OPPORTUNITIES AND CHALLENGES FOR INDUSTRIAL STATISTICIANS IN THE 21ST CENTURY

Gerald J. Hahn--Manager, Applied Statistics Program
GE Corporate Research and Development
1 Research Circle
Niskayuna, New York, 12309, USA
Email: hahn@crd.ge.com

1. The New Environment

I have had the fortunate, and somewhat unique, opportunity to work in industry, and, in fact, for the same company for nearly half of the 20th century--almost 45 years to be exact (of course, I started very young!). During this time, I have witnessed substantial changes. When I started one of our major pre-occupations was in how to reduce the intense labor required to conduct statistical analyses using noisy and very slow mechanical calculators--which, themselves had only recently replaced completely manual calculations! We focused on such tools as orthogonal polynomials, to "simplify" the fitting of curvilinear regression models. When we were all through, we had little energy left to explore the real problems of our customers and to do more ambitious, exploratory analyses. Today, the implementation of complex statistical analyses is simple and speedy, and, generally, the least of our problems. Moreover, the rate of change has accelerated in recent decades. I would not want to predict the state of our profession in another 100, 50, or even 45, years, but I do want to share my thoughts on how applied statisticians, in general, and industrial statisticians, in particular, need do to be maximally effective as we enter the 21sr century.

The changing role of the industrial statistician is mainly due to the radical changes in the environment in which we operate. I will focus on only two elements (see Hahn and Hoerl (1998) for further discussion):

- *The profound impact of computer advances*. This subject requires little elaboration. We now have at our fingertips the ability to handle massive databases and to explore (or "mine") these with ease, using highly sophisticated methods that we would hardly have dreamt of at the beginning of my career. As a result, the opportunity to base decisions on a disciplined, and timely, analysis of the relevant data, in contrast to making purely intuitive evaluations, has grown tremendously.
- The so-called "*democratization" of statistics*” (similar to the democratization of mathematics, suggested by Vere-Jones (1995)). Whether we like it or not--and, in my opinion, the advantages far out-balance the disadvantages—statistical methods are readily accessible to all. So are related techniques not necessarily develope by statisticians, such as neural nets. Statisticians no longer have a monopoly on performing, and even inventing, statistical analyses.

In some companies, including my own, formal programs, such as the Six Sigma initiative, to leverage the vast opportunities provided by the new environment, have blossomed. Six Sigma is a highly disciplined, data-oriented approach for gaining quality and operational improvement by removing "defects" from products, processes and transactions. Some aspects of Six Sigma that are especially noteworthy are that (1) it is driven passionately by top management--an essential ingredient for success, (2) it is relevant to all aspects of operations, including product design and commercial transactions—and everybody in the corporation is expected to be involved, (3) it includes formal, but relatively brief and comprehensive, training, a significant part of which is in statistics and experimental design and (4) the companies involved have claimed significant impact on their bottom lines. (See Hahn, Hill, Hoerl and Zinkgraf (1999).) Although not all companies will formally adopt Six Sigma, most will experience similar pressures for a disciplined, data-oriented approach to improvement. This is because are all subject to similar competitve pressures. In fact, there will, likely, be comparable initiatives for organizations beyond industry, such as banking, hospitals, and even schools.

2. Opportunities

The new environment poses numerous challenges to our profession. Some may be concerned that, while the awareness of statistical tools has never been higher, this has not always resulted in an increased appreciation of the value of statisticians. In fact, managers in today's leaner and meaner companies might ask," Now that statistical tools are readily accessible to all, why do we need statisticians?" Our response can, perhaps, be likened to that of computer scientists. Early in my career, computer analyses were so complicated that they were generally the exclusive domain of computer scientists. Today, most professionals use computers with relatively little guidance from computer scientists. However, computer scientists, far from becoming obsolete, are more important than ever--they have just advanced to do higher level work. So must we! But how? What we need do to succeed, can, in my opinion, be captured under two headings:

- Accept, adapt to, and leverage the new environment (see also Hoerl, Hooper, Jacobs and Lucas (1993)).
- Provide well targeted added technical value.

2.1 Accept, Adapt to, and Leverage the New Environment

Accept: One eminent statistician has been quoted as saying that making statistical tools available to non-statisticians is like handing over the keys of a car to a child. I have greater trust in the world, and must, respectfully, disagree. In any case, the "democratization of statistics" is a fait accompli. Moreover, we will make no friends by suggesting that there is no way for practitioners to learn from intensive four week Six Sigma training, or a few short courses, more than a small fraction of the knowledge that we have amassed in years of study and practice. Instead, this should become evident from our contributions. Moreover, we should applaud the fact that our colleagues have come to appreciate the value of statistics, and celebrate the fact that this allows us to go on to yet bigger and better accomplishments.

Adapt: We must *help non-statisticians be maximally effective in using statistics.* This includes getting across, in the language of our customers, basic concepts (as opposed to specific tools)--such as Deming's (1975) distinction between enumerative and analytic evaluations, stressing underlying assumptions and their importance, demonstrating the usefulness of graphical approaches, helping fill in the important gaps that inevitably exist in applying standard canned software to specialized situations, and undoing much of the not very useful stuff (such as undue emphasis on significance tests) for which we have only ourselves to blame. In short, we should act as teachers, guides, and mentors.

Leverage: However, most important is that we do much more than react to the current environment--and, instead, seize the opportunity to move forward with it. This has various implications. One is to *strategically select* a few, among the many, potential, projects in which we might get involved for more in depth participation. The criteria for selection include the business importance of the problem, its technical complexity relative to the knowledge base of other team members, and, of course, the chances that we will succeed. This includes ensuring that we have effective working partners. Secondly, we need *look at problems broadly*, or from a holistic, rather than a narrow, parochial, view. Finally, we need do our utmost to promote *a proactive approach*. Traditionally, it has been easiest for us to become involved in "crisis" situations--where we are asked to assess the magnitude of the problem and, sometimes, to help evaluate the viability of alternative solutions. We should use these situations as launching pads to help understand root causes and to establish systems to avoid the future problems. All of this calls for the skillful balancing of serving as loyal and trusted team members, and, in selected areas, providing leadership.

2.2 Provide Well-targeted Added Technical Value

The new environment provides unparalleled opportunity for us to add needed technical value. This comes in many forms and, varies from one business to the next. I'd like to suggest a few common threads. The variety of issues, briefly raised here and reflecting my personal experience, emphasize the richness of the opportunities.

Focus on Tools Targeted at Analyzing Large Databases. It used to be that statisticians concentrated on squeezing the maximum amount of information from limited data. This made us focus on statistically efficient methods. Today, more and more, the volume of available data exceeds our ability to digest and to transform such data into useful information. Thus, we need concentrate on providing tools directed at handling large data sets. For example, we have found Classification and Regression Tree (CART) and related analyses to be particularly appealing in evaluating customer profitability data. A typical application is that of establishing groupings for different marketing actions. These methods have also been found attractive because of the easy-to-understand graphical summaries that they provide.

Help Move Product Quality Improvement Up Front. Industrial statisticians' major focus has traditionally been on statistical process control (SPC) for manufactured products. However, many quality defects are due to design, rather than manufacturing, problems. Thus, a major challenge is to provide engineers improved tools to rapidly identify and remove the root causes of design problems. A related goal is to make the product maximally immune to manufacturing fluctuations and to variation in the customer use environment. In this area, we can build on the concept of robust design advocated by Taguchi and his associates (see, for example, Phadke (1989)).

Emphasize the Proactive Elements of Reliability Improvement. The most important element of product quality experienced by our customers is often quality over time, or reliability. Again, the statistician's past role has frequently been reactive, i.e., quantify the magnitude of the problem, rather than understand the root causes and address them for future improvement. Moreover, life data analysis presents some special technical problems (e.g., handling non-normal distributions and censored data) of which most practitioners, and, for that matter, many statisticians are unaware. For example, one frequently encounters such technical complex situations as life data that simultaneously involve both left and right censoring (e.g., some units have failed prior to their first inspection times; others remain unfailed). The combination of business importance, the new emphasis on building reliability into the design of products, and the technical complexity presents special opportunities for statisticians, see Hahn, Doganaksoy and Meeker (1999) for a user-oriented introduction.

Help Integrate SPC and Control Theory. When dealing with manufacturing processes, we need work towards the closer coupling of classical statistical process control (SPC) with engineering control theory, rather than regarding these as two different disciplines. Box and Kramer (1992) and Vanderwiel, Tucker, Faltin and Doganaksoy (1992) discuss this opportunity.

Emphasize the Need for Building Adequate Databases. This is, perhaps, paramount among our challenges and represents one of our greatest opportunities, as well as frustrations. Typically, the databases available to us--especially for commercial applications and reliability analysis--were developed, not for the purpose of incisive analysis, but to provide financial or other information. When we inherit such data, we find important gaps and inadequacies in meeting our needs. As a result, much of our time is spent in getting the data into reasonable shape and developing clever schemes to get useful information from limited or ambiguous data. Thus, our best opportunities often arise in helping design databases for new processes. In many situations, e.g., reliability monitoring for a household appliance, it may be impractical, or too expensive, to develop a database that includes all units in the population. In that case, our goal might be to develop a sound sampling plan. In fact, frequently, our major contribution is that of emphasizing the fact that a small amount of well targeted and complete data, when appropriately selected, easily beats a large amount of irrelevant or incomplete data.

Promote Sequentially Designed Experiments. There is today a good appreciation, at least among engineers, in industry of the criticality of conducting designed experiments--thanks, in part, to the efforts of Taguchi and his associates. However, these are frequently thought of as "one shot" affairs . The importance of designed experiments as a stepwise learning process—see Box, Hunter and Hunter (1978) and Box (1999)—needs to be better recognized and appreciated by practitioners.

Revisit the "College Admission Officer" Problem. A college admissions officer wants to use the performance of previously admitted students to guide future admissions policy. Thus, analyses are performed to relate students' grade point averages to their credentials at the time of college application--as measured by SAT scores, high school grades, etc. One problem with such analyses is that the available information is on the partial sample of students who were previously accepted by the college and who chose to enroll, while the results are to be applied to the much wide group of *all* applicants. Similar problems arise in using past performance to develop models for making credit decisions, and a variety of marketing applications. In each of these cases, we need ways to assess how the possible bias introduced by having a non-random sample impacts the results, and to correct for such bias.

Provide Tools that are Robust to Misuse by Practitioners. With the increasing use of statistical methods by non-statisticians, it becomes increasingly important for us to provide approaches that are robust to misuse. This suggests, for example, more use of graphical methods and analyses that do not rely heavily on distributional and other assumptions. It also includes providing appropriate "customer protection" warnings. These might, for example, discourage extrapolations beyond the range of the data (or, at least, raise warning flags), explain key assumptions and their consequences in the tools that we provide practitioners, and encourage sensitivity analyses.

Promote the Use of Simulation for Conducting Statistical Analyses. Practitioners who have difficulty in understanding statistical methods often find simulation analyses more convincing. Sometimes, they even build up the confidence to apply this approach directly to their own problems. Moreover, statisticians are finding these approaches—perhaps, once regarded as a "lazy person's approach" more and more palatable. A typical example is that of assessing sample size requirements, for example, in a life test. Also there is an obvious tie-in to bootstrapping and jacknifing methods. In short, we should promote the use of simulation —sometimes, even when analytic solutions exist.

3. Concluding Comments

As we approach the 21st century, industrial statisticians face unparalleled challenges and opportunities. Technology has expanded our horizons and radically changed the environment under which we operate. The new environment calls for statisticians with a much broader and more proactive outlook than ever before—and ones who recognize and seize the many new technical and non-technical opportunities. My comments have been those of an "on the firing line" industrial statistician. However, the issues raised are not limited to industry, because all of us share the new environment.

References

Box, G.E.P, Hunter, J.S., and Hunter, W. (1978). Statistics for Experimenters, Wiley Interscience.

Box, G.E.P., and Kramer, T.T. (1992). Statistical process monitoring and feedback adjustment—A discussion, Technometrics 37, 64-73.

Box, G.E.P. and Liu, P.Y.T. (1999). Statistics as a catalyst to learning by scientific method, Journal of Quality Technology, 31, 1-29.

Deming. W.E. (1975), On probability for a basis for action, The American Statistician, 29, 146-152.

Hahn, G.J., Doganaksoy, N., and Meeker, W.Q. (1999). Reliability Improvement: Issues and Tools, to appear in Quality Progress, May 1999.

Hahn, G.J. and Hoerl, R. (1998). Key challenges for statisticians in business and industry (with discussion), Technometrics, 40, 195-213.

Hahn, G.J., Hill, W.J., Hoerl, R.W. and Zinkgraf, S.A. (1999). The impact of Six Sigma Improvement—A glimpse into the future of statistics, to appear in The American Statistician, August.

Hoerl, R.W., Hooper, J.H., Jacobs, P.J. and Lucas, J.M. (1993) Skills for industrial statisticians to survive and prosper in the emerging quality environment, The American Statistician, 47, 280-291.

Phadke, M.S.. (1989). Quality Engineering Using Robust Design, Prentice-Hall, Englewood Cliffs, NJ.

Vander Wiel, S.A., Tucker, W.T., Faltin, F. and Doganaksoy, N. (1992). Algorithmic Statistical Process Control: Concepts and an Application, Technometrics, 34, 286-297.

Vere-Jones, D. (1995). The coming of age of statistical education. International Statistical Review, 63, 1, 3-23.