GEORGE E. P. BOX*

When Fred Leone, our Executive Director, was explaining to me what my presidential duties were, he told me that one of the "perks" associated with this job is that I get to give the annual address to the Association, and I have a captive audience for as long as they are prepared to sit there. Fred said to me, "George, don't give them anything too technical because this is a light occasion and there will be a lot of people that the statisticians have dragged along—husbands, wives, friends—who have had about all the statistics they can take."

Well, imagine my disappointment. I had prepared a 200-page draft of my talk. It was called "The Present Status of the One-Armed Secretary Problem: A Decision-Theoretic Approach," and it made free use of σ-fields, Hilbert spaces, and all kinds of squiggly letters with dots on. This I reluctantly set aside. (I don't think any of you would have understood it anyway.) I have had to look for an alternative. I toyed for some time with the title, "Whither Statistics?," subtitled "Perhaps We Shouldn't Start from Here," but in the end abandoned that too. Eventually it struck me that many of the issues that we face as members of the American Statistical Association are really not very different from those we face as ordinary human beings. This is what my talk is about.

THE BEST IS OFTEN NOT VERY GOOD

Some of us have had a preoccupation with optimal or best procedures. But the best, of course, is not necessarily very good. For instance, to bring in the aspect of everyday life, if ever I *had* to decide between cutting my throat with a razor blade or with a rusty nail, I suppose I would choose the razor blade. But, although not strictly relevant to the problem as posed, one question that might cross my mind would be, "Have I considered all my options?"

A principle that is being given more attention these days is that of "robustification." Here one doesn't attempt to guarantee that things will be optimal over some tractable, but perhaps very narrow, set of circumstances. Instead one tries to ensure that they will be fairly good over a wide range of possibilities *likely to happen in practice*. Look at the human hand, for example. I doubt if there is any single thing that it does that could not be done better by some special instrument, but it is very good at doing a very large number of things that come up in facing the world as it actually is.

Another way to say this is that there is really nothing wrong with optimization per se, but that we ought to try to optimize over *that distribution of circumstances which the world really presents to us*. The mistake is choosing the best over too narrow a set of alternatives, suboptimization. It is sometimes argued that by doing simplified exercises, we can at least obtain useful pointers. However, I feel that such pointers are very likely to indicate the *wrong* direction, as might be true in the case of the razor blade and the rusty nail.

WHAT IS THE REAL WORLD LIKE?

The difficulty in taking the wider robustification approach is that we cannot expect to get good results unless we are really prepared to engage in the hazardous undertaking of finding out *what the world is really like*. It requires us, as statisticians, to have some knowledge of reality.

I believe we do have members, we may have ASA fellows, possibly we have even had ASA Presidents who really do not care what the world is really like. Some years ago a friend of mine told me about his daughter who was then at Oxford University. She was a very bright girl, but she got interested in politics (it was in the 1960s); she got behind in her studies, and the time of graduation was approaching. You may know that in the English system, there are many different grades of bachelor's degree. The young lady started to worry: was she going to get a "pass" degree (which is almost like the University spitting at you), or was it to be a third class, lower second class, upper second class, or a first class honours degree? She decided to ask her tutor about it. Finding him buried somewhere in the dust of one of the Oxford colleges, she eventually got around to asking him the delicate question, "Would it matter in the outside world if I didn't get a very good degree?" He looked very startled and said, "*Outside world*? What do *I* know about the outside world?"

When the statistician looks at the outside world, he cannot, for example, rely on finding errors that are independently and identically distributed in approximately normal distributions. In particular, most economic and business data are collected serially and can be expected, therefore, to be heavily serially dependent. So is much

* George E.P. Box is R.A. Fisher Professor of Statistics, University of Wisconsin, Madison, WI 53706. This article is the text of the Presidential Address delivered at the 138th Annual Meeting of the American Statistical Association, August 15, 1978, in San Diego.

March 1979, Volume 74, Number 365
Presidential Address

of the data collected from the automatic instruments which are becoming so common in laboratories these days. Analysis of such data, using procedures such as standard regression analysis which assume independence, can lead to gross error. Furthermore, the possibility of contamination of the error distribution by outliers is always present and has recently received much attention. More generally, real data sets, especially if they are long, usually show inhomogeneity in the mean, the variance, or both, and it is not always possible to randomize.

To find out what the world is really like, we must spend more time looking for ourselves at real sets of data. For example, David Cox says that deviations from normality often occur in the direction of light-tailed distributions as well as heavy-tailed ones. Let us discover if he's right. If he is, it could seriously affect some proposed robust methods.

In order to better confront the realities of the outside world, some have urged us to abandon classical methods of estimation, such as employ likelihood and Bayes' Theorem, and resort to a *new* empiricism for each problem that arises, and for each author that writes about it. I think that notion is wrong-headed. The imperfection, of course, lies not with the estimation method, but with the model that we put into it. For example, it is true that the sample average, which is the maximum likelihood estimate of the mean on standard normal assumptions, could be a very poor estimate if we believed that the data were generated by a *contaminated* normal distribution. Nevertheless, if our model took account, not of what we did *not* believe, but rather of what we *did* believe, we could obtain excellent estimates by standard methods. The great advantage of the model-based over the ad hoc approach, it seems to me, is that at any given time we know what we are doing.

Models, of course, are never true, but fortunately it is only necessary that they be useful. For this it is usually needful only that they not be grossly wrong. I think rather simple modifications of our present models will prove adequate to take account of most realities of the outside world. The difficulties of computation which would have been a barrier in the past need not deter us now.

CHOOSING THE "BEST" DOESN'T MAKE SENSE WITH OPTIONS THAT ARE NOT ALTERNATIVES

Another difficulty with optima is a tendency to want to choose the best *one* of a set of items that are not really alternatives. The relevant question then is not, "Which is best?," but "Do these different entities have a role, and if so, what is it?" For example, it turns out that it is much better to have two sexes than one—and this not merely for hedonistic reasons. Again, if one mentions Bayesian analysis and sampling theory analysis in the same breath, one not only hears the question, "Which is best?," but also the question, "Which is right?," and religious passions are quickly aroused. Yet, to my mind Bayes theory and sampling theory are not alternatives at all.

It is widely recognized that the advancement of learning does not proceed by conjecture alone, nor by observation alone, but by an iteration involving both. Certainly, scientific investigation proceeds by such iteration. Examination of empirical data inspires a tentative explanation which, when further exposed to reality, may lead to its modification. This modified explanation is again put in jeopardy by further exposure to reality, and so on, in a continued alternation between induction and deduction.

I am continually surprised that statisticians, even good ones, still seem to ignore this iterative aspect of investigation and talk as if the movement from an initial (perhaps ill-posed) question, to design, to data collection, to analysis of the data, to "the answer" were a one-shot affair. The wise investigator expends his effort not in one grand design (necessarily conceived at a time when he knows least about unfolding reality), but in a series of smaller designs, analyzing, modifying, and getting new ideas as he goes. This iterative aspect of research has a profound influence on almost everything the investigator and the statistician do, and it has been the source of much misunderstanding. Just as the rules that govern mathematical iteration are very different from those that govern solutions in closed form, so the rules that ought to apply to the statistics of most real scientific investigations are different, broader, and vaguer than those that might apply to a single decision or to a single test of hypothesis.

Now, since scientific advance, to which all statisticians must accommodate, takes place by the alternation of *two* different kinds of reasoning, we would expect also that *two* different kinds of inferential process would be required to put it into effect.

The first, used in estimating parameters from data *conditional* on the truth of some tentative model, is appropriately called *Estimation.* The second, used in checking whether, in the light of the data, *any* model of the kind proposed is plausible, has been aptly named by Cuthbert Daniel *Criticism.*

While estimation should, I believe, employ Bayes' Theorem, or (for the fainthearted) likelihood, criticism needs a different approach. In practice, it is often best done in a rather informal way by examination of residuals or other suitable functions of the data. However, when it is done formally, using tests of goodness of fit, it must, I think, employ sampling theory for its justification.

Bayes and likelihood inferences are necessarily conditional and, therefore, should not be used alone for the same reason that the statement, "If the moon was made of green cheese, it would be a great place for mice," should not tempt a mouse to hang around Cape Kennedy.

INAPPROPRIATE DIVISION OF AN ENTITY

While we can make a mistake by looking for *one* answer when we should be looking for two or more, we can make another mistake by *dividing* an entity inap-

propriately. You will recall the story of Solomon, who determined the true mother of a child of disputed parentage by offering to cut it in two. One slicing of our subject which I think can be harmful is that into Applied Statistics and Theoretical Statistics.

I hear people saying things like, "Of course I'm a theoretical statistician myself, but I agree there should be some applied statisticians and there should even be applied statistics departments; in fact, some of my *best friends* are applied statisticians." Now, in my opinion, that isn't any good, because, if you imagine the theoretical statisticians distributed about a point on the right of a scale and the applied statisticians distributed about a point on the left, you will end up with a bimodal distribution with low density in the center. Now the people most needed are, in my opinion, those in the middle, and perhaps that's why they seem to be in such short supply. If, alternatively, we aimed at a central target, then we might achieve a single unimodal distribution. This would still, of course, allow diversity. We would have some highly theoretical people in one tail and some highly applied people in the other. But the majority, while having proper theoretical training, might also possess ability and experience in applying what they knew to the solution of scientific problems.

TRAINING OF STATISTICIANS

This suggests the question of how statisticians should be trained. It's fairly easy to see how we should *not* train them. I will make an analogy with swimming.

Swimming could be taught by lecturing the student swimmers in the classroom three times a week on the various kinds of strokes and the principles of buoyancy and so forth. Some might believe that on completing such a course of study, the graduates would all eagerly run down to the pool, jump in, and swim at once. But I think it's much more likely that they would want to stay in the classroom to teach a fresh lot of students all that they had learned.

Let me mention another distinction which is now needed, and which threatens to become an unnecessary and harmful slicing. Statistical practitioners have known for a long time that, prior to using the methods that most textbooks emphasize, there is a very important and largely neglected[1] phase of activity which Fisher called specification and which has also been called model identification. This involves informal techniques of analysis of data, many of them graphical, aimed at looking at the data in a preliminary and exploratory way in order to help understand what questions should be asked and what tentative models might be entertained. Until recent years, however, this process was regarded by the majority as not entirely respectable. Like the black art, it was widely felt that it should be conducted, if at all, only behind closed doors.

[1] An early exception was the second chapter of *Statistical Methods for Research Workers* (Fisher 1925), first published in 1925, in which Fisher discussed the use of preliminary graphical techniques.

It was a stroke of genius to realize that to render "a deed without a name" respectable, you should name it (or perhaps I should say rename it), and we are all grateful for the name "Data Analysis." This important part of our subject can now be studied without apology or shame, and courses on it are taught and may be attended by consenting adults. The elevation of Data Analysis to its proper place as a subject meriting serious study makes me as happy as I would be if some neglected but important activity of the carpenter, such as the use of the saw or the chisel, had at last received proper recognition and study. But my enthusiasm for the naming of Data Analysis does not extend to the renaming of Statisticians as "Data Analysts," any more than I should be happy to hear a carpenter described as a sawyer or a chiseler. Indeed, I am as appalled by the appearance of Data Analysts *as entities* as I would be at contemplating one half of the baby over which Solomon adjudicated, and for the same reason. There can be no feedback between the parts of a once-living thing cut in two.

Please can Data Analysts get themselves together again and become whole Statisticians before it is too late? Before they, their employers, and their clients forget the other equally important parts of the job statisticians should be doing, such as designing investigations and building models? By invention of the concept of Experimental Design, Fisher promoted the statistician from a curator of dusty relics to a valued member of a scientific team, responsible for planning and taking part in the conduct of an investigation. Let us not allow him to be relegated to his previous passive and inferior role by an injudicious choice of a name. "Our Data Analyst" is too close for my liking to "Our Tame Statistician," a poor thing if that is all he is.

THE AMERICAN STATISTICAL ASSOCIATION

Finally, I want to talk a bit about *our* Association because it is ours and it can be as good or as bad as we make it.

During my time as president I have received a number of letters, all of them interesting and some of them critical. Some members feel that we should be doing things we are not doing, some feel that we are doing things we should not be doing, some feel that the articles in our journals are not on the subjects they would like, or are not written with sufficient clarity. The suggestions made in such letters are, of course, given careful consideration not only by your president but by your Board and its committees. If you have ideas on these or other subjects, however, and want to see something more done about them, I do urge you, if you have not done so already, to volunteer for active duty. The Association is always seeking new faces and new ideas for its committees. And there are other things you can do too.

Suppose, for example, as a statistical practitioner, you feel that the journal *Technometrics* is not adequately

> papers illustrating the application of known statistical method to new or novel environments, expository or tutorial papers on particular statistical methods, and papers dealing with the philosophy and problems of applying statistical methods to research, development, design and performance.

I have it on good authority that editors have two problems in carrying out such a mandate: in the first place, articles of this kind are extremely hard to come by, and in the second place, referees tend to reject such articles for the wrong reasons (perhaps because they have not read the mandate). I urge you, therefore, to consider one or both of the following courses:

1. If you have suitable material, please write it up and send it in.
2. Please volunteer to act as a referee.

The editors are in desperate need of good referees for articles of this and every other sort. They need people who will go carefully through a paper, say encouraging words about good things, suggest how ideas could be clarified and how imperfections can be put right and who, when necessary, will firmly reject unsuitable manuscripts.

In closing, I want to say how much I have enjoyed being your president, especially because of the kindness, consideration, and help I have received from the members, board, and officers. In particular, I wish to thank Fred Leone, Ed Bisgyer, Jean Smith, and the rest of the Washington staff. The Association is indeed fortunate to be served by such accomplished and dedicated people.

REFERENCE

Fisher, R.A. (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.