hand or with a hand calculator. On the rogue plot, nine of the samples had zero infection. There was clearly no need for a calculator: the field worker had simply recorded the single non-zero value as the average value. Once this was corrected, the rogue point jumped magically onto the straight line and the relationship between $y$ and $x$ was evident. Subsequent sophistications in the "statistical" analysis seemed to me somewhat less important than the IDA phase, as regards the aim of finding out from the raw data the answer to the plant pathologists' question.

(e) I never trust any published formula, no matter how eminent the author. Here is another problem with the publication policies of statistical journals. Several editors take the view that proofs are only for mathematicians, and so they decree that results may be published but their proofs should not be. Without the proofs, how can we check the results? Moreover, as Chatfield notes in Example 7(c), proofs do tend to go hand in hand with clearly defined notation and clearly stated assumptions, both of which are too often dismissed from statistical journals as being no use to practical people.

(f) Knowing whom to ask for help and advice can be more of an asset than knowing all the techniques. A corollary is: don't be afraid to show that you have made a mistake or do not know what to do.

# Comment

**Murray K. Clayton and Erik V. Nordheim**

It is a pleasure for us to have the opportunity to comment on this timely article. As Dr. Chatfield properly points out, there are many facets of a successful statistical investigation that are not taught in most books or in most courses. Although a solid grounding in statistical methods and theory is necessary for success in solving real-world problems, it is not sufficient. An understanding of the potential pitfalls and strategies for avoiding them is a clear requirement for achieving this success.

Chatfield provides suggestions on a wide range of topics related to statistical consulting and provides a very useful bibliography. In addition to those references cited by Dr. Chatfield, we would add the volume of Boen and Zahn (1982). We find ourselves in strong agreement with virtually all of Chatfield's suggestions. We would like to point out some additional areas where our experience has shown the need for particular attention.

## 1. INTERACTING WITH THE INVESTIGATOR

It should be recognized that the active involvement of the investigator is essential in a successful statistical investigation. Too often the view is taken that once the statistician gets the data from the investigator, then the "real statistics" begins and the investigator's role is diminished. (This attitude may be reflected in the silence of the delegates in Chatfield's Example 5.) A critical reason for investigator involvement is that he/she holds the key to much information that is essential to the conduct of the analysis and that *cannot* be determined solely by looking at the data. We address two aspects of this involvement.

(a) As articulated by Chatfield, a clear statement of the objective of an investigation is necessary in order to carry out a useful statistical analysis. However, it is our experience that obtaining a clear statement is often quite difficult. If you ask the investigator early in a consulting session, "What are your objectives in this study?", you can receive a variety of responses, many of which are only of marginal use. Sometimes the investigator will attempt to abstract a statistical problem, as he/she perceives it, in order to get "right to the matter quickly." On other occasions, you will be given a superficial description of the problem to "spare you all the experimental details." In still other situations, the investigator has not thought that far. It is our experience that it is often ineffective to ask the investigator for a statement of the objective at the very beginning of a consulting session.

We find it useful to pursue two major lines of questioning early in our meetings. One line is to find out about the background of the project. We try to ask questions like, "What do you anticipate to learn from this study?", "How will you use the

*Murray K. Clayton is Associate Professor of Statistics and Plant Pathology, and Erik V. Nordheim is Professor of Statistics and Forestry, Department of Statistics, University of Wisconsin-Madison, 1210 W. Dayton Street, Madison, Wisconsin 53706.*

results of this study?" and "Have you worked on similar problems in the past?" The second line concerns learning about the specific nature of the study. As Chatfield points out, this can involve looking at the data or observing experimental procedures in the field or in the lab. Although experience teaches the statistician when the detail becomes excessive, we believe that, when in doubt, it is better to probe too deeply rather than not deeply enough. It is our belief that, after some time of following these two lines of questioning, the objectives usually become quite clear. (Perhaps one hour is a rough "guesstimate" of the median time to reach this stage with problems of substantial complexity.) As the objectives become apparent to the consultant, it is often useful to use "playback" techniques. The consultant should phrase the objective in his/her own words, "As I understand it, you are interested in determining . . . ." The response of the investigator will either confirm the consultant's view or allow it to be modified.

(b) In addition to determining the objectives, the investigator can provide critical information at various stages in the analysis. For example, in a linear regression problem, a number of tools might be used to assess independence of the errors. A first thought might be to look at an autocorrelation function, fit an ARIMA model or perform a Durbin-Watson test. However, if there are only 10 observations in the data set, then none of the suggested methods will have much power, and yet, by asking the client, it is still possible to obtain some information regarding a possible lack of independence. A direct question such as: "Are your errors independent?" is not likely to be helpful, but a more indirect line of questioning might be fruitful: "How did you do your experiment? Did you randomize the order of the $x_i$ settings? How?" etc.

## 2. CAREFUL DESIGN

There is no substitute for careful design of any experiment or study. One hour of statistical guidance at the planning stages can be worth five or more hours at the analysis stage. Chatfield very properly addresses the extreme importance of careful planning. We wish to touch on four specific points.

(a) It is altogether too common for an investigator to announce at the beginning that he/she will conduct an experiment with $n$ sample units (e.g., transistors, animals, fields) due to financial and/or time constraints. It is incumbent on the good consultant to make it clear to the investigator how much sensitivity such a study will have. If the power is too small to detect effects of interest, the investigator can either modify the experiment (e.g., reduce the number of treatments) or, if suitable modification is not possible, avoid wasting time and money on a study that has a very small chance of obtaining useful results.

(b) Substantial attention should be devoted to appropriate methods of analysis of the data at the design stage of a study. In particular we have found it most helpful to formally identify the "experimental unit" (Cochran and Cox 1957) for all treatments. In our experience, many studies involve some elements of nesting or subsampling, although these might not always be recognized. For example, based on a cursory look, data from a split-plot experiment and from a completely randomized experiment might look the same, but the analysis and conclusions might be very different. In this light, investigators may feel that they are obtaining more than enough observations but, after accounting for nesting and subsampling, they may have too few experimental units to study the factor of major interest. This issue is clearly very important in assessing study sensitivity. An experimenter may feel that 240 observations are more than enough for any reasonable study but may feel differently when told that there are only 2 degrees of freedom for error for testing the key factor.

(c) We have noticed that, even when sufficient care has been devoted to the layout and planning of a study, insufficient attention is devoted to the management of the study. Factors such as run order, time of day of particular manipulations (e.g., watering plants) and different observers can impact in a most negative way on a study. An example from our experience illustrates this nicely. This study was a $7 \times 2 \times 2$ factorial study on survival of honeybees in response to certain chemical insults. There were three containers of approximately 50 bees for each of the 28 treatment combinations. These containers were placed in a room with alternating 12 hour periods of light and dark. During the first days of the experiment, each container was examined daily to determine the number of dead bees (and to remove them from the container). The containers were counted (and the bees removed) during the light hours; the process often took several hours (perhaps 4 to 6). Thus, particularly if there was enhanced mortality during the hours of light, the containers counted near the end had substantially more time for the bees to die than those counted near the beginning. This introduces unnecessary variability, or perhaps bias if the containers were counted systematically. This variability and/or bias could easily have been reduced by blocking on experiment start-up and counting order (with three blocks).

(d) All too frequently, experimenters begin a project with limited understanding of the process under study. Even with careful thought devoted to planning and design, unanticipated results can lead to a largely wasted study. We strongly advocate the use of pilot studies to minimize the chances of such waste. Pilot studies can be used to determine background variability, to determine potential observed treatment differences, to learn of the effects of different management practices, etc. Often a careful run-through of an experimental protocol can expose unforeseen problems.

## 3. CAREFUL ANALYSIS

We strongly support Chatfield's comments on data analysis including his division of the topic into two stages: IDA and appropriate inference. We would like to add two points.

(a) Chatfield has commented on the issue of aiming for optimality and the potential conflict with avoiding trouble. We second that point, and add that there are potentially many ways in which optimality can be measured and evaluated. We might recommend a "suboptimal" analysis because such an analysis may be optimal in a different fashion: it can be more robust, easier to conduct, easier for the researcher to understand and easier for the researcher's colleagues to accept. For example, an analysis of a complex unbalanced analysis of variance might be made clearer by breaking the data into subsets and performing analyses on the various subsets. Indeed, there may be several useful ways in which data can be partitioned. Overall, we believe that in many cases it is better to be a "splitter" than a "lumper." In some extreme cases, a highly messy data set might not be amenable to any type of formal inference. Still, effective use of plots, tables and summary measures can be of value to the investigator.

(b) Many authors have discussed the handling of residuals and outliers; Chatfield's article makes a number of key points. As noted in Chatfield (1988), if it is difficult to determine whether a potential outlier is an experimentally aberrant point or a "real" observation, it is often helpful to perform an analysis with and without the outlier. If the basic conclusions are unchanged, then it is unnecessary to worry excessively about the outlier. If, on the other hand, the results differ, then it may be worth devoting extra attention to the outlier. This may mean a careful review of field or lab notes but, on some occasions, may necessitate additional experimentation. A simple example from our experience illustrates the value of careful analysis of outliers. The project was designed to study an automated dispersal system of patient food trays from the central kitchen of a hospital to several wards. The researchers were food scientists concerned with possible microbiological contamination of some foods. One component of the data was the distribution of times that each tray was in transit. The overwhelming majority of the observations were believable (less than 8 minutes), but there was a tiny fraction of outliers. These outlier values were sometimes as large as 40 minutes. A brief study indicated that there was an occasional computer malfunction in the automation. This malfunction was easily fixed.

## 4. CONCLUDING REMARKS

In order for the practice of statistics to achieve its promise, it is necessary that the advice given in this and related articles be taken to heart by practitioners. Articles such as this and specialized courses for statistics students on statistical consulting can be of help. Moreover, we would add that it is important that such advice be regularly incorporated into the introductory and methods courses that are offered to students from other fields who will be the investigators of tomorrow. In particular, these courses should incorporate real world complexity and provide students with the insights to help them avoid the potential statistical pitfalls that Chatfield has so eloquently described.