ceptually much simpler characterization of the center of a population than the mean. To introduce the concept of a confidence interval, we then simply consider the appropriateness or inappropriateness of statements such as, the population median falls between the smallest and the largest observations in the sample; the population median falls between the second smallest and second largest observations in the sample; and so on. By comparing observations that are smaller or larger than the population median to heads and tails in fair coin tosses, the random nature of the confidence interval emerges quite naturally. For example, the most extreme confidence interval does not cover the true median, if we observe nothing but heads or nothing but tails. The probability of this event, and thus the confidence coefficient, is easily found.

I have emphasized two reasons for preferring nonparametrics in an introductory statistics course, namely, greater mathematical and greater conceptual simplicity. But there is one additional reason, the more general validity of the nonparametric approach. A single extreme observation can invalidate the conclusions of a t test, not to mention nonnormality, which means very little to a student in an introductory statistics course. With a nonparametric procedure, students do not only know what they are doing, they can also feel reasonably safe that they have done the correct thing.

[Received September 1978. Revised May 1979.]

We Need Both Exploratory and Confirmatory

JOHN W. TUKEY*

We often forget how science and engineering function. Ideas come from previous exploration more often than from lightning strokes. Important questions can demand the most careful planning for confirmatory analysis. Broad general inquiries are also important. Finding the question is often more important than finding the answer. Exploratory data analysis is an attitude, a flexibility, and a reliance on display, NOT a bundle of techniques, and should be so taught. Confirmatory data analysis, by contrast, is easier to teach and easier to computerize. We need to teach both; to think about science and engineering more broadly; to be prepared to randomize and avoid multiplicity.

KEY WORDS: Exploratory data analysis; Confirmatory data analysis; Paradigms of science and engineering; Sources of ideas; Randomization; Multiplicity.

Analysis of data, with a more or less statistical flavor, should play many roles. We need to recognize this, and act upon it, without regard to the ease or completeness with which these roles can be formalized.

1. *An incomplete paradigm*. We are, I assert, all too familiar with the following straight-line paradigm—asserted far too frequently as how science and engineering function:

(*) question → design → collection → analysis → answer

Any attempt to claim that this straight-line, confirmatory pattern is more than a substantial part of the story neglects crucial questions (and their answers):

1. How are questions generated? (Mainly by quasi-theoretical insights and the exploration of past data.)
2. How are designs guided? (Usually, by the best qualitative and semiquantitative information available, obtained by exploration of past data.)
3. How is data collection monitored? (By exploring the data, often as they come in, for unexpected behavior.)
4. How is analysis overseen; how do we avoid analysis that the data before us indicate should be avoided? (By exploring the data—before, during, and after analysis—for hints, ideas, and, sometimes, a few conclusions-at-5%/k.)

I assert, and I count upon most of you to agree after reflection, that to implement the very confirmatory paradigm (*) properly we need to do a lot of exploratory work.

Neither exploratory nor confirmatory is sufficient alone. To try to replace either by the other is madness. We need them both.

2. *The origin of ideas*. Reorganizing the early stage of the last paradigm can help us understand better what is going on. What often happens is better diagrammed thus:

(*) idea → ↗ question ↘ / ↖ design ↙ → collection → analysis → answer

If we have an idea that a certain drug will help in a certain disease, and say we want to find out, we have not yet formulated a question in the sense of (*). What we have is an idea of a question—something often thought of in terms of the common language as a question—but not at all the kind of question that can have a statistically supported answer.

The kind of question that does have an answer here will be much more circumscribed—and its choice is a matter of practicality, not desire. We might, for

* John W. Tukey is Donner Professor of Science and Professor of Statistics, Princeton University, P.O. Box 37, Princeton, NJ 08544; and Associate Executive Director–Research, Bell Telephone Laboratories, Inc., Murray Hill, NJ 07974. This article was prepared, in part, in connection with research at Princeton University sponsored by the Department of Energy.

example, wish fervently to answer the question—"Of those who would otherwise die within three years from the disease, what fraction could be saved with this therapy?" But no known design will let us isolate these people for an experiment.

The best we are likely to be able to do is to formulate a question limited by constraints such as

1. Age and sex of patients,
2. A minimum set of symptoms,
3. Absence of other life-threatening disease,
4. Type of patients usually seen by participating investigators.

The formulation of the question itself involves what can in fact be asked, what designs are feasible, as well as how likely a given design is to give a useful answer. Both inchoate insight and extensive exploration (of past data) can—and should—play a role in this process of formulating the question.

Science—and engineering, which here includes agriculture and medicine—DOES NOT BEGIN WITH A TIDY QUESTION. Nor does it end with a tidy answer.

The picture of the scientist struck—as by lightning—with a question is very far from the truth.

3. *Important questions.* Some questions are important—are the sort of question about which a body of the wisest men available would say, "We all agree, this question MUST be answered, and we must be CERTAIN of the answer!" What then?

If the necessary resources of money, skill, and data management are available, we will go ahead. Our best watchwords are often these:

1. Randomize, RANDOMIZE!
2. Preplan THE main analysis (having even two main analyses may be too many)!

The solidest confirmatory analyses we have are based upon randomization theory—the best way to ensure the applicability of randomization theory is to randomize, carefully and appropriately.

Problems of multiplicity have been too little recognized. To say in advance that we will look at one of 12 analyses is to give many hostages to fortune. If the results of the 12 analyses are statistically independent, at least one will be "significant at 5 percent" a large fraction, $1 - (.95)^{12} = .54$, of the time. This is ordinarily an unacceptable Type I error. If we protect against this by going to 5%/12 as our significance level, then, IF the results are highly correlated, we shall have been highly wasteful, concluding less about our question than we should.

I see no real alternative, in most truly confirmatory studies, to having a single main question—in which a question is specified by ALL of design, collection, monitoring, AND ANALYSIS.

It may be wise, sometimes, to include alternative analyses, but we ought to regard any such case as a failure of statistical theory. To calculate two statistics and take the larger becomes a single analysis as soon as we know enough about the null distribution of "the larger of this and of that"! If neither statistical theory or computer rerandomization can tell us enough about the actual null distribution, how can we be content?

4. *The broad general inquiry.* We, as statisticians or as data analysts, have thought too little about the broad general inquiry, about this sort of question:

"Now, what do you suppose goes on in that general area?"

This is often meaningful—and sometimes extremely important. It certainly usually leads to data examination and often to data collection. Yet we have thought little about it.

Clearly, the answer of the general question will often lead us on to questions such as

"What can I try easily?"

and

"How do I study what's going on out there now?"

How do we help such an asker? Who has thought about designs for exploring a broad area, either initially, or as the next follow-up? (Clearly what we do here is exploration; if we are lucky, we may formulate questions deserving of attempted confirmation.)

Who knows about designs for exploration? Who has studied techniques such as "automatic interaction detectors" empirically?

This is one of the Roba el Khali's (cf. Section 12) we need to learn to penetrate.

5. *A maxim.* We need, as statisticians or data analysts, as well as scientists or engineers, to bear in mind that

"Finding the question is often more important than finding the answer."

6. *Exploratory data analysis.* Some have suggested that "exploratory data analysis" is just "descriptive statistics" brought somewhat up to date. Much effort, much intelligence and understanding has been devoted in recent years to convince us that "the map is not the region"! Perhaps an equal effort, at least among statisticians, is needed to persuade us of the equally true statement, "the usual bundle of techniques is not a field of intellectual activity"!

If we need a short suggestion of what exploratory data analysis is, I would suggest that

1. It is an attitude, AND
2. A flexibility, AND
3. Some graph paper (or transparencies, or both).

No catalog of techniques can convey a willingness to look for what can be seen, whether or not anticipated. Yet this is at the heart of exploratory data analysis. The graph paper—and transparencies—are there, not as a technique, but rather as a recognition that the picture-examining eye is the best finder we have of the wholly unanticipated.

7. *Confirmatory data analysis.* Whatever those who

have tried to teach it may feel, confirmatory data analysis, especially as sanctification, is a routine relatively easy to teach and, hence,

A ROUTINE EASY TO COMPUTERIZE.

This last is far from an empty point.

We are still in the very early years of the computer. But with 100,000 subscribers to *BYTE* magazine, a magazine for home computer buffs, we can—if we wish—see the handwriting on the wall.

The only way humans can do BETTER than computers is to take a chance of doing WORSE.

So we have got to take seriously the need for steady progress toward teaching routine procedures to computers rather than to people. That will leave the teachers of people with only things hard to teach, but this is our proper fate.

8. *In which order should we teach?* Experience, on various campuses, with the limited preliminary edition of *Exploratory Data Analysis* (Tukey 1977) made it quite clear that

> Students who have never been exposed to CONFIRMATORY seem to learn EXPLORATORY more easily.

So which should we teach FIRST?

9. *A question, and an answer.* A sort of question that is inevitable is

> "Someone taught my students exploratory, and now (boo, hoo!) they want me to tell them how to assess significance or confidence for all these unusual functions of the data. (Oh, what can we do?)"

To this there is an easy answer:

> TEACH them the JACKKNIFE.

10. *A rhetorical question.* "If A will not be learned unless someone teaches it, while B will be forced upon our students (either by their major professors or by the pressures of the real world), which obligation rests more firmly on our shoulders,

> 1. To teach B, OR
> 2. To teach A?"

In answering such a question we must remember that we have an obligation to teach both.

I wish exploratory data analysis were a B, not an A, but we have to take things the way they are.

11. *Conclusions.* Some of my conclusions, then, are these:

1. There is NO question of teaching confirmatory OR exploratory—we need to teach both.
2. We need to think about science and engineering more broadly than the narrow, inadequate paradigm of a straight line from question to answer.
3. When we want to do careful confirmation on important questions we need to be very careful—randomizing and avoiding multiplicity.
4. We need to teach exploratory as an attitude, as well as some helpful techniques, and we probably need to teach it before confirmatory.

12. *Geographical remark.* The Roba el Khali, or Rub' al Khali, is the part of Arabia known as "the empty quarter." It was the last part of Arabia to be penetrated by Western explorers.

[*Received November 1978. Revised May 1979.*]

REFERENCE

Tukey, John W. (1977), *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley.