proper, with its error probabilities, power considerations, and so on. Consequently, I have separated these two approaches to statistical testing.

The student's understanding is enhanced by first clearly understanding *p*-values as measuring consistency between the data and the hypothesis. The *p*-value calculation is relatively easy. Nevertheless, the real-world interpretation is not so straightforward. For instance, although small *p*-values are usually what we look for, with not small *p*-values being inconclusive, just the reverse may hold for the practical interpretaion. For cases of "discriminatory selection," for example, where the hypothesis to be challenged is "random choice," a NOT small *p*-value is quite conclusive—it's impossible to maintain an accusation of discrimination if the choice is consistent with having been random—whereas a small *p*-value may be relatively conclusive or relatively inconclusive, depending on the context. See the discussion beginning on page 245.

There are a number of advantages to presenting tests of significance as a separate topic. Certain issues concerning statistical testing are more clearly presented with reference to *p*-value calculations, unburdened by the heavy-handed and irrelevant machinery of null and alternative hypotheses, error probabilities, rejection regions, decision rules, power, and so on. For example, the test of significance already highlights the distinction between practical and statistical significance. It also reveals the asymmetry inherent in statistical testing, the asymmetry seen in the difference between "small *p*-value" (the hypothesis is challenged) and "not small *p*-value" (the data is inconclusive). Further, the test of significance, seen as a "probabilistic argument by contradiction," clarifies the underlying logic of statistical tests in general. Again, chi-squared tests are more appropriately introduced as tests of significance than as hypothesis tests because there is often no meaningful "alternative" hypothesis. Finally, by presenting tests of significance as a separate technique, instructors who prefer to spend less time on statistical testing can omit the complexity of "hypothesis tests" altogether and confine their discussion to this much simpler case.

Hypothesis tests are first studied without control of type II error. Our point of view is that when you "fail to reject H_o," you take no action based on the test itself since in that case you have exercised no control over the possible error. This asymmetry is exactly parallel, of course, to the asymmetry of "small" versus "not small" p-values. The important new idea here as compared with tests of significance is the "control of error" for type I error. Aside from the pedagogical advantages, leaving discussion of type II error until later is justified by the impossibility in some testing situations of finding a model for the alternative hypothesis. In other words, there are, indeed, situations where control of type II error is not practical.

This chapter gives an in-depth discussion of the role of hypothesis tests. For example, we see how the logic of hypothesis tests compares with classical inductive inference, leading to the distinction between null hypotheses which are sometimes true and sometimes false (in monitoring situations) and null hypotheses which are either always true or always false (the classic situation of inductive inference, where "accumulation of evidence" is the motivation for repetitions of the experiment).

There may be an objection to so much emphasis on the logic of testing. It has even been suggested that statistical tests should be omitted altogether because the same conclusions can be obtained from a confidence interval.[2] But statistical testing is too pervasive in statistical practice to justify omitting it. The student is not well served in being left ignorant of terms like "p-value," "null hypothesis," and so on. Given that we're going to teach the topic at all, surely we must teach it clearly so that common confusions and misunderstandings do not arise.

The distinction between tests of significance and hypothesis tests is certainly not artificial. Failing to make that distinction leads to a number of points of confusion. To name only one: Are you allowed to look at the data before setting up the test? For a test of significance where the question is "Does this data seem to challenge our hypothesis?" there is no "setting up" of the test. The data is part of the original question.

[2] As a matter of technical fact, this is not true in the case of proportions since the standard error will differ. If the hypothesis to be tested is false that difference could be significant.

How can you avoid looking at it in advance? For an hypothesis test, on the other hand, which is properly a monitoring procedure, the data will change from one run of the test to the next, so, OF COURSE, you have to set the test up without reference to the data. Most textbooks tell the student not to look at the data and then, in every example and problem, give the data in the problem statement. How can the student avoid looking at it? To add insult to injury, the solution—which here means simply choosing a direction for the test—will always be correct if you base it on the data. Never does the student see data which would be in the "wrong" tail of a properly determined test.

For other points of confusion which arise when the distinction between tests of significance and hypothesis tests is not made, see the text of Chapter 6. None of this says the distinction remains necessary for someone who clearly understands the entire logic of tests. But we should distinguish between what is logically correct and what is pedagogically clear.

At the end of this chapter, as at the end of Chapter 3, there is a critically important set of mixed review problems, the most challenging set of problems in the text. It should not be omitted. I give about a week of class time to these problems. It is through this set of problems that students assimilate the statistical techniques of Chapters 5 and 6.

Chapter 7

This chapter is a brief introduction to simple linear regression. In Chapter 4, anticipating the present chapter, the normal distribution is described along these lines: Suppose you have a fixed systematic "effect" for which any variation is due solely to something that "looks like" random error. Then you should expect a normal distribution. Take, for example, diameters of machine parts where the systematic "effect" is the manufacturing process itself which attempts to meet specifications (diameter 3.2 mm). The variability in diameters is purely random unless there's something wrong in the process. So diameters are $D = \mu + \epsilon$. Here $\mu = 3.2$ is the systematic part with ϵ being "like random error" so that $\epsilon = N(0, \sigma^2)$.

When we come to regression, we make ONE SIMPLE STEP FORWARD IN COMPLEXITY for the model. The "effect" which determines the mean is no longer fixed, but variable. But not variable in just any way at all; that's much too complex. Instead, the mean is determined through a linear function of the effect, a linear function being the simplest nontrivial function. So we obtain a model with a variable "effect," X, which is as simple as possible and which "affects" (not necessarily causally!) only the mean of Y, not affecting Y in any other way. The usual long, intimidating and unmotivated list of assumptions for the regression model follows quite naturally from this characterization [see problem 7.1.4(c)].

Of course, one may object, X can certainly affect more than just the mean of Y. For example, it can affect the percentiles. But that's not a *different* effect. If you tell me the effect on the mean of Y, I can determine the effect on the percentiles. This is parallel to what we say about the parameters of a model. For example, σ and σ^2 are not two different parameters; give me one I can calculate the other. For the hypergeometric model, p is not a fourth parameter because I can calculate it from two of the other three: $p = R/N$.

Here again, as in many other instances—confidence intervals in particular—I allow possibly ambiguous statements when it's convenient and natural, making a point of the ambiguity and its proper resolution. Understanding comes in being clear about the ambiguity. So, in the regression model, X affects only the mean of Y, but not in an absolute sense, rather in the sense that any other effect from X can be calculated from the effect on the mean.

With some hesitation, I will describe how I currently use this text in my classroom. I hesitate simply because I would not want to prescribe a "right" way to use the book. Indeed, I hope various instructors will find various effective ways of using the text.

At the beginning of each class, I make an assignment for the next class. The students are expected to read that material and process the problems on their own with no preliminary in-class discussion. At the beginning of the next class, there is a brief, very routine quiz on the assignment. These quizzes serve many useful purposes, not the least of which is to encourage students to actually do the assignment. Needless to say, when the students have already thought about the material, the class discussion can deal with issues in much greater depth and subtlety. For sections which meet three days a week in 50 minute classes, there are approximately 20 such quizzes per semester of which I drop the lowest three or four. These quizzes count 20% of the course grade. The rest of the course grade is determined by two 100-minute tests (each given over two successive class periods) and the final examination.

The grading policy for the quizzes is very lenient because the material is quite new at the time of the quiz. The quizzes provide an opportunity for the students to catch errors or misunderstandings without incurring a serious penalty. Thus I count off only for gross errors or completely wrong approaches which would indicate that the student did not really do the assignment. I often write "OKT" beside a mistake, meaning, "No penalty THIS time, but be alerted: this is an error!"

I allow questions before the quiz. If the question is relevant to the quiz, I am obliged to answer. If the question is not specifically relevant to the quiz, I may postpone discussion of that question until later. Attention in the class is never quite so clear and focused as during that question period before the quiz! In practice, the quiz is sometimes at the very beginning of the class, sometimes fifteen or twenty minutes into the period, sometimes not until the very end. Rarely—in the interests of time—I may omit a quiz (unannounced in advance) to spend the entire class going carefully over some topic or conducting a review.

There is always initial resistance from the students to this approach. It seems to be quite a novelty that they should be expected to read material and assimilate it on their own. But usually after a week or so they begin to accept the responsibility to work on their own and begin as well to appreciate the value of developing their skill for independent study. The structure of this text, with its complete solutions to the problems, facilitates this approach.

My experience shows that students coming through this course develop valuable skills for independent study and for critical, analytical thinking. In fact, for many students those are possibly the most valuable results of the course.

Acknowledgments

It is a pleasure to acknowledge the help of many people in the writing of this book. First of all, I would like to thank my colleague Al Schainblatt for his continuing encouragement. More than anyone else, he has been close to this project from the beginning, sharing his ideas and insights and providing very necessary critique. I would also like to thank Gerry Manning who encouraged me in this project at the outset. And I thank Jamie Eng for cheerfully responding to odd-hour phone calls and knocks on her office door with questions on arcane points. Not at all arcane have been her excellent problem ideas which I have used. Both Gerry Manning and Jamie Eng have been very cooperative as department chairs in arranging class schedules to accommodate the writing and classroom testing of this text. I would like to thank two other colleagues, Paul Schmidbauer and Rick Wing, who offered helpful and detailed suggestions on chapters which they read. I would like to thank Edward Tufte and Tim Redmond for permission to reproduce some very instructive material. My friend and colleague Bharat Sarath of New York University has been a valuable source of information and discussion. Finally, I would also like to thank David S. Moore for his help.

Of course, I must thank the many unnamed students who have been my audience over a period of six years and for whom I really wrote the book. Many other people have read parts of the manuscript providing feedback or have helped in other ways. Among them I would like to thank David Axel, Dana Dixon, James Lipka, Kenneth McCormick, and Carol Whitfield.

I would like to thank Martin Gilchrist, my editor at Springer-Verlag, and the book production staff. All authors should have editors so cheerful, patient, and helpful. Finally, I would like to thank the series editors, Stephen Fienberg and Ingram Olkin. At the very beginning, they recognized some value in my proposal before there was any text to substantiate it. But I thank them with a special sense of gratitude for their patience and encouragement at a particularly difficult time when I thought this book would never see the light of day.

I would like to acknowledge the reversal—often—of traditional gender roles of English pronouns throughout this book. If the reader finds the effect jarring, as I sometimes have, prehaps she will also find it amusing (or even instructive) to reflect on exactly why such "gender bending" seems so disturbing. I certainly have. Finally, I would like to acknowledge tireless support and encouragement throughout six years of tedium and toil from my loving helpmate, if only I had one.

Contents