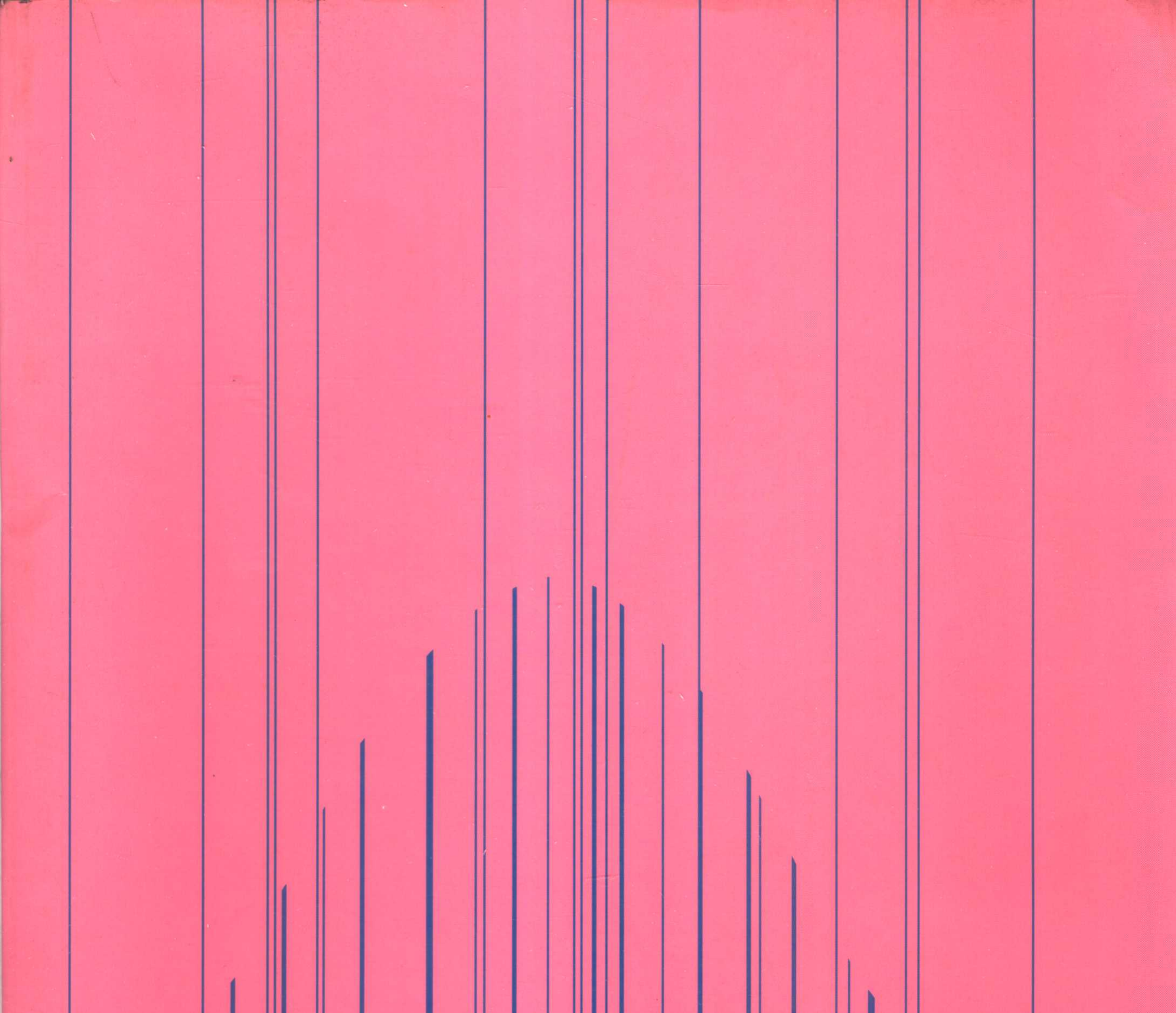


Student's Guide and Workbook

ELEMENTS OF STATISTICAL REASONING

Edward W. Minium
Robert B. Clarke

prepared by Gordon Bear

STUDENT'S GUIDE AND WORKBOOK TO ACCOMPANY

ELEMENTS OF STATISTICAL REASONING

EDWARD W. MINIUM
ROBERT B. CLARKE

prepared by

GORDON BEAR
Ramapo College of New Jersey

JOHN WILEY & SONS, INC.

New York • Chichester • Brisbane • Toronto • Singapore

ISBN 0 471 86463 3
Printed in the United States of America

10 9 8 7

Printed and bound by Malloy Lithographing, Inc.

Acknowledgments

Many people deserve thanks for their contributions to this book. I am grateful to Ed Minium and Bob Clarke for the inspiration they provided in the form of their unusually fine text *Elements of Statistical Reasoning*, to which this *Guide* is a companion, and for their valuable advice on this *Guide*; to Bob Abelson, Barry Collins, and Fred Sheffield for their conscientious instruction as my own teachers of statistics; to my teaching assistants at the University of Wisconsin, who worked diligently with me and provided helpful advice for improving my instruction; to my students at Wisconsin and at Ramapo College, from whom I have learned a great deal about pedagogy; and to my colleagues near and far Gary Beecher, Jane Goldman, Mike Lougee, Elsa and Les Wormack, and Bob Worsham, for welcome help with a variety of matters.

In spite of their many other obligations, the staff of John Wiley & Sons offered me all the personal attention I needed. I am especially indebted to Jack Burton, Elizabeth Doble, Carol Luitjens, Anita Kann, and Pam Cassell for their expertise and their good will.

I created this book via the computerized word-processing system SUPER-WYLBUR, and for help with the hardware, the software, and the operation thereof, I am grateful to the staff of the New Jersey Educational Computing Network, especially Dan Devine, Elaine Kaltenbach, Frank McArthur, Dave Norman, Eva Valentine, and Carol Yunker. At Ramapo College, Sue Wagner, Diane Coles and others in the computer center provided valuable assistance, and the College paid for my computing time. Chris Muller of Muller-Mugno Corporation and Ray Roesser of Science Typographers taught me everything I know about compositor codes, and I am particularly indebted to Daniel Breen of Anderson Jacobson, Inc. Dan supplied my hardware at a fair price, and it worked flawlessly: an AJ 880 teleprinter, an AJ 460 Micro Diskette System for offline storage, and an AJ A242A acoustic coupler.

I am also grateful to Dribble and Nibble, a.k.a. Slugger Paws and Critter Cat, who offered quiet companionship during many long nights of work.

For Evy ... you know why.

Using this Guide: How and Why

This *Guide* can help you learn the material in Edward Minium and Robert Clarke's textbook *Elements of Statistical Reasoning.* To get the most from the text and this *Guide*, do the following when you set out to study a chapter in the text. (But you need not do this all at one sitting; in fact, you probably shouldn't.)

1. First read the **Objectives** listed in the corresponding chapter of this *Guide*. They'll give you a preview of what's in the chapter and some specific goals to work toward.

2. Also before you begin studying the text, you *may* find it helpful to inspect the **Map** in the corresponding chapter of this *Guide*. The map is an illustration showing how some of the concepts in the chapter interrelate. It too will give you a preview of the chapter—but it may provide too much detail at this time. If it looks complicated, don't let it scare you. When you've worked through the chapter, you'll understand the map.

3. Study the **Chapter** in the text. And as you do so, keep referring back to the objectives. Try to see how each section of the chapter furthers your progress toward one or more of the objectives. In fact, you may find it helpful to construct a kind of index by adding next to each objective the number of any relevant page from the text.

4. Return to this *Guide* and reread the **Objectives**, to fix your goals in your mind.

5. Then do the **Review**. The review will lead you through the important points of the chapter, giving you the opportunity to practice recalling them. Such practice in retrieving information from memory is an important technique for learning the information. You'll also be constructing a summary of the chapter that's more detailed than the one in the text.

In working through a review, fill in the blanks by writing words in the spaces provided in the margin. Consult the list of answers in the back of this Guide, *and make any corrections in the margin. Later on, if you wish to re-review, just cover the margin. This time, jot your answers on a piece of paper, and only after committing them to writing, uncover the margin to check yourself.*

Warning: you will not be able to "cram" successfully by filling in the blanks in a review. Doing a review cannot substitute for careful attention to the text.

6. Study the **Map** for the chapter. As noted above, this is an illustration showing how important concepts in the chapter interrelate. It's essentially a series of sentences; the end of a sentence is shown by an arrowhead. As you inspect a map, then, pay attention to the arrowheads; they tell you where the sentences end.

You may be able to improve on the map, adding information or rearranging what's there. If you can, do so. Working actively with ideas is a fine way to learn them.

Beware: because of limitations on space, the map does not cover all the important points in the chapter. It will help you learn whatever it shows, but there will be more that you should learn from the chapter. Again, you may wish to create your own map of the omitted information.

7. Study any **additional material** in the *Guide* for the chapter you're working on. There are such things as notes on memorization and tricks for catching your errors before your instructor does.

8. Back in the text, work at least some of the **Exercises** the authors have helpfully provided. By giving you practice in using concepts and applying techniques, they solidify your knowledge of the chapter.

9. Return to this *Guide* for the **Self-Test**. *Mark your choices only in the margin, and after you check the answers in the back of the* Guide, *make any corrections only in the margin. In this way, you can test yourself again — as you might wish to do when you review for an examination. Just cover the margin.*

If you do poorly on the self-test, you know you need to work at the chapter some more. But be careful: with its 10 items, the self-test cannot cover all the important points in the chapter. So doing well on a self-test means only that you know *some* of the important points in the chapter—not necessarily all of them. (To identify the important points, use the Objectives and the Review.)

Also available in this *Guide* are one or two sheets of **Homework Exercises** for each chapter. Your instructor may assign these. If not, you can use them as exercises supplementing those in the text. Ask your instructor for the answers.

You may find it helpful to keep track of your progress by making notations on the lists of objectives. To indicate how far you've come in meeting an objective, you could write in next to it something like "Solid," "Okay," "Unsure," or "Ask about this."

Contents

CHAPTER 1

INTRODUCTION

OBJECTIVES

Ch. 1 describes the purposes of statistical techniques and introduces some basic concepts and principles. In studying the chapter, work to understand these matters:

- Descriptive Statistics
- Inferential Statistics
- Relationship as a Problem of Description and Inference
- Prediction as a Problem of Description and Inference
- Population
- Sample
  - Random Sample
    - Random Sampling Variation
- Variable
  - Quantitative Variable
    - Numerically Scaled Variable
  - Qualitative Variable
- Substantive Question • Substantive Conclusion
- Statistical Question • Statistical Conclusion
- Criticisms of Statistics • Rebuttals to Criticisms
- Tips on Studying Statistics
- Exact Number • Approximate Number

REVIEW*

What Is (Are?) Statistics?

In ordinary speech, the term *statistics* refers to statements involving numbers, as in the expression "unemployment statistics." The word is plural in this sense ("The statistics *are* due to be released soon"). In Ch. 1, however, the text uses the term in a different sense, to refer to a specialization within the discipline of mathematics. Statistics, in this

*For advice on how to do a Review, see the section on "Using this Guide" at the beginning of this book.

sense, is like arithmetic, geometry, and algebra—a set of mathematical techniques for solving problems. In this sense, the word is singular ("Statistics *is* a field of mathematics").

Descriptive and Inferential Statistics

The techniques of the field called statistics are helpful for solving problems that arise in research when an investigator collects observations. There are two kinds of statistical techniques. One kind is called descriptive statistics, and the other kind is called _1_ statistics. 1.

The purpose of _2_ statistics is to organize and summarize observations 2.
(data). Calculating an average and drawing a graph are examples of the techniques called _3_ statistics. 3.

The techniques of inferential statistics are helpful when a researcher goes to make inferences from the observations. Chance factors usually play a role in any investigation, and the purpose of _4_ statistics is to 4.
take account of the _5_ factors when drawing conclusions. Inferential 5.
statistics is also called _6_ statistics. 6.

Relationship and Prediction

Two important problems that frequently arise in research have to do with relationship and prediction. A researcher may wish to know how one characteristic of his or her subjects is related to another characteristic of the subjects. For example, the researcher may wish to know how strongly a person's score on a certain test is related to that person's score on the same test when it is readministered four months later. If the score is influenced by a stable characteristic like the person's intelligence, there will be a strong relationship between the first score and the second. That is, people who scored high the first time will generally score high the second time, while people who scored low the first time will generally score low the second. But if the score is markedly influenced by an unstable characteristic such as the subject's current mood, there will be little _7_ between the first score and the second. To assess the extent of 7.
the relationship between one characteristic and another, a researcher will collect data and apply _8_ statistics to organize and summarize the data. 8.
To reach a conclusion, the researcher will apply _9_ statistics. 9.

Questions about prediction arise when a researcher knows one characteristic of the subjects and wants to use that knowledge to estimate another characteristic. For example, a personnel office may administer an

aptitude test to people seeking clerical work, and the office may hire only
people who earn a sufficiently high score on the test. But does the score
really allow an accurate prediction of how well the person will do on the
job? To answer such a question, a researcher must collect observations,
use descriptive statistics to _10_ and _11_ the observations, and then 10.
apply inferential statistics to draw a(n) _12_ from the observations. 11.

Problems involving relationship and prediction are thus matters to 12.
which an investigator can still apply the two kinds of statistical tech-
niques, namely _13_ and _14_ statistics. 13.
14.

Populations, Sampling, and Samples

A population is the complete set of observations about which a re-
searcher wishes to draw conclusions, and a sample is a _15_ of a 15.
population. Techniques of description can be applied either to a _16_ or 16.
to a sample, but techniques of inference are applied only to a sample, in
order to derive from that sample a conclusion about the population from
which it came. The inferential techniques take into account the _17_ 17.
factors that operated when the sample was selected from the population.
If we could inspect a complete population, there would be no need to
apply inferential techniques, because we could learn everything we would
want to know about the population by simply applying the other
techniques, those of _18_. 18.

The process of selecting a sample from a population is called sampling.
There are many possible ways to select a sample, but in statistical work,
one particular kind of sampling is especially important. This is random
sampling, and it is sampling in which _19_ does the selecting of the 19.
observations. The chance factors that produce a given _20_ sample will 20.
vary from sample to sample, of course, and thus the samples themselves
will vary. This chance variation from sample to sample is known as _21_ 21.
sampling _22_. 22.

Suppose there is a population of observations from which we draw a
random sample, and suppose the sample is small in size (consisting of
just ten observations, say). If we draw additional samples of this same
small size from the given population, the random _23_ variation across 23.
these samples will probably be relatively large. That is, there will
probably be major differences among the samples.

But now suppose that from the same population we draw a random
sample that is large in size (consisting of, say, 100 observations). If we
select additional samples of this same large size from the given popula-

tion, the random sampling 24 across the samples will probably be relatively 25: large or small? . That is, there will probably be only minor differences among the samples. 24. 25.

In general, then, when the size of several random samples from a given population is larger, the random sampling variation among those samples is 26: larger or smaller? . 26.

Variables

In the course of an investigation, a researcher will make observations on subjects, focusing on one or more of their characteristics (for example, their political affiliation—Democrat, Republican, or Independent—or their age). A characteristic that may vary from subject to subject, taking on different values, is called a 27 . 27.

There are different kinds of variables. A 28 variable is one whose values differ in quality, or kind, rather than in amount. Political affiliation is an example of a 29 variable, in that the values of this variable—Democrat, Republican, or Independent—differ in kind. A Democrat is not more or less of something than a Republican; a Democrat is a person who differs in quality from a Republican. In contrast, a characteristic whose values differ in quantity or amount is called a 30 variable. Age is an example of a 31 variable. 28. 29. 30. 31.

There are some quantitative variables whose values are numbers that represent the quantities of the characteristic that varies. Such variables are said to be 32 scaled. Age is an example. 32.

Other quantitative variables have values that differ in quantity but are not represented precisely by numbers—for example, rank in a military hierarchy. The rank of sargeant is more of something (it's more distant from the bottom of the hierarchy) than the rank of corporal, but there are no numbers here to tell us the exact degree of the varying characteristic.

The textbook is concerned mostly with quantitative variables that are numerically 33 , but it also offers information on the treatment of variables whose values differ in kind (34 variables). 33. 34.

How Does Statistics Figure Into Research?

Any investigation begins with a question about the subject matter of the investigation. Such a query is called a 35 question. During the investigation, the researcher will collect observations, and she or he will use the techniques of 36 statistics to organize and summarize the observations (the data). The researcher may calculate an average of the scores that she 35. 36.

or he collected, for example. The researcher now asks a question about
this average or whatever statistical index is in use. For example, the
researcher may ask, "What is the likely value of the average score in the
whole population from which I sampled?" The answer to such a query is
a 37 conclusion. But such a conclusion is not yet the end of the study. 37.
The researcher must still do some careful reasoning to determine the
larger meaning of the statistical 38 . On the basis of the statistical 38.
conclusion, but also on the basis of other considerations, the researcher
will reason out a conclusion about the subject matter. Such a conclusion
is called a 39 one. 39.

Exact Numbers and Approximate Numbers

Some numbers are exact. A careful count of the number of subjects in an
experiment, for example, is a(n) 40 number, since it contains 41: 40.
some or no? margin of error. But most numbers generated by a measur- 41.
ing procedure, such as heights and weights, lack this kind of accuracy.
They are known as 42 numbers. 42.

An approximate number can be calculated to many decimal places,
and the question arises of how many decimal places to figure and to
report. No simple answer to this question is possible. You should not
calculate a(n) 43 number to many decimal places more than the 43.
number in the figures that entered into the calculation. But if you do
excessive rounding, you will lose the accuracy to which you are entitled.
The examples in the text will provide good models. Follow them.

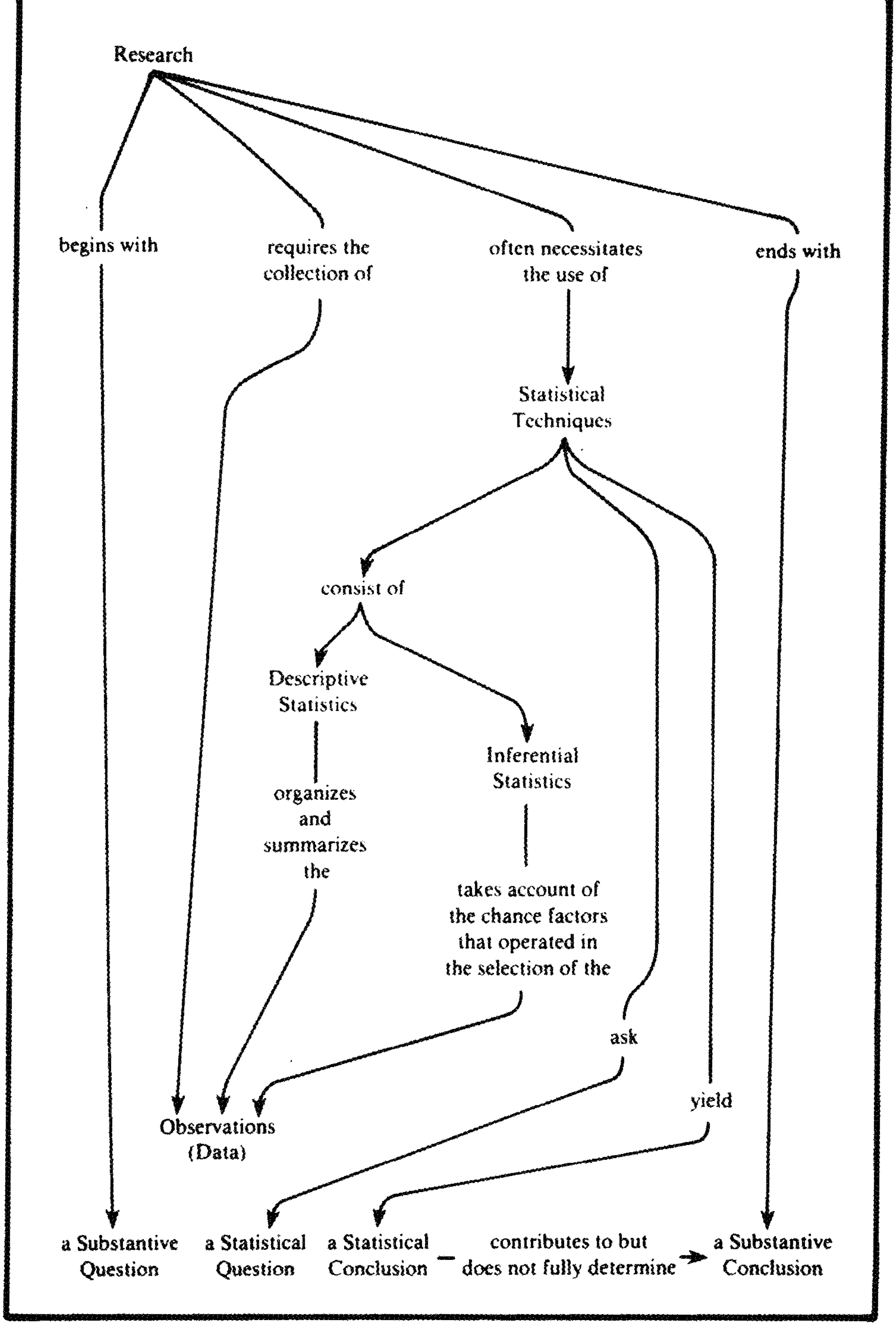
ARE YOU WORRIED ABOUT THE MATHEMATICS IN THIS COURSE?

As of this writing, I have led almost 700 students through an introductory course in statistics. Nearly all of them, I'm sure, were initially worried about the mathematics facing them. One young woman even dreamed of being attacked by numbers on the night before her first class with me. But 95% of these students passed the course with a grade of *C* or better, in spite of my high standards, and the person who had the nightmare earned a strong *A*. She also acquired good self-confidence. *You can be a success too.*

To put your mind at rest about the mathematics in this course, consider these facts:

- Your text emphasizes the *logic* of statistics, not the theorems, formulas, and proofs that mathematicians work with. The title of the text is *Elements of Statistical* ***Reasoning***, and it's reasoning, not mathematics, that's important here.

Statistics and its Role in Research



- The mathematics employed in the text is only simple algebra. You covered this in high school, and if you need to relearn it, you can do so easily. Appendix A in the text will help.

Furthermore, look what you've got going for you:

- Your text presents much more than the bare facts. It also provides the "big picture," so you can see how the facts fit together. It also adds the details, so you can gain insight into the facts.
- Right under your nose at this very moment is a study guide that offers you a lot of help with the text.
- Your instructor (and your teaching assistant, if you have one) will go over the material in the text and will answer any questions you have.

Moreover:

- *This course itself provides a leisurely review of mathematics.* The math in the course begins with counting (tallying up observations). It goes on to proportions and percentages, and more complex matters come up only later. So you can gradually relearn whatever you're uncertain of—and you'll be relearning it in a context that makes it vivid and useful.

The mathematics in this course is thus fully within your comprehension. If you've got adequate time to devote to the course, you can learn absolutely everything in your text and feel really good about it.

TIPS ON BUYING A CALCULATOR

A miniature calculator would be a good investment for this course, and you'll probably find other uses for it too.

You don't need anything fancy. You'll have no use for the special features of the "scientific" calculators that are meant to replace a slide rule—no use for the keys for pi (π), logarithms, exponentiation, or the trigonometric functions sine, cosine, tangent, and cotangent. (As noted above, this course does not require sophisticated mathematics.) You do want the following features:

- An add-on memory, which permits you to add a number to another number already in storage. The key for doing this is usually labeled $M+$. A machine with add-on memory typically has other keys labeled $M-$ (for subtracting something from what's in memory), *MR* or *RM* (for recalling what's in memory), and *MC* or *CM* (for clearing what's in memory). These keys will all be useful.
- Automatic constant for multiplication (for which there's no special key), or a key for squaring a number, labeled x^2. Automatic constant works like this: to square

a number (to multiply it by itself), you first enter the number and then press the × key. Instead of entering the number again, though, you just press the = key. See if this works on any calculator you're trying: $2\times=$ should get the calculator to read 4 (2×2); $3\times=$ should get the calculator to read 9 (3×3); and so on. A key labeled x^2 is even better for this same purpose.

You should also look for a machine with:

- Positive-action keys, which click or change in amount of resistance to the touch when they work. Some machines give you no feedback on whether a key has functioned when you depress it, and you then must constantly check the display, which is a nuisance.
- Keys that are big enough for you to press easily and accurately. Too much miniaturization is a liability.
- A square-root key. It'll be labeled $\sqrt{\ }$ or $\sqrt{x}$.
- A display that you can read easily from a variety of angles.

Be a smart shopper:

- Compare models, guarantees, and prices, and try several stores.
- Ask about the stores' policies on defective merchandise. What will they do if your purchase malfunctions after you get it home? The store should agree that if it proves defective within 30 days, they will replace it with a new machine from their own stock, rather than send the old one to the factory for repair. Ask the salesperson to write "30-day exchange" on your receipt and sign it.

You should be able to get what you want for less than $20.

If you're willing to spend a little more money, look for a calculator with features expressly intended for statistical use. In advertising, these are often called statistical functions. Such a calculator usually has keys labeled Σ (a Greek symbol for summation) or $\Sigma+$, $\bar{X}$ ("eks bar," a symbol for a type of average) or MEAN, and σ ("sigma," a symbol for an index of variability). These keys can save you some drudgery.

Even fancier—and more expensive—calculators offer features useful in studying relationship and prediction. The relevant keys are usually labeled r (a symbol for an index of correlation) or CORR (for *correlation*) or L.R. (for *linear regression*—you'll understand it when you get to Ch. 8). These features too can spare you some tedious work.

If you buy a calculator, take the time soon to learn the basics of it: how to add, subtract, multiply, and divide, and how to use the memory. If the machine has additional features that now seem puzzling, don't worry. You can learn to use them later, when they'll be helpful in this course.

HOW TO DO YOUR BEST ON A MULTIPLE-CHOICE TEST

1. Underline the important words in the questions and again in the possible answers. Be sure to underline negatives such as *none* and *not*.

2. Mark each choice for a given question *T* (for *true*), *F* (for *false*), or *?* (for *unsure*).

3. If you find an item especially difficult, leave it for later.

4. Answer all items. Guess if you have to.

5. If you must guess, eliminate answers you are sure are wrong before you do so.

6. If you have second thoughts about your answer for a given item and are considering changing the answer, do change it if you think you ought to.

The last point deserves comment. A persistent myth among students says that you're better off sticking with your first choice on a multiple-choice item, but educational psychologists have demonstrated that a majority of students generally improve their scores when they change their answers.*

Now, are you one of the majority who're better off changing, or one of the minority who should stick with the first choice? Right now, you don't know—but it's clear what you should do: change an answer whenever you think you ought to. The odds are in your favor if you change answers, because you're more likely to be one of the majority than one of the minority. This is an illustration of the point on p. 8 of your text: "The 'group' approach can often be turned to the probable advantage of an individual."

But keep track of what happens each time you make a change: did you change a wrong answer to a right one, a right to a wrong, or a wrong to another wrong? After you have a good-sized sample of cases in which you changed the answer—about 25 such cases—you can tell with some confidence whether you're really in the majority, or whether you're among the minority of students who generally lower their scores when they change answers.

And illustrated here is another principle from Ch. 1 of the text: "Large samples will provide a more precise estimate of what is true about the population than can be expected from small samples" (p. 5). The population in this instance is the complete set of answer changes you might ever make, and every time you change an answer, you add another case to the sample you're collecting. In accord with the principle just quoted from the book, don't try to generalize from a small sample of answer

*See, for example, "Answer-Changing Behavior and Grades" by J. J. Johnston in *Teaching of Psychology* for February 1978 (Vol. 5, No. 1).